Escherichia coli responses to acid stress: Signal Transduction and Gene regulation.

by

Hrishiraj Sen

A thesis submitted to the University of Birmingham for the degree of DOCTOR OF PHILOSOPHY

Institute of Microbiology and Infection

School of Biosciences

College of Life and Environmental Sciences

University of Birmingham

September 2017

UNIVERSITY OF BIRMINGHAM

ABSTRACT

Microbial lab-based evolution is a fundamental technique to study evolutionary theory. It is a powerful method which can provide insights into the ability of a microbe to adapt to a biological process such as low pH. To investigate pathways that could lead to an acid resistant phenotype in *E. coli*, we evolved six independent lines or populations of *E. coli* K-12 MG1655 by iterative growth and dilution experiments for approximately 740 generations at pH 4.5. Clones isolated from these evolved were significantly fitter than the ancestor at pH 4.5. They also showed greater resistance to a range of stresses as well as better survival during macrophage invasion. Five of the six evolved strains had acquired an identical mutation in *rpoA*, and mutations in *cytR* in addition to other mutations. Molecular analysis of the fossil record of the evolved populations showed that the *arcA* mutations always arose first followed by the *rpoA* mutations. Investigating the genetic basis of adaptation showed that the mutations in *arcA* were loss of function in nature and conferred caused an intermediate increase in fitness. Transcriptional analysis of the evolved strains showed a global change in their transcriptional signatures with the significant upregulation of the *arcA* regulon. One of the key components of the *arcA* regulon is *rpoS*. Our investigation showed a significant increase in the RpoS activity of the acid evolved strains. The evolved phenotype was completely *rpoS* dependent. The elevation in the RpoS activity was a result of a combination of the *arcA* and *rpoA* mutation. Our study showed how *E. coli* can utilise its existing network of stress responses to climb its fitness peak at pH 4.5.

Importance: Enteric bacteria encounter a variety of stresses in the human gut. For enteric pathogens to cause infections, they must be able to combat different stresses for transmission and survival. Using model bacterial systems, we can use lab-based evolution to study how bacteria employ their stress response systems to adapt to various stresses. It has proven to be an effective platform to study the activation of pre-existing and new pathways in response to stress. Integration of fitness, genome and transcriptomic data provides the potential for predictive and quantitative understanding of genotype-phenotype relationships that emerge from lab-based evolution experiments.

Index

ACKNOWLEDGMENTS

List of Tables

List of figures

Escherichia coli responses to acid: Lab-based evolution of acid resistance

Chapter 1-General Introduction

1.1 Preface

Most living organisms, ranging from bacteria to multicellular eukaryotes, live in an ever-changing environment. They must be able to cope with a range of different environmental conditions to grow, survive and reproduce. This is true and applicable to bacteria as they habituate some of the most inhospitable environments on this planet. Most bacterial species live in dynamic environments and are constantly faced with changing environmental factors such as nutrient availability, salt, oxidation, pH etc. Some of these changes may be harmful and detrimental to their ability to propagate and survive.

These environmental changes can be defined as stress conditions or stressors. To survive rapidly changing environmental changes, bacteria have dedicated sensory systems which enable them to detect these changes. These changes or environmental stimulus are relayed across regulatory gene networks which control the expression of groups of genes. Depending on the type of environmental change or stimulus, a dedicated response is generated which allows the bacteria to adapt and cope with the stress (Foster *et al.,* 1995).

The extent or level of response is dependent on the type and magnitude of the stress and is accompanied by phenotypic changes. The process by which bacteria sense environmental stresses and respond by expressing protective features allowing them to survive the stress conditions is known as stress response (Neidharth *et al.,* 1983). The ability to modify gene expression and alter their phenotype is a characteristic feature of bacterial stress responses.

One of the first attempts to study gene networks and regulatory systems which control bacterial stress responses was done by O' Farrell *et al.,* 1975. The study showed the identification of a large group of protein which were simultaneously expressed when *E. coli* was exposed to heat stress. There were several other studies which were focused identifying stress response regulons using molecular techniques (Lin *et al.,* 1995), (Small *et al.,* 1991). Further studies using microarray analysis led to the identification of specific regulons which could be activated in response to specific types of stresses. A regulon can be defined as a fixed set of genes which are expressed under the same control pattern (Schimdt *et al.,* 1989). Ability to respond to a variety of environmental stresses relies on the activation of specific regulons which lead to the expression of protection against the stresses. These regulons constitute the stress response systems in bacteria.

Stress response systems show a high degree of conservation amongst prokaryotes (Baird *et al.,* 2010). For example, Wong *et al.,* 2012 showed that the molecular pathways which regulated

the expression of genes involved in acid stress response, showed great degree of similarity between *Escherichia coli* and S*almonella Typhimurium*. However, the conditions under which these pathways could be activated differed significantly. Several studies have shown that inspite of a high degree of structural and organisational similarity in the stress response networks amongst different bacterial species, the conditions under which these systems can be induced differ significantly (Graur *et al.,* 2007, Gottesman *et al.,* 2009, Vogel *et al.,* 2011).

With the advent of the genomics era and the availability of cost effective next generation sequencing, technologies to understand bacterial stress responses have significantly increased. The availability of fully sequenced genomes of model bacterial systems has propelled the development of tools to investigate different aspects of bacterial behaviour, such as bacterial stress responses. One such experimental tool which can be used to the study bacterial stress responses is laboratory-based evolution.

Lab-based evolution can be used as a platform to understand how bacteria employ pre-existing and new pathways to generate appropriate responses to variety of stresses. Lab based evolution is a useful technique to provide fundamental insight into:

- study of evolutionary process

- study of adaptation to cellular processes and stress

Adaptation can be defined as an alteration in the genotype or phenotype of an organism in a environment that results from natural selection and by which the organism becomes fitter to survive and reproduce in that particular environment. Lab-based evolution can be used to study the ability of bacteria to adapt to different stresses. Adaptation would require the activation of different stress response systems and lab-based evolution can serve as a useful tool to decipher the conditions under which different stress response systems can be activated. Lab-based evolution can be used to evolve populations of bacteria under specific stress conditions and probe the genotypic and phenotypic changes that result from evolution of resistance to the stress. The integration of fitness, genome and transcriptomic data would provide the potential for predictive and quantitative understanding of genotype-phenotype relationship. We can use lab-based evolution as a tool to understand three important questions related to bacterial stress response

1) How are bacterial stress response systems activated?
2) What are the driving factors?

3) Can we elucidate mechanisms and specific pathways that respond to specific stresses?

One of the primary objectives of this study was to integrate these questions into a strategy to study bacterial response to acid using lab-based evolution as the tool.

pH is an important environmental stressor which bacteria must combat and stress responses to low pH is relevant from a physiological point of view. Acid is the first line of defence employed by the human body against invading enteric pathogens. Radiotelemetry analysis has showed that pH varies across the human alimentary canal or gut. Enteric pathogens must overcome the acid test to be able to transmit across the human gut and cause infection. Hence, studying how bacteria respond to acid stress is of importance to medical microbiology and bacterial physiology.

To study such a complex question through lab-based evolution requires a well-suited model system. A model bacterial system is a bacterial species which has been widely studied with an exhaustive knowledge base. It has simple nutritional requirements, rapid growth rate and ease of maintenance (Lenski et al., 1992). *Escherichia coli* satisfies all the above requirements and a significant amount of experimental data is available on stress responses of this organism. This study focused on evolving independent *E. coli* populations at low pH and investigating the phenotypes that emerged from lab-based evolution. The strain of *E. coli* used in this study was the *Escherichia coli* K12 MG1655 or commonly also known as the Blattner strain. Before I move on to the specifics of *E. coli* acid stress response and lab-based evolution, I will briefly introduce some key characteristics of the *E. coli* K12 MG1655 and the relationship between responses to low pH and lab-based evolution.

1.2 *Escherichia coli*: A snapshot of history

It was in 1884, that Theodor Escherich, a paediatrician, first discovered that a rod-shaped bacterium was responsible for causing diarrhoea in infants. He was the first person to conduct a study on infant gut microbiota and discovered this rod-shaped bacterium which he named *Bacillus communis coli* (Hattfield *et al.,* 1963). His research findings suggested that this bacterium was a constituent of lower gut of human beings and warm bloodied animals and in some cases was also responsible for causing disease (Shulman *et al.,* 2007). This bacterium was later named after Theodor Escherich as *Escherichia coli* or *E. coli*

1.3 *Escherichia coli* K-12 MG1655

E. coli K-12 MG1655 has a genotype of F⁻ lambda⁻ rph-1. This strain was sequenced by the Blattner laboratory and the genome sequence of this strain was published in 1997 (Blattner *et al.,* 1997). The temperate lambda bacteriophage and F plasmid were cured of this strain prior to sequencing (Blattner *et al.,* 1997). MG1655 was derived and isolated by Mark Guyer from the strain W1485, which was derived in Joshua-Lederberg's lab from the original stab culture of the K-12 strain (Clifton *et al.,* 1963). The original K-12 strain was derived from the stool sample of a dysenteriae patient in 1922 (Escherichia *coli* and *Salmonella*: Cellular and Molecular Biology, ASM Press).

1.4 *Escherichia coli*: **Characteristics**

Escherichia coli is a gram-negative bacterium, facultative anaerobe and non-sporulating in nature. It has mean diameter of 0.25 to 0.37 μm in diameter and mean length of 2.0μm. Some strains of *E. coli* are flagellated while some are non-motile. It grows optimally at 37ºC and has a generation time of approximately 30 minutes, when grown in rich medium. It propagates through binary fission and its multiplication can be driven by growth on simple carbon sources (El Mansi *et al.,* 1986).

1.5 *Escherichia coli*: **Strains and serotypes**

Many members of the *E. coli* species differ from each other on some key aspects such as utilization of nutrient source, antimicrobial resistance or colonization of a niche environment. The strain typing is based on these differences amongst the members of the *E. coli* species. Up to present date approximately 1210 non-pathogenic and 623 pathogenic strains of *E. coli* have been isolated from various sources. Further evolutionary classification of strains can be done based on the type of antigens displayed on the surface of the bacterium. This is known as serotyping. The antigens are classified as O-antigens, K-antigens, type of flagellin and capsule. The K-12 MG1655 strain of *E. coli* which is also the strain of interest in this study has a mutation in the LPS pathway which prevents the synthesis of O-antigen in this strain. Till date 200 serotypes of *E. coli* have been catalogued (Chaudhuri and Henderson, 2012).

1.6 *Escherichia coli*: **The genome.**

Escherichia coli species exhibit a significant amount genotypic and phenotypic diversity amongst them. This diversity between the members of the *E. coli* species at the genotypic level influences key aspects of their behaviour and molecular processes. 20% of the core genome is similar amongst all the strains and isolates of *E. coli*. The core genome comprises the genes

present in all the strains and isolates of *E. coli*. The pangenome of *E. coli* which comprises the complement of gene sets present across the entire clade of *E. coli*. The pangenome includes gene sets which may typically differ amongst different strains and isolates of the *E. coli* species. The pangenome of *E. coli* approximately contains around 12000 genes as modelled by Gordienko *et al.,* 2013.

1.7 *Escherichia coli*: Relationship with the human host

Escherichia coli is an important constituent of the human gut microbiota, the human gut microbiota is composed a diverse (500+ taxa) group of microbes. *E. coli* is a facultative anaerobe found in the human gut in addition to Bacteroides and Firmicutes. The latter comprise 90% of the population of the human gut microbiome (Tenaillon *et al.*, 2010). *E. coli* constitutes around 1-1.5 % of the total human microbiome. It is postulated that the niche environment where *E. coli* houses itself in the human gut, is the thin layer of mucus in the mid-gut (Katouli *et al.,* 2007). This enables *E. coli* to have free access to the nutrients being absorbed by the micro-villi. *E. coli* has a mutualistic relationship with the host, *E. coli* helps the host by breaking down complex nutrients and by producing Vitamin K and Vitamin B12 vital for development in humans (Gritsenko *et al.,* 2001).

The human gut is a complex environment and *E. coli* has to face constantly changing conditions such as temperature, oxygenation, pH etc. In addition to these factors, it also has to combat competition of other microbial species that colonize the human gut. These factors act as external environmental stressors and the ability to survive in these conditions will depend on *E. coli*'s ability to generate appropriate responses to against these stresses (Schluter and Foster, 2012). Pathogenic strains of *E. coli* lacking the *rpoS* gene showed reduced ability to colonize mouse intestines and cause infection. RpoS is the regulator of the general stress response system of *E. coli* and has been postulated to have a role in pathogenesis in the human gut (Mata *et al.,* 2017).

1.8 *Escherichia coli*: Life outside the host

E. coli is found in large numbers in many environments but being part of the gut microbiome, it is most abundantly present in human and animal excreta. It is introduced into environment mainly through faecal transmission (Savagueau, 1983). Life outside the host is a constant combat for *E. coli,* it is exposed to range of environmental conditions which can be detrimental to survival. It faces competition from other microbial species, must overcome drastic changes in temperature, pH, oxygen, osmolarity etc. (Groisman 2003, Foster 2001). *E. coli* survives the

environmental stressors or signals by employing dedicated systems it possesses that aid in sensing these signals. The signals it senses are processed through a complicated network of pathways which generate the required stress response.

This scenario is true when *E. coli* is present in the gut or in the natural environment. It must combat a range of environmental conditions to be able to survive and propagate. The ability to overcome these hurdles is provided by the existence of stress response systems that can be activated in response to different stressors.

1.9 *Escherichia coli*: Epidemiology

As described in section 1.7 it has a role to play in human digestive system. But some members of the *E. coli* species are harmful to human beings. *E. coli* is the common causative agent of gastro-intestinal infections in humans. In addition to this, it also causes urinary tract infections, colitis, peritonitis, bacteremia, meningitides and blood-septicaemia (Wei *et al*., 2010). These infections are caused by pathogenic *E. coli* strains. These strains show different phenotypic characteristics compared to standard laboratory strains. For example, toxin production is a characteristic feature seen in pathogenic strains but not in standard laboratory strains (Patwa *et al.,* 2011).

Pathogenic strains have special molecular mechanisms via which they cause infections (Frankel et al., 2012). Some of the important pathotypes of *E. coli* are EPEC (entero-pathogenic), EHEC (entero-haemorrhagic). APEC (avian-pathogenic), UPEC (uropathogenic) (Leimbach *et al*., 2012). The Centre for Disease Control reported 73300 cases of *E. coli* infections caused by enteropathogenic *E. coli* O15:H7. The two most notorious pathogenic *E. coli* strains are O5:H17 EPEC and CFT031 UPEC. The WHO and CDC estimates these two strains and its variants are responsible for 35% of the *E. coli* infections caused worldwide (UNFPA, 2016).

1.10 *Escherichia coli*: candidate for lab-based evolution.

Over the last two decades microbial sciences saw a significant expansion in next generation sequencing OMICS, bioinformatics and the requisite capabilities of these platforms. With the advent of the genomics era, significant emphasis was applied on whole genome sequencing of some microbial genomes. The availability of the complete genome sequence of *E. coli* K-12 MG1655 made working with this organism much easier in the laboratory setting. It propelled further research into understanding important properties of this strain of E. coli. A significant

repository of knowledge has been accumulated on this strain. With respect to this study, *E. coli* MG1655 is a well-suited model system for lab-based evolution. The stress response network in this strain has been studied and defined to a considerable extent. The tools available to perform genetic manipulations in this strain are well defined. This strain was also the ancestor for the previous evolution experiment on acid stress conducted in the laboratory. These reasons make this strain a suited candidate to study lab-based evolution of acid resistance.

1.11 *Escherichia coli*: Responses to low pH

Escherichia coli K-12 MG1655 has different systems which are responsible for acid stress resistance. Stationary phase *E. coli* is very resistant to acid shock of pH 2.5. Gordon and Small, 1992 showed that stationary phase *E. coli* could survive up to 2 hours at pH 2.5 whereas 99% of exponentially growing *E. coli* experience cell death within 20 minutes of exposure to pH 2.5. Pathogenic strains of *E. coli* showed much lesser sensitivity to extreme acid shock compared to standard lab-strains of *E. coli*. Small *et al.,* 1993 showed that enteropathogenic *E. coli* isolates survived significantly better at pH 2.5 compared to the standard lab-strain W3110. The authors suggested that enteropathogenic *E. coli* strains need to colonise and infect the human gut. Since pH is an important innate defence barrier employed by the human body, pathogenic strains of *E. coli* have increased resistance against low pH. They need to overcome this barrier to be able to colonise the host and cause infection.

The increased tolerance of stationary phase *E. coli* was attributed to the increased level of expression of the stationary phase sigma factor RpoS or σ^{38}. RpoS is involved in regulating a host of genes which are involved in elevating the general stress resistance of the cell. Strains lacking *rpoS* showed increased sensitivity to acid stress in stationary phase (Foster *et al.,* 1994). RpoS is one of the key stress sigma factors of *E. coli*, under conditions of stress the level of RpoS increases in the cell. RpoS competes with other sigma factors to associate with the core RNA polymerase and thereby increase the expression of RpoS specific genes (Li *et al.,* 1997). One of the ways E. coli shows elevated resistance to acid is through the increased expression of RpoS.

The sensitivity to acid shock in exponential phase can be overcome by pre-exposing *E. coli* cells to a mild acid stress or shock of pH 5.5. Lin *et al.,* 1995 showed that *E. coli* MG1655 cells pre-exposed to a pH 5.5 survived better at pH 2.5 than cells which were directly exposed to extreme acid shock. Exposing cells to a mild acid shock caused the induction of acid resistance systems which confer protection to acid stress in exponential phase. The most important of

them is the glutamate decarboxylase system (GAD) or the AR2 (Foster, 2004). An overview of this system will follow later in the introduction. The induction of the GAD system causes leads to the hierarchical transcriptional induction of the GadE. GadE in turn activates the expression of the decarboxylases GadA and GadB and the glutamate-GABA antiporter GadC. GadE also activates the expression of two key periplasmic chaperones, HdeA and HdeB.

As cells progress into stationary phase, the level of RpoS increases in the cell, thereby elevating the general stress resistant phenotype. Increased RpoS levels in the cell, causes overexpression of GadE which in turn increases the expression of the GadA and GadB. This increased expression of the Gad genes is seen even in absence of acid shock and accounts for the highly acid resistant nature of *E. coli* in stationary phase (Cornet *et al.,* 1999).

This sensitivity to acid shock during exponential phase was the basis of the previous evolution experiment Johnson *et al.,* 2014, conducted in the laboratory. In this study, seven independent cultures of *E. coli* MG1655 were exposed to extreme acid shock of pH 2.5 for 20 minutes followed by recovery at pH 7 overnight. This cyclical propagation regime was repeated for three weeks and the acid resistance of the populations was tracked over this period. Approximately at the end of three weeks these populations became very resistant to pH 2.5. Clones isolated from these populations too showed the same phenotype. Initial hypothesis suggested that this phenotype was a result of increased RpoS levels in these strains.

But preliminary experiments showed that the phenotype was completely RpoS independent. Whole genome sequencing of these strains showed that all of them had mutations in the sensor kinase *evgS*, which regulates the AR2 system. The authors showed that these mutations caused constitutive activation of the AR2 genes which conferred this extreme acid resistant phenotype. The constitutive activation of the AR2 genes was seen even in the absence of a drop in external pH. This lab-based evolution study led to another project studying the mechanism of activation of EvgS, part of which is a component of my PhD and discussed in the paper included as chapter 7 of this thesis.

E. coli has different systems which are responsible for responding to acid stress. The molecular mechanisms by which these systems are activated need to be understood more clearly. Especially because many of these acid resistance systems have also been implicated to be important in pathogenesis. One of the ways we can study how these systems are activated is through laboratory based evolution. It can prove to be a powerful tool in understanding which

genes are selected when *E. coli* is evolved under conditions of acid stress. Lab-based evolution can be used to decipher which are the key genes which enable E. coli to adapt to low pH.

1.12 Microbial laboratory based evolution.

Microbial laboratory-based evolution is an experimental technique used to investigate the ability of microbes to adapt to environmental conditions. Microbes are constantly evolving in the natural environment through the process of natural selection. Evolution is a slow process and is driven by evolutionary phenomenon such as natural selection and random drift (Karl and Woese, 1967). Lab-based evolution can be used to study this phenomenon in real time in the laboratory amidst a controlled setting. Since evolution is a lengthy process, microbes are well suited subjects of experimental evolution.

This is mainly due to their shorter generation times and simple nutritional requirements. One of the key components of evolution is mutations. Mutations occur spontaneously and are selected through natural selection (Finn *et al.,* 2001). Lab-based evolution can be used to study the ability of microbes to adapt to different environmental conditions. With the advancements in molecular techniques, it is possible to determine what are the genetic changes or mutations that enable adaptation to an environment. Using these techniques, it is possible to find out how the mutations work and their role in adaptation to the environment.

In addition to studying the standard evolutionary phenomenon using lab-based evolution, the technique can also be used to study the ability of microbes to adapt to a particular biological process, such as acid stress response. Lab-based evolution can be very helpful in understanding bacterial stress responses. The approach allows the experimenter to investigate how different stress response systems are activated under different stress conditions. Whole genome sequencing cam be used to identify the mutations which enable the activation of different stress response mechanisms. Using lab-based evolution the experimenter can identify the genes which are responsible in conferring fitness in a particular environment. This approach may reveal different stress response mechanism which may not be possible using standard molecular biology techniques. It also enables the experimenter to determine if evolution of stress response in one environment comes at a fitness price in another.

To investigate how *E. coli* adapts to low pH, we have used lab-based evolution as the experimental approach. The experimental strategy was to evolve independent populations of *E. coli* in a particular pH value over a period of time. This process is known as a long-term evolution experiment (LTEE). LTEE involves replicate populations derived from a single

microbial lineage are cultured and diluted repeatedly under defined conditions. Samples are frozen down at definite intervals to create a fossil record of the experiment. The conditions under which the evolution experiment is performed deviates from optimal growth conditions. This allows selection to operate and evolution of mutants which are fitter at the evolution setting. This study has followed a similar principle to understand evolution of acid stress resistance against mildly acidic pH using *E. coli* K-12 MG1655. Before I explain the specific details of acid stress response in *E. coli*, I will introduce some key aspects of laboratory based evolution.

1.13 Advantages of using bacteria in lab-based evolution experiments

There are many advantages of using bacterial model systems for lab-based evolution experiments. Bacteria have shorter generation times which makes conducting long term evolution experiments easier. They have simple nutritional requirements and easier maintenance needs. They are easy to count and can be differentiated using auxotrophic markers. This aspect can be exploited when designing fitness experiments to compare fitness of evolved lineages and their ancestor. Since asexual reproduction is the dominant route of propagation, clonal proliferation increases the linearity and precision of replication. We can replay the evolution tape from various points in course of an evolution experiment to check if a genetic change was contingent on the occurrence of a prior event (Portnoy, Bezdan and Zengler, 2011)

1.14 Laboratory selection methods

There are three different methods of conducting long -term evolution experiments in the laboratory.

Batch culture: This the most commonly used method in microbial laboratory based evolution experiments. In a batch culture system, cultures are grown overnight. A defined volume of culture is transferred into fresh medium from cultures grown overnight. This cycle is repeated continuously and parallelly many cultures can be cultivated at one time. By changing the volume of the culture vessel, the throughput of the experiment can be increased considerably. For e.g. Slonczewski *et al.,* 2014 used 96 well plates to evolve 25 replicate populations of *E. coli* W3110 at pH 4.6.

It is easy to control environmental factors such as temperature, pH, growth medium, volume etc. using batch culturing. Since populations being evolved using batch culturing go through the entire growth cycle, it allows selection to operate in exponential and stationary phase.
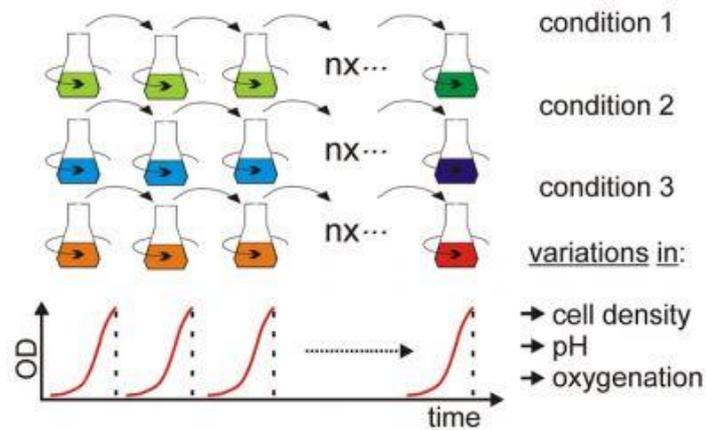
Figure 1.1: Batch culture system in lab-based evolution experiments. This figure has been taken from Barrick and Lenski, 2013. In a batch culture system, different environmental factors can be spatially regulated, and population sizes are not maintained as a function of time.

Continuous culture: Continuous culturing is an alternative method to batch culturing. Some research papers also refer to continuous cultures as chemostat cultures. Continuous culture systems are established in chemostats and bioreactors. These vessels are process controlled and automated, which adds tight regulation on the environmental factors inside the vessel and nutrient availability(Larsson, Enfors and Pham, 2009). The control systems are monitored using probes, which supply continuous information to a control unit. The control unit is fed with a regulatory program which keeps parameters such as dissolved oxygen, pH, nutrients at a constant level (Fig 2). The advantage of using a continuous culture system is that cells can be maintained a constant population size and at the same growth phase.
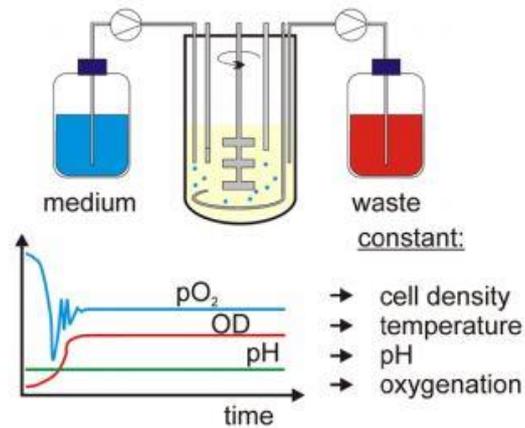
Figure 1.2: Continuous culture system in lab-based evolution experiments. This figure has been taken from Barrick and Lenski, 2013. In a continuous culture system, different environmental factors and population size are maintained and kept constant as a function of time.

Mutation accumulation: This method of laboratory selection is employed when studying mutation rates in a microbe and usually conducted on model organisms such *Escherichia coli* K12 MG1655. The method involves growing a population of cells in chosen environment and periodically bottlenecking the population diversity such that evolution proceeds only through random drift. Bottlenecking involves randomly choosing members of the evolving population to be propagated. The mutations present in the chosen individual gets arbitrarily fixed in the further descendants. Since selection operates arbitrarily in mutations accumulation experiments, only beneficial mutations accumulate in the evolved lineages. These experiments are used to study mutation rates and substitution rates in bacteria.

Laboratory selection criteria will vary experiment to experiment, but one common theme which acts as a guiding beacon on the appropriateness of the selection method is fitness. Enhanced fitness in an evolved lineage compared to its ancestor is usually the desirable end goal of any evolution experiment. Fitness can be easily measured using various methods of competition thereby by competing the evolved lineage with the ancestor.

1.15 Fitness and competition

Fitness or biological fitness can be described as a quantitative property which can be measured as the contributory effect of an organism or a genotype to future generations that emerges due to reproduction and differential survival. Fitness is an important parameter in microbial laboratory evolution experiments and direct measure of evolutionary adaptation when comparing evolved lineages and their ancestor.

Fitness is relatively easier to measure in case of bacteria. The most commonly used method is the competition experiment. The method involves introducing two competing strains in equal proportions into a competition setting and sampling their relative abundance over time. The outcome of the competition experiment is determined by calculating the fitness index (Materials and methods, sec. 2.5). Fitness index is a dimensionless number and denotes the fitness of the two competing entities at a time point, relative to each other (Kurland *et al.,* 1992).

Competition experiments can be carried out using different methods such as the conventional auxotrophic marker assays, quantitative PCR fragment analysis, resistance to antibiotics etc. Each of these methods utilises a form of a marker which can differentiate between the ancestor and the evolved lineage. Effective controls and replication is required to prove that the markers are competitively neutral and unbiased (Lenski *et al.,* 1998).

Enhancement in fitness is a direct consequence of adaptation. Adaptation to a particular environment is driven by different genetic changes which alter the phenotype of an organism. Since most microbes reproduce asexually, the dynamics of adaptation are largely under the control of mutations or genetic changes occurring in the genomes of the evolving lineages. These changes or mutations must survive the stochastic process of genetic drift to ensure transmission to the next generation (Cooper *et al.,* 2003). Mutations occur at a low rate in course of natural selection. There have been controversial claims in some bacterial lab-based evolution experiments which suggested that bacteria can sense and direct beneficial mutations which give fitness advantages (Bull et al., 2015). This has been largely refuted by several studies which focused on the role of mutations and their effects on evolving lineages. Mutations play an important role in influencing the phenotypes that emerge from microbial laboratory-based evolution experiments. The next section will briefly review the commonly observed mutations in lab-based evolution experiments.

1.16 Mutations and their role in lab-based evolution.

Mutations are classified according their nature: Single nucleotide polymorphisms (SNP), small scale insertions and deletion (indels), intragenomic recombination, movement of transposable elements (insertion sequences), amplifications and large deletions (Schneider and Lenski, 2004) . A recent review on different mutations that are selected in lab-based evolution experiments showed that SNP's or single nucleotide polymorphisms account for 61% of the total genetic changes that emerge from a laboratory based evolution experiment (Dragosits *et*

al., 2013). One of the mechanisms by which gene function is altered (loss and gain, altered or diminished) is through SNP's. 19% of the genetic changes in laboratory experiments were accounted by small deletion [indels] 6%, large scale deletions 2%) and insertion elements (3%) (Dragosits *et al.*, 2013).

A model example on how different mutations play a role in lab-based evolution experiment is the long term evolution experiment conducted by Richard Lenski. The study has been ongoing where 12 replicate populations of *E. coli* have been evolved for 50000 generations under glucose limited conditions. These population adapted to glucose limitation quickly within 5000 generations of batch culture. Investigations revealed clonal isolates from these populations harboured mutations (SNPS's) which caused complete inactivation of *rbs* or ribose utilisation and uptake operon. The SNP's were present in 11 out 12 clones isolated from the 12 evolved populations. The 12th clone has a largely insertion led deletion of the *rbs* operon, the *rbs* operon is located next to an insertion sequence known as the IS50 transposase. This IS element underwent a selection driven recombination event which led to the deletion of the *rbs* operon (Schneider and Lenski, 2004). This is a classic example of convergent evolution where two different types of mutations selected through lab-based evolution lead to a similar phenotype.

From the perspective of evolution, for any mutation to establish itself clonally amongst a heterogenous evolving population, it must be able to complement the strength and nature of selection. Even if it is successful in getting selected, it must not be lost due to genetic drift and should be able to substitute itself to the next generation (Kishimoto *et al.*, 2010). The dynamics of adaptation are not simple and in an evolution experiment there can be many routes by which beneficial mutations and are substituted in a evolving population. The fundamental process of transmission of mutants from one generation to the other depends on the combination of the

1.17 Mutation rate and substitution rate.

The rate of appearance of new mutations spontaneously during DNA replication and transmission of genetic information to offspring's is known the mutations rate. It is the compound rate of polymerase reading errors, recombinations, transpositions, unrepaired DNA damage and all other genetic processes which incorporate mutations (Lenski et al., 1992). A dominant conclusion several lab-based evolution experiments is that relative frequencies spontaneously occurring mutation rates are very low (fig 1.3)
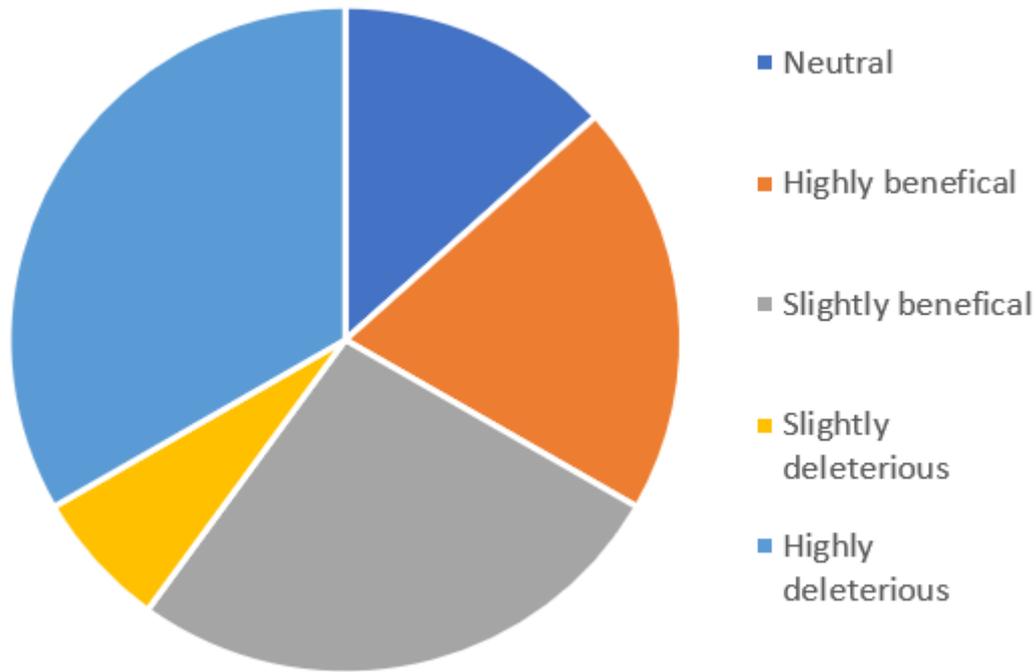
Figure 1.3: Relative frequency of different mutations in a laboratory-based evolution experiment. The figure was made based on data adapted from Gresham et al., 2008, Atsumi et al., 2009 and Gabriel et al., 2016.

The substitution rate of a mutation is the likelihood of that mutation being randomly selected through selection and being transmitted to the offspring. In evolution experiments, evolving populations are mixed and heterogenous. The populations consist of a mixture of mutants many of them having beneficial mutations which may enable adaptation. But most of these mutations are lost due to random drift and their inability to create clonal interference (Lenski *et al.,* 2010) to be able to propagate and outcompete others in the population. Even with lineages which harbour mutations in DNA repair systems creating mutator phenotypes, substitution rates of beneficial mutations are low. In a recent laboratory based evolution study conducted with Escherichia coli K12 MG1655, the experimenters, observed that evolved lineages with mutations in the base-pair excision repair genes and DNA repair genes such as *mutS* and *mutT* had 40%-60% higher mutation rates compared to other lineages which didn't, but only 6% of those mutations were finally substituted and available in the end-point phenotypes that emerged from the evolution experiment (Woods *et al.,* 2011). Mutation rates alone don't determine the evolvability of an organism, it is also determined by the ability of those mutations to substitute themselves. The percentage of beneficial mutations which get substituted is very low (fig 1.4).

In long term laboratory based evolution experiments the highest number of substituted mutations are neutral in nature.

Mutation accumulation experiments are an effective experimental avenue to study the accumulation of beneficial mutations, but since the populations are repeatedly forced through a bottleneck, lethal and deleterious mutations tend to be lost with a bias for mutations which only give a fitness advantage (fig 1.5). This one of the biggest drawbacks of mutation accumulation studies conducted over a short period and generations compared to long-term evolution experiments (Bachmann *et al.,* 2012).

An ideal evolution experiment is one which runs for several thousands of generations and large numbers of offspring are sequenced over many time points to create an exhaustive pool of information on mutation rates and substitution rates. This is necessary as the dynamics of adaptation is under their direct impact. Mutational robustness is important for the development of functions which enable an organism to climb a fitness peak in an evolutionary landscape. In evolution experiments as evolving acquire mutations, they traverse a certain path which leads them to their fitness peak (Barrick *et al.,* 2012).
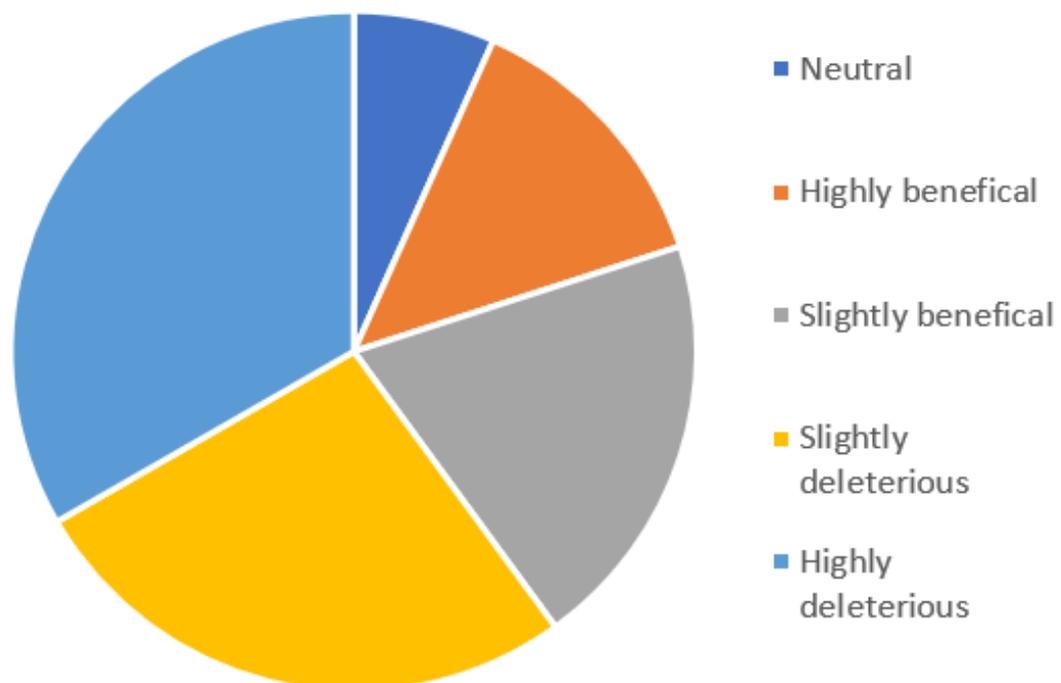
Figure 1.4: Substitution rates for different types of mutations in a laboratory-based evolution experiment. The figure was made based on data adapted from Gresham et al., 2008, Wood et al., 2011 and Bachmann et al., 2012)
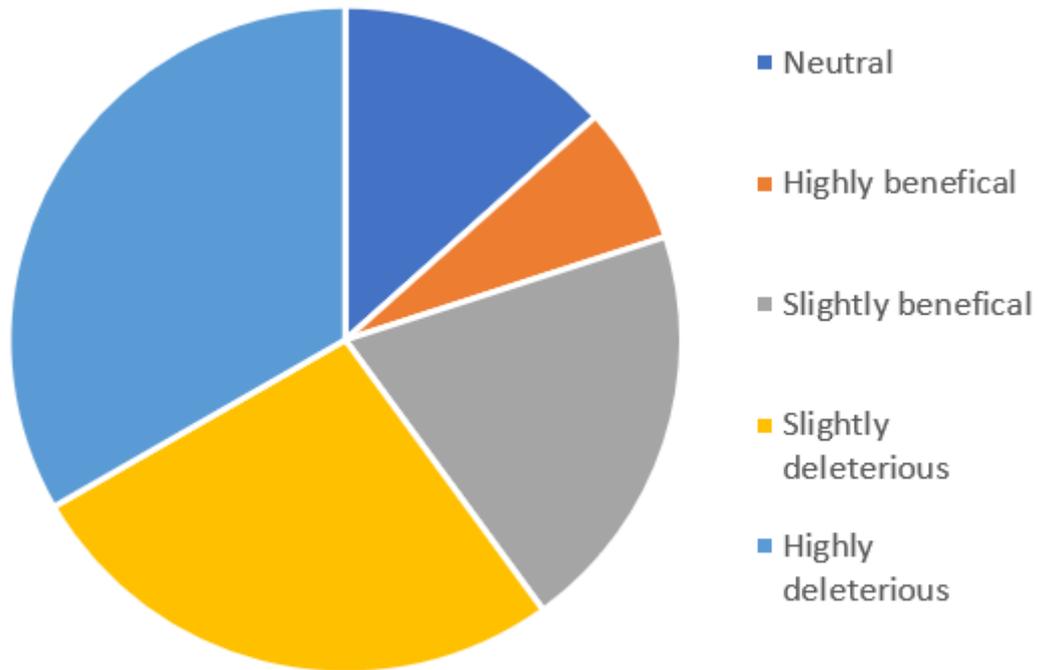
Figure 1.5: Substitution rates for different types of mutations in a mutation accumulation evolution experiment. The figure was made based on data adapted from (Barrick and Lenski, 2013)

1.18 Fitness landscapes and fitness peaks

One of the ways to study adaptation over a defined scale is to use a fitness peak (Conway and Morris, 1998). A fitness peak is a graphical depiction of fitness on the Y-axis plotted versus genetic change of an evolving lineage on the X-axis (Lenski *et al.,* 2015). Fitness peaks can be smooth in nature, showing a degree of parallelism of genetic changes (fig 1.6a) and rugged where the optimum shows a divergence from a common genetic lineage to different fitness peaks on the same genetic plane (fig 1.6b) (Conrad, Lewis and Palsson, 2011).

Different evolutionary studies have suggested that bacteria face a mix of smooth and rugged evolutionary landscapes in laboratory-based evolution experiments (Palsson et al.,2008). The landscapes they acquire largely depends on the type of mutations and the order of those mutations in the genetic space (Weinrich *et al.,* 2006. Rugged landscapes often result due beneficial mutations which creates a large effect early on in an evolution, followed by occurrence of other mutations which get selected due to varying genetic drift creating fitness peaks of unequal height (Wang *et al.,* 2010).

a   b
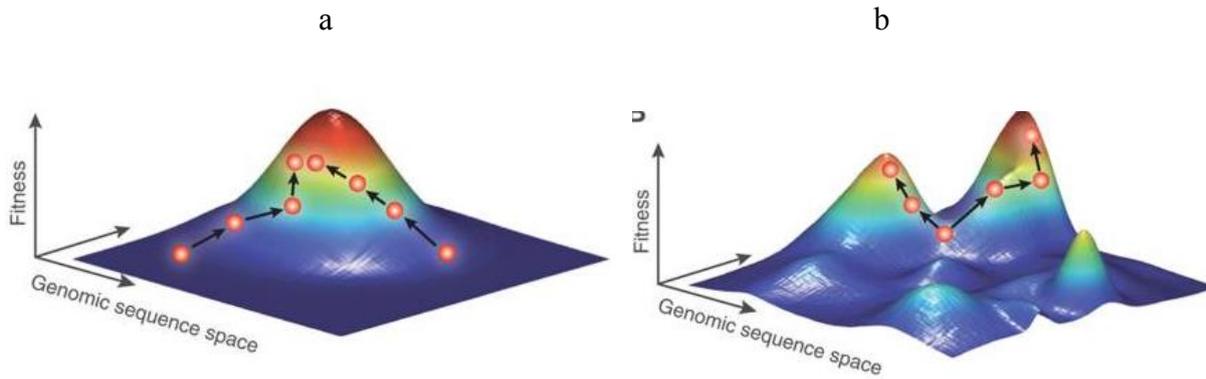
Figure 1.6: Two types of fitness peaks seen in evolution experiments. The two figures depict smooth (6a) and rough (6b) evolutionary fitness landscapes have been taken from Conrad, Lewis and Palsson, 2011. Smooth landscapes show convergence of genetic routes to a single fitness peak. Rough fitness landscapes depict the divergence of genetic routes from a common event. The arrows represent different trajectories found in replicate populations which are being evolved from a single common ancestor across the genetic space.

The entire concept of fitness landscapes depends on an important assumption that cells traverse a route which leads to an optimum state of functioning. A desired optimum can be defined as a state where the cells survive and grow with a growth rate that balances the production bacterial biomass and metabolic flux (Orth *et al.,* 2010). Several mathematical models have been constructed to predict and trace such optimum routes (Scott *et al.,* 2010). But in the evolutionary context the state of optimum can be adjusted based on the strength of selection. Varying the degree of selection can create varying routes to reach the same fitness peak. One such factor which influences this state of optimum is stress. In nature, bacteria constantly face varying conditions such as pH, osmosis, oxygenation, hydrostatic pressure, nutrient availability, competition for nutrients from other microorganisms (Nam, Conrad and Lewis, 2011). Stress responses can be broadly classified into two major categories, nutrient stress and environmental stress.

1.19 Nutrient Stress and the Lenski experiment

Nutrients are essential for survival, growth, energy and reproduction. It is a universal requirement across all domains of life including bacteria. Unavailability of nutrients causes reduction in cellular energy or ATP required for essential biological processes and homeostasis (Manch et al., 1999). This creates a selection pressure and bacteria which can adapt to this

selection pressure survive and ride out the phase of nutrient limitation through nutrient stress response.

One of the landmark lab-based experiment's is the Long-term evolution experiment (LTEE) conducted by Richard Lenski and colleagues. The experiment is ongoing has completed 65000 generations in January 2017. This LTEE has addressed some fundamental motivating questions regarding lab-based evolution such as repeatability of evolution and evolutionary hypotheses.

The Lenski experiment comprises 12 replicate populations of *E. coli* being grown on glucose-limited media on a daily serial transfer regime. The experiment was started in February 1988 and is still ongoing. *E. coli* cells are grown in batch mode, in DM medium supplemented with 25mg/l of glucose as the primary carbon sources. After 24 hours, an aliquot of cells is transferred into fresh medium. Every 75 days approximating to 500 generation population samples are frozen down and stored at -80⁰C to create a fossil record. Every time the cells are frozen down, fitness and growth rate of the mean-population samples are compared to the ancestor (Lenski *et al.,* 1992).

The experiment was aimed at understanding the dynamics of adaptation, the genotypic changes that enables *E. coli* to adapt to nutrient stress. The experiment focused on the repeatability of evolution by having 12 replicate populations. The experiment created a platform for 12 independent biological events to evolve independently and a scope to investigate the outcome of this evolution experiment. The Lenski experiment aimed at understanding the genetic changes that occurred due to adaptation to nutrient limitations. Were these changes similar across all the 12 evolving populations? Did these genetic changes show convergence? How does *E. coli* climb its fitness peak under conditions of nutrient limitation?

In the first 2000 generations of serial transfer, the replicate evolving populations exhibited bigger cell size and morphology and reduced final carrying capacity compared to the ancestor. the evolving populations had a quick gain of fitness in the first 2000 generations and fitness gain gradually slowed down leading to 20000 generation. This was seen across all the evolving populations with a high degree of parallelism (Lenski *et al.,* 1998).

Whole genome sequencing revealed the 11 populations had mutations in the ribose utilisation operon, while the 12th population had an insertion sequence driven deletion of the ribose *(rbs)* operon. Both types of mutations caused the inactivation of ribose operon. The evolved populations had comprised their ability to grow on ribose while gaining fitness on glucose limited conditions (Lenski *et al.*, 2002). The 12 replicate population also had different

mutations in the stringent response regulator *spoT*. SpoT is a universal stress response protein and a key regulator of RelA which in turn regulates the expression and cellular levels of a secondary messenger ppGpp (Guanosine 5' diphosphate 3' diphosphate). The stringent response system is a universal stress response system activated when *E. coli* experiences nutritional starvation and limitation. ppGpp controls the expression of $1/3^{rd}$ of the total genes in the cell and is involved in redirection of energy from development and propagation to ribosome biosynthesis and amino acid production (Philippe *et al.*, 2007). The production of ppGpp is catalysed by RelC (ribosomal control protein 11) and ppGpp synthetase 1 RelA. The deacetylation of transfer RNA during cellular stress is the activating signal for synthesis of ppGpp. RelA catalyses the production of ppGpp, which is in turn converted to ppGpp by SpoT releasing a molecule of inorganic pyrophosphate in the cell.

GTP+ATP $\longrightarrow$ pppGpp +AMP (catalysed by RelA)

pppGpp $\longrightarrow$ ppGpp (catalysed by SpoT)

ppGpp is a universal transcriptional regulator, its effects are global, and it is involved in capturing RNA polymerase and redirecting RNA polymerase to transcription of genes which are helpful during stress. WGS of the populations also revealed mutations in *nadR* and *pykF*. Transcriptional analysis showed that these mutations altered the cellular pool NAD and phosphoenolpyruvate, enabling cells to grow faster exponentially. Some populations also had mutations in global regulators of DNA supercoiling genes *topA* and *gyrAB*.

In contrast to these results when this experiment was conducted in oxygenated chemostats, the evolved populations had a quick gain of fitness over 5000 generation. WGS of the populations revealed multiple lineages harbouring different mutations in the regulator of the maltose operon *malT*. Additional investigations revealed that these mutations caused inactivation of the maltose operon and in ability to grow on maltose.

Glucose-limited adaptation under conditions of anaerobiosis did not showed different genetic changes. WGS of the populations being grown anaerobically showed mutations in global stationary phase regulator *rpoS* and *ptsG,* an enzyme which regulates glucose transport. RpoS was involved in the direct upregulation of approximately 600 genes and a host of stress response factors. While mutations in *ptsG* increased the flux of glucose uptake in the cell (Charusanti *et al.,* 2010).

The Lenski LTEE has highlighted some of the fundamental principles of evolutionary hypotheses. Evolution under glucose limited conditions showed high degree of parallelism with respect to fitness. Independently evolved populations attained different trajectories to converge to the same fitness peak. This degree of parallelism and convergence was also seen in case of the genes that were selected in the LTEE. Different mutations in the same genes were repeatedly selected in the evolution experiment. Interestingly, mutations in the ribose operon and maltose operons showed that gain of fitness under glucose limited conditions was accompanied by trade-off in the ability to grow on other sugar sources. The Lenski LTEE has addressed some fundamental questions regarding evolutionary hypotheses and its underlying principles. But what we need to understand further using LTEE is the repeatability of lab-based evolution. To use LTEE as a method to construct a comprehensive genotype and phenotype relationships that emerge from experimental evolution. These questions are an important part in driving the overall goals of this study. Table 1.1 lists some of the key lab-based evolution experiments aimed at studying nutrient stress in *E. coli.*

Table 1.1: Prominent Laboratory-based evolution experiments on nutrient limitation stresses. This figure has been taken from (Dragosits et al., 2013)

Strain	Nutrient Stress	Selection setting	Selection Time	Reference
E. coli REL606	DM-minimal medium, glucose limitation	batch	>50,000 generations	Lenski et al., 1992. 1998, 2010
E.coli MG1655	Minimal medium, glucose-limitation	chemostat	217 generations	McRobb and Ferenci, 1999
E.coli MG1655	Minimal medium,glucose limitation	chemostat	280 generations	Weikert et al., 1997
E.coli MG1655	Minimal medium, glycerol limitation	batch	700 generations	Ibara et al., 2002
E.coli MG1655	Minimal medium, lactate limitation	chemostat	900 generations	Dekel and Akon, 2002
E.coli MG1655	Minimal medium, adaptation to 1,3-propanediol	chemostat	900 generations	Lee and Palsson, 2010
E.coli MG1655	Minimal medium, phosphate limitation	chemostat	750 generations	Wang et al., 2010
E.coli MG1655	LB medium, xylose supplemented	batch	1000 generations	Zhao et al., 2013
E.coli MG1655	Minimal medium, adaption to *n*-butanol	batch	700 generations	Dragosits et al., 2013

1.20 Environmental stress and the general stress response of *E. coli*

Bacteria constantly face changing environmental conditions such fluctuating pH, oxygenation, heat, cold, high salt etc. These changing environmental conditions can be unfavourable for survival. To combat such environmental stresses, *E. coli* has different stress response systems. Some stress response systems are specific in function, for e.g. the heat stress response system and the general stress response system. The general stress response system of *Escherichia coli* is under the regulation of the stationary-phase sigma factor (σ^{38}) or RpoS. RpoS is a common target repeatedly selected in lab-based evolution experiments corresponding to environmental stress such as high osmolarity, acidic and basic pH, high and low temperature etc.

RpoS is involved in regulating a host of genes during stationary phase, starvation and general cellular stress (Hengge *et al.,* 2002). Although not an essential gene (Hengge *et al.,* 1995), *E. coli* cells which lack *rpoS* are sensitive to a variety of stresses and have reduced growth vigour (Ferenci *et al.,* 2005). RpoS expression is tightly regulated in the cell and it is expressed at very low levels during exponential phase. RpoS expression is induced as cells cross mid-exponential phase, depletion of nutrients and reduction in growth rate are said to be the two dominant cues for the increased expression on RpoS (Atlung *et al.,* 2002).

RpoS regulates the expression of approximately 600 genes directly and 1500 genes indirectly during stationary phase. During normal growth conditions, the transcription initiation process involves the assembly the core RNA polymerase holoenzyme consisting of the α, $\alpha^{'}$ β, $\beta^{'}$and $\acute{\omega}$ subunits. To initiate transcription this core RNA polymerase holoenzyme must associate with the house keeping sigma factor σ^{70} (Savery *et al.,* 1992). During stationary phase, increased expression of RpoS or σ^{38}, leads to accumulation of rpoS in the cell. Transcription initiation patterns are altered as RpoS outcompetes σ^{70} to associate with the core RNA polymerase (Tanaka *et al.*, 1993). RpoS has affinity for different genes which are required for stationary phase growth. It increases the expression of genes involved in secondary and fermentative metabolism. RpoS creates a global change in energy distribution in the cell (Ron, 2013).

During cold stress, RpoS expression is induced by DsrA and RprA binding to the promoter region of *rpoS*. DsrA represses translation of HNS and promotes expression of RpoS. During salt stress and high osmolarity, transcription of *rpoS* is activated via a small RNA Hfq (Bak *et al.,* 2013). RpoS is activated by unorthodox two component system ArcA/ArcB during anaerobic stress. ArcA is the response regulator component of the two-component system, while ArcB is the sensor kinase. ArcB is a membrane spanning sensor and is activated by

changes in the redox potential of the cell membrane (Mascher *et al.,* 2010). ArcB transphosphorylates ArcA. Transphopshorylated ArcA activates expression of *arcZ*, which in turn binds to the promoter region of *rpoS* to initiate transcription (Gottesman *et al.,* 2006). Acid stress response is also regulated by *rpoS* at many levels. RpoS is involved in feedback loop regulation of the Gad genes. The Gad or the Glutamate decarboxylase system is the primary low pH response unit of *E. coli.* Elevated cytoplasmic pH causes a decrease in the cyclic AMP levels. Reduction in cellular cyclic AMP acts as a signal for incremental expression of RpoS (Sittika *et al.,* 2008). RpoS binds to the AraC-like transcriptional regulator *gadX*, GadX activates transcription of *gadE*. GadE activates the expression of the glutamate decarboxylases GadA and GadB which alleviate acid stress. A review on this system will follow in the coming sections of the introduction.

Protection against oxidative stress is also conferered by *rpoS*. In situations where the cellular concentration of hydrogen peroxide and free radical oxygen ions increase, oxidation responsive genes SoxR and OxyS undergo a confirmational change. This activates the expression of the gene SoxS, which binds to the promoter of *rpoS* to initiate its expression. RpoS then turns on the expression of genes such as *sodA* and *catA* (catalases) and *gstH* (glutathionetransferases) which confer the cell protetcion against oxidative stress through free radical scavenging (Chiang *et al.,* 2012).

rpoS is essential for the regulation of virulence genes in pathogenic *E. coli*. A recent study in 2017 showed that enteropathogenic (EPEC) *E. coli* strain E238/69 and LRT9 isolated from tissue biopsies taken from patients having lower gastro-intestinal infections showed increased survival at low pH (fig 1.7), oxidative and salt stress. Western blot analysis of key global regulators showed significantly higher levels of expression of RpoS when compared to the commensal strain of *E. coli* MT88. Deletion of *rpoS* in the pathogenenic strains reduced adherence to epithelial cells and their survival at low pH. Trancriptomic analysis showed increased expression of the adhesin *bfpA* in the pathogenenic strain, this increased expression was diminished on deletion of *rpoS* from this strain. On competing the $rpoS^+$ and $rpoS^-$ version of the strain E238/69 in a tissue invasion model, the $rpoS^-$ strain showed extremely reduced fitness (Gardenia, Garson and Beny, 2017).

A similar study conducted on 36 different highly infectious isolates of entero-haemorrhagic (EHEC) showed increased *rpoS* expression in these strains was responsible for elevated levels

of resistance to low pH, oxidative stress, nalidixic acid and sodium hypochlorite. These EHEC strains were viable even after 20 days of starvation (Torterrello et al., 2009).
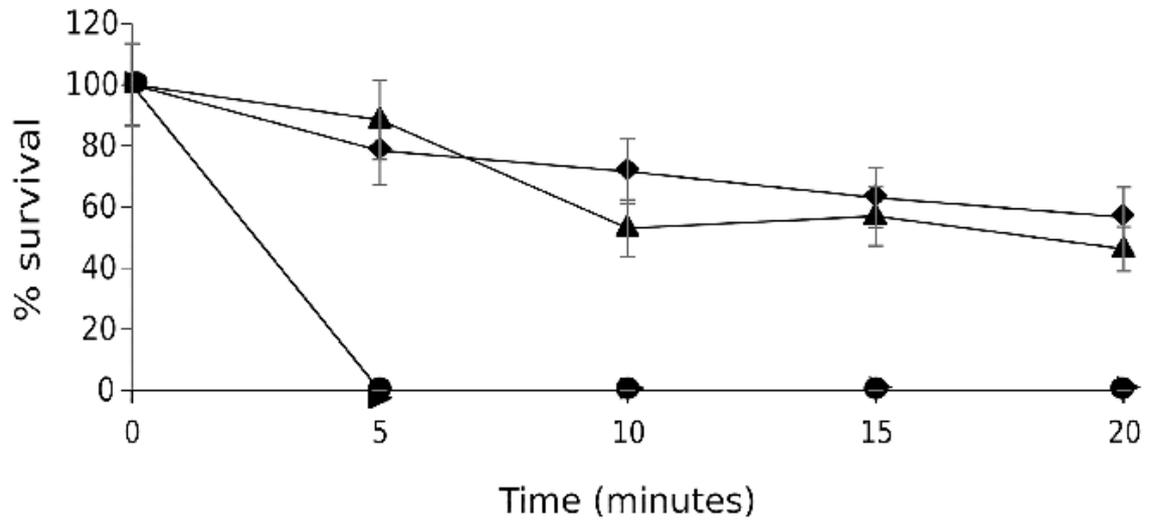
Figure 1.7: Percentage survival analysis during extreme pH shock. Enhanced survival of the two enteropathogenic strains E238/69 (▲) and LRT9 (♦) during low pH shock compared to the commensal strain MT88 (▶). This figure has been taken from Mata et al., 2017.

All of the above information directs us to the importnace of *rpoS* in *E. coli*. RpoS plays an important role in combating various types of environmental stresses.So why is RpoS so important in the cell?. The answer lies with the ability of *rpoS* to shape the SPANC or the stress protection and nutritional balance in the cell (Ferenci, 1995). The expression of RpoS causes a dynamic shift in the selctivity of the genes that are transcribed, genes which aid in stress management take the lead while nutrional competence is down regulated (Schneider *et al.*, 2006).

rpoS regulation is complex and pleotropic, but RpoS resposne can be catalogued in a simplified flow chart for our understanding (fig. 1.8). Disentangling the regylatory elements in the *rpoS* regulon has been diffucult. The ability to study such complex global regulatory circuits requires global platforms. Lab-based evolution is such a platform which may enable us to identify routes to further our understanding of complex regulatory networks in bacteria. Laboratory-based evolution gives us the experimental power and potential to study these gains and trade-offs that are made to alter the global profile of an *E .coli* when combating stress. Table. 1.2 enlists some of the key laboratory-based evolution experiments done to study environmental stress.
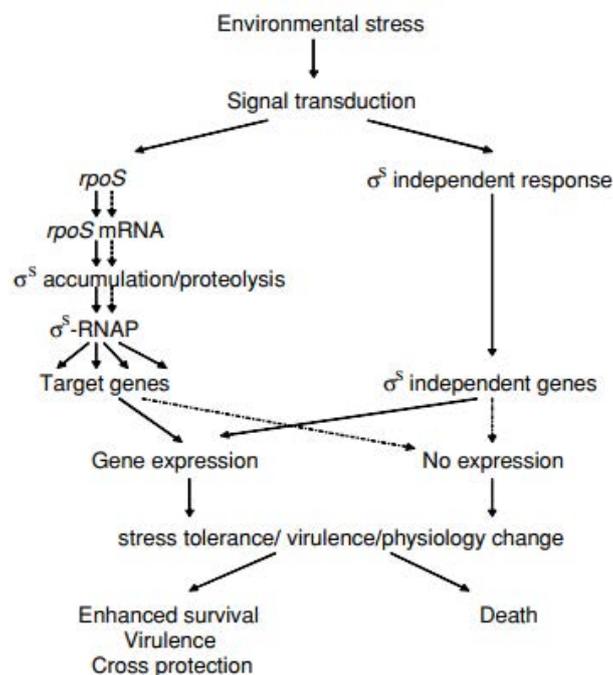
Figure 1.8: General stress response route of E. coli to cope with environmental stress. This figure has been adapted from Drake et al., 2016.

Table 1.2: Prominent Laboratory-based evolution experiments on environmental stresses. This figure has been adapted from (Dragosits et al., 2013)

Strain	Environmental Stress	Selection setting	Selection Time	Reference
E.coli MG1655	Adaptation to salinity (1.2M NacL)	batch	3000 generations	Finckel et al., 1996
E. coli MG1655	Adaptation to high temperature (41.5°)	batch	2000 generations	Rhiele et al., 2003
E.coli MG1655	Adaptation to ultraviolet light	solid-liquid combination	80 ultraviolet light cycles	Diez et al., 2004
E.coli MG1655	Adaptation to cold stress (freeze and thaw)	batch	200 cycles	Lenski and Sleight, 2007
E.coli MG1655	Adaptation to ethanol	batch	1000 generations	Horinouchi et al., 2010
E.coli MG1655	Adapation to high temperatur (42°)	batch	2500 generations	Palsson et al., 2010
E.coli MG1655	Adaptation to low pH (pH 5.5)	batch	500 generations	Dragosits et al., 2013
E.coli MG1655	Adaptation to extreme acid shock (pH 2.5)	batch	22 days	Johnson et al., 2014
E.coli W3110	Adaptation to mild acid shock (pH 4.6)	batch	2000	Slozwenscki et al., 2015

1.21 Acid stress

Gastro-intestinal pathogens have to face the gastric acidity as a innate defence prior to reaching the the intestines to cause infection (Smith J L, 1992). The acid attack is severe and enteric bacteria have to survive pH as low as 2, primarily in the stomach. pH is one of the factors which varies across the length and breadth of the human alimentary canal. Bacteria have to adjust to the changing pH and reach the lower gastro-intestinal tract, the common area of infections (Lee and Hao, 2004). The mucosal surface of the lower GI is around 200-300 mm^2 and approximately 1014 species of bacteria (fig 1.9) house the human gastro-intestinal tract (Bueno, Silva and Oliver, 2004). The types of bacteria that colonise different sections of the gut is dictated by a number of factors primarily pH, nutrient availability, mucin secretion, bacterial cooperation, bacterial antagonism, peristaltic speed. The pH of the stomach is extremely low compared to the other parts of the intestine (Lin et al., 1993).

The exit from the stomach is very swift due to fast peristalsis, a innate defence mechanism which is regulated by pH. The pH on the exit from the stomach and entry into the duodenum is of the range 3.5-6, many acid-tolerant lactobacilli house themselves in this region. This region is also a known site for upper GI infections caused by *Helicobacter pylori* and *Vibrio cholera* (Jordan *et al.,* 1999). The pH progressively becomes mildly acidic towars the ileum and jejunum. The dynamics change very drastically in the lower GI compared to the upper GI. The environment turns anaerobic, nutreints and there is an influx of short chain fatty acids and inorganic acids (Arnold *et al.*, 2001). *E. coli* causes infecrtions around this location in the intestine (Bueno, Silva and Oliver, 2004).

For pathogenic *E. coli* to reach the site of infection, it must be able to adapt and survive these varying conditions imposed by the human body. Bacterial pathogens have to breach this barrier and combat the acid stress for eventual transmission. Most enteric pathogens don't cause any infections in the human stomach, as it has the most inhospitable pH of 2-2.5.*Helicobacter pylori* is the only pathogen which cause infection in the stomach and it is one of the causative agents for stomach ulcers and in some cases cases stomach cancer cancer (Wrowbleski *et al.,* 2009). The ability to survive acid shock also depends on the infectious dose of the organims, *Shigella dysenteriae* has an infectious dose of around 10-500 cells and (30-80)% of these cells can survive acid shock at pH 2.5 (Dupont et al., 1971). Pathogenic *E.coli* have an infectious dose at the range of 10^4-10^6 cells, out of which 500-600 cells can survive an acid shock at pH

2.5 for more than 2 hours (Lin *et al.,* 1995). Salmonella on the other hand has a very low acid survivability, at an infectious dose of 10^6-10^9
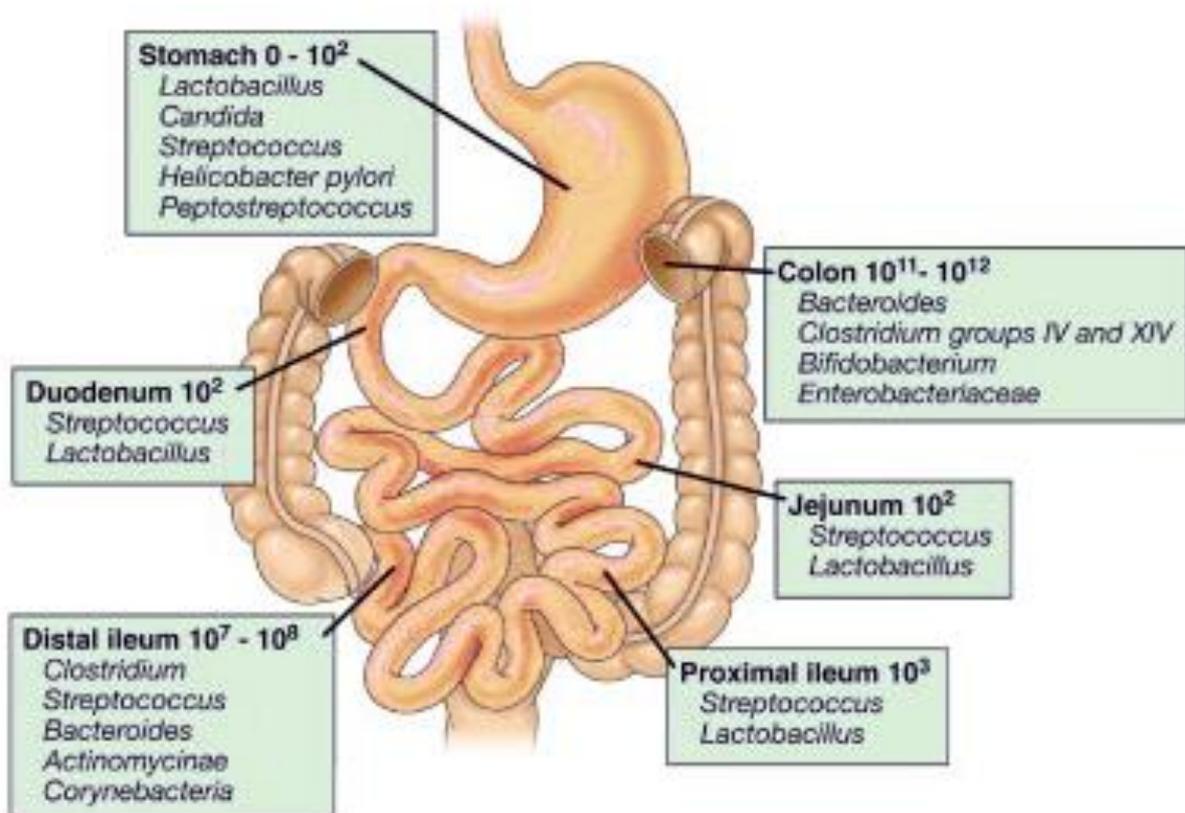
Figure 1.9: Distribution of commensal bacteria in the human gastrointestinal tract. This figure has been adapted from Microbial influences in inflammatory bowel disease, Gastroenterology, Sator, 2008.

Once *E. coli* is transmitted in the human gut it will encounter a complex mix of organic and short chain fatty acids and not just hydrochloric acid (Lin et al., 1995). From the perspective of medical microbiology, acid resistance is a phenotype of clinical relevance. If enteric pathogens have to cause infection, they should be able to override the acid stress barriers that they encounter in the human host. Studying the systems which they employ against acid stress is importance in clinical microbiology.

Before moving on to the details of acid resistance, I would like to highlight the concept of Proton Motive Force. Maintaining cellular homeostasis is fundamental for the survival of *E. coli*. One of the driving factors of cellular homeostasis is proton motive force. The transport of H^+ or hydrogen ions across the cell membrane generates an electrochemical gradient, this gradient creates a force by which protons can be attracted or repelled from the cell membrane, this force is termed as the proton motive force (Bradbeer, 1993). The transport of H^+ ions creates an electrochemical reaction in the cell, which converts ADP (adenosine di-phosphate) to ATP (adenosine tri-phosphate) through the F_o-F_1, ATP synthase (fig 1.10). This releases energy in the cell, which it can harness to perform a biochemical process, such as transport of a compound across the cell membrane (Divies et al., 1999).

The proton motive force (PMF) is made up of the cumulative sum of two components, the pH gradient or the chemical difference in the concentration of protons on either side of the cell (ΔpH) and difference in the electrical charge across the inner and the outer membrane, the membrane potential($\Delta\Psi$). Both these entities are represented in millivolts. We can mathematically represent PMF= $\Delta pH + \Delta\Psi$. The concentration of protons outside the cell is always greater than inside. ΔpH of the Escherichia coli is around -140 to -180 millivolts. This chemical difference in concentration contributes the membrane potential $\Delta\Psi$. Membrane potential is the total sum of the charges in the cell. If the ΔpH of the cell changes, the cell must compensate for this change by changing the $\Delta\Psi$ (Foster, 2004). A functional membrane potential keeps its integrity intact and all channels of trans-membrane transport functional. If the membrane potential change is radical, due to an influx or efflux of charge, it can cause membrane depolarisation and lead to cytoplasmic leakage causing cell death (Galiano, Fernandiz and Campa, 2001). The cell maintains a constant ΔpH, to prevent frequent fluctuations to the membrane potential. Changes in the ΔpH leads to an imbalance in the H^+ ion concentration in the cell. This imbalance is one of the causative reasons for acid stress. For example, extremophiles, such as *Thiobacillus acidophilus*, which lives an environment of pH 3 maintain an internal pH of 6.6 ± 0.2. They do so by employing the dedicated systems which

alters the membrane potential drastically to maintain cellular homeostasis (Richard and Foster, 2001). *E. coli* can survive exposure to pH 2.5 while it maintains an internal cellular pH of 4.6±03, it can do so by utilising specific systems which enable it to cope with acid stress.

Over the last two decades, significant advances have been made in understanding *E. coli*'s acid resistance mechanisms, the signal transduction network and the regulatory network involved in controlling the expression of for the purposes of this study I am going to give a brief review of the acid resistance systems of *Escherichia coli*.
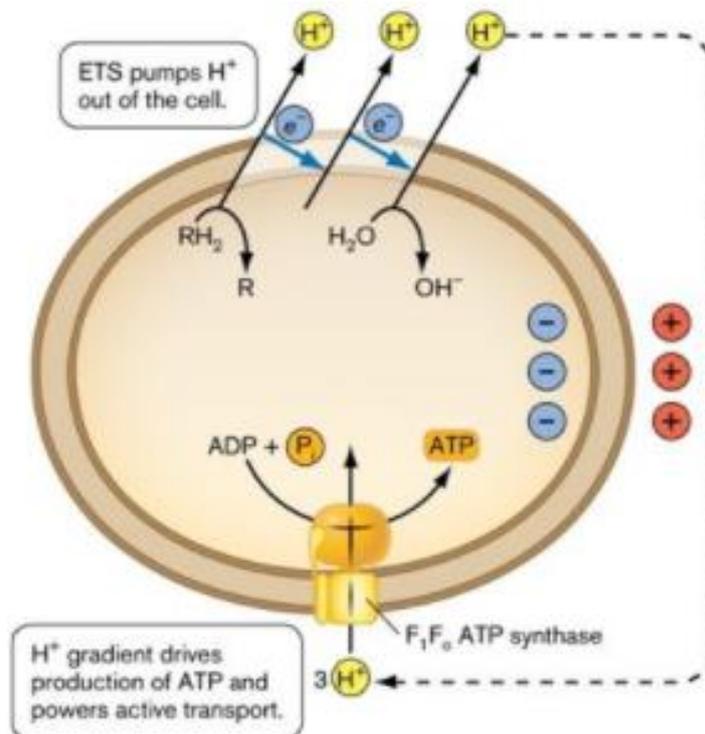
Figure 1.10: Membrane potential and proton motive force of the cell. This figure demonstrates the energy based transport of hydrogen ions across the cell membrane. Uptake of ions is dependent on the force generated by the difference in concentration (ΔpH), which contributes to maintaining a stable membrane potential ΔΨ.

1.22 Acid resistance systems of *Escherichia coli*.

Stationary phase *E. coli* is very resistant to extreme acid stress and as described in section 1.11, this resistance can be attributed to the increased levels of RpoS in the cell. Exponential phase *E. coli* on the other hand is sensitive to extreme acid stress (pH 2.5). The sensitivity to acid shock can be can be overcome by pre-exposing *E. coli* to mild acid shock pH 4.5-pH 5.5. Cells pre-exposed to a mild acid shock could survive an extreme acid shock of pH 2.5 for up to 2 hours (Miller *et al.,* 1992). 99% of the cells which were directly exposed to extreme acid stress experienced cell death within 10 minutes of exposure (Miller *et al.,* 1992).

Lin *et al.,* 1994 showed that this increased extreme acid survival phenotype of *E. coli* was dependent on the presence of some amino acids in the growth medium. Further investigations on the molecular mechanisms of extreme acid survival showed that, pre-exposing exponentially growing cells to mild acid shock lead to the activation of acid resistance (AR) systems. The impact of these systems in exponential phase was *rpoS* independent and required the presence of amino acids in the growth medium (Gordon and Small, 1992).

Escherichia coli has four distinct acid resistance systems (AR):

   a) RpoS dependent or acid response (AR1)
   b) Glutamate decarboxylase or Gad system (AR2)
   c) Arginine decarboxylase or Adi system (AR3)
   d) Lysine-cadaverine decarboxylase or Cad system (AR4)

These four acid stress response systems are the response units of *E. coli* to acid stress. They have different molecular mechanisms of activation and confer protection to different pH values (Smith *et al.,* 1996).

The *rpoS* dependent system is part of the general stress response of *E. coli*. RpoS is involved in activating the glutamate decarboxylase system during stationary phase growth, this is one of the factors that drives increased acid resistant nature of stationary phase *E. coli* cells (Shin *et al.,* 2001). RpoS also activates the expression of acid responsive periplasmic chaperones HdeA and HdeB, which aid in alleviating extreme acid stress. *rpoS* mutants show reduced tolerance to acid stress in stationary phase. This reduction in acid tolerance was also seen in *rpoS* mutants of pathogenic *E. coli* strains such as O15:H7 (Davis e*t al.,* 2002).

Exponentially growing *E. coli* on the other hand depend on the remaining three acid resistance systems to combat acid stress. These systems are amino acid dependent as their mechanisms

of action depend on importing the requisite extracellular amino acid and reductive decarboxylation of the amino acid using cytoplasmic protons. This process helps in reducing the cytoplasmic proton stress during acid stress. For the point of view of this study, I am going to present a brief review of the glutamate decarboxylase system and the evolution experiment conducted by Johnson *et al.,* 2014 which showed the role of this system in combating extreme acid stress. This study also helped direct the aims and objectives of this study.

1.23 Glutamate decarboxylase system or Gad system (AR2)

The glutamate decarboxylase system is one of the important acid resistance systems of *Escherichia coli*. This acid resistance system comprises the two decarboxylases GadA and GadB and a membrane antiporter GadC (fig 1.11). The Gad system is complex hierarchy of genes and figure 1.11 depicts the pathway that regulates the activation of the Gad genes. The Gad system is under the control of the AraC/XylR like transcriptional activator GadE. *gadE* is indirectly under the control of the unorthodox two-component system EvgS/EvgA (Yamashino *et al.,* 2001). EvgS is an unorthodox two component system which sits at the top of the regulatory circuit of the Gad system. Two-component systems (TCS) are tools which bacteria use to sense external environmental signals using a membrane bound histidine kinase (HKS) and intracellular response regulator (RR) which is involved in transcriptional activation of its target genes.

EvgS is known to be activated on exposure to mildly acidic pH of 5.5 (Burton *et al.,* 2010), (Kao *et al.,* 2010). Activated EvgS transphosphorylates EvgA, which in turn activates the promoter for *ydeP*. The regulatory role of ydeP is not clearly understood. EvgA binds to YdeO and YdeO activates the expression of GadE (fig. 1.11). YdeO binds to the transcriptional activating region of *gadE*, known as the Gad box. The Gad box is stretch of DNA 63 base pairs upstream of the transcription start site of *gadE* to which YdeO binds to activate *gadE* transcription.

GadE activates the transcription of *gadA, gadB* and *gadC.* GadC, the membrane antiporter imports glutamate from outside the cell into the cytoplasm. GadA and GadB reductively decarboxylate glutamate to GABA (gamma-amino butyric acid) by utilising a cytoplasmic proton (fig. 1.12). The decarboxylated product is extruded out of the cytoplams by GadC. This process continues inside the cell until pH homeostasis has been achieved and cytoplasmic proton stress is alleviated. This process is also followed by AR3 and AR4 which utilise arginine and lysine as the substrates for decarboxylation.
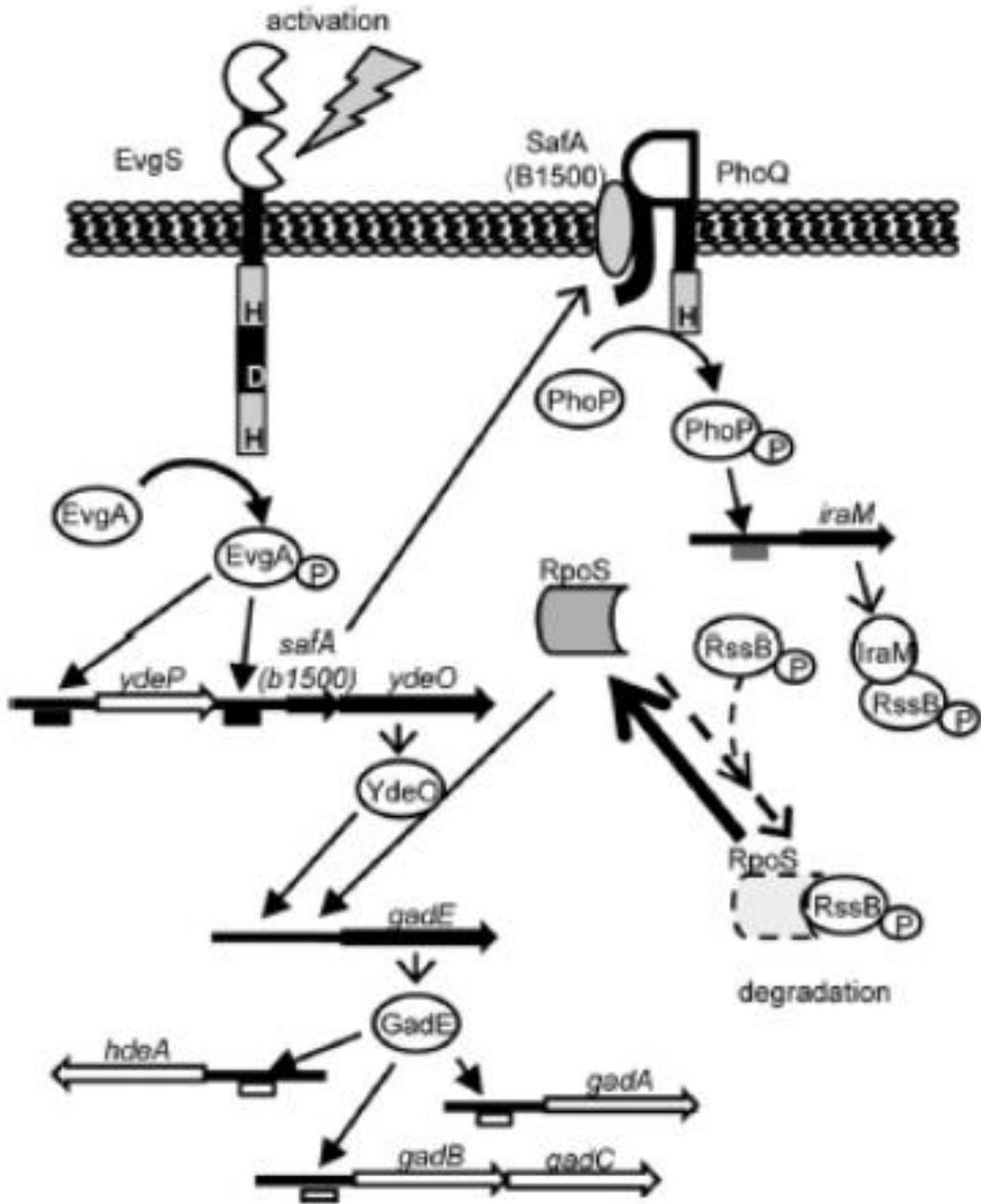
Figure 1.11: Activation of the glutamate decarboxylase system of E. coli. This figure has been taken from Eguchi and Utsumi, 2014.
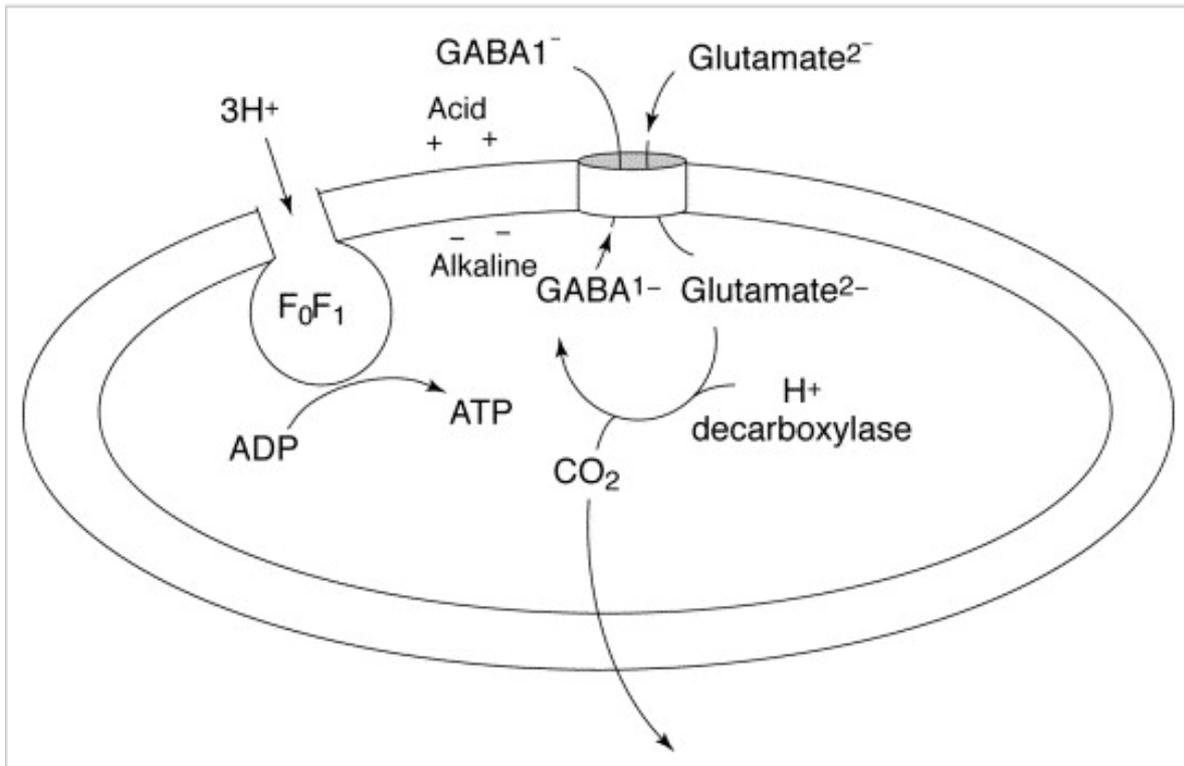
Figure 1.12: Reductive decarboxylation of glutamate to GABA (gama-aminobyutric acid). The figure has been adapted from Watermann et al., 2002.

1.24 EvgS/EvgA: Unorthodox two component system of *E. coli*

EvgS is the histidine sensor kinase component of the unorthodox two component system EvgS/EvgA. It is involved in the detection of low pH and activation of the glutamate decarboxylase system. Bacteria use two component systems to sense external environmental stimuli. The detection of an external signal leads to the phosphorylation of a dedicated histidine residue in the sensor kinase. This process is known as the activation of the kinase. The activated kinase transfer the phosphate molecule to the cognate response regulator. This process is known as transphosphorylation. The phosphorylated response regulator now acts as transcriptional regulator of its target genes. The mechanism described here is followed by conventional or orthodox TCS (fig 1.13 A). This one step relay has been observed across many two component systems of *E. coli* such as the EnvZ/ompR and KdpE/KdpD system (Mascher *et al.,* 2011).

Some TCS don't follow this single step phosphorylation mechanism. These TCS comprise histidine sensor kinases which have a different domain organisation compared to the orthodox sensor kinases. Unorthodox sensor kinases have multiple domains (Jourlin *et al.,* 1996) (fig. 1.13 B). These multiple domains are involved in an intricate internal phosphorelay before the activated kinase transphosphorylates its cognate response regulator. EvgS is one of the unorthodox sensor kinases of *E. coli*.

The focus and interest on EvgS in our laboratory started when it was selected repeatedly as a target gene in an evolution experiment aimed at understanding survival under extreme acid stress. EvgS is a sensor kinase which consists of two large periplasmic solute binding domains (fig 1.14). This constitutes the periplasmic regions of the kinase. The cytoplasmic region of the kinase has the internal phosphorelay domains and the PAS domain (Johnson *et al.,* 2014). PAS domains (Per-Arnst-Sym) are biologically ubiquitous (Christie et al., 2010) and play a role in signal transduction and dimerization. The structure of EvgS has been modelled by comparative studies with the virulence regulator sensor kinase of *Bordetella pertussis* BvgS.

EvgS is a key regulator of the Gad system and has been one of the subjects we have investigated in this study. Recent experimental investigations in the lab have led to the modelling of the structure of the periplasmic domains of this kinase based on the structure of BvgS. We have also identified key residues which are important in signal perception and structural integrity of the protein located in the periplasmic domain. We have also presented a model on the activation of kinase and the role of the cytoplasmic domain of the kinase in acid perception. The results

of this section of the study has been attached as a paper in chapter 8 of this study (Sen *et al.,* 2017). The details of this study which were investigated as part of my PhD project will be detailed in the aims and objectives section of this thesis.
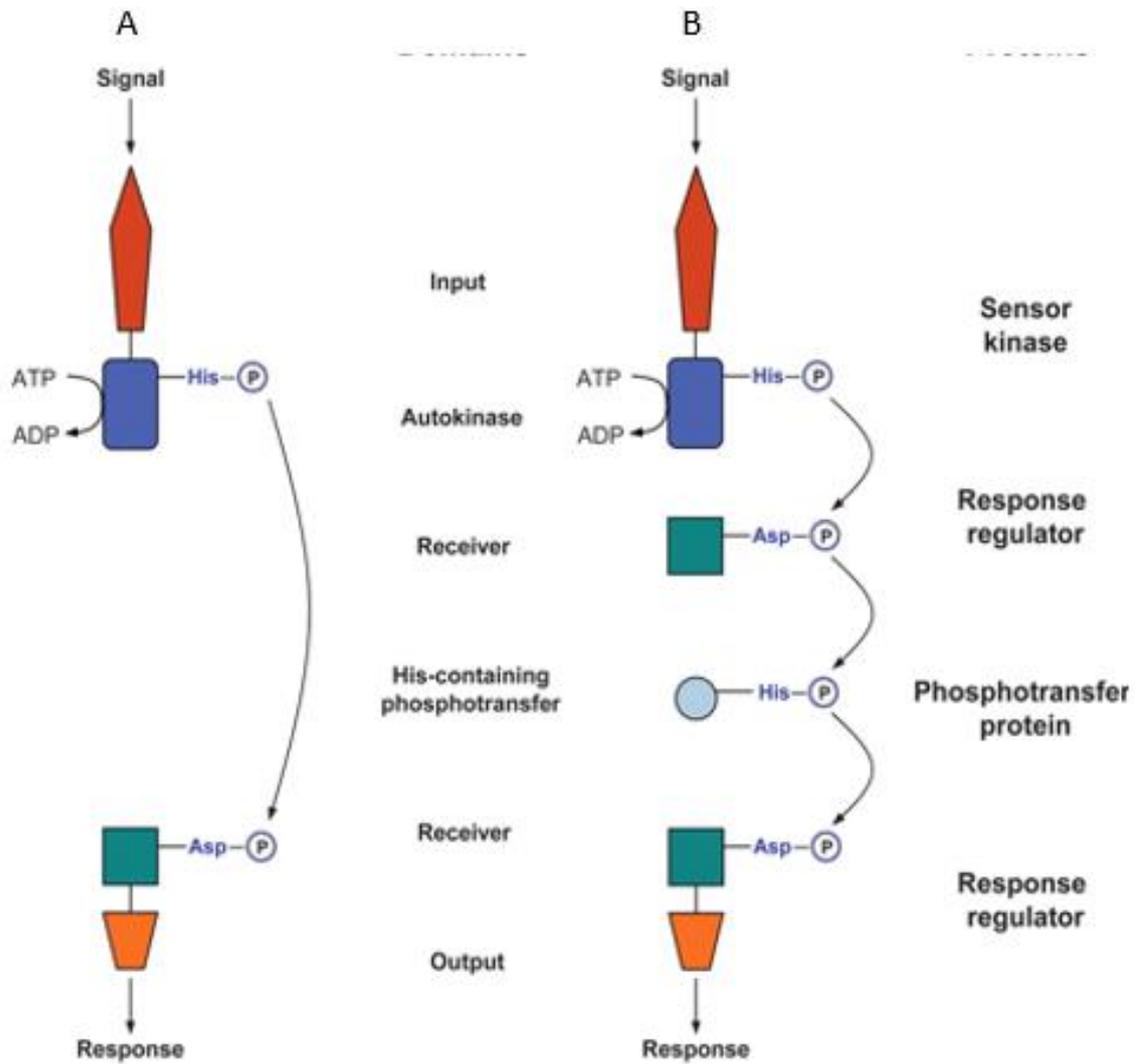
Figure 1.13: Two types of two component systems. This figure was taken from Herrou et al., 2007. The figure shows the functional characteristics of an orthodox (A) and unorthodox (B) two component system.
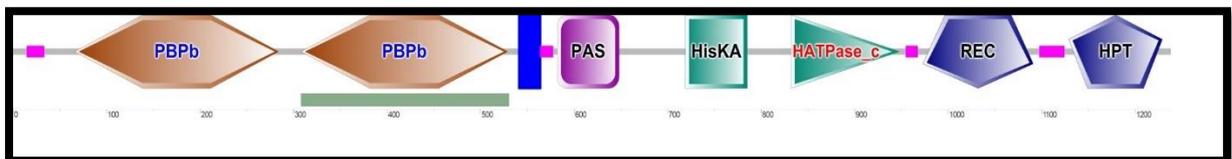
Figure 1.14: Domain organisation of the sensor kinase EvgS. This figure was generated using the domain prediction and organisation tool of the SMART database (http://smart.embl-heidelberg.de)

1.25 Evolution of acid resistance under extreme acid stress pH 2.5 (Johnson *et al.,* 2014).

Lab-based evolution is a powerful tool to study how bacteria employ pre-existing stress response systems to adapt to different stress conditions. Exponential phase *E. coli* is very sensitive to extreme acid stress (XAR). Exploiting this sensitivity to extreme acid stress, Johnson *et al.,* 2014 setup an evolution experiment aimed at evolving survivors of XAR through experimental evolution. Lab based evolution provided the scope to connect the molecular mechanisms which respond to extremely low pH and their effect on the phenotype of *E. coli.*

The evolution experiment comprised 5 independent populations of *E. coli* MG1655 derived from a single common ancestor. In this study, five independent cultures of *E. coli* were grown to mid log phase of OD_{600nm} of 0.2-0.3. These mid log cultures were serially diluted in LB broth at pH 2.5 and incubated for 20 minutes at 37ºC. These cultures were then recovered overnight at optimal growth pH 7. The overnight grown cultures were diluted and grown to mid log phase and exposed to pH 2.5. This cyclical propagation was repeated for around 25 days. Initially the strains were acid sensitive but after 14 days of propagation a acid resistant phenotype emerged (fig 1.15A). This extreme acid resistant phenotype increased over the next few days of the regime. The propagation regime was stopped as population attained significantly better survival at pH 2.5 compared to the ancestor (Johnson et al., 2014). This extreme acid resistant phenotype at pH 2.5 was also displayed by clones or extreme acid resistant strains isolated from the five populations.

Initial predictions and hypothesis suggested that the phenotype was a result of increased expression of RpoS. Increased expression of RpoS was known to confer resistance to extreme acid stress (Lin et al., 1993). But deletion of *rpoS* from the strains did not alter the extreme acid resistant phenotype. This indicated that the phenotype was completely rpoS independent and was acting through a different molecular mechanism (Johnson et al., 2014).

Whole genome sequencing of the clones evolved at pH 2.5 showed mutations in different genes. But amongst the pool of mutations, the gene that was repeatedly selected across all the evolved strains was *evgS*. The authors tested the contribution of *evgS* to phenotype by deleting evgS from the evolved strains. Deleting *evgS* completely abolished the extreme acid resistant nature of the evolved strains. This showed that phenotype was significantly *evgS* dependent. As described in the previous section, EvgS is an unorthodox sensor kinase which is involved in acid perception and regulation of the glutamate decarboxylase system.
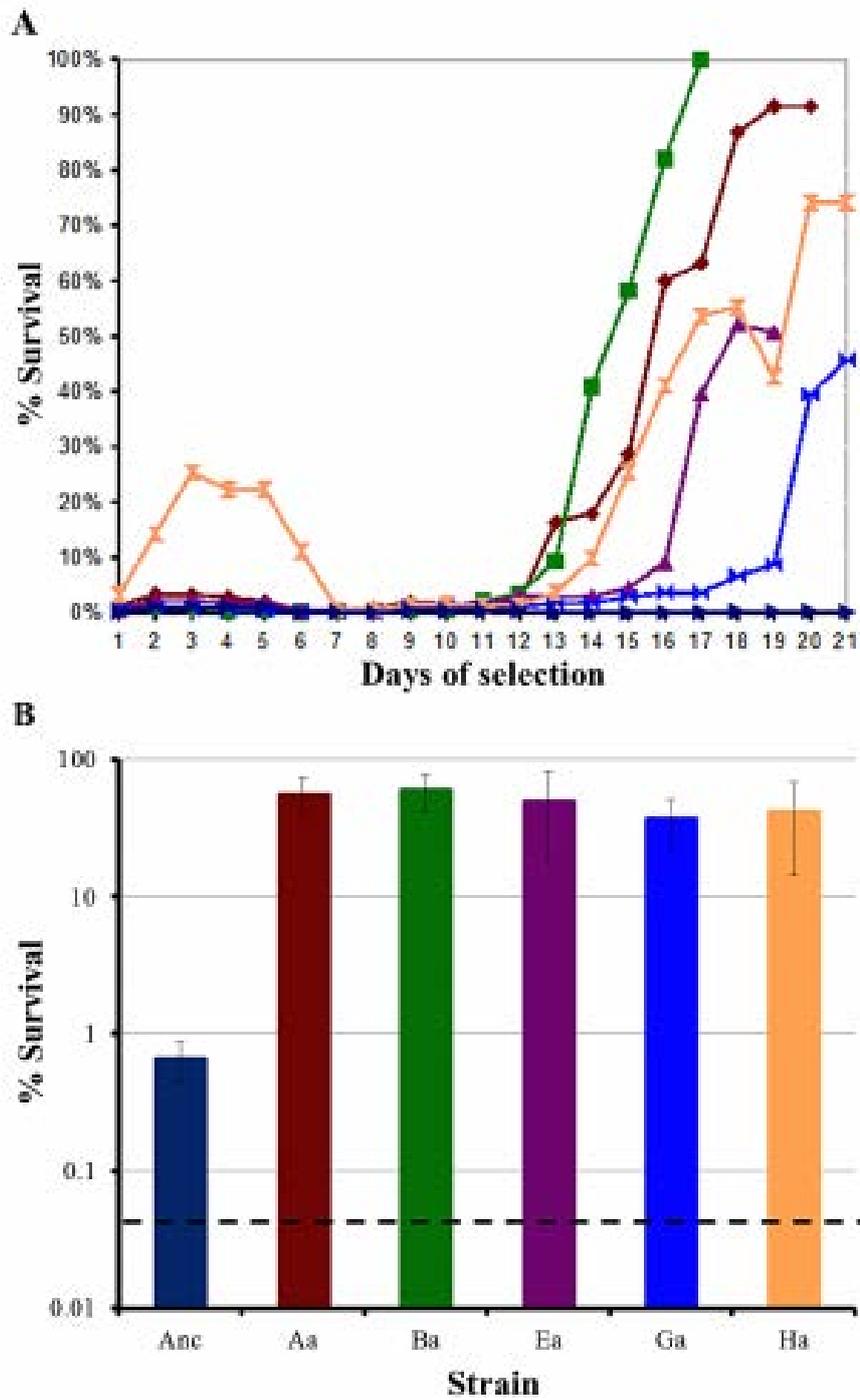
Figure 1.15: Evolution of extreme acid resistant phenotype in E. coli MG1655. The emergence of acid resistant phenotype of 5 evolved populations and increased survival at pH 2.5 (fig 1.15A). Clones isolated from the populations showed similar survival phenotype at pH 2.5 (fig 1.15B). This figure has been taken from Johnson et al., 2014.

The mutations selected in the evolution experiment were located in the PAS domain of the sensor kinase (Johnson et al., 2014). EvgS was a known homolog of the senor kinase BvgS in Bordetella pertussis. Structural studies on BvgS had shown mutations in the PAS domain altered the functional state of the kinase (Perraud *et al.,* 2009). Hence, the authors further investigated the role of this mutations in conferring the extreme acid resistant phenotype to the strains evolved at pH 2.5. Expression analysis of the evolved strains showed a massive upregulation of the genes located in the glutamate decarboxylase system (fig. 1.16). The Gad system is one of the key acid resistance systems of *E. coli* and regulated by evgS. It was not surprising that the Gad system was upregulated in these strains, as upregulation of this system would protect against cytoplasmic and periplasmic acid stress.

To analyse the role of the mutations in isolation, the authors moved the individual mutations to the ancestral background and studied their acid resistant phenotype. Each of the *evgS* mutations conferred similar survival phenotype at pH 2.5 to the ancestor as seen in the fully evolved strains (fig 1.17A). Promoter probe analysis showed that these mutations upregulated the expression of the genes located in the Gad system. This upregulation was seen in absence of any acid shock as well. The authors showed that thee mutations constitutively turned on the Gad system in all the evolved strains (Johnson et al., 2014).

The authors hypothesised that these mutations were involved in causing loss of function of EvgS. The authors suggested that EvgS was a strong dimer when not sensing acid and transitioned to a weak monomeric state when activated. The *evgS* mutations selected from the evolution experiment disrupted this strong dimer to weak monomer transition, thereby changing the conformation of the kinase to the activated form leading to the constitutive activation of the Gad genes. This hypothesis was backed by structural modelling which showed that the mutations located in the PAS domain were likely to disrupt a possible interaction in the dimer interface (Johnson et al., 2014).

The authors concluded that their study was able to show, how experimental evolution can be used to understand the activation of stress response system under specific conditions and the molecular mechanism responsible for it. The study showed that integrating various approaches such as experimental evolution, whole genome sequencing and post genomic tools would help us understand genotype-phenotype relationships.
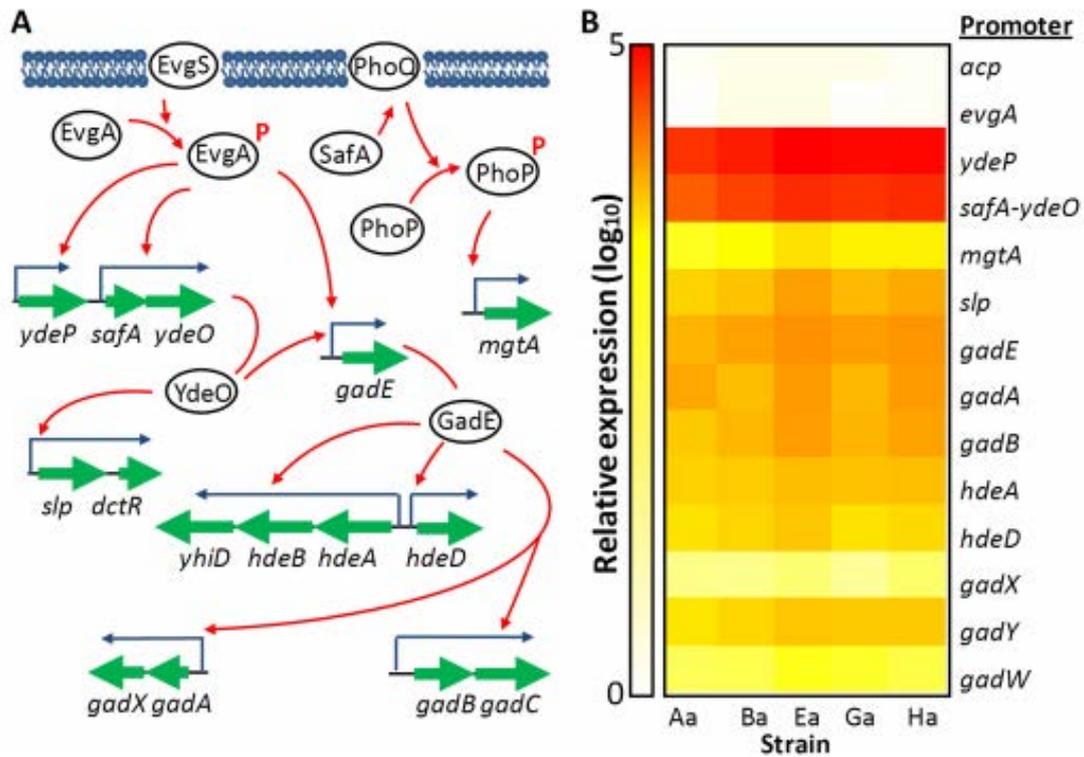
Figure 1.16: Promoter probe analysis of the glutamate decarboxylase system. EvgS is a key regulator of the glutamate decarboxylase (Gad) system. Figure A shows the different genes which constitute the Gad system. The red arrows indicate the process of activation. Figure B shows the expression of the promoters of the Gad genes in all the evolved strains using a luciferase promoter reporter probe. The gene acp is the internal reference control and is not part of the Gad regulon. This has been taken from Johnson et al., 2014.
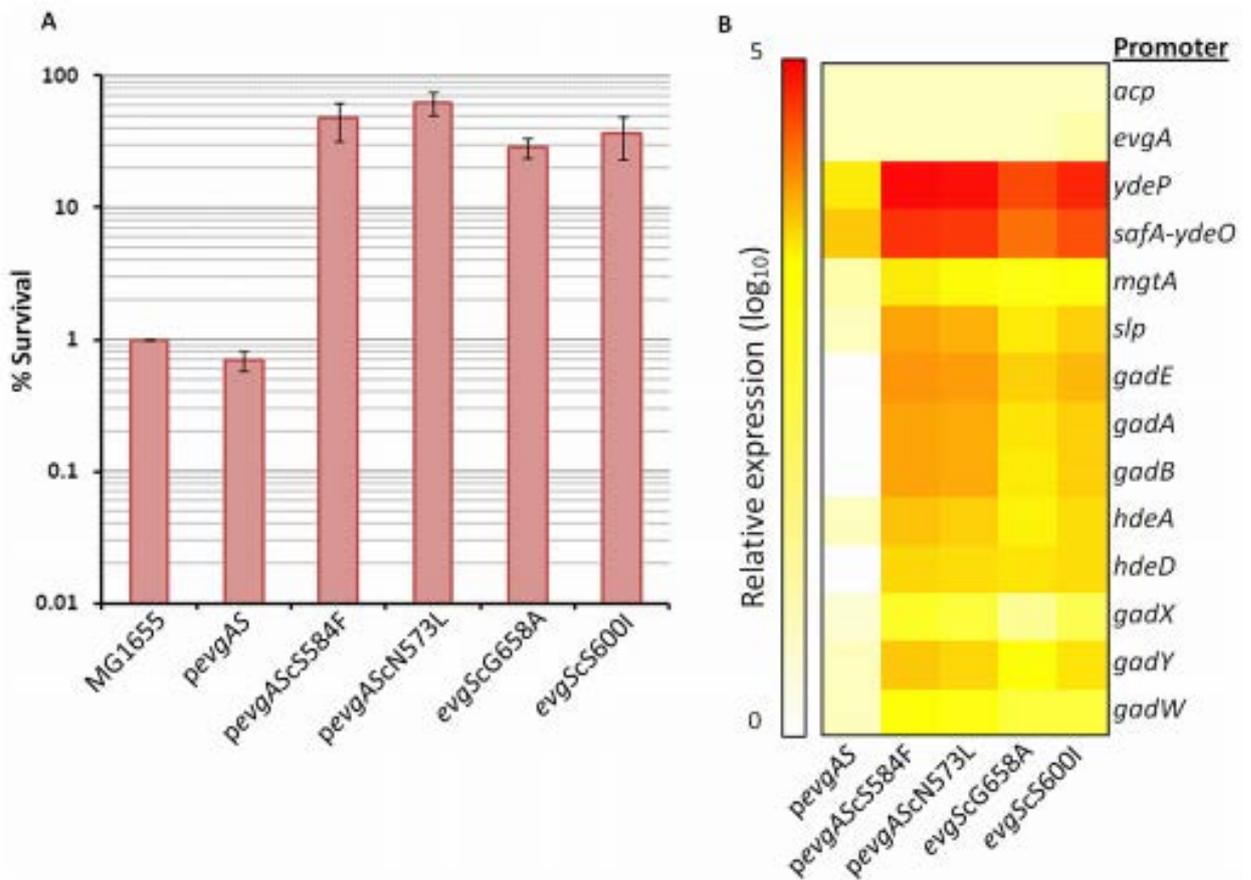
Figure 1.17: Extreme acid survival and promoter probe analysis of the evgS mutations. Figure A shows the comparison of survival at pH 2.5 between the ancestor and the evgS mutants being expressed on a low copy plasmid in the ancestor or inserted into the chromosomal copy of evgS. Figure B shows the promoter probe analysis of the wild type evgS and the evgS mutants. This figure has been taken from Johnson et al., 2014.

1.26 Inspiration for this study.

The experimental evolution study described in the previous section showed that, lab-based evolution can be used as a technique to understand how bacteria adapt to stress conditions. It also showed how stress response systems can be activated under specific conditions. This lab-based evolution experiment led to another project studying the mechanism of activation of EvgS, part of which is a component of my PhD and discussed r in chapter 8 of this thesis. This section has been presented as the paper published this year in July 2017 and is bound at the end of this thesis. This included investigating role of the cytoplasmic domain of *EvgS* in acid perception and biochemical evaluation of the dimer-monomer transition of the kinase. Since majority the research work in my PhD focused on the lab-based evolution of acid resistance, this thesis completely focuses on that part of my PhD project.

Taking inspiration from Johnson *et al.,* 2014, we wanted to conduct a lab-based evolution experiment with a different evolutionary approach. We wanted to evolve *E. coli* under mild acid stress conditions. We hypothesised that changing the pH value of the evolution experiment would help us understand a different aspect of *E. coli* acid resistance. Study their genotype and phenotype and use an integrated approach of whole genome sequencing, transcriptomics and molecular techniques to establish a link between the two.

1.27 Aims and objectives

Project 1: Lab-based evolution of acid resistance against mild acid stress in *E. coli* K-12 MG1655.

a) Evolve *E. coli* in long term propagation under mild acid stress
b) Genotypic and Phenotypic characterisation of the acid evolved strains
c) Understanding the transcriptional landscape and cross stress resistance of the evolved strains
d) Constructing a functional genotype and phenotype relationship of the acid evolved strains
e) Deducing a molecular mechanism by which E. coli climbs its fitness peak under mildly acidic conditions

Project 2: Structure-function analysis of the *E. coli* acid sensing kinase, EvgS: (Sen et al., 2017) This part of the PhD has been published and the paper is bound at the end of the thesis

a) Expression, purification and characterisation of the cytoplasmic domain of EvgS
b) Expression, purification and characterisation of the cytoplasmic domain of EvgS with an evgS activating mutation
c) Biochemical comparison of the wild type cytoplasmic domain and its mutant version

Chapter 2
Materials and Methods

2.1 Bacterial Strains and Plasmids

Table 2.1: Bacterial strains used in this study.

Strain	Genotype	Source/Reference
E. coli K-12 MG1655	Λ^-, F$^-$, rph-1 (Ancestral strain for the evolution experiment)	Blattner *et al.,* 1997
E$_1$A	Evolved strain of MG1655	This study
E$_2$A	Evolved strain of MG1655	This study
E$_3$A	Evolved strain of MG1655	This study
E$_4$A	Evolved strain of MG1655	This study
E$_5$A	Evolved strain of MG1655	This study
E$_6$A	Evolved strain of MG1655	This study
KHOO1	*E. coli* MG1655 with a deletion of the lacZ gene.	Kerry Host, 2013
BW25113$\Delta arcA$	*E. coli* K-12 BW25113 with *arcA* inactivated by a kanamycin cassette	KEIO library, Baba *et al.,* 2006
BW25113$\Delta rpoS$	*E. coli* K-12 BW25113 with *rpoS* inactivated by a kanamycin cassette	KEIO library, Baba *et al.,* 2006
BW25113$\Delta cytR$	*E. coli* K-12 BW25113 with *cytR* inactivated by a kanamycin cassette	KEIO library, Baba *et al.,* 2006
BW25113$\Delta flhD$	*E. coli* K-12 BW25113 with *flhD* inactivated by a kanamycin cassette	KEIO library, Baba *et al.,* 2006
BW25113$\Delta flhC$	*E. coli* K-12 BW25113 with *flhC* inactivated by a kanamycin cassette	KEIO library, Baba *et al.,* 2006
F$_1 arcA$	Intermediate evolved strain of MG1655 isolated from the fossil record of the strain E$_1$A. It has the M39I mutation in *arcA*.	This study
F$_1 arcA\Delta cytR$	F$_1 arcA$ with the chromosomal cop*y* of *cytR* inactivated by a kanamycin cassette	This study

Strain	Genotype	Source/Reference
F$_1$arcAΔflhC	F$_1$arcA with the chromosomal copy of flhC disrupted by a kanamycin cassette	This study
F$_1$arcAΔflhD	F$_1$arcA with the chromosomal copy of flhD inactivated by a kanamycin cassette	This study
F$_1$arcAΔcytRΔflhD	F$_1$arcA with a cytR deletion and the chromosomal copy of flhD inactivated by a kanamycin cassette	This study
F$_1$arcAΔcytRΔflhC	F$_1$arcA with a cytR deletion and the chromosomal copy of flhC inactivated by a kanamycin cassette	This study
E$_2$AΔrpoS	E$_2$A with the chromosomal copy of rpoS inactivated by a kanamycin cassette.	This study
E$_4$AΔrpoS	E$_4$A with the chromosomal copy of rpoS inactivated by a kanamycin cassette.	This study
Salmonella Typhimurium SL1344	*Salmonella enterica* subspecies serovar Typhimurium	Way et al., 1978
F$_1$arcA::rpoA N294H	F$_1$arcA with the N294H mutation in rpoA on the chromosome and marked with tetracycline resistance (Tn10)	This study
MG1655::rpoA N294H	*E. coli* M1G655 with the N294H mutation in rpoA on the chromosome and marked with tetracycline resistance (Tn10)	This study

Table 2.2: Plasmids used in this study.

Plasmid	Description	Source/Reference
pRW50	Broad range lac fusion cloning vector with multiple cloning site. Tet^R	Lodge *et al.,* 1992
pSynP21- *rpoS::sfGFP*	pXG10 derivative GFP reporter plasmid. It contains a highly selective *rpoS* responsive promoter-GFP fusion. Amp^R	Carrier and Keasling, 1999
pRW50 *estA1::lacZ*	93 base pair fragment of the ETEC stable toxin promoter *estA1* cloned into pRW50. Tet^R	Haycocks *et al.,* 2015
pRW50 *estA2::lacZ*	96 base pair fragment of the ETEC stable toxin promoter *estA2* cloned into pRW50. Tet^R	Haycocks *et al.,* 2015
pCP20	Temperature sensitive plasmid which has the FLP recombinase. Used in removal of kanamycin cassettes. Amp^R	Datsenko and Wanner, 2000

2.2 Culture and growth conditions.

Cultures were grown in LB (lysogeny broth medium) containing 1% w/v NaCl (Fisher Scientific), 1% w/v Bactotryptone (BD Biosciences) and 0.5% w/w Yeast extract (BD Biosciences). Unless specified bacterial cells were grown aerobically at 37ºC. For growth experiments involving a single carbon source, bacterial cells were grown in M9-minimal medium. M9-medium was composed of 12g/l NaH_2PO_4, 6g/l KH_2PO_4, 1g/l NH_4Cl and 0.5 g/l NaCl. M9-medium was supplemented with 0.4% Glucose, 0.2% Casamino acids, 625mM KCl, 250mmM $MgSO_4$ and 0.1M $CaCl_2$. Depending on the required pH value, the pH of the growth medium was adjusted using 2M HCl or 2M NaOH. For growth on solid media 15g/l of bacteriological grade agar (Sigma) was added to lysogeny broth to make LB agar. All overnight cultures were grown in a 20ml universal glass container with 5ml volume of culture in them. All growth experiments were conducted in 250ml Erlenmeyer flasks with 100ml volume of culture in them.

2.3 Evolution experiment.

The lab-based evolution experiment comprised six populations isolated from a single common ancestral strain of *E. coli* K12 MG1655. The evolution experiment was started on 17/03/2013 and completed on 08/09/2014. The cultures were propagated using a 1:100 dilution regime where 0.1ml of culture is inoculated in 4.9mls of unbuffered LB medium pH 4.5. The cultures were grown in batch mode in 25 ml glass universal bottles. The cultures were incubated at 37ºC with shaking at 180rpm for 24 hours. After 24 hours, these cultures were re-suspended in 2ml LB and 0.1ml of the resuspended culture was transferred into fresh medium. Every 15 days' cells were frozen down and stored at -80ºC to form the fossil record of the experiment. The evolved populations have been named E_1P, E_2P, E_3P, E_4P, E_5P and E_6P. Six purified clones were isolated from the six populations on 08/05/2014. These clones are referred to as the acid evolved strains in this study. They were named E_1A, E_2A, E_3A, E_4A, E_5A and E_6A

2.4 Competition experiment

Fitness measurements in this study were made using the standard competition experiment. The experimental principle of the competition assay was adapted from Lenski *et al.,* 1991. The competitors were distinguished based on an auxotrophic marker (lac^+ and lac^- phenotype on MacConkey lactose agar medium). A lac- derivative (KH001) of *E. coli* MG1655 was used as a proxy for the ancestor. To ensure that the marker conferred no competitive bias, KHOO1 and the ancestor were competed at pH 4.5 and pH 7.

To start the competition experiment, 5mls of the lac- and 5mls of the lac$^+$ strain was grown overnight. Unless specified, the overnight cultures were then diluted and introduced in equal proportions in 100mls of unbuffered LB medium pH 4.5. The starting OD_{600nm} for every competition experiment in this study was 0.05. Samples were taken after two competing strains were introduced into the competition setting and serially diluted in unbuffered LB medium. Then they were plated on to MacConkey agar (BD Biosciences) supplemented with 10% lactose (Sigma). Unless specified cultures were incubated for 24 hours at 37ºC and samples were taken again, serially diluted and plated. For most of the competition experiments in this study, the fitness measurements were made at 24 hours.

2.5 Measurement of fitness or competitive index (W).

The formula to measure fitness in this study was adapted from the formula described in Lenski *et al.,* 1991. The method is based on measuring the ratio of the relative populations sizes of the two competing strains. As stated in the previous section, samples were taken at the start and end of the competition experiment and plated on to MacConkey lactose plates. The colony counts of the lac$^+$ (red) strain and lac$^-$ (white) strain was recorded at timepoints (0 and 24) hours. The colony counts were used to compute their respective population sizes at these time points. The fitness measurement was made using the formula shown below.

$$W=[\ln(R_{T24}/R_{T0})/\ln(V_{T24}/V_{T0})]$$

Fitness or competitive index is denoted by W. R and V represent the population sizes of lac$^+$ (red) strain and lac$^-$ (white) strain at different time points. T_0 denotes the time of commencement of the competition experiment and T_{24} denotes the time point after 24 hours of growth of the mixed culture in the competition setting. In this study, in case of some experiments competitive indices were also calculated over a time course (Chapter 3). But following the standard practise observed across most lab-based evolution studies, fitness was measured at 24 hours of the competition experiment. This measurement method allowed calculation of competitive indices up to a value of 3. Fitness indices greater than 3 were not represented in this study as the colony count of the outcompeted strain in those cases caused massive statistical variation.

The formula used to calculate the competitive index is based on the growth equation given by

$$x(t) = x_0 . e^{kt}$$

The equation defines the size of a population growing exponentially at time t equals the size at time zero multiplied by e^{kt}, where k is the growth constant.

Mathematically, this equation can be used to derive the growth constant of an exponentially growing population.

$$x(t) = x_0 . e^{kt}$$

$$[x(t)/x_0] = e^{kt}$$

$$\ln[x(t)/x_0] = kt$$

$$k = \ln[x(t)/x_0]/t$$

This formula can be used to compare any two strains being competed together and the ratio of their growth constants is the same as the formula used to measure fitness index (W) (Lenski *et al.,* 1991). Thus, the relative fitness of two strains in a competition experiment can be measured by calculating the ratio of their relative growth constants.

Although, there is a small caveat in the formula used to measure fitness as it is based on a clear assumption that two competing strains are always exponentially growing. This would be true if the two strains in a competition experiment are growing exponentially all the time. Based on the assumption then competitive index would remain constant if measured an any point in growth curve. But bacterial cultures don't grow exponentially all through their growth cycle and the formula neglects the fact that stationary phase growth follows a completely different dynamic compared to exponential phase.

This was seen clearly when fitness measurements were made over a time course in this study. Our results showed that competitive index didn't remain constant over time. As results in the next chapter would show the competitive indices of strains fitter at pH 4.5 increased through the course of the growth curve. The fitness advantage was seen in exponential and stationary phase (section 3.7, chapter 3).

2.6 Molecular genetics techniques

2.6.1 Isolation of plasmid DNA

In all circumstances, plasmid DNA isolation was carried out using the Qiagen miniprep kit (Qiagen, Hilden, Germany). Plasmid DNA was isolated from bacterial cells grown overnight from a single colony inoculated in 5ml of growth medium. The principle of operation of the kit involved permeabilization of the bacterial cells through SDS treatment and digestion of cellular RNA by RNase treatment. Purification of plasmid DNA was achieved by selective extraction, where the plasmid DNA was bound to silica gel membrane. Bound plasmid DNA was eluted using elution buffer which contained TE (Tris-EDTA) buffer at pH 7.5. Plasmid DNA was quantified using the high sensitivity double stranded DNA Qubit molecular probes (Fischer chemicals). The 260/230 and 260/280 ratios were estimated using the Nanodrop (Thermoscientific)

2.6.2 Isolation of genomic DNA

In all circumstances, genomic DNA isolation was carried out using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Hilden, Germany). Protocol no 3.8 from the kit was followed to isolate DNA from bacterial cells. Unless specified genomic DNA was isolated from bacterial cells grown overnight from a single colony inoculated in 5ml of growth medium. The kit used proteinase K and lysozyme treatment to permeabilize the cells. The lysed cells were treated with the chaotropic salt guanidine thiocyanate to selectively destabilise nucleases and proteins. Cellular RNA was digested by a RNase treatment. Purification of genomic DNA was achieved by binding isolated DNA to the QIAmp genomic DNA purification column and elution in TE buffer (Integrated DNA technologies). Genomic DNA was quantified using the high sensitivity double stranded DNA Qubit molecular probes (Fischer chemicals). The 260/230 and 260/280 ratios were estimated using the Nanodrop (Thermoscientific)

2.6.3 Polymerase Chain Reaction for sequencing purposes.

Template DNA (plasmid and genomic) was isolated as per the protocol stated above. For purposes of sequencing PCR fragments and plasmid DNA, Phusion high fidelity PCR master mix was used in all cases. All PCR reactions were conducted in 50µl reaction volume, 2µl of genomic DNA or 5µl of plasmid DNA was added to 25µl of the master mix, primers were adjusted to a final working concentration of 0.5µM. For each reaction, the final volume was

adjusted to 50 µl using nuclease free water (Gibco). The PCR reaction was run in the Eppendorf thermocycler per the program shown in the table.

Step 1	Initial denaturation	95°C	1 minute	1 cycle
Step 2	Denaturation	95°C	30sec	
Step 3	Annealing	Primer based	30 sec	20 cycles
Step 4	Extension	72°C	30sec/kb	
Step 5	Final extension	72°C	5minutes	1 cycle.

2.6.4 Mismatch amplification mutation assay (MAMA)

The primer design for the MAMA assay was done as per guidelines of Birdsell *et al.,* 2012. The MAMA primers were designed to probe the fossil record of each evolved strain to determine the order of emergence of some of the key mutations. The order of emergence was tabulated based on their date of first detection. A brief description of the primer design is shown here below.

The evolved strain E3A has a mutation in the gene *prfB* which causes a tyrosine to alanine change at the 246th amino acid (T246A). The MAMA forward primer is designed to incorporate the mismatch at 3' end, with a random change in the penultimate base of the primer. The MAMA forward primer terminates at the mismatch. An example of a MAMA primer is shown here below.

ATCAACCCGGCGGATCTGCGCATTGACGTTTATCGCACG- Wild type sequence

ATCAACCCGGCGGATCTGCGCATTGACGTTTATCGCGCG- Mutant sequence (T246A)

MAMA primer: ATCAACCCGGCGGATCTGCGCATTGACGTTTATCGTG

For genes, which had more than one SNP detected across the evolved strains, a common reverse primer was used. For example, all the evolved strains have mutations in *arcA*, hence the forward primer was designed specific to each mutation while a common reverse primer was used in all the PCR assays (section 4.2, chapter 4).

Unless specified, all MAMA assays were conducted using the BiolineTM Taq polymerase master mix. Template DNA was isolated by melting 10µl of overnight culture of the mutant at 95ºC for 15 minutes. 1µl of the melted lysate was used as the template DNA for the MAMA assay. Each MAMA reaction mixture was 50µl in volume containing 25µl of the PCR master mix, forward and reverse primer to a final working concentration of 20µM and 1µl of template DNA. For each reaction, the final volume was adjusted to 50 µl using nuclease free water (Gibco). The annealing temperatures was calculated using the NEB (New England Biolabs) Tm (**https://tmcalculator.neb.com**) calculator and validating the predicted temperature through a gradient PCR

Initial validation of the MAMA assay was conducted testing the sensitivity of the assay. This was done by mixing different ratios of the mutant and wild type DNA in the PCR reaction mixture and analysing the appearance of the PCR product for the mutant. This work was done by Dorothy Ling, a PhD student in the lab. Each MAMA primer was tested for its annealing specificity. This was done by conducting a gradient on the primers using wild type DNA and mutant DNA as the template and analysing the strength of the signal from both the reactions (section 4.2, chapter 4). The primers used in the MAMA assay have been listed in table 2.3. The MAMA PCR assays were conducted as per the program shown below

Step 1	**Initial denaturation**	**98ºC**	**2 minute**	**1 cycle**
Step 2	Denaturation	98ºC	1 minute	
Step 3	Annealing	Primer based	1 minute	35 cycles
Step 4	Extension	72ºC	30sec/kb	
Step 5	Final extension	72ºC	6 minutes	1 cycle.

2.6.5 Agarose Gel Electrophoresis

PCR products, plasmid DNA, genomic DNA and DNA fragments were analysed by agarose gel electrophoresis and separated using 1%(w/v) gel made from agarose powder (Fisher-scientific) dissolved in TAE buffer (50X TAE buffer contains 0.05M EDTA, 1M acetic acid and 150mM Tris). DNA samples were mixed with 6X blue loading dye (Bioline) in a 1:6 ratio. The exception to this was for PCR products obtained by thermocycling using BiolineTM Taq polymerase master mix which contains a loading dye. The DNA ladder used in the agarose gels was the Hyper ladder 1kb from Bioline. Agarose gels were run in Biorad 25 gel tanks with sufficient amount of 1X TAE as running buffer. The gels were run at 100 volts for (45-60) minutes. Gels were viewed using the Geneflow gel doc (Syngene) and images were captured using the Quantity one software. In cases where gel extraction had to be performed gels were viewed using the non-UV blue light illuminator (Nippon) to prevent any DNA damage.

2.6.6 Site-directed mutagenesis

All site directed mutagenesis on vector DNA was carried out using the QuickchangeTM Lightning mutagenesis kit (Santa Clara, United States) per manufacturer instructions. Mutagenic primers were designed according to guidelines of the Agilent primer design tool. Reaction setup consisted of 10μl of reaction buffer, (50-75)ng of double stranded vector DNA template, 1.5μl of quickchange solution, 125ng of forward and reverse primer, 1μl of dNTP mix and 1μl of quickchange lighting polymerase. The final volume of the reaction was adjusted to 50μl using nuclease free water (Gibco). The PCR reaction was run in the thermocycler as per the program shown below.

Step 1	Initial denaturation	95ºC	30 sec	1 cycle
Step 2	Denaturation	95ºC	30 sec	
Step 3	Annealing	55ºC	1 minute	18 cycles
Step 4	Extension	72ºC	30sec/kb	
Step 5	Final extension	72ºC	1min/kb	1 cycle.

The amplified PCR product was digested using *DpnI* enzyme and incubated at 37ºC for 1 hour. This was done to eliminate any methylated DNA from the parental vector. The digested PCR product was then transformed into XL_{10}-gold supercompetent cells supplied with the mutagenesis kit. Vector DNA was prepped from successful transformants and validated by sequencing.

2.6.7 Isolation of RNA and ribodepletion

All RNA isolations in this study were conducted using the RNeasy miniprep kit (Qiagen, Hilden, Germany). All RNA extractions in this study were carried out following this protocol. Three independent replicates of the bacterial strains were grown overnight in LB broth. The overnight cultures were diluted to a starting OD_{600nm} of 0.05 in LB broth at pH 4.5 or pH 7. The cultures were grown until they attained OD_{600n} of 0.8 ±0.04. For RNA mini preps the number of cells from which RNA was to be harvested was adjusted to a maximum (2×10^{11}- 4×10^{11}) CFU/ml (per manufacturer guidelines). Briefly, the protocol involved cell lysis by a 20mg/ml of proteinase K (Qiagen) and 100μg/ml of lysozyme treatment. During lysis the mixture was incubated at 75 degrees for 30 minutes. The lysate was then added to the silica membrane for selective binding of RNA to the column.

To remove cellular DNA from the lysate, an on-column DNA digestion was performed using the Qiagen 250 DNase Kit (Qiagen, Hilden, Germany). RNA was eluted using DNase and RNase free water. RNA samples were stored at -80ºC.

For estimation of RNA yield, RNA was quantified using the high sensitivity RNA Qubit molecular probes (Fischer chemicals). The 260/230 and 260/280 ratios were estimated using the Nanodrop (Thermoscientific). The samples were analysed using the Agilent 2100 bioanalyzer.

The ribosomal RNA content of the RNA samples was depleted using the Turboprop Ribozero rRNA removal kit, as per manufacturer guidelines. Post rRNA depletion, the RNA samples were again quantified using Qubit and analysed using the bioanalyzer (Agilent) to confirm the depletion.

Table 2.3: MAMA primers used in the study (for further explanation refer to sec.3.11)

Primer Name	Mutation	Strain	Sequence (5'-3')
arcA_1	M39I	E_1A, E_6A	GGCTATGATGTTTTCGAAGCGACAGATGGCG CG GAAACT
arcA_2	N106K	E_2A	GGCTATGATGTTTTCGAAGCGACAGATGGCG C
arcA_3	G78V	E_3A	GGTAAGAACGGTCTTCTGTTAGCATA
arcA_4	R67S	E_4A	GCGCTACGATGAAGGCGATATGCTGCATTCG C
arcA_5	D98A	E_5A	GAAGTCGATAAAATTCTCGGCCTCGAAATCG GT GCATCT
cytR_1	D34/T*	E_2A	CGACCGTCTCCCGAGCATTAATGAATCCCGT CT
cytR_2	P291L	E_3A	CTTTAGCGAAATTATTCGCGGTATCGAAGTTA CATA
cytR_3	F82L	E_4A	CTGGTGATTGTCCCGGATATCTGCGATCCCTG AC
cytR_4	A93E	E_5A	GATAACATCGACCTGACGCAATTTTGTGACA T
cytR_5	L179/*	E_6A	TGCTTGATCTTCCACAACCGCCTACTGCTGTC TTCTGCCATA
arcB_1	E343*	E_4A	GCGCTACGATGA AGGCGATATG CTGCATTGCA
cadC_1	L381*	E_4A	5'GTTCAATCATCCCCAGAATTTACCTACGCG AGAGCAGAAAAAGCGCG
sspA_1	D80N	E_4A	GAGCGTCACATCCACCACCAGGTGCGGCGTG AGCTGGTTATCCAGC
fimE_1	R107/C*	E_6A	GGCGCTGACCGGACTGACGCTATATTTATTTC TCACGCGGGAGT
tnaC_1	F132L	E_1A	TGCATTATGAATATCTTACATATATGTGTGAC CTCAAAATGGTGA
prfB_1	T246A	E_3A	ATCAACCCGGCGGATCTGCGCATTGACGTTT ATCGTA
rpoA_1	N294H	E_1A, E_2A, E_3A, E_5A and E_6A	CCGAGGTTGAGCTCCTTAAAACGCCTCACCTT GGTAAAAAATCTC

Table 2.4: Primers used for validating knockouts and site-directed mutagenesis

Primer Name	Sequence (5'-3')	Description
arcA forward	CCGAAAATGAAAGCCAGTAAAGAAGTTACAACGGAC	Forward flanking primer to validate *arcA* deletion
arcA reverse	GTTGTTGACGTTGATGGAAAGTGCATCAAGAACGC	Reverse flanking primer to validate *arcA* deletion
cytR forward	CCACGACCTACAACTAGGCTAACTCCCAATTTG	Forward flanking primer to validate *cytR* deletion
cytR reverse	CGAAGGTGCCGGTTGGCTGCGGCGTACATAATC	Reverse flanking primer to validate *cytR* deletion
flhD forward	GACGGATGGCTGGCCGCTGTATGAATCCCG	Forward flanking primer to validate *flhD* deletion
flhD reverse	TAAGCTGCAGGCAAAGCTGCCAACAGGCTGG	Reverse flanking primer to validate *flhD* deletion
flhC forward	GCGACATCACGGGGTGCGGTGAAACCGC	Forward flanking primer to validate *flhC* deletion
flhC reverse	CGCTTTCGTGCGTCTCAATCTCTTCATCCATCAG	Reverse flanking primer to validate *flhC* deletion
rpoS forward	GCCGCGTTGTTTATGCTGGTAACGCGCTGCGCGG	Forward flanking primer to validate *rpoS* deletion
rpoS reverse	CCGATTCAGTGCGTATTGGTGACGCTGGCAGCCTG	Reverse flanking primer to validate *rpoS* deletion
Kan-pKD4 forward	GATTGAACAAGATGGATTGCACGCA	Forward flanking primer to validate the insertion of the kanamycin cassette
Kan-pKD4 reverse	TCAGAAGAACTCGTCAAGAAGGCGATAGAA	Reverse flanking primer to validate the insertion of the kanamycin cassette
rpoAN294H forward	GGTTGAGCTCCTTAAAACGCCTCACCTTGGTAAAAAATCTCTTAC	Forward primer for site directed mutagenizing the pLAW2 *rpoA* plasmid
rpoAN294H reverse	GTAAGAGATTTTTTACCAAGGTGAGGCGTTTTAAGGAGCTCAACC	Reverse primer for site directed mutagenizing the pLAW2 *rpoA* plasmid

2.6.6 Gel extraction

To eliminate unutilised PCR reagents, small oligonucleotides (<40bp) and protein, PCR products were purified using the QIAquick gel extraction kit (Qiagen, Hilden, Germany). Briefly, the gel encased DNA is dissolved in an appropriate buffer containing a pH stabilizer. This was done by incubating the gel slice and buffer at 50ºC. The pH stabilizer is present to ensure that the pH of the mixture is appropriate to facilitate DNA binding to the silica membrane. The solubilised gel mixture is then added to the QIAprep column. DNA bound to the silica membrane is eluted using elution buffer or TE (Tris-EDTA) buffer at pH 7.5

2.7 Bacterial transformation: Making chemically competent cells.

The strain of interest was grown overnight from a single colony inoculated in 5ml lysogeny broth (LB). The overnight culture was then diluted 100 fold to a starting OD_{600nm} of 0.05 in LB broth. The cultures were grown to an OD_{600nm} of (0.3-0.4). The culture was then incubated on ice for 45 minutes. Post incubation on ice 40ml of culture was centrifuged at 8000g for 10 minutes in 50ml falcon tube (Greiner). The centrifuged pellet was re-suspended in 25nl of ice cold 0.1M $CaCl_2$ and incubated on ice for 30 minutes. The cells were again centrifuged and re-suspended in 1ml of ice-cold 0.1M $CaCL_2$. The re-suspended cells were aliquoted in 100µl volume into 1.5ml eppendorfs and stored at -80ºC.

For transformation (50-75)ng of plasmid or vector DNA was added to 100µl of competent cells in a 1.5ml Eppendorf. The mixture was incubated on ice for 30 minutes and heat shocked at 42ºC for 45 seconds. The heat shocked mixture was then placed on ice for 5 minutes after which 800µl of L-SOC medium (Sigma) pre-warmed at 37ºC was added to it. This mixture was incubated aerobically at 37ºC under shaking conditions for 1 hour. For temperature sensitive vectors the incubation was performed at 30ºC. The recovered cells were centrifuged at 5000g and re-suspended in 1 ml LB broth. The cultures were serially diluted and plated onto LB agar plates with the appropriate antibiotics.

2.8 Bacterial transformation: Making electrocompetent cells

The strain of interest was grown overnight from a single colony inoculated in 5ml lysogeny broth (LB). The overnight culture was then diluted 100-fold to a starting OD_{600nm} of 0.05 in LB medium. The cultures were grown to an OD_{600nm} of (0.3-0.4). The culture was then stored on ice for 45 minutes. Post incubation on ice 40ml of culture was centrifuged at 8000g for 10 minutes in 50ml falcon tube (Greiner). The pellet was washed four times by centrifugation and

resuspension using decreasing volumes (45mls, 25mls, 10mls and 5mls) of ice cold 10% glycerol. The cells were finally resuspended in 1ml of 10% glycerol and aliquoted in 100µl volume into 1.5ml eppendorfs to be stored at -80ºC.

For transformation, 100µl of electrocompetent cells was aliquoted in 1.5ml electroporation cuvette (Cell projects) and 100ng of plasmid or vector DNA was added to the cells. The cells were electro-pulsed in an Eppendorf miniporator at 2100mV. 1ml of L-SOC medium was immediately added to the pulsed mixture and recovered aerobically at 37ºC (30ºC for temperature sensitive vectors) under shaking conditions for 2 hours. The recovered cells were centrifuged at 5000g and re-suspended in 1X phosphate buffered saline. The cultures were serially diluted and plated onto LB agar plates with the appropriate antibiotics.

2.9 Beta-galactosidase assay.

All strains used for the assay were grown in minimal M9 medium supplemented with 0.2% casamino acids and 0.4% glucose. For the assay, three independent replicates of each strain were grown overnight in LB broth. The overnight cultures were diluted to a OD_{600nm} of 0.05 in M9-cas medium and grown aerobically at 37ºC under shaking conditions. In experiments where low pH shock was necessary the pH of the media was adjusted to the required pH value using 2M HCl. Cells were grown to a OD_{600nm} of 0.2-0.4 and harvested using 10% sodium dodecyl sulphate (SDS) and chloroform. The cell lysate was incubated with Z-buffer (Miller, 1972) and ONPG (*ortho*-Nitrophenyl-β-galactoside, Sigma) was used as the chromogenic substrate for the reaction. The reaction mix was incubated for 30minutes and the reaction was stopped using 2M Na_2CO_3. The optical density of the reaction end-product was read at 420nm using a spectrophotometer. The beta-galactosidase activity was calculated using the formula **(1000 * OD$_{420}$) / (OD$_{600}$ * culture volume used * time)**

2.10 Measuring RpoS promoter activity

All strains used for measurements of RpoS activity were transformed with the pSynP21-*rpoS::sfGFP* plasmid. For every experiment the *E. coli* K12 MG1655 was used as a positive control and MG1655Δ*rpoS* was used as the negative control. Three independent replicates of each strain were grown overnight in LB medium. The overnight cultures were diluted to OD_{600nm} of 0.05 in 100 ml LB broth at pH 4.5 and pH 7 in 250ml conical flasks. The cultures were grown aerobically at 37ºC for 480 minutes. Samples were collected every 60 minutes and their optical densities and fluorescence intensities were recorded. For reading the fluorescence, 200µl of sample was taken from the culture in the conical flask and added to a 96-well

polypropylene plate (Greiner Bio). The 96-well plate was then inserted into the Fluoroscan where a program was set according to the specific plate set-up to measure the fluorescence at 515nm. The program used to measure the fluorescence was

Step 1: Incubate for 2 minutes at 37⁰C

Step 2: Shake at 150 rpm for 30 seconds

Step 3: Measure fluorescence at 515nm with excitation set at 485nm.

In parallel to the fluorescence intensity measurements, optical density measurement was made by retrieving 1ml of culture from the conical and measuring the optical density at 600nm using the spectrophotometer. When multiple strains were to be assayed, sample retrieval was staggered to account for any time delays. The absolute RpoS activity or fluorescence was calculated by dividing the OD_{600nm} for each time point by the corresponding fluorescence measurement at OD_{515nm}.

2.11 P*1*-transduction or generalised transduction.

P*1*-transduction was used to engineer chromosomal inactivation or deletions in strains constructed in this study. Unless specified, this method was used to engineer deletions by P1 mediated transduction of a kanamycin cassette from a KEIO library strain to a strain of interest. The protocol was adapted from the original protocol described in Thomartom *et al.,* 2008. In all cases in this study P1 transduction was used to transduce a kanamycin marker from a donor strain to engineer chromosomal deletions in a recipient strain.

The donor strain was grown overnight in 5ml LB medium. The overnight culture of the donor was diluted to a starting OD600nm of 0.05 in 10ml LB medium containing 0.4% glucose and 0.1mM $CaCl_2$. The donor strain was grown to an OD600nm of 0.5 and infected with a freshly prepared P1 lysate of *E. coli* K12 MG1655. The fresh P1 lysate was made as described by Atterson *et al.,* 1994.

The P1 lysate of the donor strain was prepared by adding 100μl of P1- lysate of *E. coli* K12 MG1655 to 500μl of the donor strain. The mixture was incubated at 37⁰C in a water bath for 45 minutes. On completion of incubation 200μl of chloroform was added to eliminate any un-lysed bacteria. The lysate was then centrifuged at 10000*g* for 15 minutes at 4⁰C. The supernatant was then filtered through a 0.22μm coring filter and stored at 4⁰C. This P1 lysate was used for transduction of the requisite marker to the appropriate recipient strain.

For the transduction of the marker a single colony of the recipient strain was inoculated in 1ml LB medium and grown overnight aerobically at 37ºC. The overnight culture was then added to 4ml of LB broth and grown to a OD6$_{00nm}$ of 0.5. The appropriate P1 lysate was serially diluted ($10^{0,}$ 10^{-1}, 10^{-2} and 10^{-3}) in LB broth. 100µl of the serially diluted lysate was added to 500µl of the recipient strain. The recipient-lysate mixture was incubated at in a water bath at 37ºC for 45 minutes. Post incubation the mixture was centrifuged at 6000g for 10 minutes. The supernatant was discarded and the cells were washed with 1X PBS supplemented with 50mM sodium citrate to remove any remnant lysate particles. This step was repeated three times to wash away residual phage particles. Then the infected cells were re-suspended in 1ml of L-SOC medium (pre-warmed at 37ºC) ad recovered at 37ºC under shaking conditions (180rpm) for 2 hours.

The recovered cultures were again centrifuged at 6000g for 10 minutes, the supernatant was discarded and the cell pellet was re-suspended in 1ml of sterile 1X PBS(Gibco). The resuspended culture was serially diluted and plated onto LB agar containing the appropriate antibiotic and 5mmM sodium citrate. Successful transductants were usually selected from the plate onto which cells were plated from the culture infected with the least amount of phage.

2.12 Removing antibiotic resistance marker using pCP20

Antibiotic cassettes were removed from marked strain following the protocol as described in (Datsenko and Wanner, 2000). In this study, only kanamycin cassettes were removed using this method. Temperature sensitive plasmid pCP20 was transformed into the required strains using electroporation. The transformants were selected on LB agar + ampicillin plates after overnight incubation at 37ºC. Ten transformants were randomly chosen from the successful transformants and grown overnight in 5ml LB broth at 43ºC to activate the expression of the flipase recombinases from the plasmid. The overnight cultures were then plated onto LB agar and the colonies were then replicated plated onto LB + ampicillin, LB + kanamycin and LB agar only. Loss of the kanamycin marker and pCP20 plasmid was confirmed by sensitivity to ampicillin and kanamycin and successful growth on LB agar. The successful candidates were screened suing colony PCR. A representative example is shown in figure 2.1
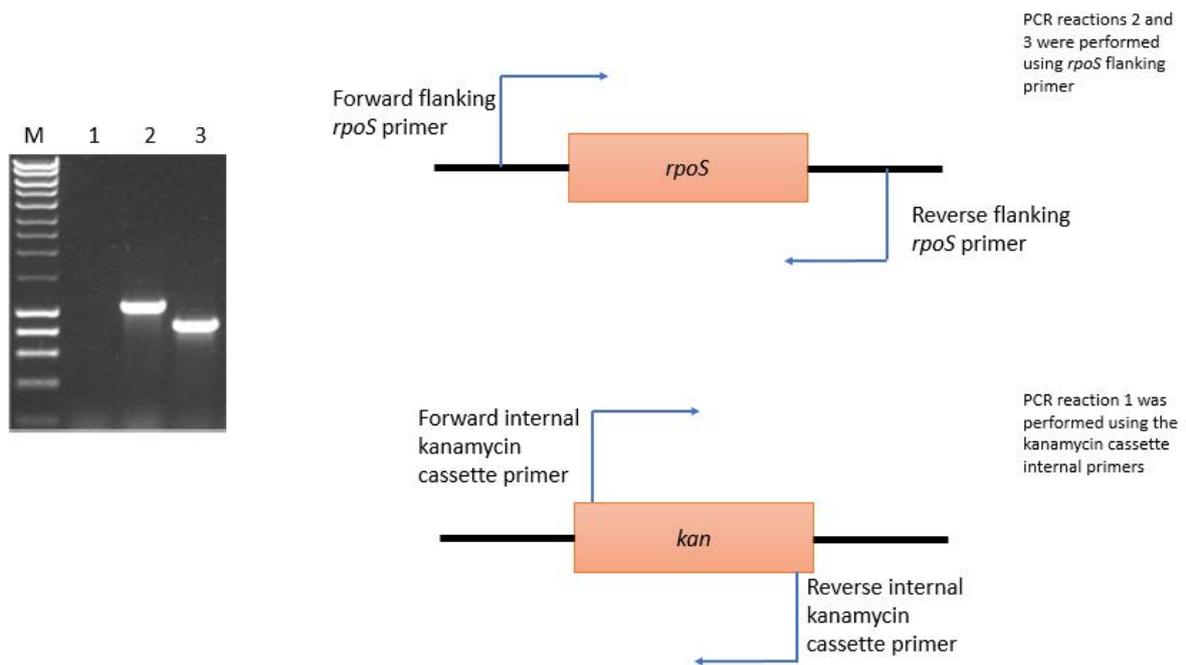
Figure 2.1: Colony PCR analysis of the rpoS deletion in E. coli K12 MG1655. Using P1 mediated transduction of a kanamycin marker the rpoS gene was deleted in MG1655. PCR reactions 2 and 3 were performed to confirm the deletion. PCR reactions 2 and 3 were performed using rpoS flanking primer (table 2.4). PCR reaction 2 showed an expected 1.5kb band when DNA from the knockout was used as the template and PCR reaction 3 showed an expected 1kb band when DNA from the ancestor was used as the template. PCR reaction 1 was performed to confirm the removal of the kanamycin cassette from the rpoS deletion strain of MG1655. The PCR reaction showed no product when DNA from the cured strain was used the template. The PCR was performed using primers specific to the kanamycin cassette (table 2.4).

2.13 Whole genome sequencing.

Genomic DNA for whole genomes sequencing were prepped using the protocol described in section 2.6.2. Genomic DNA was submitted to the sequencing facility at the University of Birmingham, Microbes NG. The service from Microbes NG included the delivery of the trimmed bam of the sequences genomes, variant analysis, zero coverage and quality analysis of the sequenced genomes. The whole genome sequencing of all strain sequenced in the study was done using the Illumina Mi-seq platform.

2.13.1 Data analysis of whole genome sequencing

The trimmed bam files containing the sequenced reads of the sequenced genomes were provided to us by MicrobesNG. The assembly of the bam files and quality analysis was done by Joshua Quick and Dr. Emily Richardson from MicrobesNG. The assembled bam files were aligned against the reference strain *E. coli* K12 MG1655. The genome identifier of the reference strain was NC00913.3. This version of the reference strain was used for all genome alignments and analysis in this study. The alignments of the bam files and the reference strain was done using the Illumina recommended Integrated Genome Viewer (**https://software.broadinstitute.org**) (IGV) (sec. 3.13, chapter 3). A variant file for all the sequenced strains was generated by Dr. Emily Richardson, the variant file was loaded on to IGV and the SNP's were analysed using the SNP calling tool in IGV. A zero-coverage file was also provided to us by Dr. Emily Richardson and this file was used to detect large deletions in the sequenced genomes (sec, 3.14, chapter 3). One of the issues encountered while using IGV was the detection of small deletions and insertions. To ensure we had retrieved all possible mutations from the sequenced genomes, they were gain analysed using the Breseq pipeline (http://barricklab.org/twiki/pub). The Breseq analysis was done by Mathew Milner in 2017. Since the reanalysis of the genomes showed some new results work is progress to understand some of the new mutations found from the Breseq analysis. The results of the analysis will be incorporated in this study in the future.

2.14 RNA sequencing

RNA samples were prepped here in our laboratory and the quality control of the prepped samples was done under the supervision of Dr Thippesh S. Prepped and processed samples were shipped to the Centre for Genomic Research (CGR) at the University of Liverpool. RNA sequencing was done at CGR and all the analysis of the RNA sequencing data was done by Dr

John Herbert. A representative example of the pipeline involved in sequencing and analysis prepared by Dr. John Herbert is shown here below.

RNAseq pipeline
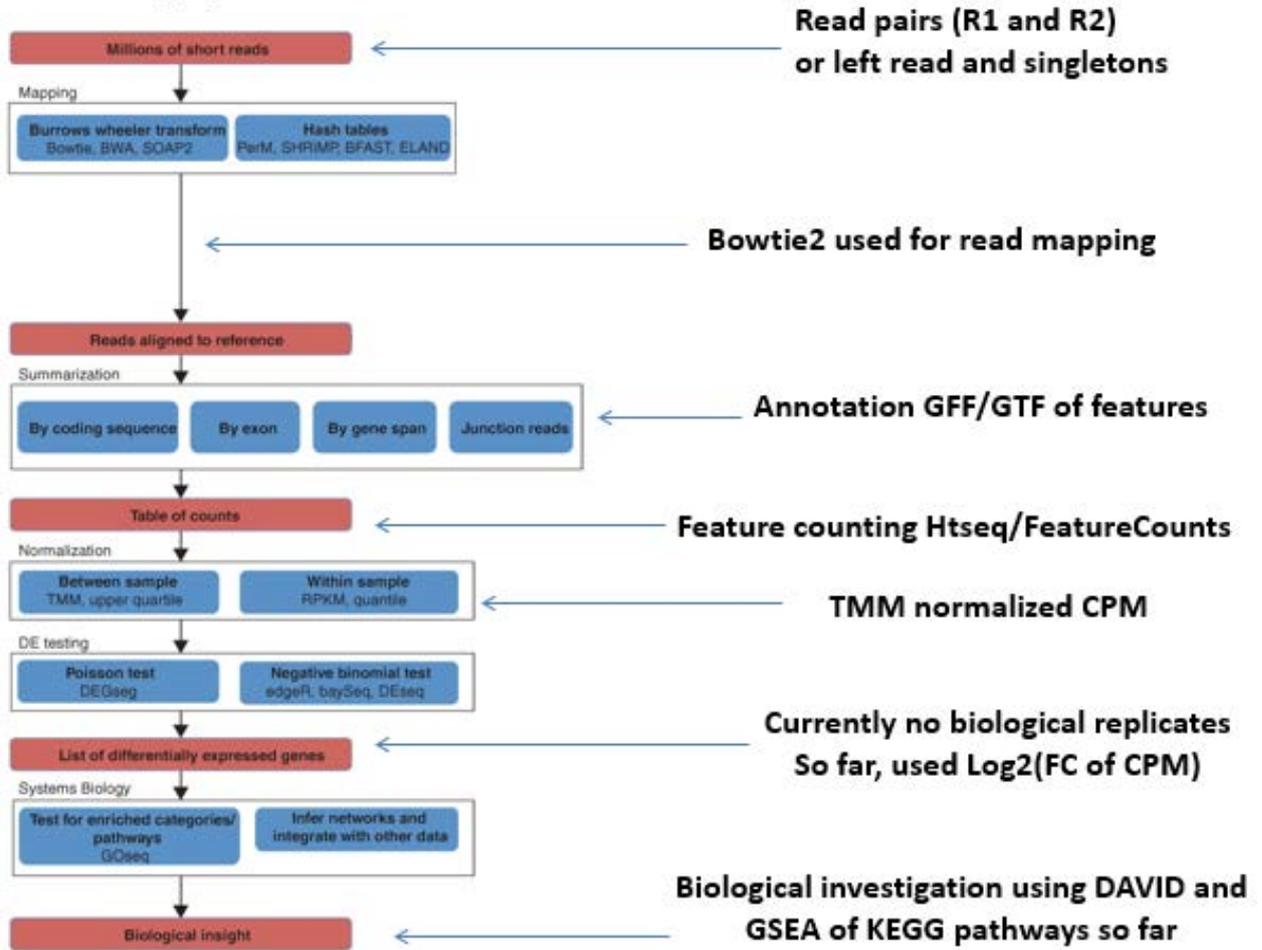
Figure 2.2: RNA sequencing analysis pipeline. The flow chart was prepared by Dr. John Herbert and depicts the process followed in analysis of the RNA sequencing data.

2.15 Quantitative reverse transcriptase polymerase chain reaction. (Q-RTPCR)

Q-RTPCR was performed to analyse the relative expression of two genes *flgM* and *fimC*. pepA was used as the internal reference control. The reasons for choice of the three genes have been explained in section 6.2 chapter 6. Q-RTPCR was performed using the Taqman probe assay. The guidelines for the experiment were taken from the Primetime qPCR application guide (*https://www.idtdna.com/pages/docs/default-source/catalog-product-documentation/qpcr-guide-2015-102115_4thedition_v5a.pdf?sfvrsn=25*). The experiment was performed by me and Maria Massoura, PhD student in the lab under the supervision of Dr Thippesh S. The dyes for the three genes have been listed below.

Gene	Reference dye	Excitation(nm)	Emission(nm)
flgM	Hex (phosphoramitide)	550	564
fimC	ROX (Rhodamine-X)	588	608
pepA	6-FAM (6' Fluorescein)	495	520

RNA isolation was performed following the protocol in section 2.6.7, except for ribodepletion. The isolated RNA was quantified using Qubit and cDNA synthesis was done using the tetra-cDNA synthesis kit (Bioline) per manufacturer's guidelines. The synthesised cDNA was validated for any DNA contamination using primers specific to the gene *evgS*.

The specificity of the primers and probes were analysed using the BlastN tool (NCBI) and the primer quest design tool recommended by Integrated DNA technologies. The primers and probes used in the experiment have been shown in table 2.4. The QPCR reactions were set up as per manufacturer guidelines. Quantification of the expression of gene was done using the AriaMIX quantitative real time PCR system (Agilent, Santa Clara, United States). Normalisation and efficiency of gene expression was calculated using the guidelines recommended in section 6.4.2b of the Primetime qPCR application guide (fourth edition, 2015)

Table 2.5 Q-RTPCR primers and probes

Primer name	Sequence 5-3	Length	Amplicon length
pepA_F	GGTCTGATGGCGAACCATAAT	21	
pepA_R	GGACTCCAGTTGTTCCTGATAC	22	117
pepA_probe/Hex	TGAACTGATTGCCGCGTCTGAACA	24	
flgM_F	AGCACCAGTGTGACGTTAAG	20	
flgM_R	CCGTTACGAATCGCCAGTT	19	107
flgM_probe/6-FAM	ATATCACTGCTGCCGGGTTGCATC	24	
fimC_F	CTGATTAACCCGACACCCTATT	22	
fimC_R	AGGCAATTTAACCGTGCTTTC	21	111
fimC_probe/ROX	ACGGTAACAGAGTTGAATGCCGGA	24	

The QPCR reaction settings used for quantification of gene expression using AriaMix have been detailed below in the table

Step 1	**Denaturation**	**95⁰C**	**15 mins**	**1 cycle**
Step 2	Annealing	65⁰C	5 mins	25 cycles
Step 4	Extension	60⁰C	2 mins	1 cycle

2.16 Preparation of L-cells for the differentiation and culture of murine cells to murine bone marrow macrophages and culture of J774 macrophages.

L-cell media is used as the primary media for the differentiating murine cells isolated from the mouse bone marrow into bone marrow derived macrophages. The bone marrow derived macrophages were used for the macrophage invasion assay of the acid evolved strains described in section 3.12 of chapter 3. L-cells produce granulocyte macrophage stimulating factor which drives the differentiation of murine bone marrow cells to either dendritic cells or bone marrow macrophages. So, the experiments conducted using the bone marrow derived macrophages were mixed in nature. For the preparation of L-cell media, 1ml of L-cells were thawed by incubation in a water bath at 37ºC.

Thawed cells were re-suspended in 1ml of RPMI (Rosewell Park memorial institute) medium supplemented with 10% (v/v) foetal bovine serum (FBS) and 1% (w/v) penicllin-streptomycin. Supplemented RPMI is known as cRPMI. The 1ml mixture was then mixed with 9mls of cRPMI and centrifuged at 1000g for 5mins at 4ºC. The supernatant was discarded and the cells were gently re-suspended in 10ml of cRPMI. The 10ml mixture was then gently transferred to a 25ml tissue culture flask (Coring) and an additional 15ml of cRPMI was added to the flask. The flask was incubated overnight at 37ºC. The following day the flask was examined under a focusing microscope to examine the L-cells. The spent medium was discarded and dead cells were removed using an aspirator. 25ml of fresh cRPMI was added to the L-cells in the flaks and the flaks was incubated at 37ºC overnight. This process was repeated for four days until confluence of cell growth was achieved.

Once confluent the cells were washed with 10ml of cRPMI and detached from the flaks by a 1X trypsin (Gibco) digestion. The detached cells were mixed with 10ml cRPMI and seeded into a 175ml tissue culture flask (Coring). The incubation and reseeding process mentioned above was repeated until the scale up culture attained confluence. Once confluent the cells were washed and re-suspended in 10ml cRPMI. The confluent cells were detached using a trypsin digestion and filtered through a 0.22μm filter. A small volume of the cells was counted using a haemocytometer. In total 2×10^6 were harvested from the scale up culture. The filtered medium was stored at -80ºC. The filtered medium contained the L-cell granulocyte macrophage stimulating factor which was used for seeding the bone marrow derived macrophages.

2.17 Macrophage Invasion and survival assay.

The harvested L-cell medium was used to differentiate three independent replicate lines of bone marrow derived macrophages and standard lab cell line J774. This work was done by Dr. Amanda Rossiter, and the confluent macrophages were given to me which I used for the invasion assay. For the invasion assay 15MOl of confluent (bone marrow and J774) cells were centrifuged at 6000g for 10mn at 4ºC. The supernatant was discarded and the macrophages were re-suspended in 5ml of cRPMI medium pre-warmed at 37ºC. The cells were counted using a haemocytometer and the cell were diluted to 5×10^5/ml using cRPMI. 200µl of each macrophage line was seeded into a 96-well round bottom plate and incubated overnight at 37ºC. The amount of macrophage cells to be seeded was based on the instructions in Sukumaran *et al.,* 2001.

Three independent replicates of each bacterial strain were grown overnight at 37ºC. Bacterial cells from the overnight cultures were adjusted to 5×10^6 cells/ml using plain RPMI medium. The number of cells used in the experiment was based on the optical density estimation that an OD$_{600nm}$ of 1 equals 1×10^9 cells/ml. Prior to the invasion, cRPMI medium was removed from the 96-wel plate and washed with 200µl of PBS twice. For each invasion, 200µl of bacterial cells were added to the macrophage cell seeded into the 96-well plate. To initiate the invasion the 96-well plate was centrifuged at 1000g for 5 min. The plates were incubated for 4 hours at 37ºC to allow for the invasion and to evaluate survival upon invasion.

On completion of incubation the spent medium in each well was removed and 200µl of RPMI medium with 100µg/ml of gentamycin was added to each infected well. This was done to eliminate ant free moving bacteria. The wells were washed with 1X PBS three times and to lyse the macrophages 50µl of 1% Triton X-100 was added to each infected well. The plate was incubated for 30 minutes at 37ºC. An additional 150µl of 1X PBS was added to each well. Cells from each well were serially diluted and plated on LB agar plates and incubated overnight at 37ºC. Plates were counted and the colony forming units were calculated to determine the outcome of the invasion and survival assay (sec. 3.12, chapter 3)

2.18 Cross-stress protection analysis of the evolved strains.

The cross-protection analysis of the evolved strains was done to evaluate their ability to resist other forms of stresses other than pH 4.5. To define test conditions for each form of stress, the wild type ancestor was first grown in a range of conditions for each cross stressor. For each cross stressor, the corresponding condition which showed an overall reduction in growth was chosen as the test condition for the evolved strains. The selected conditions were

oxidative stress (100mM Menadione),

membrane Stress (8% SDS),

osmotic Stress (400mM NaCl),

pH dependent biofilm formation,

weak organic acid stress (pH 5.5, 5mM acetic acid),

acid (pH 3.5),

alkali (pH 9.5).

Except for the biofilm assay, all other cross stress experiments were conducted through growth curve analysis. For the cross-stress protection growth curves three colonies of each strain were grown overnight as starter cultures. The starter cultures were diluted to a starting OD 600 of 0.05 in 250ml flaks containing 100mls of the stressor and the optical density of these cultures were monitored for 540 minutes (sec. 3.10, chapter 3)

2.19 Biofilm Assay

Biofilm assay was done as per the protocol described in Nakao et al., 2005 with certain modifications. The biofilm assay was done in LB broth and supplemented minimal A medium. The assay was done at pH 4.5 and pH 7 to study the influence of pH on biofilm forming capabilities of the bacterial strains. For, the biofilm assay 1×10^6 cfu/ml cells from overnight cultures inoculated in 200µl of growth medium in round bottom 96-well plates (Coring, 3596). The bacterial strains were grown overnight in static conditions aerobically at 37ºC. Post incubation spent medium and planktonic cells were removed and the wells were stained with 0,1% crystal violet for 45 minutes. The wells were washed three times with sterile distilled water to remove any excess stain. The plates were air-dried for 30 minutes and any excess liquid was absorbed onto tissue paper. 95% ethanol was added to each cultured well and the optical density of the wells were read at OD_{595nm}.

2.20 Anaerobic growth experiment.

All strains were grown in lysogeny broth (LB) at pH 4.5 and pH 7. Three replicates of each strain were grown overnight aerobically at 37°C. The overnight cultures were then diluted to a starting OD_{600nm} of 0.05 in 100ml volume of growth media in a 250ml Erlenmeyer glass conical flask. Post inoculation the flaks were transferred to an anaerobic gas cabinet (Model no A808316, Whiteley scientific limited). The cultures anaerobic cabinet is sparged with a gas mix of 4% CO_2 and 95% N_2. The cultures were grown for 480 minutes and their OD600nm was monitored every sixty minutes.

Evolution of acid resistance to pH 4.5 in *E. coli* K12 MG1655

Chapter 3

Phenotypic and Genotypic Characterisation

3.1 Evolution of acid resistance in *Escherichia coli*

Stationary phase *E. coli* K12 is very resistant to extreme acid shock (pH 2.5). This resistance is seen in supplemented minimal medium and undefined rich growth medium. Exponentially growing *E. coli* is very sensitive to extreme acid shock. Lin *et al.,* 1995, showed that 99% of exponentially growing *E. coli* experience cell death within 20 minutes of exposure to pH 2.5. Pathogenic *E. coli* on the other hand show increased resistance to extreme acid shock during exponential phase compared to standard laboratory strains. Enteropathogenic *E. coli* O5:H17 showed increased survival rates than *E. coli* K12 MG1655 and *E. coli* K12 W3110 after 2 hours of exposure to pH 2.5 (Gordon and Small, 1993).

This acid sensitivity during exponential phase growth can be overcome by pre-exposing the cells to a mild acid shock at pH 5.5. A study in 1996 showed that *E. coli* K12 cells pre-exposed to a pH 5.5, were significantly more tolerant to extreme acid shock pH 2.5 than cells which were directly exposed to pH 2.5 (Small et al., 1996). Pre-exposing exponentially growing cells to a mild acid shock leads to the activation of the acid stress response systems in *E. coli*. The most important of all the systems is the Glutamate decarboxylase system (GAD) or the AR2. The activation of the GAD system leads to the hierarchical expression of the transcriptional activator GadE. GadE in turn activates the expression of the decarboxylases GadA and GadB and the glutamate-GABA antiporter GadC. GadE also activates the expression of two key periplasmic chaperones, HdeA and HdeB.

As cells go into stationary phase, the level of the general stress sigma factor RpoS, which leads to increased general stress resistance of the cell. Increased RpoS levels in the cell, causes overexpression of GadE which in turn increases the expression of the GadA and GadB. This increased expression of the Gad genes is seen even in absence of acid shock and accounts for the highly acid resistant nature of *E. coli* in stationary phase (Cornet *et al.,* 1999).

This sensitivity to acid shock during exponential phase was the basis of the previous evolution experiment Johnson *et al.,* 2014, conducted in the laboratory. In this study, seven independent cultures of *E. coli* MG1655 were exposed to extreme acid shock of pH 2.5 for 20 minutes followed by recovery at pH 7 overnight. This cyclical propagation regime was repeated for three weeks and the acid resistance of the populations was tracked over this period. Approximately at the end of three weeks these populations became very resistant to pH 2.5. Clones isolated from these populations too showed the same phenotype. Initial hypothesis suggested that this phenotype was a result of increased RpoS levels in these strains. But

preliminary experiments showed that the phenotype was completely *rpoS* independent. Whole genome sequencing of these strains showed that all of them had mutations in the sensor kinase EvgS, which regulates the AR2 system. The authors showed that these mutations caused constitutive activation of the AR2 genes which conferred this extreme acid resistant phenotype. The constitutive activation of the AR2 genes was seen even in the absence of a drop in external pH. This lab-based evolution study led to another project studying the mechanism of activation of EvgS, part of which is a component of my PhD and the published paper on this part of my PhD is attached at the end of the thesis.

Taking inspiration from Johnson *et al.,* 2014, we wanted to conduct a lab-based evolution experiment with a different evolutionary approach. We wanted to evolved *E. coli* under mild acid stress conditions. We wanted to evolve or select for improved phenotypes which would grow better at mildly acidic pH than being resistant to extreme acid stress. We hypothesised that changing the pH value of the evolution experiment would help us understand a different aspect of *E. coli* acid resistance. The objective of the experiment was to evolve *E. coli* at a mildly acidic pH. Study their genotype and phenotype and try to create a link between the two. In this chapter I will discuss how the evolution experiment was conducted, and the initial phenotypic and genotypic characterisation of these acid evolved strains.

3.2 Choosing the right model for evolution

In the introduction, I have discussed in the detail the important arguments that makes *E. coli* K12 MG1655 the organism of choice to conduct a laboratory based evolution experiment. MG1655 (F$^-$,γ$^-$, rph-1) is also known as the Blattner strain. The analytical tools to scrutinize its genotype and technologies available to perform genetic manipulations in MG1655 make working with this organism easy in the laboratory. It is a non-pathogenic strain with safety concerns that can be easily managed in the laboratory, it is sensitive to regularly used antibiotics in the laboratory. Extensive research data on acid stress response regulation of MG1655 has been collected over the last three decades, creating vast potential for comparison of data generated from this evolution experiment. As this strain was the ancestor for the previous evolution experiment conducted in our laboratory, we wanted to do the same with this experiment. All these reasons were the driving factors that led us to choose *E. coli* K12 MG1655 for this evolution experiment.

3.3 Choosing a pH value for the evolution experiment.

E. coli can grow comfortably over a range of different pH values. The aim of this project was to evolve *E. coli* to grow rapidly under mild acid stress. To do so we were required to choose a pH value for the evolution experiment. This pH value would need slow down the overall growth of *E. coli* compared to neutral or optimal pH. We did not want a pH value which would subject *E. coli* to killing conditions but a mild acid shock. This would allow selection to operate and cells to adapt and grow better at the chosen pH.

To choose the pH value for the evolution experiment, we grew *E. coli* in unbuffered LB medium at different pH values for 480 minutes. The growth conditions are described in the Materials and methods chapter. The results are shown in Figure 3.1. I have plotted the optical densities in the linear scale to ensure the differences in the growth rate at the different pH values is visible. Decreasing the pH caused reduction in the growth rate leading to no growth at pH 4 and pH 3.5. This is in accordance to previously published data. pH 4.5 was chosen for the evolution experiment.

3.4 Lab-based evolution experiment to study adaptation to mild acid stress pH 4.5.

The lab-based evolution experiment comprised six populations isolated from a single common ancestral strain of *E. coli* K12 MG1655. The evolution experiment was started on 17/03/2013 and completed on 08/09/2014. The cultures were propagated using a 1:100 dilution regime where 0.1ml of culture is inoculated in 4.9mls of unbuffered LB medium pH 4.5. The cultures were grown in batch mode in 25 ml glass universal bottles. The cultures were incubated at 37 degrees with shaking at 180rpm for 24 hours. After 24 hours, these cultures were resuspended in 2ml LB and 0.1ml of the resuspended culture was transferred into fresh medium. Every 15 days' cells were frozen down and stored at -80ᵒC to create a fossil record of the experiment. The evolved populations have been named E_1P, E_2P, E_3P, E_4P, E_5P and E_6P. Six purified clones were isolated from the six populations on 08.05.2014. These clones are referred to as the acid evolved strains in this study. They were named E_1A, E_2A, E_3A, E_4A, E_5A and E_6A

3.5 Calculating the number of generations.

As described in the previous section, the evolved populations were propagated using a 1:100 dilution regime. Lenski et al., 1991 showed that batch cultures of *E. coli* being propagated using this dilution regime in their study completed 6.6±0.05 generations in a 24-hour incubation period. The number of generations was calculated following the experimental procedure

described in McCartney *et al.*, 1997. Briefly the initial and final population sizes of each evolved population was estimated by calculating the number of colony forming units at the start and end of a 24-hour incubation cycle. The ratio of the natural logarithm of the two counts was used to determine the number of generations. The table below shows the mean populations sizes of each evolved population at pH 4.5.

Strain	Mean population size at T_0 (CFU/ml)	Mean population size at T_{24} (CFU/ml)	Number of generations=$\ln(T_{24})/\ln T_0$
E_1P	1.57×10^3	2.52×10^{11}	3.51
E_2P	2.83×10^3	4.96×10^{11}	3.39
E_3P	2.66×10^3	3.27×10^{11}	3.35
E_4P	1.88×10^3	3.53×10^{11}	3.52
E_5P	2.19×10^3	4.43×10^{11}	3.46
E_6P	3.21×10^3	5.5×10^{11}	3.34

The number of generations per incubation cycle, was taken as the average of the six measurements seen in the table above. The average number of generations the evolved populations completed per incubation cycle was 3.43±0.079. The total number of daily dilutions completed in this evolution experiment was 215. Hence, the total number of generations these populations were propagated for is approximately 740.
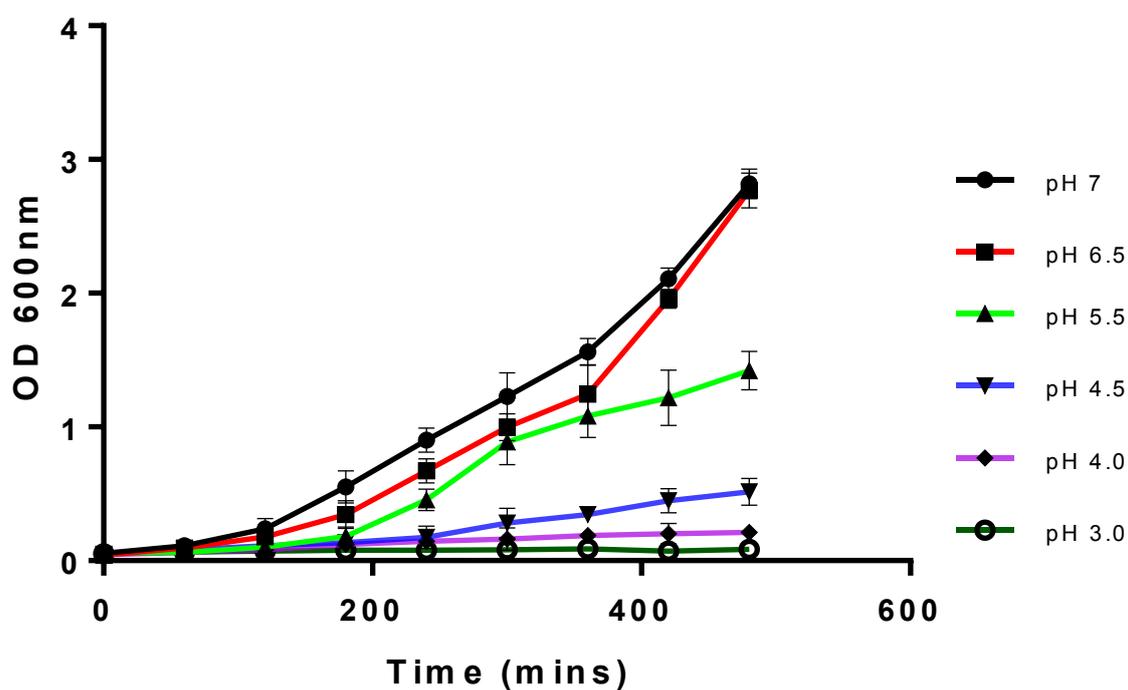
Figure 3.1: Growth curves of E. coli MG1655 at different pH values. Experiments were done as described in Materials and Methods (Chapter 2). Values are the mean of three biological replicates and the error bars are standard error of mean.

3.6 Comparing fitness of ancestor MG1655 and KH001.

To compare the fitness of the evolved strains and ancestor, a method for measuring the relative fitness of the two strains was required. A common method used to measure relative fitness is by competition experiments. In a competition experiment two competing strains are introduced in equal proportions at the start of the experiment and the relative numbers of the two strains are recorded through the course of the competition experiment. Several lab-based evolution studies use the conventional auxotrophic marker method, where one of the strains has a marker making the two strains easy to distinguish. To be able use this marked strain, the marker should not influence the outcome of the competition experiment. It should be competitively neutral.

For the competition experiments in this study, we use a lac- derivative of MG1655, KHOO1 as a proxy for the ancestor. Lac+ and Lac- colonies can be easily distinguished on MacConkey lactose medium through red-white differentiation. To test the competitive neutrality of KHOO1, we competed KHOO1 against the ancestor. We counted their relative numbers and calculated the competitive index of KHOO1 relative to the ancestor at different time points of competition experiment (materials and methods). As can be seen from figure 3.2, the competitive index of KHOO1 remains constant throughout the experiment. Competitive index calculated at 24 hours (Fig. 3.3) showed no difference in the relative fitness of KHOO1 at pH 4.5 and pH 7. Hence, KHOO1 can be used as a proxy for the ancestor for fitness measurements.
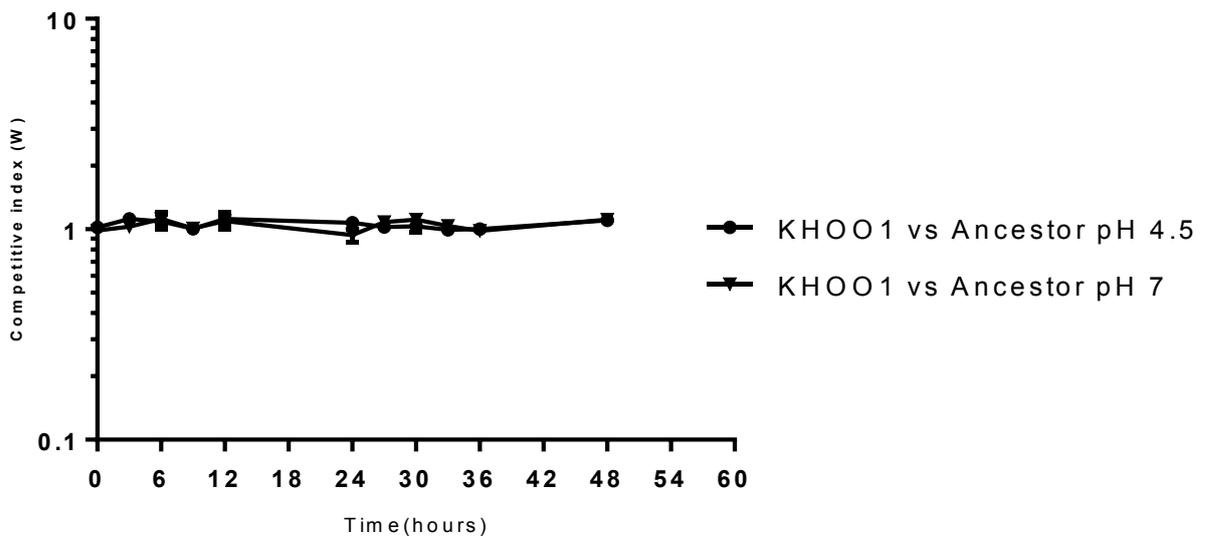
Figure 3.2: Competitive index of KHOO1 versus the ancestor against time. The values represent the average of three independent biological replicates and the error bars the standard error of the mean.
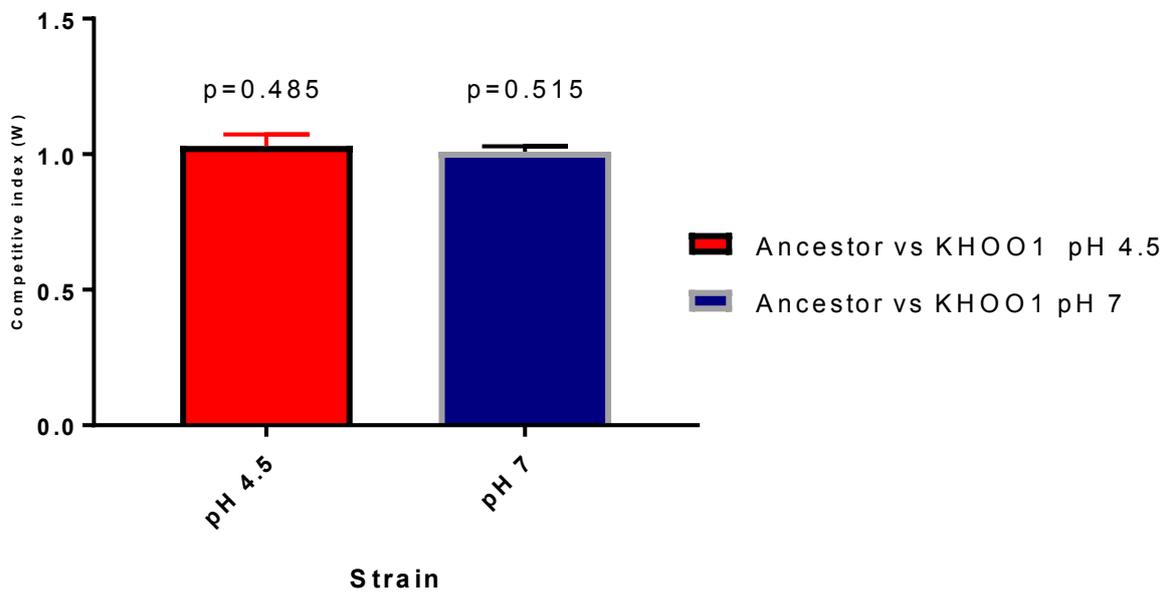
*Figure 3.3: Competition index of KHOO1 relative to the ancestor (MG1655) at pH 4.5 and pH 7. Values are an average of three biological triplicates and the error bars are standard error of mean. Significance was measured **using an unpaired student's t**-test, comparing the*

3.7 Fitness measurements of the evolved populations at pH 4.5

The evolving populations had been propagated for 4 months starting November 2013. At this point we wanted measure the fitness of these evolved populations relative to their ancestor. To do so we competed the six evolved populations E_1P, E_2P, E_3P, E_4P, E_5P and E_6P against KH001 and upon completion counted their relative numbers to generate competition indices (chapter 2). The competition index is a measure of the relative performance of a strain when competing with another for the same pool of resources. The competition index is calculated using the equation given below. It is denoted by W

$$W=[\ln(R_{TX}/R_{T0})/\ln(V_{TX}/V_{T0})]$$

Where R and W represent the two competing strains, TX denotes the time of calculation of the index and T0 represents time zero of the competition experiment.

Reviewing results of competition experiments of other-lab based evolution studies on *E. coli*, brought forth an important point. Several of these studies assume that fitness remains constant throughout the growth cycle. Hence, they make fitness measurements only at 24 hours. To test whether this assumption was true we calculated the fitness the competitive indices of the evolved populations at different time points of the experiment (Fig. 3.4)

The competitive index of the evolved populations changed with time. It showed a progressive increase over the course of the experiment (Fig. 3.4). The competitive advantage was seen both in exponential and stationary phase (Fig. 3.5, data shown for E_1P).

We also calculated the competitive index of the populations at 24 hours (Fig. 3.6). The evolved populations are fitter than the ancestor at pH 4.5
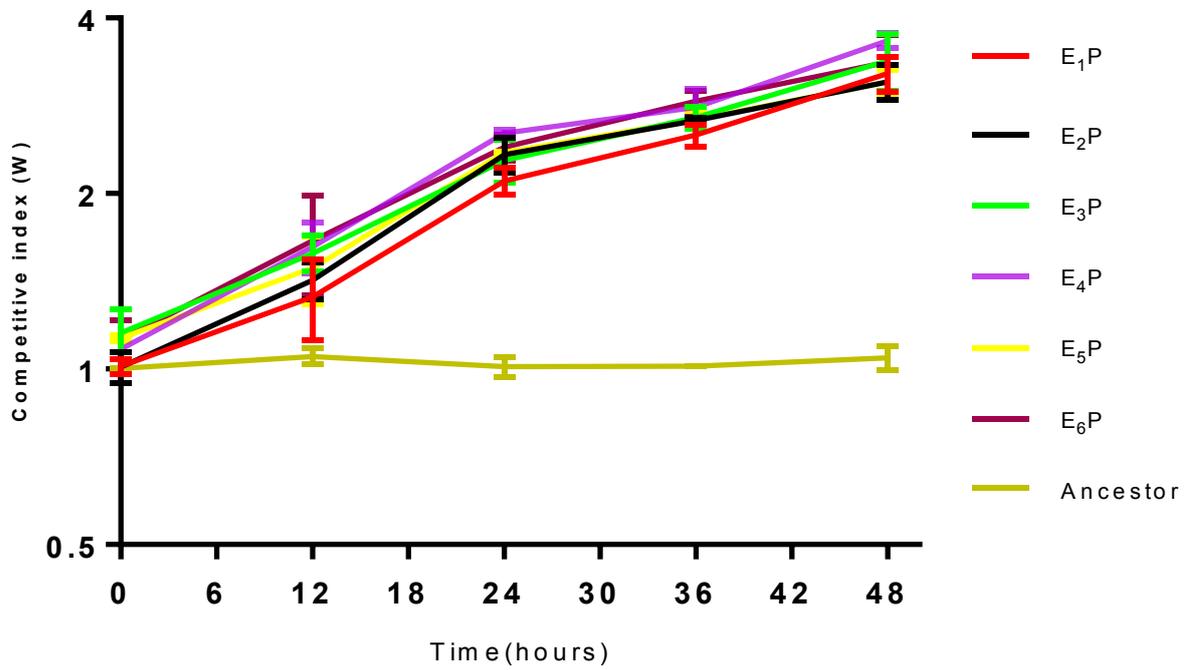
Figure 3.4: Competitive indices of the evolved populations versus time. The values are an average of three biological triplicates and the error bars are standard error the mean.
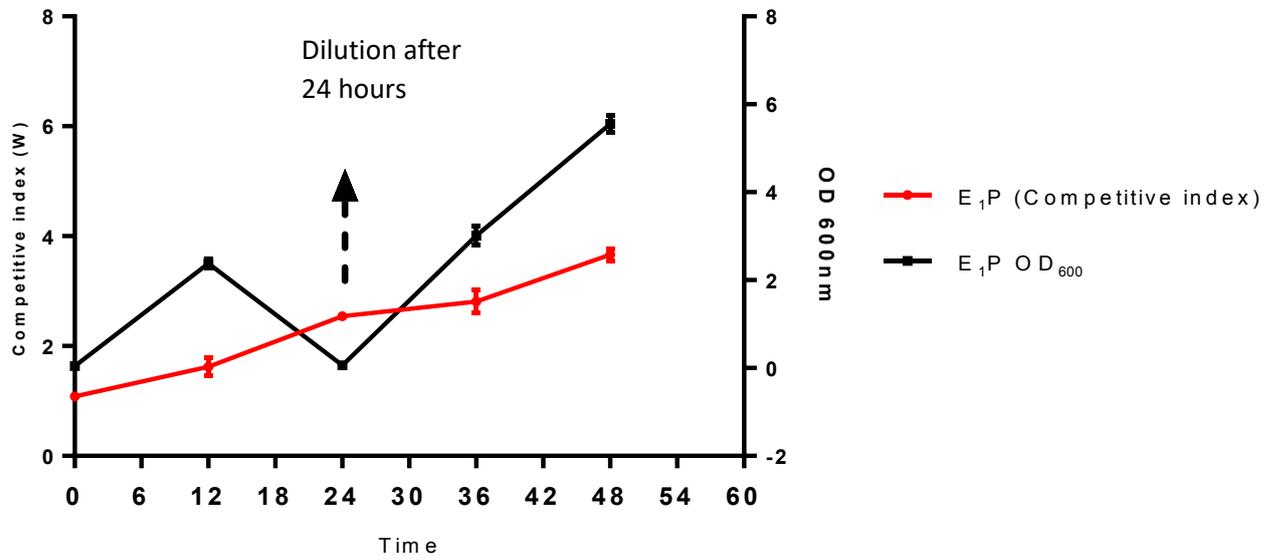
Figure 3.5: Three variable graph of the competitive index and optical density versus time. The left Y-axis shows the competitive index and the right Y-axis shows the OD 600. The values are average of three biological replicates and error bars are the standard error of the mean.
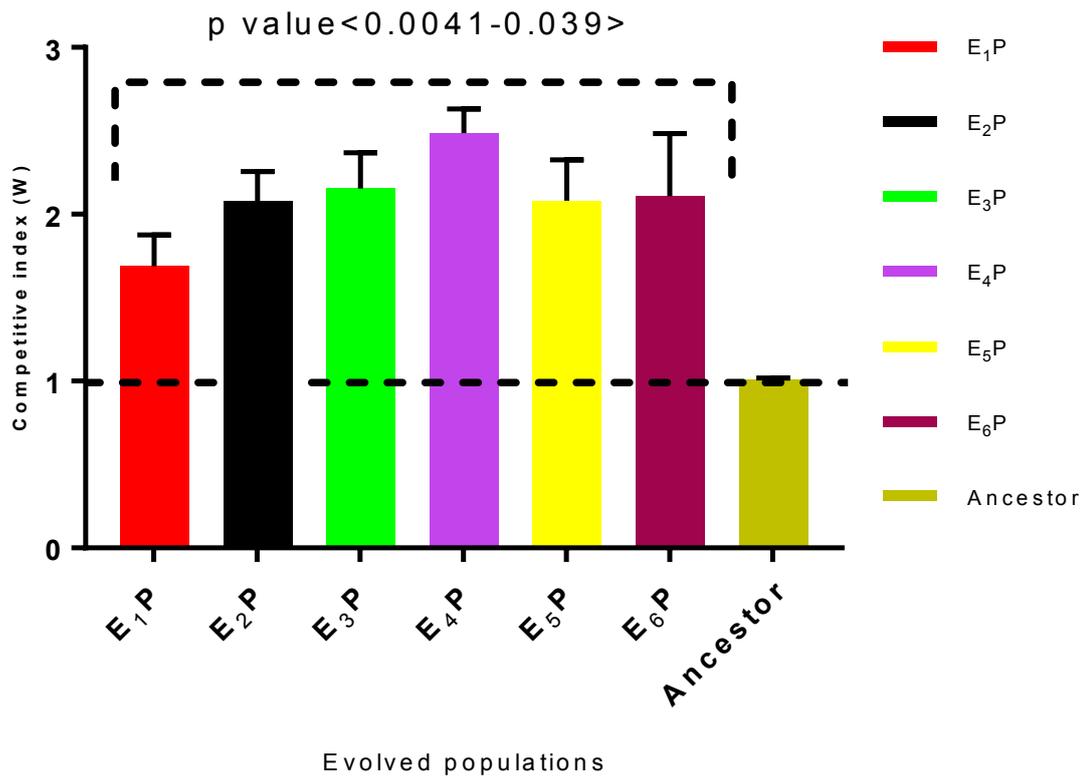
*Figure 3.6: Competitive index of the evolved strains and ancestor at pH 4.5. Values are average of three biological replicates and errors bars are standard error of mean of three independent competition experiments. P-**values were calculated using a student's T**-test by comparing the competitive index to a value 1. Value of 1 denotes competitive neutrality.*

3.8 Fitness measurements of the evolved populations at pH 7

Lab-based evolution experiments help us understand and test fundamental principles of evolutionary theory. One such principle is the occurrence of trade offs. Gain of fitness in one environment is often accompanied by trade offs in another. For example, Johnson *et al.,* 2014 demonstrated that *E. coli strains* evolved to resist extreme pH (2.5) had compromised their ability to grow at pH 7. To test whether the populations evolved at pH 4.5 had any fitness losses or trade offs at optimal growth pH, we repeated the competition experiment at pH 7. The results showed no significant differences in fitness between the evolved strains and ancestor (Fig 3.7). Hence, we concluded the evolved populations have not traded of their ability to grow at pH 7.

3.9 Fitness measurements of the acid evolved strains at pH 4.5

As stated in section 3.3, six purified clones were isolated from the six evolved populations. These clones are referred to as the acid evolved strains in this study.

As explained for the populations in section 3.7, we wanted to test the fitness of these evolved clones at pH 4.5. By doing so we could also test, if these clones displayed the same fitness phenotype as their parent populations. To test this, we competed the six evolved clones or acid evolved strains against KHOO1.

As done with the populations, we calculated the fitness indices at different time points of the competition experiment at pH 4.5 (Fig. 3.8). The competitive index of these strains increases through the course of the experiment (Fig. 3.9). The competitive advantage is seen both in exponential and stationary phase (Fig. 3.9, data shown for E_1A).

From the same data set we extracted the competitive indices of the strains at 24 hours. The evolved clones or acid evolved strains outcompete the ancestor at the selection pH. The acid evolved strains were significantly fitter than ancestor at pH 4.5(Fig. 3.10)
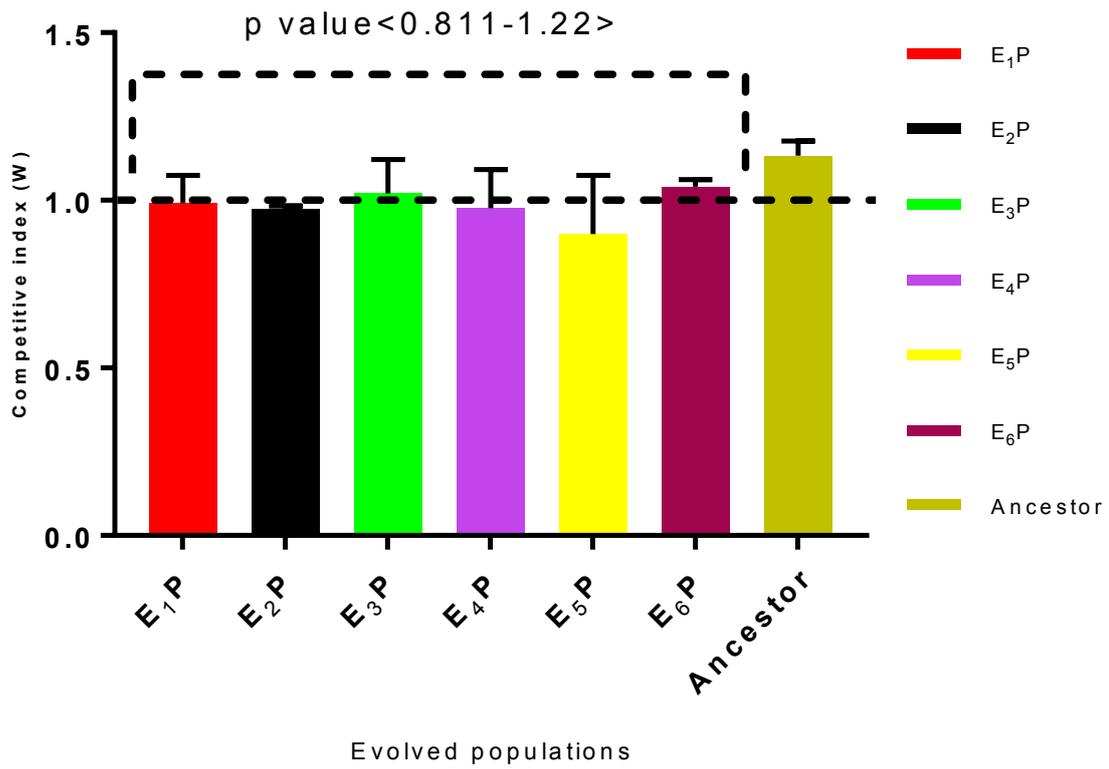
Figure 3.7: Competitive indices of the evolved populations at pH 7. Values are average of three biological replicates and errors bars are standard error of mean of three independent competition experiments. P-values were calculated using a student's T-test by comparing the competitive index to a value 1. Value of 1 denotes competitive neutrality.
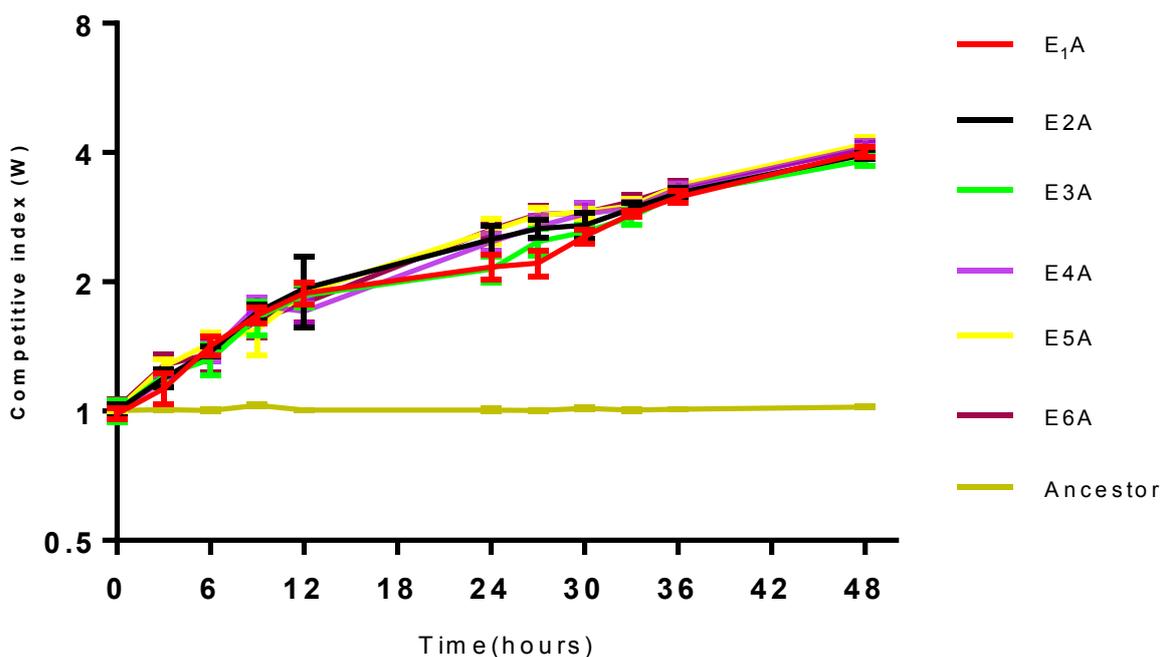
Figure 3.8: Figure 3.3: Competitive indices of the acid evolved strains versus time. The values are an average of three biological triplicates and the error bars are standard error the mean.
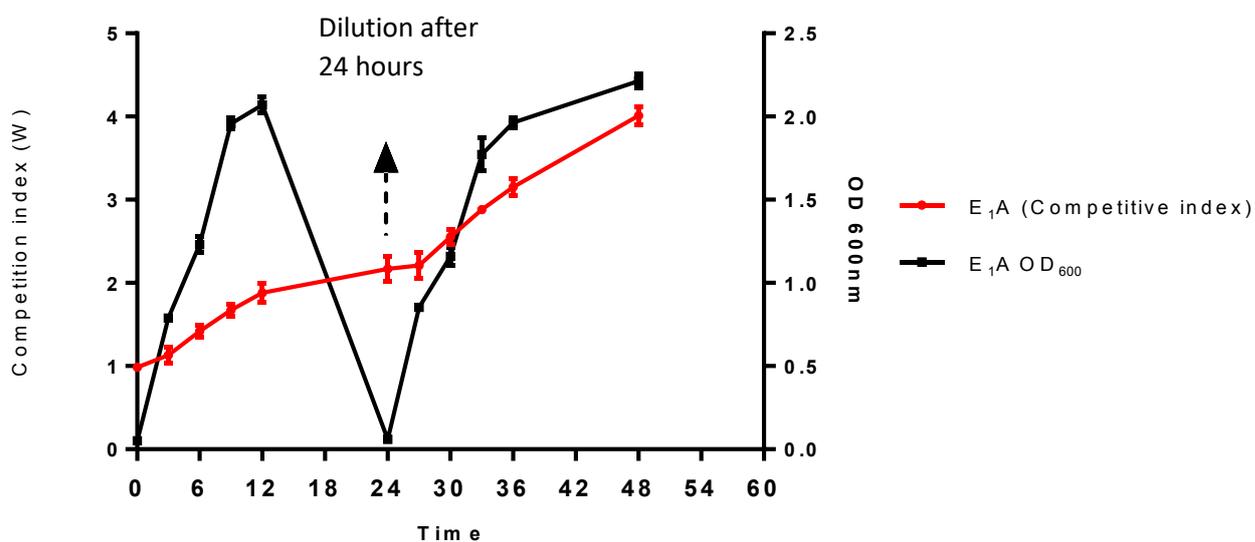
Figure 3.9: Three variable graph of the competitive index and optical density versus time. The left Y-axis shows the competitive index and the right Y-axis shows the OD 600. The values are average of three biological replicates and error bars are the standard error of the mean.
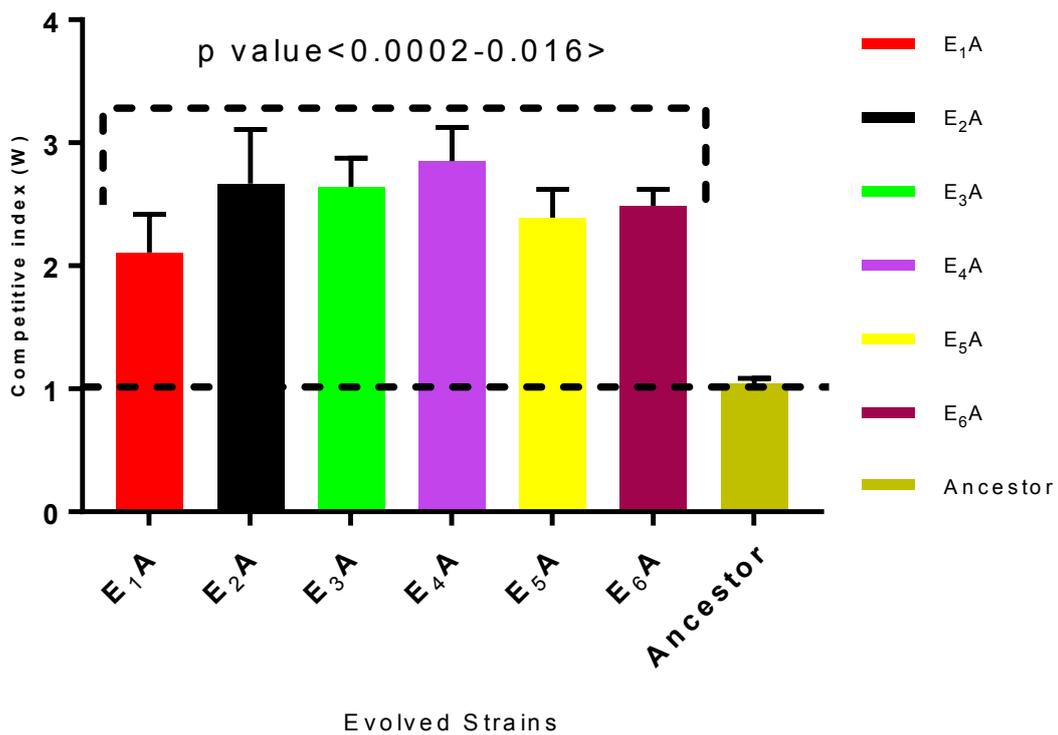
*Figure 3.10: Competitive indices of the evolved strains at pH 4.5. Values are average of three biological replicates and errors bars are standard error of mean of three independent competition experiments. P-**values were calculated using a student's T**-test by comparing the competitive index to a value 1. Value of 1 denotes competitive neutrality.*

3.10 Fitness measurements of the acid evolved strains at pH 7

As described for the populations in section 3.8, we competed the acid evolved clones or strains against KH001 at pH 7. The results showed no significant difference in fitness between the evolved strains and the ancestor (Fig. 3.11). Hence, we concluded the acid evolved clones or strains had not traded off their ability to grow at neutral pH.
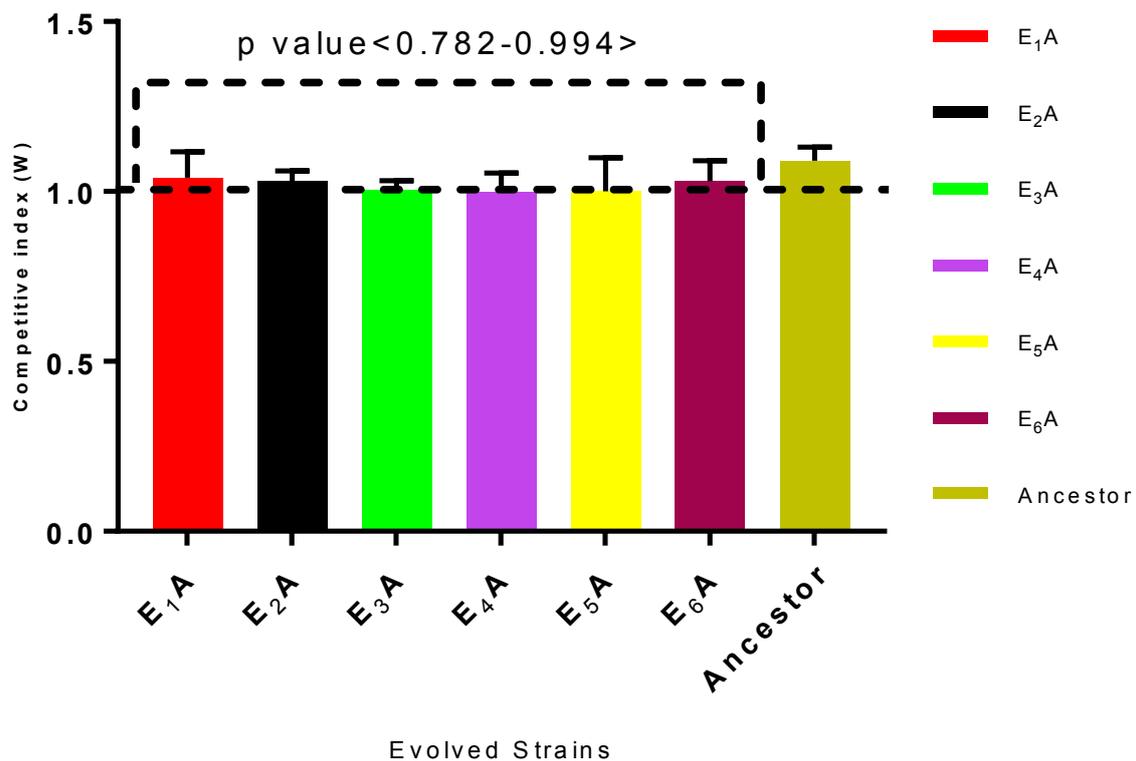
Figure 3.11: Competitive indices of the evolved strains at pH 7. Values are average of three biological replicates and errors bars are standard error of mean of three independent competition experiments P-values were calculated using a student's T-test by comparing the competitive index to a value 1. Value of 1 denotes competitive neutrality.

3.11 Cross-stress phenotyping of the evolved strains

One of the common outcomes of lab-based evolution experiments to study stress response is the evolution of cross stress resistance. Evolution of resistance induced by exposure to one form of stress also induces resistances to other stresses. For example, Rudolph *et al.,* 2010 showed that *E. coli* populations evolved to resist high temperature (45ºC), also displayed resistance to ethanol stress and peroxide stress. Dragosits *et al.,* 2013 showed that *E. coli* populations evolved to resist high salt concentration showed resistance to low pH, butanol stress and oxidative stress.

We therefore wanted to test the if evolution of resistance to pH 4.5 also conferred enhanced resistance against other stresses. To test the cross-resistance of the evolved strains, I assayed the growth characteristics of the six evolved strains and ancestor in the conditions listed here below. The weak organic acid stress experiment was done by Francesca Bushell.

a) oxidative stress (100mM Menadione)
b) membrane Stress (8% SDS)
c) osmotic Stress (400mM NaCl)
d) pH dependent biofilm formation
e) weak organic acid stress (pH 5.5, 5mM acetic acid)
f) Extreme acid (pH 3.5)
g) Extreme alkali (pH 9.5).

All the evolved strains exhibited a varying degree of cross protection to other stresses compared to the ancestor. Most of them showed increased resistance to osmotic stress (Fig. 3.12a). membrane stress (Fig. 3.13 a) and organic acid stress (Fig. 3.13 b). The evolved strains grew better than the ancestor at extreme pH 3.5 (Fig. 3.15) but showed poor growth at extreme alkaline pH 9.5 (Fig. 3.12b). The evolved strains showed enhanced biofilm formation capabilities at pH 4.5 compared to the ancestor (Fig. 3.14).

The data generated from the cross-stress phenotyping of the acid evolved strains and ancestor was fit into a logistic growth model program in MATLAB. The program was designed to calculate parameters such as specific growth rate and carrying capacities. The input data set comprised growth curve data of the evolved strains and ancestor obtained from the cross-stress experiments. All calculations were made relative to the performance of the ancestor. The output data ranks the evolved strains and ancestor based on the parameter chosen by the user. The output of the program depicted a heat map which ranked all the strains against each cross

stressor based on specific growth rate or carrying capacity. I chose carrying capacity as the output parameter because choosing specific growth rate skewed the output data plot. The program needed some improvements that could account for big differences in growth rate which would prevent skewing of the output data plot. Figure 3.16 shows a heat map which ranks the strains based on carrying capacity. Based on overall carrying capacities, E4A was the most cross-stress resistant amongst all the evolved strains (Fig. 3.16)

a

b
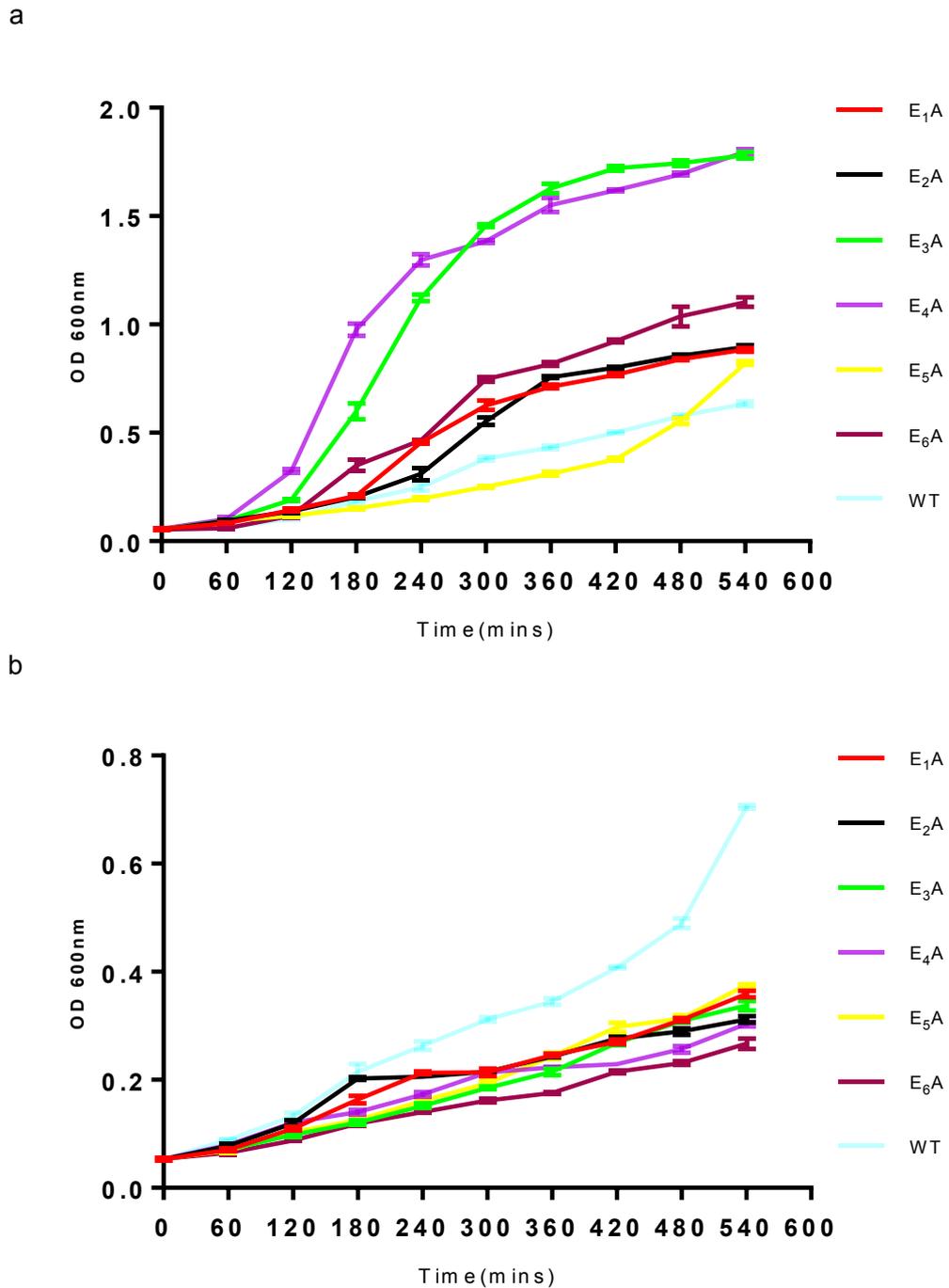
Figure 3.12: Cross stress analysis in osmotic and extreme alkali stress. Growth curves of the ancestor (WT) and the evolved strains in LB medium containing 400mmM NaCl(a) and LB medium pH 9.5(b). For the cross-stress protection growth curves three colonies of each strain were grown overnight as starter cultures. The starter cultures were diluted to a starting OD 600 of 0.05 in 250ml flaks containing 100mls of the stressor and the optical density of these cultures were monitored for 540 minutes. Values are the mean of three biological replicates and the error bars are standard error of the mean.
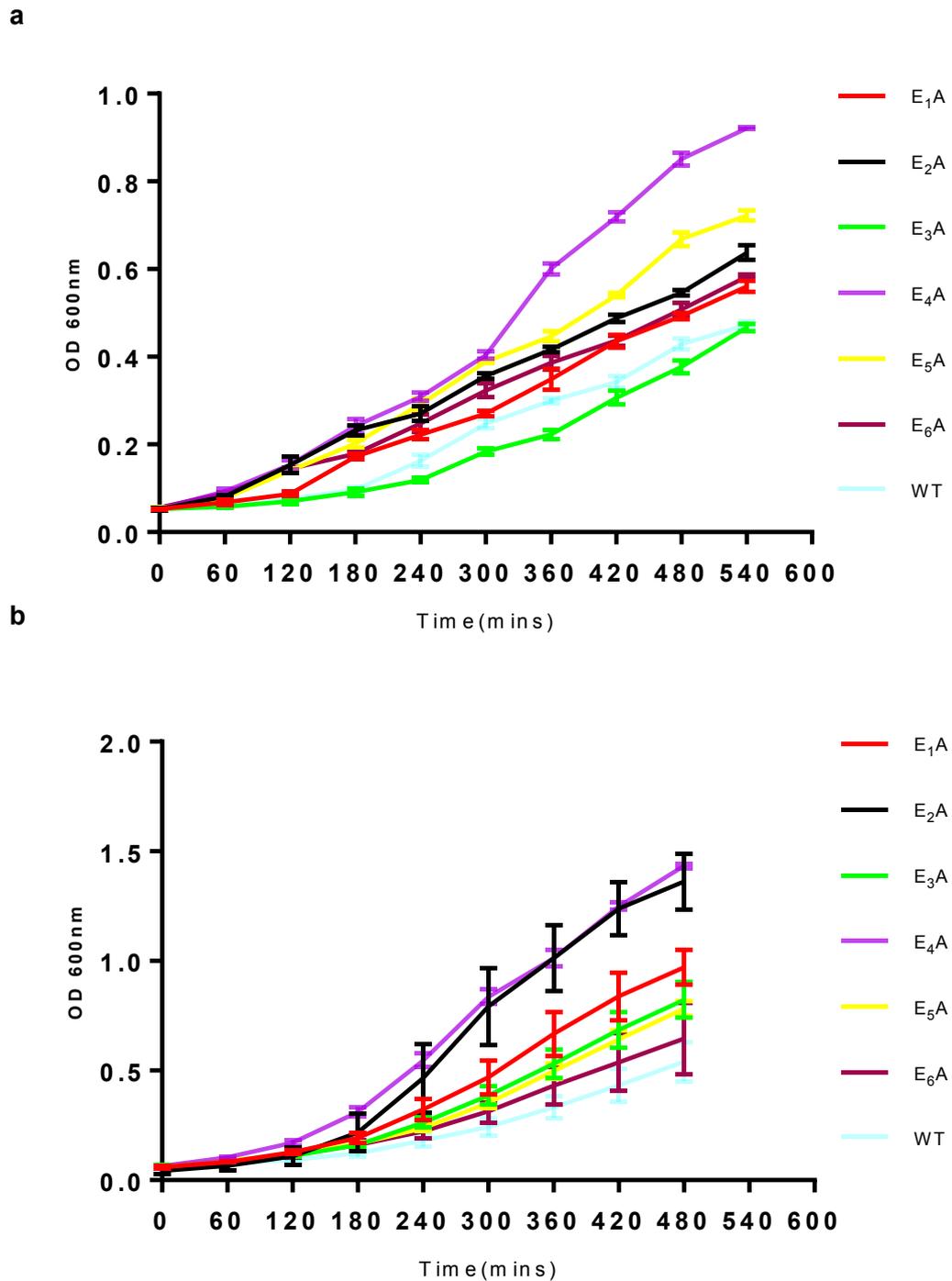
Figure 3.13: Cross stress analysis in membrane and organic stress. Growth curves of the ancestor (WT) and the evolved strains in LB medium containing 8% SDS(a) and M9-Cas medium pH 5.5, 5mM acetic acid(b). For the cross-stress protection growth curves three colonies of each strain were grown overnight as starter cultures. The starter cultures were diluted to a starting OD 600 of 0.05 in 250ml flaks containing 100mls of the stressor and the optical density of these cultures were monitored for 540 minutes and 480

minutes (acetic acid stress, 3.10b). Values are the mean of three biological replicates and the error bars are standard error of the mean.
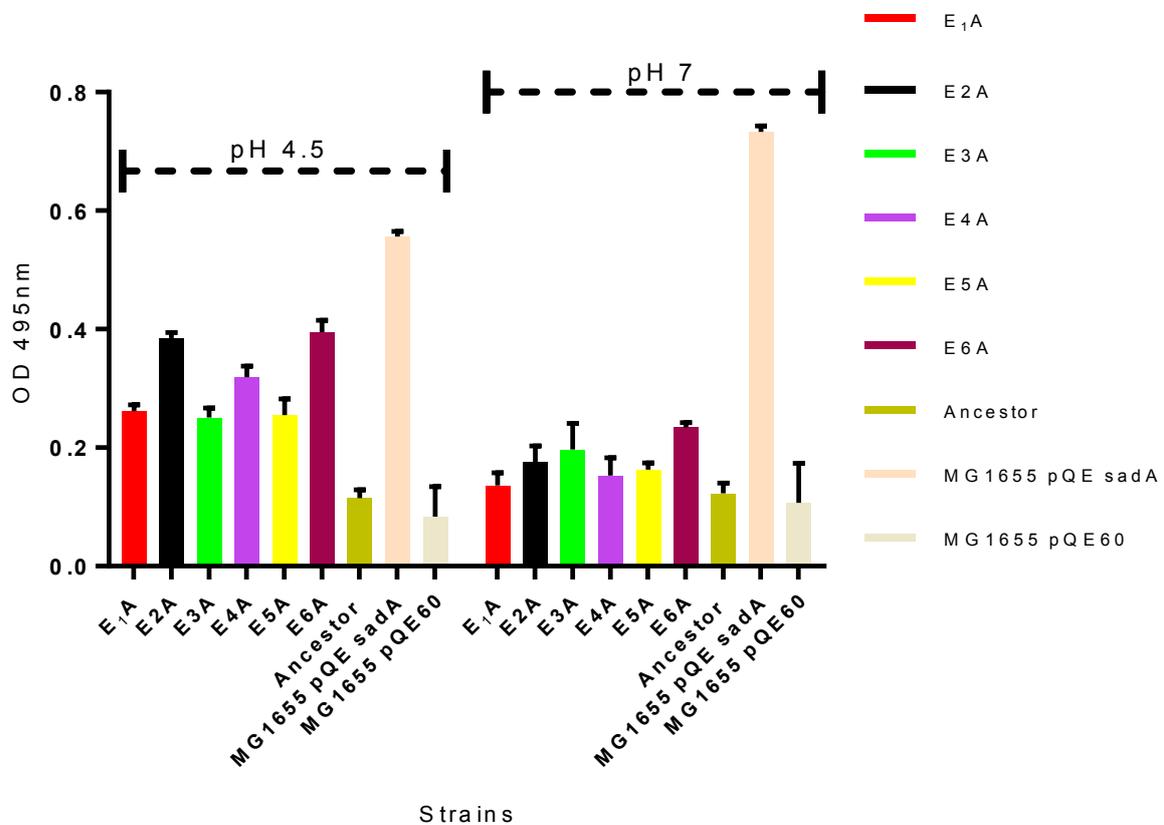
Figure 3.14: Biofilm development of the ancestor and evolved strains at pH 4.5 and pH 7 in M9-Cas medium, 37°C. MG1655 pQE-sadA and MG1655 pQE60 are the positive and empty vector control. SadA is the salmonella adhesion protein which is being overexpressed in E. coli from the pQE60 vector. Over expression of sadA in E. coli confers a hyper biofilm forming phenotype. Values are average of three independent biological replicates and error bars are standard error of the mean.
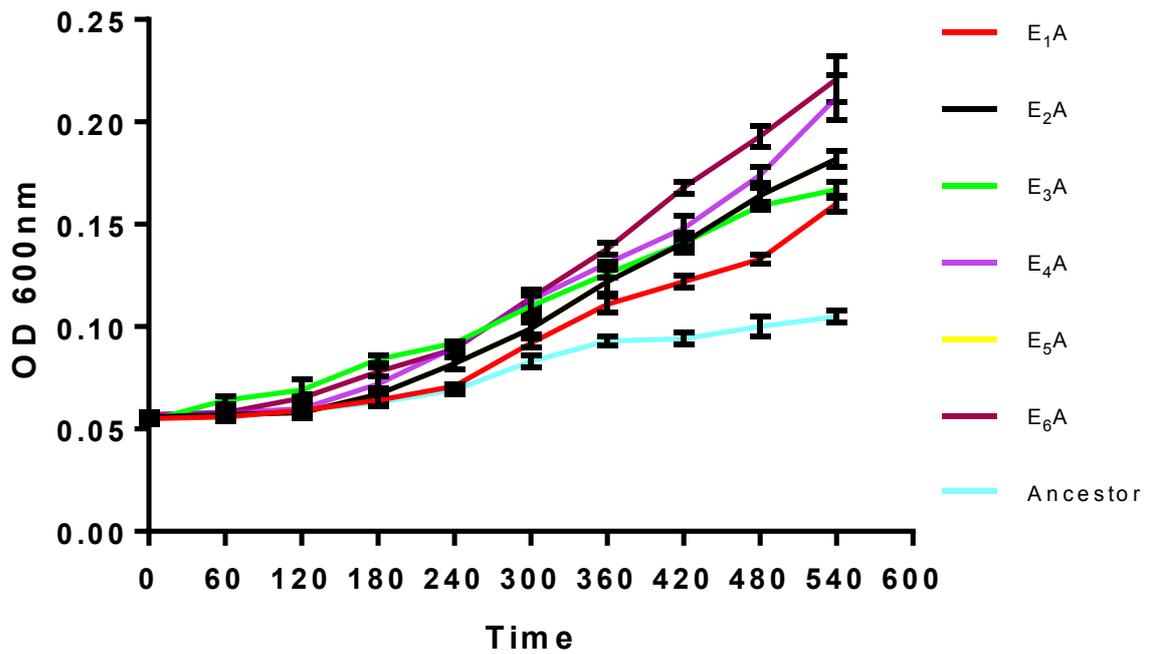
Figure 3.15: Cross stress analysis in extreme acidic stress. Growth curves of the ancestor (WT) and the evolved strains in LB medium pH 3.5. For the cross-stress protection growth curves three colonies of each strain were grown overnight as starter cultures. The starter cultures were diluted to a starting OD 600 of 0.05 in 250ml flaks containing 100mls of the stressor and the optical density of these cultures were monitored for 540 minutes. Values are the mean of three biological replicates and the error bars are standard error of the mean.
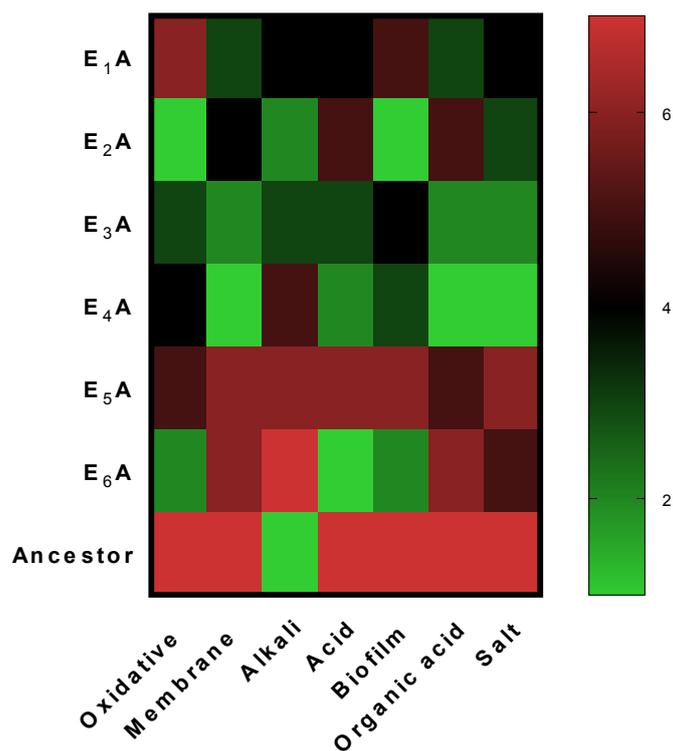
Figure 3.16: Cross-stress map of the evolved strains and ancestor based on estimated carrying capacities. The colour coding represents the ranking of the strains computed by the MATLAB program. The ranking is based on estimated carrying capacity of each strain calculated by the program from the growth curve data fed as the input. Input data was average of three independent biological replicates. The program was designed by Dr. Sara Jabbari and the execution was carried out with assistance from Francesca Busheli.

3.12 Macrophage invasion and survival assay of the evolved strains

Several of the cross-stressors we tested are also employed by macrophages against invading pathogens. Being a non-pathogenic strain without any resisting mechanisms, *E. coli* MG1655 is very sensitive to macrophage attack. We wanted to test whether the evolved strains could survive a macrophage invasion assay. To do so we performed a macrophage invasion assay (Materials and Methods) on E_2A and E_4A and the ancestor. The assay involved infecting macrophages with bacteria and counting the number of survivors at the end of the invasion assay. Using this data, the total bacterial yield is calculated. This represents the number of bacteria which successfully survived the macrophage invasion.

Since *Salmonella typhimurium* 1344 is known to survive a macrophage invasion, we used this strain as a positive control for the experiment. The results (Fig. 3.17) showed that the evolved strains survived significantly better than the ancestor in the invasion assay. The improved survival was seen against J774, a standard lab cell line of macrophages and bone marrow derived primary macrophages.
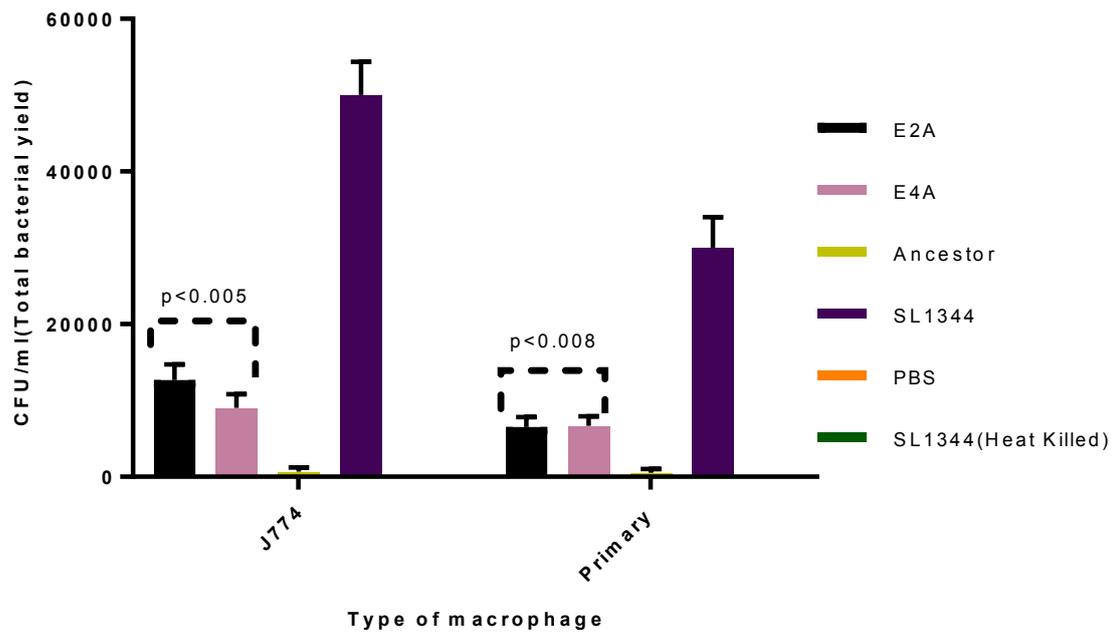
Figure 3.17: Macrophage invasion assay of E₂A, E₄A and ancestor against J774 and bone marrow derived primary macrophages. Salmonella typhimurium 1344 was used as the positive control for the invasion. A mock invasion was conducted using phosphate buffered saline and heal killed SL1344 was used as the negative control. The macrophages used were derived from three independent cell lines. Column values are average of three independent invasions and error bars are standard error of the mean.

3.13 Whole-genome sequencing of the acid evolved strains.

The lab-based evolution experiment at pH 4.5 yielded six evolved populations resistant to pH 4.5. Six clones or acid evolved strains were isolated from these populations which were also resistant and fitter at pH 4.5 compared to the ancestor. The acid evolved strains were phenotypically cross stress resistant. One of the objectives of the of this evolution experiment was to elucidate a genotype-phenotype link. The first step to elucidating this link was to sequence the genomes of the acid evolved strains to identify the genetic basis of the improved fitness. The genomes of the evolved strains and ancestor were sequenced using the Ilumina Mi-seq platform at the sequencing facility Microbes NG.

Before the sequenced reads were mapped to the reference genome NC00913.3, a preliminary quality check analysis was done using the FastQC program. All the sequenced genomes had an average Q40 value of (38 ± 2). The quality control revealed poor indexes for E_1A and E_5A (Table 3.1), hence the two genomes were resequenced. The reads from the both the sequencing runs for E_1A and E_5A were assimilated together. The new Q40 values the resequenced genomes of E_1A and E_5A were 38 and 39 respectively.

The sequenced reads of each strain were aligned and mapped to the reference genome using two independent genome analysis platforms

IGV or the Integrated genome viewer
(http://software.broadinstitute.org/software/igv)

Breseq pipeline
(http://barricklab.org/twiki/bin/view/Lab/ToolsBacterialGenomeResequencing).

IGV was found to be more suited in mutation calling for single nucleotide polymorphisms and Breseq to map genome reorganisations, large deletions and transposition events. The Breseq analysis was done by Mathew Milner. The Breseq analysis was done later in this study and showed some key results which the IGV analysis had missed. The Breseq analysis is ongoing and results of the analysis would be incorporated when completed.

Genome sequencing of the ancestor revealed almost identical sequence to the reference genome, except for a single nucleotide polymorphism in the gene *bioH* and *cbtA*. Since the two mutations were already present in the ancestor and seen in the evolved strains as well, we excluded the two mutations from our analysis.

Mutation calling was carried out on all the evolved strains. We identified different types of mutations in the evolved strains (Table. 3.2). A total of 16 mutations were identified in the coding regions, 4 intergenic insertions, two large deletions, 9 small deletions, two single base pair deletions. The ratio of the synonymous to non-synonymous mutations is low compared to other studies.

Interestingly every evolved genome had a missense mutation in the global transcriptional regulator *arcA* (Table 3.3). All the evolved strains except E4A had an identical missense mutation in the RNA polymerase alpha subunit *rpoA*. E_2A and E_5A had missense mutations in the dual transcriptional regulator *cytR*. E_1A, E_3A, E_6A has missense mutations in *cytR* and an intergenic insertion of a IS5 68 base pairs upstream of cytR. E4A had a missense mutation in cytR and the same IS5 insertion in the reverse orientation 68 base pairs upstream of *cytR*.

The evolved strains also have mutations in known regulators of acid stress response in *E. coli* such as *cadC, sspA* and *arcB*. The functional annotations of the mutated genes are listed in Table 3.4.

Both E_1A and E_4A have a large deletion each in their genomes. E_1A has lost the entire DLP12 prophage island in a large deletion of 9.7kb (Table 3.5). E4A has lost an entire operon (9.4kb) known as the Type 1 regulator of motility. This operon consists of the sensory and regulatory genes required for systematic expression of the flagella genes (Table 3.5) (fig. 3.18).

The small deletions are listed in the table 3.6. The small deletions are mostly located in intergenic regions. We excluded the small deletions from our overall analysis as further chapters will show that their effect is minimal in the overall phenotype of the acid evolved strains.

Table 3.1: Quality control results of the sequenced genomes.

All statistics are based on contigs of size >= 500 bp, unless otherwise noted (e.g., "# contigs (>= 0 bp)" and "Total length (>= 0 bp)" include all contigs.)

☑ Show heatmap
Worst   Median   Best

Statistics without reference	E1A	E2A	E3A	E4A	E5A	E6A	WT
# contigs	291	120	95	83	81	92	82
Largest contig	91 120	261 984	265 718	314 971	315 048	265 718	315 149
Total length	4 528 305	4 572 615	4 581 414	4 577 373	4 586 597	4 581 922	4 587 014
Total length (>= 1000 bp)	4 498 392	4 565 569	4 574 915	4 570 482	4 578 006	4 574 076	4 579 477
Total length (>= 10000 bp)	4 027 035	4 423 978	4 472 224	4 486 208	4 506 616	4 494 647	4 499 553
Total length (>= 50000 bp)	1 318 642	3 321 939	3 639 949	3 874 581	3 974 321	3 737 775	3 908 683
Misassemblies							
# misassemblies	10	2	5	7	3	3	3
Misassembled contigs length	132 685	61 538	125 356	217 953	352 361	160 833	204 295
Mismatches							
# mismatches per 100 kbp	4.34	2.96	3.83	1.2	1.22	1.47	0.94
# indels per 100 kbp	0.38	0.22	0.2	0.22	0.17	0.18	0.2
# N's per 100 kbp	12.41	11.79	11.57	16.490	11.58	11.59	11.58
Genome statistics							
Genome fraction (%)	97.428	98.421	98.573	98.51	98.709	98.524	98.749
Duplication ratio	1.002	1.001	1.002	1.001	1.001	1.002	1.001
# genes	4103 + 135 part	4216 + 67 part	4222 + 66 part	4228 + 57 part	4238 + 51 part	4222 + 61 part	4243 + 49 part
# operons	806 + 59 part	841 + 34 part	842 + 34 part	843 + 32 part	849 + 28 part	844 + 32 part	847 + 30 part
NGA50	32 797	83 592	100 580	133 066	133 147	95 901	133 147
Predicted genes							
# predicted genes (unique)	4265	4264	4288	4272	4280	4283	4280

Table 3.2: List of the different types of mutations in the evolved strains.

Types of mutations	E$_1$A	E$_2$A	E$_3$A	E$_4$A	E$_5$A	E$_6$A
Neutral	1	-	-	-	-	-
Nonsense	-	-	-	2	-	3
Missense	3	3	6	3	3	
Insertion elements	1	-	1	1	-	1
Small deletions	1 (16bp)	2 (10bp and 23bp)	2 (18bp and 11bp)	-	2 (40bp and 54bp)	2 (23bp and 73bp)
Large deletions	1(9.7kb)			1 (9.3kb)		

Table 3.3: List of mutations and insertions in the evolved strains

Strain	Gene	Mutation	Amino acid change	Nature of mutation	Comments
E₁A	*arcA*	ATG/ATT	M39I	Missense	Non-synonymous
E₁A	*rpoA*	AAC/CAC	N294H	Missense	Non-synonymous
E₁A	*tnaC*	TTC/TAC	F132L	Missense	Non-synonymous
E₁A	*IS5: cytR*	-	-	Insertion	Intergenic modifier
E₂A	*arcA*	AAC/AAA	N106K	Missense	Non-synonymous
E₂A	*rpoA*	AAC/CAC	N294H	Missense	Non-synonymous
E₂A	*cytR*	GAT/GA*	D34/T*	Single-base pair deletion	Frameshift
E₃A	*arcA*	GAT/GCT	G78V	Missense	Non-synonymous
E₃A	*rpoA*	AAC/CAC	N294H	Missense	Non-synonymous
E₃A	*cytR*	CGC/CGT	P291L	Missense	Non-synonymous
E₃A	*ttdT*	GGA/GCA	G353E	Missense	Non-synonymous
E₃A	*prfB*	ACG/GCG	T246A	Missense	Non-synonymous
E₃A	*IS5: cytR*	-	-	Insertion	Intergenic modifier
E₄A	*arcA*	CGT/AGT	R67S	Missense	Non-synonymous
E₄A	*cytR*	TTT/CTT	F82L	Missense	Non-synonymous
E₄A	*IS5: cytR*	-	-	Insertion	Non-synonymous
E₄A	*sspA*	GCG/GTG	D80N	Missense	Non-synonymous
E₄A	*cadC*	TTA/TGA	L381/*	Non-sense	Stop-codon gained
E₄A	*arcB*	GAA/TAA	E434*	Missense	Non-synonymous

E₅A	*arcA*	**GAT/GCT**	**D98A**	**Missense**	**Non-synonymous**
E₅A	*rpoA*	AAC/CAC	N294H	Missense	Non-synonymous
E₅A	*cytR*	GCG/GAG	A93E	Missense	Non-synonymous
E₆A	*arcA*	ATG/ATT	M39I	Missense	Non-synonymous
E₆A	*rpoA*	AAC/CAC	N294H	Missense	Non-synonymous
E₆A	*cytR*	TTA/TGA	L179/*	Nonsense	Stop-codon gained
E₆A	*fimE*	CGC/CG*	R107/C*	Single-base pair deletion	Frameshift
E₆A	*IS5: cytR*	-	-	Insertion	Intergenic modifier

Table 3.4: Functional annotation of genes identified with mutations in the WGS

Gene	Annotation
arcA	Response regulator component of the two-component system arcB/arcA, global transcriptional regulator of anaerobic stress
rpoA	Alpha subunit of RNA polymerase
cytR	Cytidine regulator and CRP associated transcriptional repressor
cadC	Transcriptional regulator of the lysine-cadaverine acid resistance system
sspA	Starvation response protein
fimE	Recombinase, regulator of fimbriae synthesis and phase variation
tnaC	Operon leader peptide regulator of tryptophan metabolism
arcB	Histidine sensor kinase, repressor of *rpoS* or σ^{38}
prfB	Translation polypeptide chain release factor.
ttdT	Tartarate-succinate antiporter
IS5	Insertion element, transposase, trans-activator and trans-repressor.

Table 3.5: Large deletion in E₁A and E₄A

Strain	Deletion	Origin	Termination
E₁A	*(icd, ymfD, ymfL, ymfE, lit, intE, ymfL, ymfG, cohE, ymfD, croE, ymfL, ymfM, owe, aaaE, ymfR, bee, jayE, ymfQ, strP, tfaP, tfaE, strE, pine)*	1196135	1211413
E₄A	*(tap, tar, cheW, cheA, motA, motB, flhD and flhC)*	1969225	1978505

Figure 3.18: 9.7kb deletion in the genome of E₁A, the deletion comprises of genes located in the DL12 prophage island. The deleted region is highlighted in black. Sequenced reads were aligned against the reference genome NC00913.3 (E. coli K12 MG1655). A zero-coverage script was run on the aligned reads to detect regions of large deletions.

Strain	Deletion	Origin	Termination
E$_1$A	Upstream of *nirB*	3486537	3486553
E$_2$A	Upstream of malK	4235365	4235375
E$_2$A	Intergenic deletion between *ompT* and *envY*	574380	574403
E$_3$A	Upstream of *cysG*	1100934	1100945
E$_3$A	Downstream of *panC*	148977	148989
E$_5$A	Upstream of *sfmD*	547707	547657
E$_5$A	Upstream of *yfgA*	316052	316106
E$_6$A	Downstream of *folK*	147496	147569
E$_6$A	Upstream of *galR*	2965137	2965160

3.14 Discussion for chapter 3.

In this chapter I introduced the evolution experiment, which comprised six populations evolved for 740 generations at pH 4.5. This chapter also discussed the preliminary phenotypic and genotypic characterisation of the six clones or acid evolved strains that were isolated from the six evolved populations.

Following 740 generations of lab-based evolution at pH 4.5, the six evolved populations displayed significant fitness gain at this pH. The six clones or acid evolved strains isolated from the populations showed a similar fitness profile. A high degree of phenotypic convergence was seen at fitness level with respect to the populations and clones. Neither the populations nor the clone showed any differences in fitness at pH 7. The fitness improvements observed in this evolution experiment were greater than observed in another study, where the authors evolved 24 replicate populations of *E. coli* W3110 at pH 4.6 for 2000 generations (Harden *et al.,* 2015). A careful comparison of the fitness profiles of clones isolated in both the studies showed that the evolved strains in this study attain higher competitive indices at 24 hours than the clones studied in Harden *et al.,* 2015. Although both the evolution experiments were conducted at similar pH values, such subtle differences may be a result of using a different strain or different propagation conditions etc. Moreover, the comparison of fitness indices is only limited to 24 hours in case of both the experiments, as the authors don't show a time course calculation of fitness indices in their competition experiments.

This is true in case of several lab-based evolution studies, where fitness indices have been calculated and reported only at 24 hours. Most of the available literature of lab-based evolution states that fitness remains constant throughout the competition experiment. But our data suggests that fitness indices improve and change as cells move from exponential to stationary phase. The conditions under which this evolution experiment was done must have selected for improved phenotypes in exponential and stationary phase.

This improvement in fitness seen in the six acid evolved strains at pH 4.5 is accompanied by increased resistance to other stresses. These evolved strains showed broad cross stress resistance. This is not surprising as another study Dragosits *et al,* 2013 showed that *E. coli* cells propagated for 500 generation in high salt medium showed increased resistance to other stress such as osmotic stress, membrane stress and acid stress. The acid evolved strains in this study are not only resistant to pH 4.5 but also show a significant resistance to weak organic acids. It

is possible that exposure to acid stress caused by HCl maybe inducing resistance mechanisms which confer protection against organic acids.

This aspect has not been seen in other studies focusing on lab-based evolution of acid resistance. For example, the authors of Harden *et al.,* 2015 don't report broad cross protection abilities of their acid evolved strains, although both the evolution experiments were done under fairly similar conditions. It is possible that these differences in phenotype may be result of differences in experimental design and the capabilities of the ancestral strain.

But such broad cross stress resistance also comes with trade-offs in other environments. The six evolved strains have traded off their ability to grow at alkaline pH. Improvement in fitness at low pH comes with a correlated change in fitness at alkaline pH. It may be due to a shift in the optimum range of pH values within which these evolved strains can comfortably grow.

The cross-stress resistant nature of these acid evolved strains also enable them to survive better in a macrophage invasion assay. The pH of the phagosome is around 4.6 ± 0.3 (Rotstein *et al.,* 1998) and relevant to the pH of the evolution experiment. Evolved mutants from the previous evolution experiment which were evolved to resist pH 2.5 didn't show protection capabilities against a macrophage invasion (Johnson *et al.,* 2014). These differences seen between strains evolved to resist different pH values shows that the specificity of adaptation. Evolution at extreme pH 2.5 yielded survivors which could resist extreme acid shock but evolution under mildly acidic conditions yielded strains which were generally more stress resistant.

Such differences can be attributed to the differences in the underlying genetic basis of adaptation. Whole genome sequencing of the six acid evolve strains showed that all the evolved strains had mutations in the global transcriptional regulator *arcA*. ArcA is involved in regulating large gene regulatory networks. Mutations in global regulators cause pleiotropic effects and may be one the factors contributing to the broad stress resistance of the six acid evolved strains. One of the key genes *arcA* regulates is the stationary phase sigma factor *rpoS* or $\sigma^{38.}$ The following results chapters will highlight the role of *arcA* and *rpoS* in the six acid evolved strains

Interestingly five out the six evolved strains have the identical mutation in the RNA polymerase (α) subunit or *rpoA*. RNA polymerase mutations are commonly seen in evolution experiment (Harden *et al.,* 2015), (Palsson *et al.,* 2012). Although the initial thoughts on this result suggested a cross contamination event early in the evolution experiment, the coming chapters will show that this mutation arose independently across the five evolved strains. Another study

showed that 10 *E. coli* clones isolated from populations evolved under NADPH stress had the same identical mutation in the gene *nuoF.* The mutation in rpoA is a classic example of parallel and convergent molecular evolution, where strains evolved independently have identical mutations (Woods *et al.,* 2006)

All the evolved strains have mutations in the *cytR* locus either as SNP's or insertions elements or both. *CytR* plays a key role in CRP-associated transcriptional regulation of genes involved in nutrient uptake and metabolism. The next chapters show the effect of *cytR* on the overall phenotype of the evolved strains. Since CytR is involved in CRP associated regulation, this led us to measure CRP levels in the evolved strains, results of which will follow in the further chapters.

In addition to these three keys genes our analysis of the evolved genomes showed mutations in other genes such as *cadC* and *sspA*. Both these genes are involved in the regulation and expression of genes in response to acid stress. CadC or the cadaverine antiporter is involved in the transcriptional regulation and expression of the lysine decarboxylases or AR4 (refer to chapter 1, section xxx) (Richard and Foster,1998). *sspA* is involved in relieving HNS mediated suppression of the glutamate decarboxylases or AR2 in stationary phase (Ming *et al.,* 2001). All the above genes are involved in regulating large gene networks which could be responsible for the cross-resistant phenotype of these strains.

Two of the evolved strains have large deletions in their genomes. Large deletions often occur in lab-based evolution experiments and are very important as they can lead to chromosomal rearrangements and altered gene expression which may not be possible through point mutations (Raeside *et al.,* 2014).

The results obtained in this phase of the study posed many important questions. Owing to limitations of time we had to focus on the most important ones, which were

a) What is the nature of mutations in arcA and what percentage of the phenotype depends on *arcA*?

b) What role does *cytR* play in generating this phenotype. Is the effect of *cytR* contingent on the presence of *arcA* mutations?

c) Why is the same locus mutated in *rpoA*? Is it a case of cross-contamination?

These questions form the basis of the next results chapter. In the next chapter we further investigate the role of these genes and their link to the observed phenotype.

3.15 Summary of results for chapter 3:

a) We successfully conducted a lab-based evolution experiment, where we evolved six independent populations in LB medium pH 4.5 for 740 generations.

b) The evolved populations display enhanced fitness compared to their ancestor at pH 4.5, but didn't show any trade-offs at optimal pH.

c) Clones isolated from these populations known as the acid evolved strains (AES), displayed the same fitness profile as that of the populations.

d) The acid evolved strains display cross-protection against other stresses.

e) These acids evolved strains successfully survive macrophage invasion attack compared to the ancestor.

f) WGS of the acid evolved strains identified three key targets for further analysis- *arcA, cytR* and *rpoA*.

Evolution of acid resistance to pH 4.5 in *E. coli* K12 MG1655

Chapter 4

Identifying the genetic basis of adaptation to pH 4.5- Part 1

4.1 Identifying the genetic basis of adaptation

Chapter 3 introduced the evolution experiment, I described the phenotypic and genotypic characterisation of the six acid evolved strains. The evolved strains had different types of mutations in their genomes. We identified three genes which were repeatedly selected in these evolved strains, *arcA*, *cytR* and *rpoA*. This chapter will focus on understanding the role of some these genes in conferring the observed phenotype. But before I move on to the results, I will present a brief review about the three genes. This will help in understanding the results of the current and coming chapters.

4.1.1 *arcA*-Phosphorylated DNA binding transcriptional regulator.

ArcA is one of the seven global regulators of gene expression in *Escherichia coli* (Perrenoud *et al.,* 2005). It is also synonymously known as *dye* and *sfrA*. It belongs to OmpR/ PhoB family of transcriptional regulators. ArcA is the response regulator component of the hybrid sensor kinase ArcB. The primary role of the ArcB/ArcA two component system is to regulate the switch between aerobic to anaerobic growth in *E. coli*. Exposure to microaerophilic and anaerobic conditions causes the oxidation of key residues in the membrane spanning domain of ArcB (Colloms et al., 1998) which leads to the activation of the kinase.

Activation of ArcB starts a cascade of phosphorylation reactions which leads to transfer of a phosphate molecule from the sensor kinase to its cognate response regulator ArcA (Davies *et al.*, 1999). This process is known as transphosphorylation of ArcA. Phosphorylated ArcA acts as a transcriptional regulator. The primary role of ArcA as a transcriptional regulator is the switching of the carbon oxidation pathways during anaerobic stress (Lin and Kwon, 1999). Although the dedicated role of *arcB* and *arcA* is to regulate the response to anaerobic stress, they also regulate the expression of various genes during exponential phase. Battesti *et al.,* 2002 showed that ArcA expression is required for the expression of genes involved in magnesium and calcium metabolism in exponential phase. The role of ArcA in exponential phase growth is not very well understood but discussions in various research papers on ArcA suggest that oxygen may be a limiting factor for exponentially growing cells. Cells may experience micro-aerobic conditions which may lead to the activation of ArcA.

ArcA represses the gene regulatory networks responsible for controlling the aerobic respiratory metabolism and citric acid cycle. It activates some key genes which control fermentative metabolism in *E. coli* (Gottesman *et al.,* 2001*)*. Chromatin immunoprecipitation analysis showed that *arcA* regulates 146 operons in *E. coli*. Out of these 146 operons, 88 of them are

under the direct regulation of arcA. 74 of these operons are repressed while 11 of them are activated by *arcA*. (Park *et al.,* 2013). ArcA interacts with 17 small RNA's in *E. coli*, the role of these small RNA's is to act as intermediary vehicles which allow ArcA to interact with other global regulators such as HNS and FNR during conditions of stress (Gottesman *et al,* 1998).

The regulatory mechanisms by which *arcA* causes a global shift in the transcriptional landscape of the cell in response to stress is not well explained in the available literature. But what is known is the role of ArcA is regulating the general stress sigma factor RpoS.

One of the key roles of ArcA in *E. coli* is to control the levels of the general stress sigma factor RpoS or σ^{38}. RpoS is the master regulator of the general stress response in *E. coli*. The ArcB/ArcA two component system controls RpoS at the level of transcription, translation and post-translation (Mike and Hengge, 1992). There are four known routes through which this two-component system regulates RpoS.

The ArcB sensor kinase under goes auto-phosphorylation when the cell experiences high sugar and phosphate levels, during exponential growth. Autophosphorylated ArcB, transphosphorylates the response regulator RssB. RssB is a post-translational regulator of RpoS. RssB binds to RpoS and delivers it to ClpXP protease complex. The ClpXP protease complex proteolytically degrades RpoS in the cell (Matshushika and Mizuno, 1998). ArcB mediated phosphorylation of RssB and proteolysis of RpoS is responsible for maintaining low levels of RpoS in rapidly growing cells (Hengge *et al.,* 1995).

ArcA regulates the levels of RpoS by three known routes, by preventing transcription initiation of *rpoS* and increasing RpoS instability. I will describe the three mechanisms in brief. *rpoS* has two promoter binding sites in the transcription activating region denoted as p1 and p2 (Jeon et al., 2001). The p1 site is where CRP-cAMP binds to act as the Class 1 transcriptional activator of *rpoS*. *arcA* has two binding sites on this promoter p1. Phosphorylated ArcA forms an oligomer and binds to promoter p1 to block transcription initiation by CRP-cAMP (Yamanato et al., 2005).

Phosphorylated ArcA interacts with a host of small RNA's one of them being *arcZ. arcZ* is a Hfq like protein and acts as transcriptional repressor. In exponentially growing cells, phosphorylated *arcA* activates the expression of ArcZ which binds to the promoter p2 of *rpoS* to block transcription initiation by CRP-cAMP. *arcZ* mutants show elevated RpoS levels during exponential phase (Levanon *et al.,* 2002).

IraM is one of anti-RpoS adaptor proteins and is involved in controlling the post-translational stability of RpoS in *E. coli*. IraM prevents, RssB mediated proteolytic degradation of RpoS. IraM prevents ClpXP assembly on RpoS by blocking the active site of the proteolytic chaperone ClpX *(Mickveski et al., 2002)*. This process takes place when the *E. coli* experiences nutrient stress or is in stationary phase growth and requires active RpoS to regulate the general stress response pathway. During normal cellular conditions, phosphorylated ArcA binds to the promoter of *iraM* and inhibits expression by blocking RNA polymerase assembly for transcription initiation. Repression of IraM causes an increase in the RssB mediated proteolytic degradation of RpoS *(*Welch *et al.,* 2014*)*. Further discussions on the role and contribution of arcA would follow later in this chapter and the coming chapters.

4.1.2 *rpoA*-RNA polymerase subunit alpha (α)

RpoA or the RNA polymerase subunit alpha (**α),** is one of the subunits of the *E. coli* RNA polymerase holoenzyme. *rpoA* consists of two independently folded domains connected through a flexible linker. The N-terminal domain or the (αNTD) plays a key role in the overall assembly of the RNA polymerase core complex, whereas the C-terminal domain or the (αCTD) is involved in contacting the DNA binding regions. The C-terminal domain interacts with transcriptional activators during the formation of the activation complex (Bell *et al.,* 1990). Structural studies have shown that mutations in the C-terminal domain of RpoA alters the cooperativity the RNA polymerase core complex and the DNA binding regions (Blatter *et al.,* 1994).

Much of the available literature on *rpoA* is based on its role in CRP-associated transcriptional activation. CRP or cyclic AMP receptor protein is a transcriptional activator which binds to conserved DNA binding regions upstream of the transcription start site in presence of cyclic AMP or cyclic adenosine monophosphate (Savery et al., 1992). CRP is structurally homodimeric and binds to specific locations on the DNA. Based on the binding patterns of CRP, the promoters CRP activates can be grouped into Class I and Class II. Class I CRP promoters contain a CRP binding site centred at *-61.5* from the transcription start site, whereas Class II CRP promoters contain a binding site at *-41.5*.

CRP interacts with the C-terminal domain of RpoA to facilitate binding of the RNA polymerase core complex to the DNA segment adjacent to the CRP binding site. This interaction between CRP-αCTD takes place through the surface exposed loop comprising residues 145 to 156 of CRP (Gourse *et al.,* 1996). In case of the Class I promoters this interaction takes place between

the downstream section of the C-terminal domain and *CRP* whereas for Class II CRP promoters this interaction takes places between the upstream section of the C-terminal domain of RpoA and CRP (Attey *et al.*, 1992).

Site-directed mutagenesis and alanine substitution studies showed that the interaction between CRP and the C terminal domain of RpoA at Class I *CRP* promoters depends on 6 key residues of the C-terminal domain. These residues are 281, 285, 294, 301, 303 and 307. Residues 281, 285, 301, 303 and 307 are involved in maintaining the stability of the interaction between CRP-αCTD. Residue 294 is critical in the driving the interaction between the CRP- αCTD complex and DNA. Attey *et al.,* 1992, showed that mutating the 294 residue alters the transcription profile of the Class I CRP promoters.

This locus is also relevant to this study, as five out six evolved strains have mutations at the 294 position in the RNA polymerase α subunit. The mutation causes an asparagine to histidine substitution at position 294. The coming chapters will show that this mutation has a crucial role in conferring the acid resistant phenotype to these strains.

In this chapter I establish that although the same mutation arose in five independent strains this was not a result of cross contamination and the further chapters outline the interaction between *arcA* and *rpoA* in conferring the acid resistant phenotype of the evolved strains. The functional mechanism by which mutations in RNA polymerase subunits create a global change in the cell are not very well understood. This study too will pose some important questions about the mechanisms by which mutations in the RNA polymerase subunits effect global transcription in *E. coli*.

4.1.3 *cytR -cytidine dual transcriptional regulator.*

CytR is the regulator of nucleoside transport and metabolism in *E. coli*. It is a helix-turn-helix (HTH) like transcriptional repressor. It acts as the repressor of the genes located in the *cytR* regulon and as a co-activator of transcription in some cases. The *cytR* regulon consists of nine unlinked genes which are all involved in nucleoside transport and metabolism (Hammer and Dandadell, 1989). The key genes which constitute the cytR regulon are *cdd* (cytidine deaminase), *deoA* (thymidine phosphorylase), *deoB* (deoxyribomutase), *deoC* (deoxyriboaldolase), *deoD* (purin nucleoside phosphatase) and *udp* (uridine dephosphorylase). In addition to these genes, *cytR* also regulates the expression of *nupG*, *nupC*, and *cytX*. These genes are involved in the membrane transport of nucleosides (Pederson *et al.,* 2003). In addition to the regulating the expression of the genes located in the cytR regulon, it is also involved in

the repression on the genes involved in biofilm formation. It is one of the key negative regulators of the *fim* operon.

E. coli uses ribonucleosides for the biosynthesis of nucleotides but also as sources of carbon and nitrogen when it experiences stress. Most of the genes which are involved in the regulation and uptake of ribonucleosides are negatively regulated by *cytR,* unless cytidine is present in the growth medium. The presence of ribonucleotide cytidine in the growth medium causes downregulation of *cytR* and thus increase in the nucleoside metabolism of the cell (Hansen et al., 1998). Downregulation of *cytR* also promotes biofilm formation, which in turn increases resistance to stress.

CytR is a member of the lacI family of transcriptional regulators. It primary role in the cell is to regulate transport of nucleosides and other essential components such as phosphate and calcium. However, another key role of CytR in *E. coli* is to disrupt cAMP-CRP associated transcription, by preventing *t*he formation of the CRP-cAMP activation complex. CRP is dimeric in structure and has two conserved binding sites, it forms a dimer creating an empty cleft on the DNA. The empty cleft is occupied by cAMP to form the CRP-cAMP transcription activation complex.

Invitro-transcription assays have shown that CytR binding sites are usually situated in between the CRP binding sites. The mechanism by which *cytR* acts as a repressor is often referred to as anti-activation rather than direct repression. This anti-activation mechanism involves CytR binding to the region of DNA in-between the CRP binding sites. This causes the CRP-DNA complex to form but the cleft which should be accessible to cAMP is blocked by CytR. This prevents the CRP-cAMP activation complex formation and thereby stalls transcription activation. This process is referred to as cytR directed inactivation. The inactivation by CytR is relieved when cytidine binds to CytR and this releases CytR from its binding sites and enables CRP to interact with cyclic AMP (Barbier *et al.,* 1997).

This multistage interaction between CytR and CRP makes *cytR* an important transcriptional regulator. Studies have shown *cytR* mutants have high levels of CRP and increased expression of CRP associated transcription. At the cellular level this is very relevant to stress responses as CRP regulates several gene networks which regulate responses to a variety of stresses (Haugo and Whitnick, 2008).

Regarding this study, genome sequencing of the acid evolved strains showed that five out the six evolved strains had mutations in *cytR*. Review of the literature on *cytR* shows that CytR

plays a key role in regulating systems dedicated to stress responses. Hence, it is an important target to study and analyse in this study. As the next chapter, will show CytR has an important contribution in altering the transcriptional landscape of these acid evolved strains and further sections of this study will detail the role of ArcA and CytR together in conferring the phenotype of the evolved strains.

arcA, rpoA and *cytR* are all global regulators of gene expression in *E. coli*. The general stress resistance exhibited by the six evolved strains may be a result of alterations in the global transcriptional landscape caused by mutations in these three genes. *arcA*, *rpoA* and *cytR* interact with other key global regulators such as RpoS, HNS and CRP and mutations in three former three genes may lead to changes in the levels of the other global regulators. In the next chapter I will discuss results of the experiments looking at RpoS, HNS and CRP. This chapter will focus on understanding the order in which mutations arose in these strains, the nature of mutations in *arcA and cytR* and some key evidence on the role of these two genes in attaining the gain in fitness at pH 4.5.

4.2 Determining the order of emergence of mutations in the evolved strains.

Whole genomes sequencing of the acid evolved strains generated a list of genes with single nucleotide polymorphisms or point mutations. We wanted to determine the order in which these mutations arose in the evolved strains. To determine the order, we would have had to screen the fossil record of each strain. This could have been done using whole genomes sequencing but we were interested in using a rapid detection method.

Mismatch Amplification Mutation assay is a cost-effective technique used for single nucleotide polymorphism profiling. This technique exploits the fidelity of Taq polymerase to generate an amplicon despite a mismatch at the 3'end of the primer. (Materials and Methods). Having done a thorough survey of the experimental features of this method, we found this method suitable to profile the SNP's in this study.

The objective was to use MAMA to determine the order of emergence of the mutations in each evolved strain. To use this method, the MAMA primers should be able to distinguish between the mutant and wild type locus. The method involves designing primers which incorporate the SNP at the 3' end of the primer and a random change in the penultimate base (Materials and Methods).

For example, the *arcA* mutation in E_1A causes a Methionine to Isoleucine change at the 39th amino acid. Since the genomes of these strains had been sequenced we could determine at exact SNP which causes this change. The mutation causes a base pair change from ATG to AT==T==, the SNP locus is highlighted in yellow. The MAMA primer incorporates this SNP and a random change at the penultimate base.

The sequence the primer would be 5'- GGC TAT GAT GTT TTC GAA GCG ACA GAT GGC GCG GAA ACT-3'

This unique design capitalises on the principle of that primers with such mismatches at the 3' end will destabilise the primer-template annealing temperature, thereby reducing PCR efficiency at the wild type locus due to poor extension capabilities of Taq polymerase.

To test the specificity of the MAMA, we designed primers specific to the arcA mutation (M39I) in E_1A and set-up PCR reactions using genomic DNA isolated from E_1A and the ancestor.

As seen in figure 4.1 the MAMA gave a positive PCR product only in case of E_1A and not the ancestor. A positive PCR product can be seen only when the template DNA has the mutant

SNP locus, and a faint band in case of the wild type. The method was sensitive enough to detect a mutant locus even if the template DNA containing the mutant locus was diluted 100-fold compared to the wild type DNA (data not shown, work done by Dorothy Ling)

Prior to performing MAMA on the fossil record, the specificity of all MAMA primers used in this study was checked. Since the genomes of the evolved strains and ancestor had been sequenced, for every MAMA assay the corresponding evolved strain was used as the positive control and ancestor as the negative control to validate the PCR results.
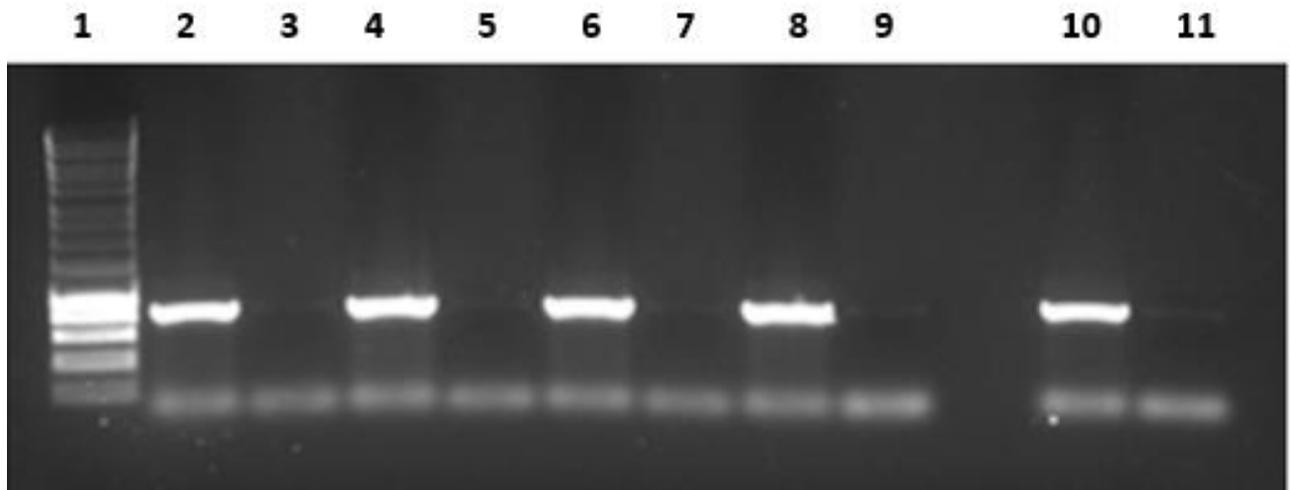
Figure 4.1: Agarose-gel electrophoresis of MAMA PCR products of E1A and ancestor for the detection of mutation in arcA(M39I). 1 is the ladder, 2,4,6,8 and 10 correspond to PCR reactions at 51.9℃, 58.9℃, 61.7℃ and 65.3℃ for E₁A and 3,5,7,9 and 11 correspond to PCR reactions at 51.9℃, 58.9℃, 61.7℃ and 65.3° for the ancestor

4.3 Timeline of emergence of mutations in the evolved strains.

As described in the previous section, the order of emergence of the mutations was to be determined using MAMA. Knowing the order of the mutations was important for two reasons. First to establish if the identical *rpoA* mutation in five out the six evolved strains was a result of a cross contamination event and second to determine if we could trace a consistent pattern in which the mutations arose. The results in the further chapters will show that determining the order of emergence helped us design experiments to elucidate the link between some mutations and their effect on the overall fitness of the six evolved strains.

Determining whether the *rpoA* mutation was a result of a cross contamination event was important. Since all the strains had different mutations in *arcA*, we hypothesised that if the *arcA* mutations occurred independently prior to the emergence of the *rpoA* mutation, it was impossible that the *rpoA* mutations were a result of cross-contamination during the experiment. But if the *rpoA* mutation emerges first followed by the *arcA* mutations, we could not the rule out the possibility of a cross contamination.

To do so we MAMA checked the fossil record of all the evolved strains using primer specific to the *arcA* mutations and the *rpoA* mutation. I will highlight one of the MAMA results as an example.

E_1A and E_2A have two different mutations in *arcA* but the identical mutation in *rpoA*. *The* arcA mutation in E_1A is M39I (methionine to isoleucine) and E_2A is N106K (asparagine to lysine). We MAMA checked the fossil record of the two strains and found that both the *arcA* mutations emerge on 30/11/2013, 14 days after the start of the evolution experiment (fig. 4.2)

However, the *rpoA* mutation in E_1A arose on 30/01/2014 (fig. 4.3) while the *rpoA* mutation in E_2A arose on 15/01/2014 (fig. 4.3). To validate this PCR result, 4 clones were isolated from the fossil tube of E_1A frozen on 15.01.2014 and whole genome sequenced. WGS of these clones showed they only had the *arcA* mutation but not the *rpoA*.

MAMA on the fossil records of each strain showed that in each case *arcA* mutation arose first followed by the *rpoA* mutation (Table 4.1). The MAMA results suggest that the different *arcA* mutation were selected and fixed early in the evolution experiment across all the evolved populations. Based on the experimental evidence we concluded that the N294H mutation in *rpoA* was not a result of cross contamination.

After having found the order of emergence of *arcA* and *rpoA*, we performed MAMA on the fossil records of the evolved strains to determine the order of emergence of the remaining SNP'S (Table 4.1). Each MAMA was repeated twice to confirm the validity of the results.

Knowing the order in which the mutations arose in the six strains showed an interesting and consistent pattern. The first mutation to arise in each case was in *arcA*, followed by the one in *rpoA* and then in *cytR*. It is highly likely that the phenotype of the six strains largely depends on these three genes. The results in the coming sections will show the nature of mutations in *arcA* and *cytR*.
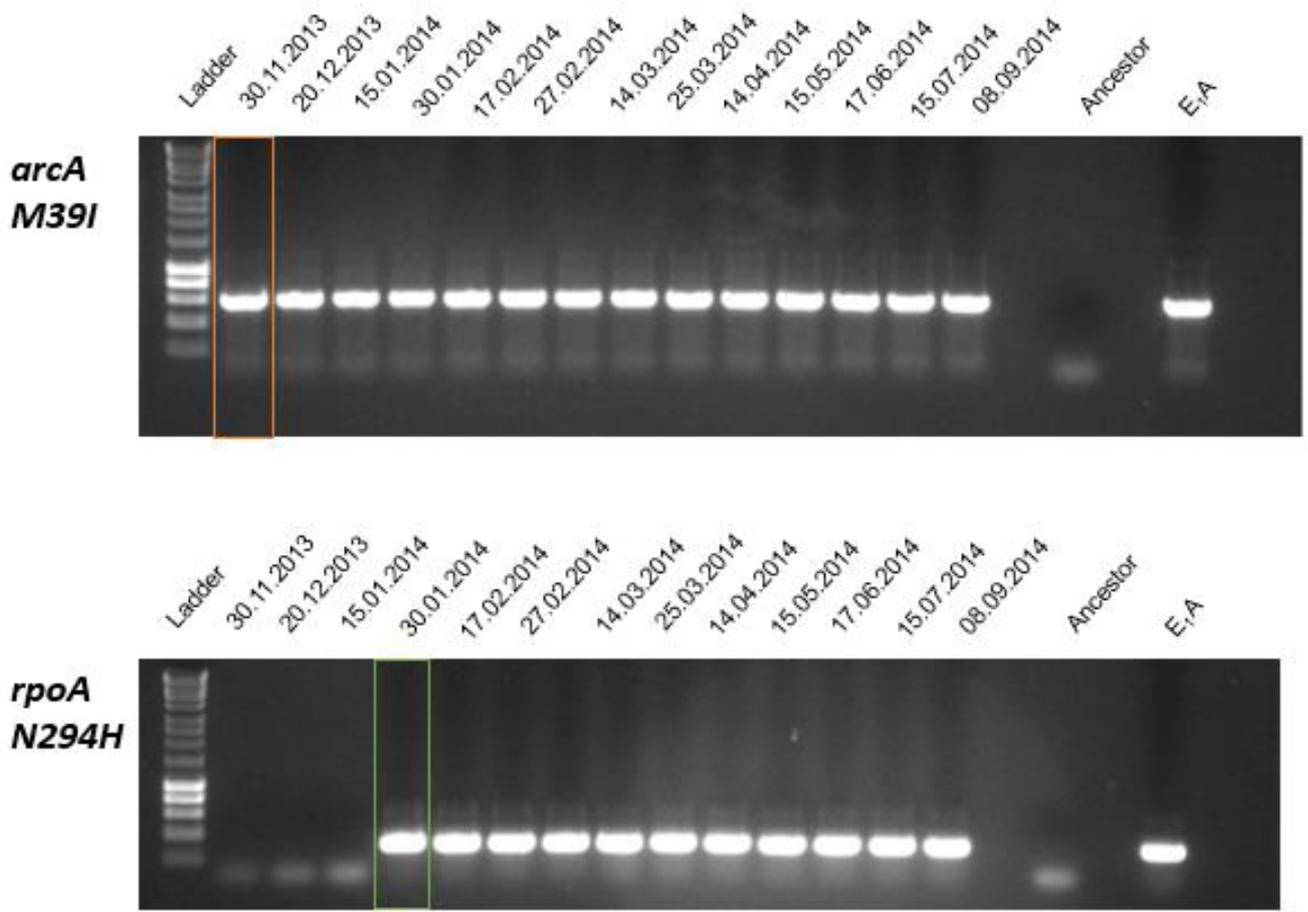
Figure 4.2: Agarose-gel electrophoresis of MAMA PCR check analysis of the fossil record of E_1A. The experiment was conducted as explained in Materials and Methods. Each lane corresponds to the date when cells were frozen during the evolution experiment. DNA from E_1A and the ancestor was used as the positive and negative control for the PCR reaction.
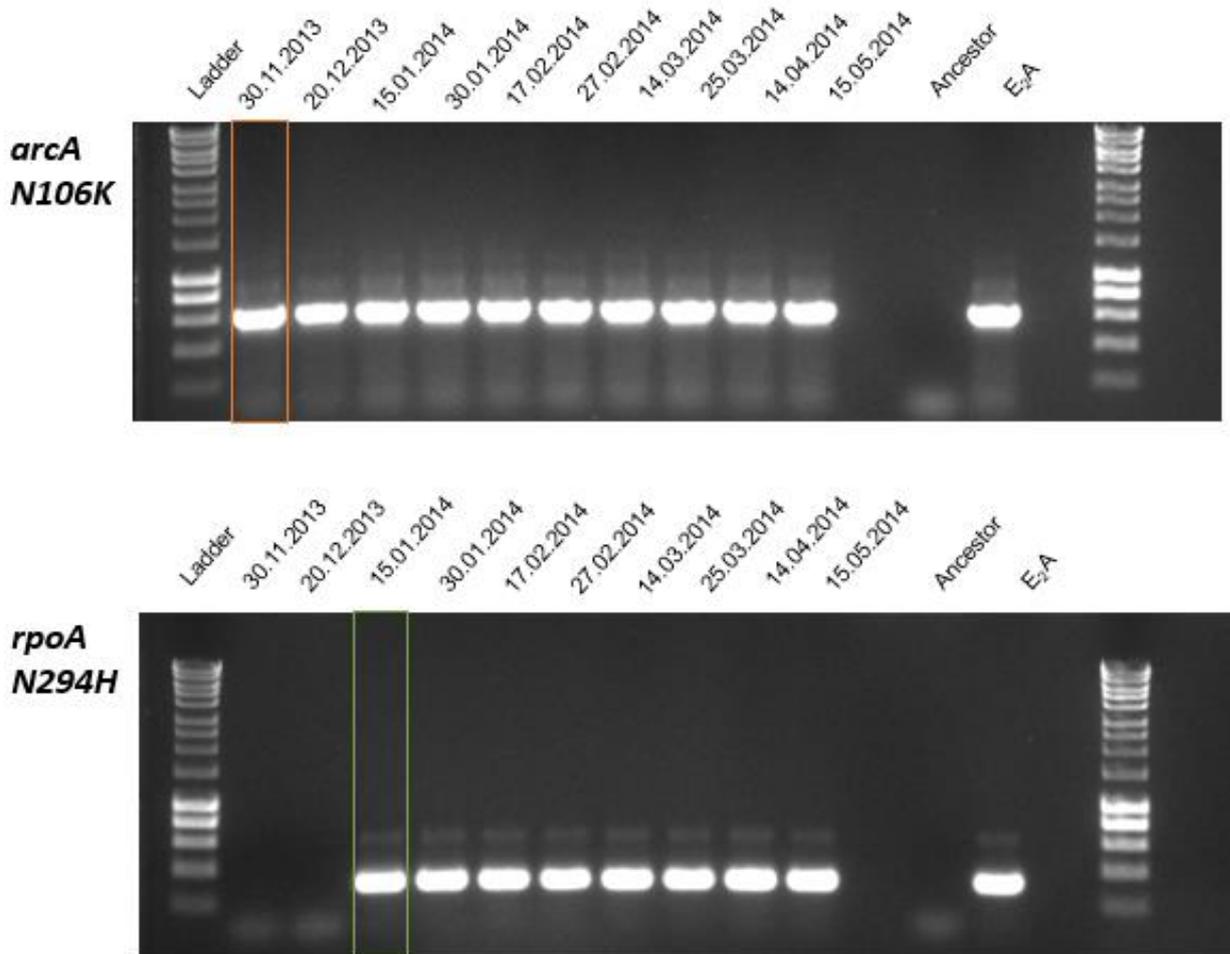
Figure 4.3: Agarose-gel electrophoresis of MAMA PCR check analysis of the fossil record of E₂A. The experiment was conducted as explained in Material and Methods. Each lane corresponds to the date when cells were frozen during the evolution experiment. DNA from E₂A and the ancestor was used as the positive and negative control for the PCR reaction.

Table 4.1 The order of emergence of mutations in the acid evolved strains according to the date of first detection.

Fossil Record	E₁A	E₂A	E₃A	E₄A	E₅A	E₆A
30.11.2013	*arcA*	*arcA*	*arcA*	*arcA*	*arcA*	*arcA*
20.12.2013				*IS5:cytR*		
15.01.2014	*IS5::cytR*	*rpoA*			*rpoA*	*rpoA*
30.01.2014	*rpoA*		*rpoA*			*IS5:;cytR*
17.02.2014						
27.02.2014			*IS5::cytR*			
14.03.2014		*cytR*		*cytR*	*cytR*	
25.03.2014			*cytR*			*cytR*
14.04.2014						
15.05.2014	*tnaC*			*sspA*		
17.06.2014				*cadC*		*fimE*
15.07.2014		*prfB*		*arcB*		
08.09.2014						

4.3.1 Timeline of emergence of the IS5: *cytR* insertion in E$_1$A. E$_3$A, E$_4$A and E$_6$A.

The previous chapter details the different mutations present in the genomes of the evolved strains. As described in section 3.11, chapter 1, the sequenced genomes were analysed using two platforms. First using the integrated genome viewer (IGV) and second using the Breseq pipeline. The Breseq analysis was done by Mathew Milner. Getting the genomes analysed using Breseq helps us determine if there were any cases of large scale genome reorganisations and transpositions.

The Breseq analysis showed that E$_1$A, E$_3$A and E$_6$A had a IS5 insertion 65 base pairs upstream of the *cytR* gene, while E$_4$A had the same IS5 insertion in the reverse orientation 58 base pairs upstream of the *cytR* gene. Since we already had determined the order of emergence of some of the mutations using MAMA, we also wanted to know the order of emergence of the IS5 insertion. This was important because some of the evolved strains also had point mutations in cytR. It was imperative that we knew when the insertions arose relative to the point mutations.

To do so the genome sequences of the four strains were analysed and primers were designed to amplify the region containing the insertion sequence. These primers were used to check the fossil record of the four strains. This work was jointly done by Mathew Milner and Dorothy Ling.

As seen in figure 4.3, the PCR check of the fossil record of E$_4$A showed that the insertion of the IS5 element upstream of *cytR* occurs on 20/12/2013 (fig. 4.4). The insertion of the IS5 element took place right after the *arcA* mutation emerges in E$_4$A (Table 4.1). In case of E$_6$A the IS insertion took place on 30/01/2014. the *arcA* mutation precedes the insertion event in E$_6$A as well (fig. 4.5).

The order of emergence of the IS insertions upstream of cytR have been listed in the Table 4.1 along with order of the other mutations. Interestingly except E$_1$A all the five evolved strains also have point mutations in *cytR*

The IS:cytR insertions were at a later time in the study and experimental analysis of the effects of this insertion element on the fitness of the evolved strain is in process and being conducted by Dorothy Ling.
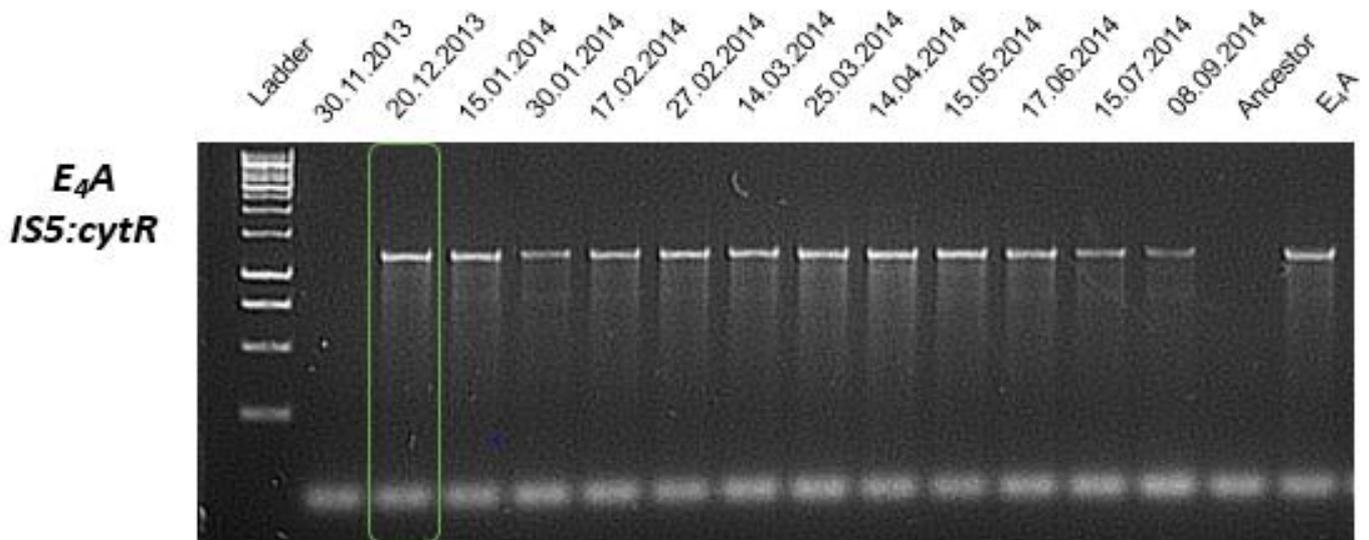
Figure 4.4: Agarose-gel electrophoresis of PCR check analysis of the fossil record of E_4A. Each lane corresponds to the date when cells were frozen during the evolution experiment. DNA from E_4A and the ancestor was used as the positive and negative control for the PCR reaction.
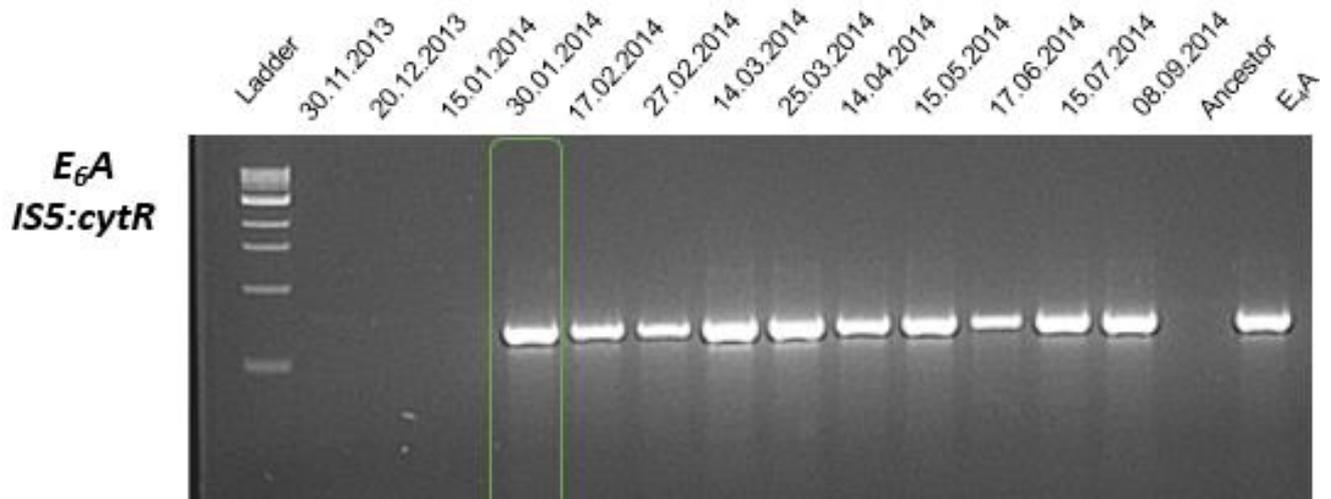
Figure 4.5: Agarose-gel electrophoresis of PCR check analysis of the fossil record of E_6A. Each lane corresponds to the date when cells were frozen during the evolution experiment. DNA from E_4A and the ancestor was used as the positive and negative control for the PCR reaction.

4.4 Mutations in *arcA* cause loss of function.

One of the key objectives of this lab-based evolution experiments was to create a link between the genotype and phenotype of the six acid evolved strains. To create the link, it was necessary that we understood the nature of the mutations in *arcA*, *rpoA* and *cytR*. For example, Goodarzi *et al.,* 2010 showed that clones isolated from populations of *E. coli* evolved to resist high ethanol concentrations had mutations in the gene *arcA*. These mutations caused a loss of function of *arcA,* which lead to derepression of pathways involved in ethanol breakdown and assimilation.

Taking cue from such experimental evidence, we wanted to test the nature of mutations in *arcA*. Since all the evolved strains had mutations in arcA and it was the first mutation to be selected for, it was logical to evaluate this gene first. To test whether the mutations in *arcA* caused a loss or gain of function, we deleted the chromosomal copy of *arcA* in these strains using P1 mediated transduction of the kanamycin marker from BW25113Δ*arcA*::kan of the KEIO library. (Materials and methods). The knockouts were PCR validated using primers flanking *arcA* (fig. 4.6)

We hypothesised that if the nature of the mutations resulted in gain of function, deleting *arcA* in the evolved strains would lead to loss of the improved fitness observed at pH 4.5. If the mutations were loss of function, inactivation of *arcA* would not alter the fitness profile at pH 4.5. Moreover, if the mutations were loss of function in nature, then we predicted that the deletion of *arcA* in the ancestor would cause an improvement in fitness under acid stress.

To test our hypothesis, we competed the (E_1A, E_2A, E_3A, E_4A, E_5A and E_6A)Δ*arcA* and ancestorΔ*arcA* against KHOO1 at pH 4.5. I calculated the fitness indices of the competition experiments at 24 hours and compared the results with the fitness indices of the evolved strains and ancestor at pH 4.5.

As can be seen from figure from figure 4.7, the Δ*arcA* version of the evolved strains showed no difference in fitness compared to their parent strains. But deleting *arcA* from the ancestor improves its fitness at pH 4.5. This was an interesting finding, as it was consistent with the hypothesis that deletion of *arcA* would cause an intermediate gain in fitness at pH 4.5.

ArcA acts a repressor of many operons in *E. coli*, many of which are involved in key aspects of cellular metabolism and general stress resistance. Inactivation of ArcA may lead to the derepression of such pathways which would confer a general stress resistance to the cells. This

maybe one of the reason why *arcA* mutants are recovered at such high frequencies in lab-based evolution experiments. Further discussion on this point will follow in the last section of this chapter.
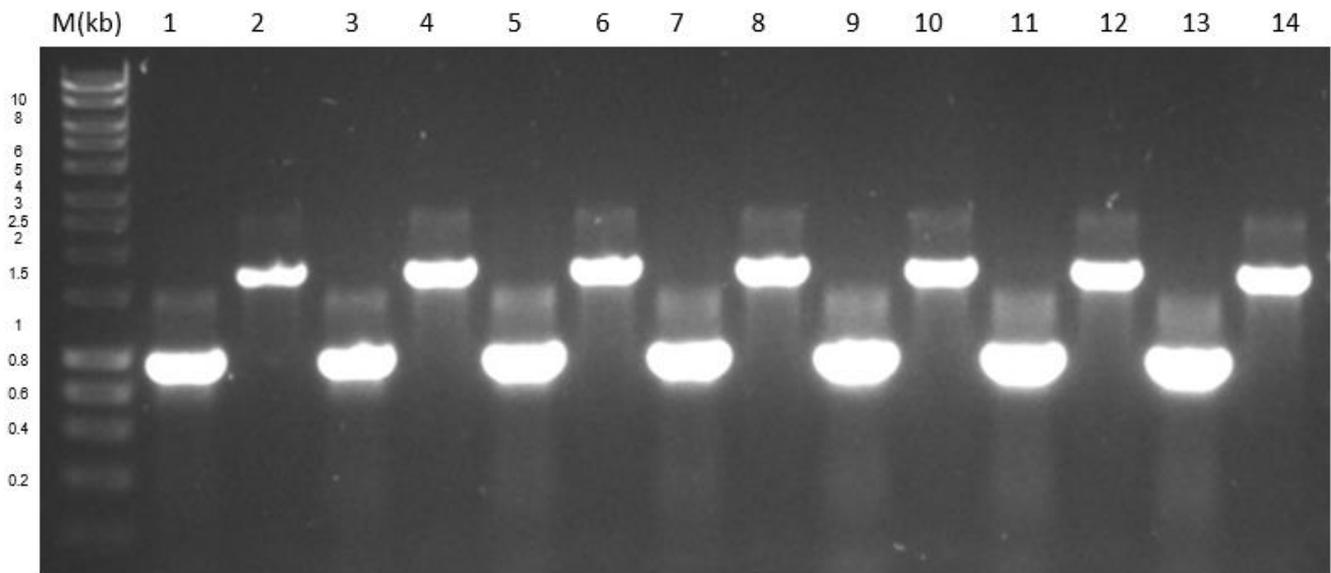
Figure 4.6: PCR validation of the arcA knockouts in the six evolved strains and ancestor. A kanamycin cassette replaces the arcA gene in the six evolved strains and the ancestor. Knockouts were created using generalised transduction and validated by PCR amplification using primers flanking the arcA gene. The expected PCR product size of WT arcA was 0.787kb and that of the kanamycin cassette was 1500bp. 1(E_1A), 3(E_2A), 5(E_3A), 7(E_4A), 9(E_5A), 11(E_6A) and 13(ancestor)showed the presence of a 0.787 kb fragment when DNA from the ancestor was used as the template. 2($E_1A\Delta arcA$), 4($E_2A\Delta arcA$), 6($E_3A\Delta arcA$), 8($E_4A\Delta arcA$), 10($E_5A\Delta arcA$), 12($E_6A\Delta arcA$) and 14(ancestor$\Delta arcA$), showed the presence of a 1.5kb band when DNA from the knockout strains was used a template.
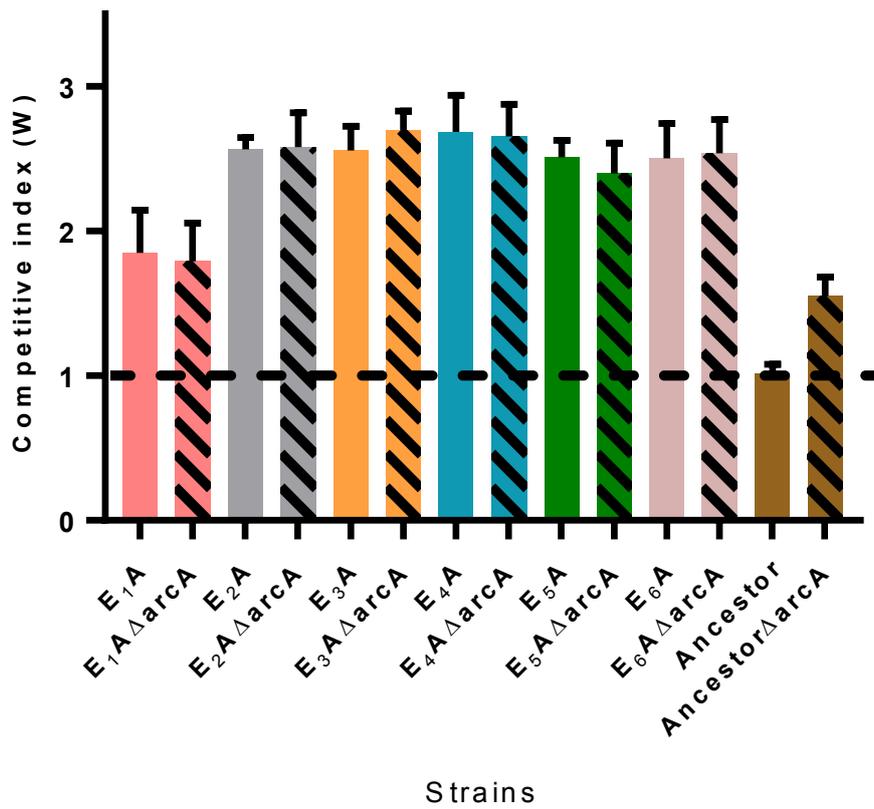
Figure 4.7: Competitive indices of the acid evolved strains, ancestor and arcA-knockouts of the evolved strains and ancestor at pH 4.5. Values are an average of three independent biological replicates and error bars are the standard error of the mean. The dashed represents a competitive index of 1 which denotes competitive neutrality.

4.5 Confirming the loss of function of *arcA*.

As stated in the previous section, one of the primary roles of *arcA* is to act as a repressor of different genes. One of the key genes it represses during anaerobic growth is *icdA* or isocitrate dehydrogenase. IcdA is a key enzyme component of the citric cycle and glyoxylate cycle. It enzymatically converts isocitrate to 2-oxoglutarate.

We could therefore use measurement of IcdA expression as method to conform the loss of function of ArcA. To do so we P1-transduced an *icdA* promoter-*lacZ* fusion (Materials and Methods), into the acid evolved strains, ancestor and *arcA* knockout of the ancestor. Beta-galactosidase activity of all the strains was recorded and a repression index was calculated for each strain. The experiment was done under aerobic and anaerobic conditions.

The ancestor is successfully able to repress IcdA expression under both the conditions. It shows a 4-fold repression in aerobic and 15-fold repression in anaerobic conditions approximately (Fig. 4.7). Our data are similar to published data on IcdA repression by ArcA (Park and Kiley, 2014). The evolved strains on the other hand were unable to repress IcdA expression under either of the conditions (Fig. 4.6). The repression profile of the evolved strains was similar to that of a Δ*arcA* strain.

Based on this data we confirmed and concluded that the mutations in *arcA* were indeed loss of function in nature.
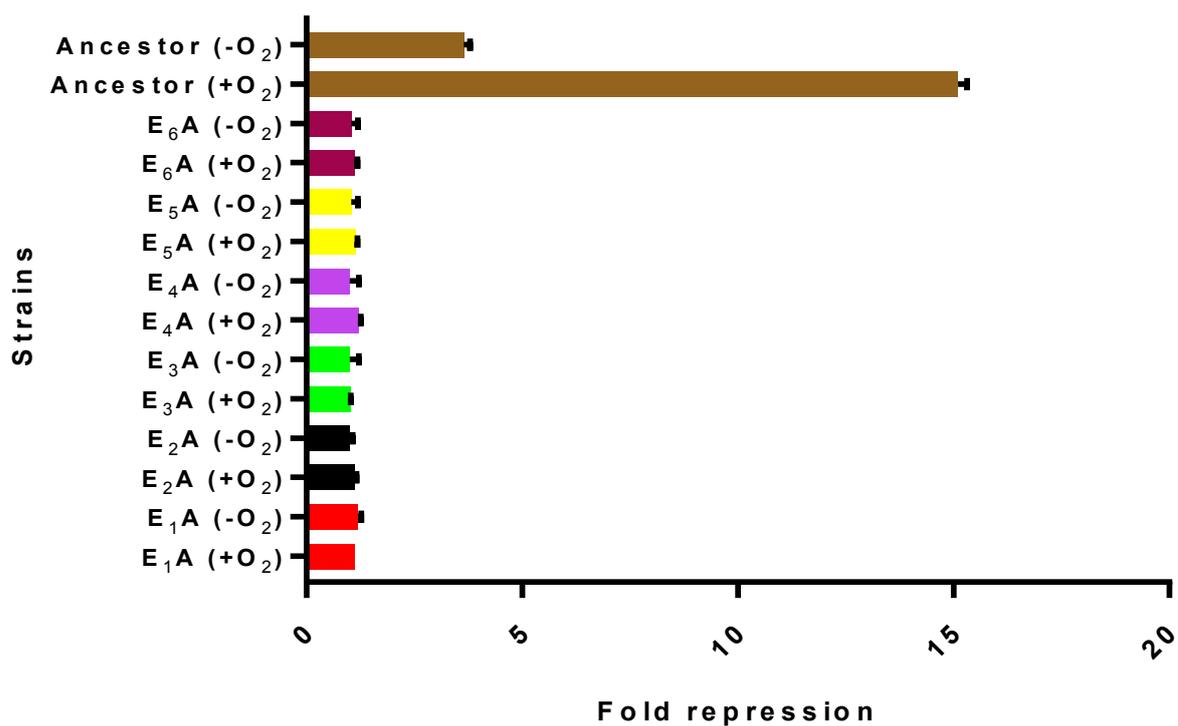
*Figure 4.8: Fold repression of P$_{icdA\text{-}lacZ}$ in the acid evolved strains and ancestor. Values are an average of three independent biological replicates and error bars are the standard error of the mean. (+O$_2$) and (-O$_2$) refer to aerobic and anaerobic conditions. The fold repression for each strain is calculated by diving the beta-galactosidase activity of **Δ**arcA strain (for e.g., 900 miller units for ancestor **Δ**arcA in -O$_2$) by the beta-galactosidase activity of a arcA$^+$ strain (for e.g. 57 miller units for the ancestor in -O$_2$).*

4.6 Evolved strains have traded off their ability to grow anaerobically.

ArcA regulates *E. coli's* ability to transition from aerobic to anaerobic growth. Since the evolved strains had lost ArcA function, we thought it would be interesting to analyse their phenotype under anaerobic conditions. To test this, we grew two of the evolved strains E_2A and E_4A and the ancestor anaerobically at pH 4.5 and pH 7.

Both E_2A and E_4A show poor growth under anaerobic conditions at pH 4.5 and pH 7 compared to the ancestor (fig. 4.9). Anaerobic stress impedes the growth of the ancestor at both the pH values, but the growth of the evolved strains is severely attenuated relative to the ancestor. Even though these strains are acid adapted to pH 4.5, 480 minutes of growth at pH 4.5 under anaerobiosis, yielded a final OD_{600} of approximately 0.3 ± 0.02 while the ancestor still managed to grow to a final OD_{600} 0.5 ± 0.03.

This supports the hypothesis that the acid evolved strains had traded-off their ability to grow anaerobically in exchange to the gain in fitness at pH 4.5. Interestingly, another study where *E. coli* populations were evolved under similar conditions to this study, didn't report loss of function of *arcA* and the trade-off to grow anaerobically (Harden *et al.,* 2015). This maybe because they evolved their strains in microtitre plates. Microtitre plates are known to experience anaerobiosis and higher oxygen depletion (Kram *et al.,* 2014). Since their cultures were evolved in microtitre plates, ArcA may be required for optimum growth and selected against in their evolution experiment. Whereas in the conditions this evolution experiment was done, loss of function of *arcA* was beneficial to the evolved strains. Such differences throw light on the fact that, conditions under which selection operates has considerable effect on the outcome of lab-based evolution experiments.
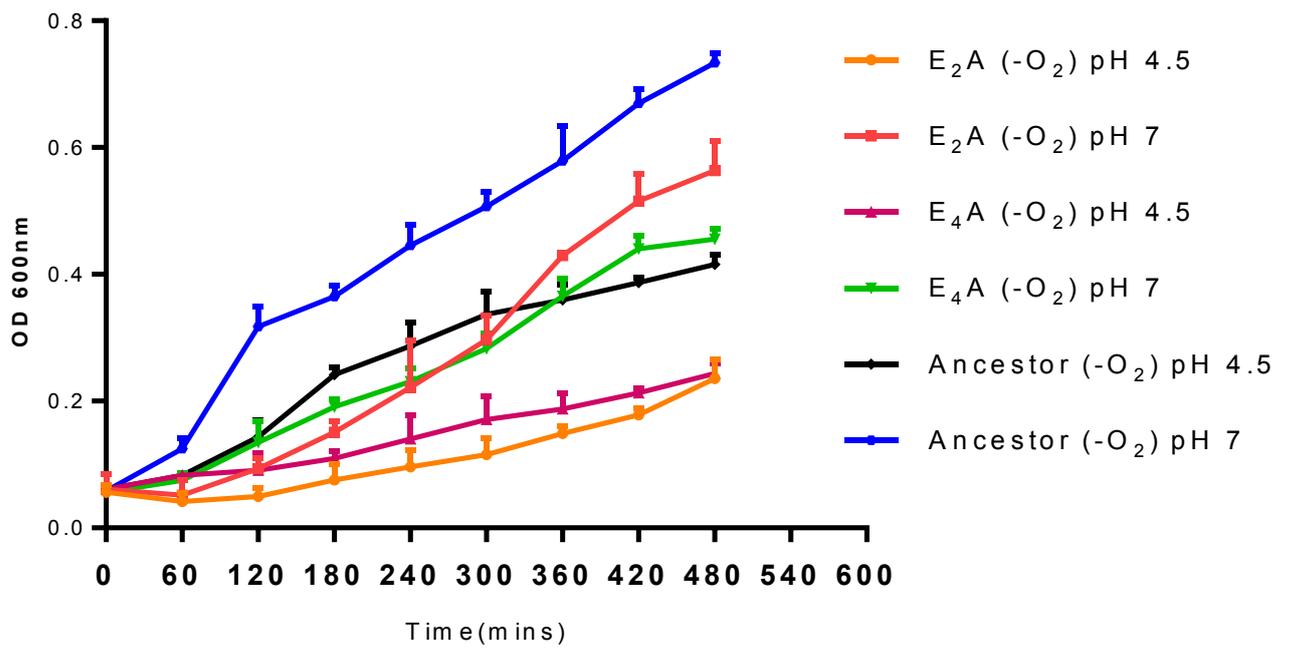
Figure 4.9: Growth curves of E2A, E4A and ancestor at pH 4.5 and pH 7 in (-O₂). Values are an average of three independent biological replicates and error bars are the standard error of the mean. (-O₂) denotes anaerobic growth conditions

4.7 *CytR* mutation are also loss of function in nature.

The results of the previous sections showed that the mutations in *arcA* caused loss of function of the gene in the six evolved strains. The loss of function of *arcA* was accompanied by the trade off to grow anaerobically. The key conclusion from section 4.5 was that loss of function of *arcA* improves the fitness of the ancestor at pH 4.5. One of the other genes which had mutations in five out of the six evolved strains was *cytR*. As described in the introduction of this chapter *cytR* is an important transcriptional regulator and mutations in *cytR* are known to alter global transcriptional profiles. Hence, it was imperative that we analysed the nature of mutations in cytR and understood the contribution of *cytR* in the overall phenotype of evolved strains.

We wanted to test whether the mutations in *cytR* were a gain of function or loss of function in nature. Just as in the case of *arcA*, we hypothesised that if the mutations were gain of function in nature, deleting *cytR* would cause a loss of fitness at pH 4.5 but if they were loss of function in nature, their fitness profiles would remain unaltered.

To test this, we deleted the chromosomal copy of *cytR* in all the evolved strains using P1 mediated transduction of the kanamycin marker from BW25113Δ*cytR*::kan of the KEIO library. (Materials and Methods). The knockouts were PCR verified using primers flanking the *cytR* gene. Upon confirming the knockouts, they were competed against KHOO1 at pH 4.5 and their fitness index was calculated. This work was jointly done by Assel Nurmagambetova, an MSc student under my supervision and me. These experiments were done before the reanalysis of the genomes using the Breseq pipeline which showed that some of the evolved strains had an IS insertion upstream of the cytR gene.

The results showed that the deletions didn't alter the fitness profiles of the evolved strains at pH 4.5, except for E_1A (Fig. 4.10). The results showed that the *cytR* mutations in E_2A, E_3A, E_4A, E_5A and E_6A were loss of function in nature. The experimental evidence showed that deleting *cytR* in E_1A improved its fitness at pH 4.5. E_1A consistently showed poorer cross stress resistance and fitness compared to the remaining five evolved strains. The key difference between them was that all the other five evolved strains had mutations in *cytR* except E_1A. Loss of function of *cytR* was seen to be beneficial for E_1A at pH 4.5 (fig. 4.10). Deletion of *cytR* in the ancestor didn't enhance its fitness at pH 4.5 when compared to the result which showed that deletion of *arcA* improved its fitness at pH 4.5 (fig. 4.10).
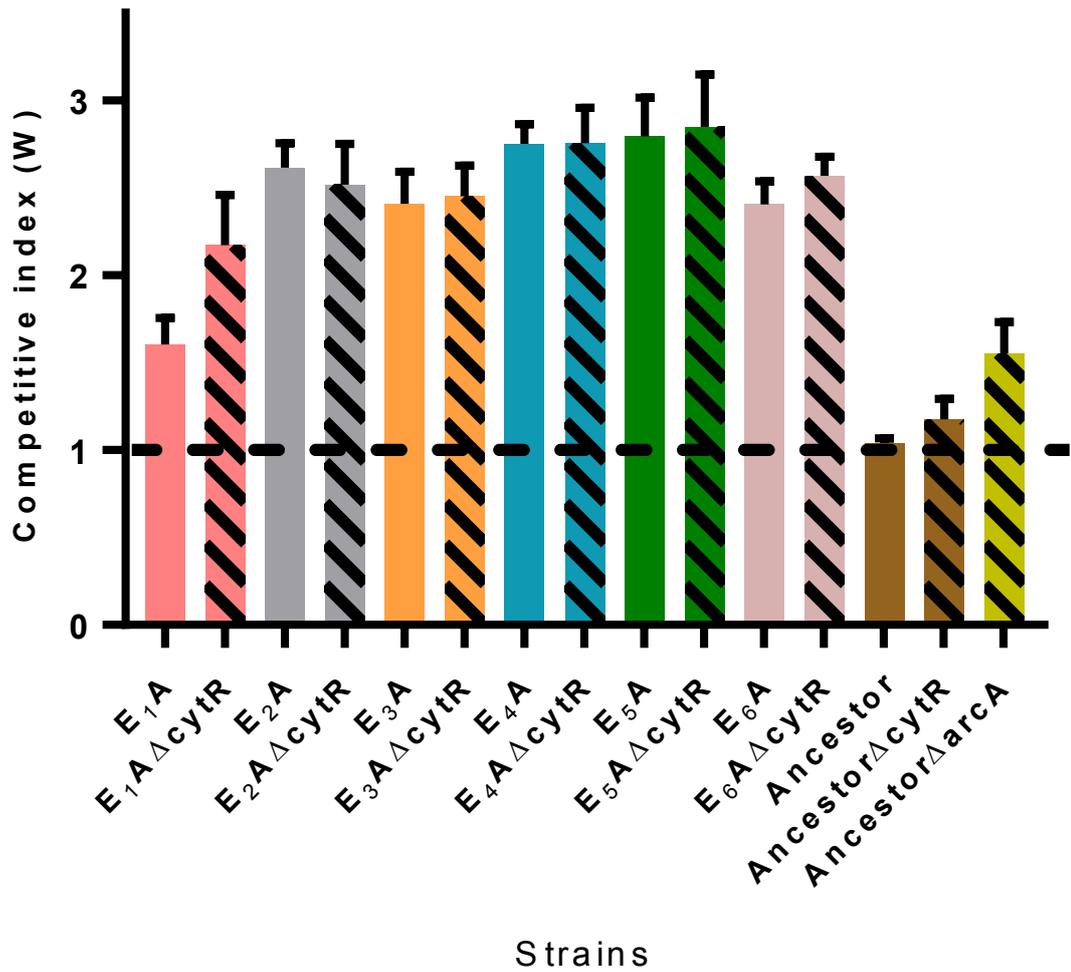
Figure 4.10: Competitive indices of the acid evolved strains, ancestor and cytR knockouts of the evolved strains and ancestor at pH 4.5. Values are an average of three independent biological replicates and error bars are the standard error of the mean. The dashed line represents a competitive index of 1 which denotes competitive neutrality.

4.8 Determining the fitness phenotype of a strain with an *arcA* only mutation.

The six evolved strains have mutations in *arcA* which cause a loss of function of the gene. As described in section 4.3, loss of function of *arcA* in the ancestor improved its fitness at pH 4.5. This result led us to conclude that it would be interesting to move some these *arcA* mutations individually to the ancestral background and measure the fitness of those strains. By doing so we could test the contribution of the *arcA* mutations to the overall phenotype of the evolved strains.

At that point we realised that we could screen strains which had only the *arcA* mutations in their genomes following the timeline or order of emergence of the mutations (section, 4.2). Since we knew which mutations arose at what point in the evolution experiment we could screen mutants from the fossil record. Moreover, since the *arcA* mutations arose first, it was possible to isolate an *arcA* only strain from the fossil record of any of the six evolved strains.

As described in section 4.2, to verify the validity of the MAMA method, we had isolated four strains from the fossil record of E_1A which were genome sequenced. These four strains all had the M39I mutation in their genomes. We randomly chose of one them for further analysis. The objective was to measure the fitness of this strain at pH 4.5 and compare its phenotype to the ancestor and the ancestor$\Delta arcA$ strain. This strain was named $F_1 arcA$

We hypothesised that if loss of function of *arcA* caused an improvement in fitness of the ancestor, then a strain with only the *arcA* mutation ($F_1 arcA$) would also display a similar fitness profile at pH 4.5.

As can be seen from figure 4.11, the $F_1 arcA$ strain shows an intermediate fitness phenotype at pH 4.5 compared to the fully evolved strain (comparison to E_4A, fig. 4.10). The phenotype is similar to ancestor$\Delta arcA$ (Fig. 4.10). This intermediate gain in fitness showed that *arcA* had a significant contribution in the overall phenotype of these six evolved strains.

Having established the contribution of ArcA to overall fitness of the evolved strains we moved our focus to *cytR*. We wanted to test the effect of loss of function of *cytR* on the fitness of the ancestor. As described in the previous section to do so we deleted *cytR* from the ancestor. We hypothesised that loss of function of *cytR* would improve the fitness of the ancestor at pH 4.5, this hypothesis followed on from the result which showed loss of function of *cytR* improved the fitness of E_1A. But the surprisingly, the loss of function of *cytR* showed minimal improvement in fitness of the ancestor at pH 4.5 (fig. 4.11)

The improvement in fitness was negligible when compared to the improvement in fitness observed for the loss of function of *arcA* (fig. 4.11)
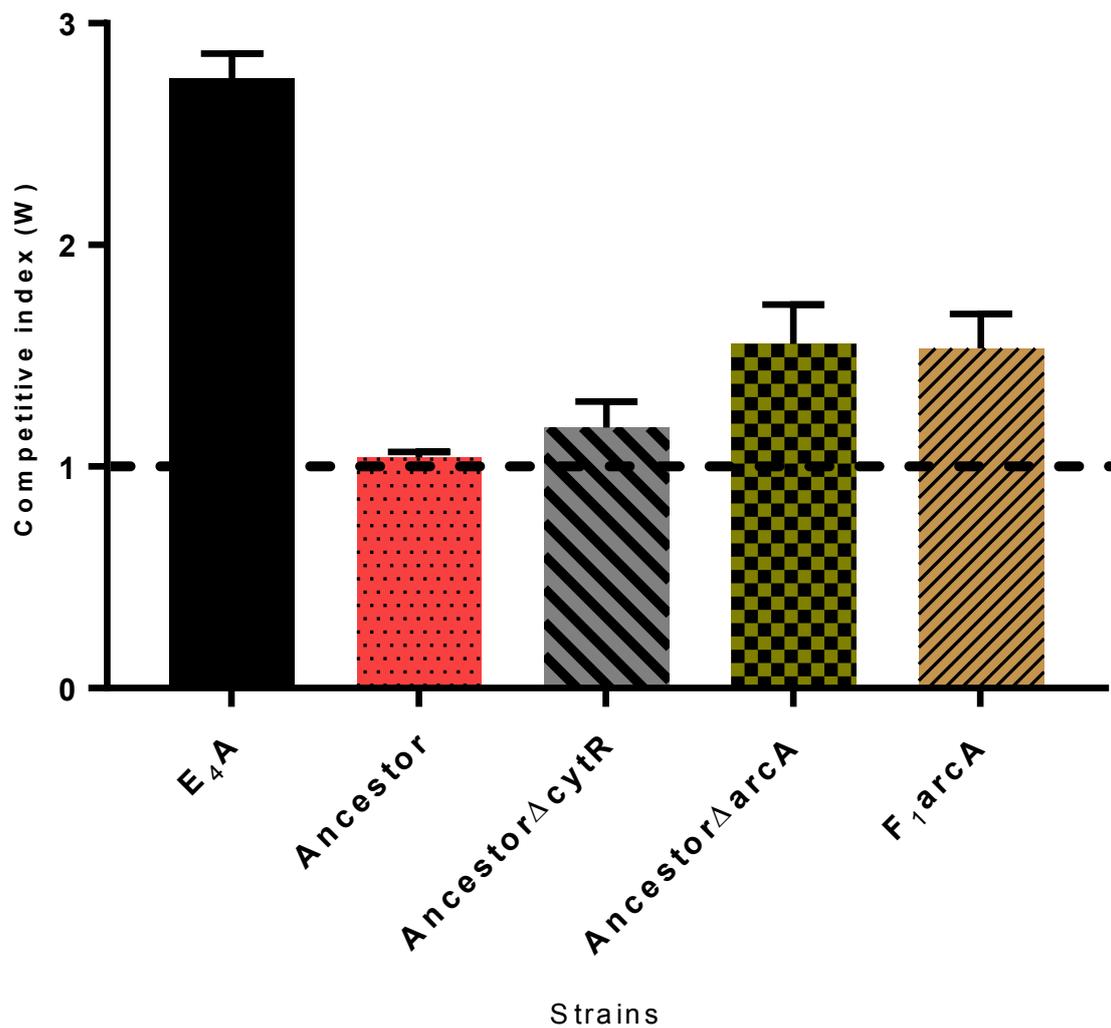
Figure 4.11: Competitive indices of the E4A, ancestor, ancestorΔcytR, ancestorΔarcA and F_1 arcA at pH 4.5. Values are an average of three independent biological replicates and error bars are the standard error of the mean. The dashed line represents a competitive index of 1 which denotes competitive neutrality. Statistical analysis was done using an unpaired Students t-test, by comparing the mean of each sample to 1.

4.9 The effect of *cytR* is contingent on the presence of the *arcA* mutation.

Fitness measurements on the strain isolated from the fossil record of E_1A with the arcA only mutation in its genome, showed an intermediate phenotype at pH 4.5. But no such increase was seen when *cytR* was deleted alone. This led us back to the result where deletion of *cytR* in E_1A improved its fitness at pH 4.5. We proposed a scenario where the loss of function of cytR could be contingent on the loss of function of *arcA*. It was possible that the mutations in *arcA* and *cytR* had an additive effect where the fitness benefit conferred by loss of function of *cytR* was contingent on *arcA*.

This kind of contingency is often reported in the literature on lab-based evolution. For example, Palsson *et al.,* 2012 showed that *E. coli* strains isolated from 11 independent populations, evolved at 42ºC for 200 generations, had different mutations in *rpoB* and *rpoH*. Genetic analysis of the mutations showed that, the *rpoH* mutation could confer the benefit in fitness only when the *rpoB* mutation was present in the background. Transcriptional analysis showed that the two mutations worked coherently to elevate expression levels of genes involved in heat stress response.

This made us hypothise that the fitness advantage conferred by *cytR* could be contingent on the *arcA* mutation. To test this hypothesis, we deleted the *cytR* gene in the F_1arcA strain and measured the fitness of the resultant strain at pH 4.5.

As can be seen from figure 4.12, the $F_1arcA\Delta cytR$ strain too shows an intermediate phenotype. The intermediate phenotype showed a slight improvement in fitness compared to the F_1arcA strain alone. The result did show that the fitness effect of loss of function of *cytR* is contingent on the presence of arcA in the background but it also showed that this combination was not successful in restoring the overall phenotype as seen in case of the fully evolved strains (fig. 4.12, comparison with E_4A). This was a case of diminishing-returns, a phenomenon which has been reported regularly in lab-based evolution studies. The results of this experiment indicated that, there was a missing link. This missing link could possibly be *rpoA*. Five out of the six evolved strains have mutations in *rpoA*. We proposed a reasonable hypothesis based on the experimental evidence that the *rpoA* mutation may have a significant contribution to fitness This fitness contribution maybe an effect of the *arcA* and *rpoA* combinations. As further chapters, will show this combination is very crucial to the fitness of the evolved strains.
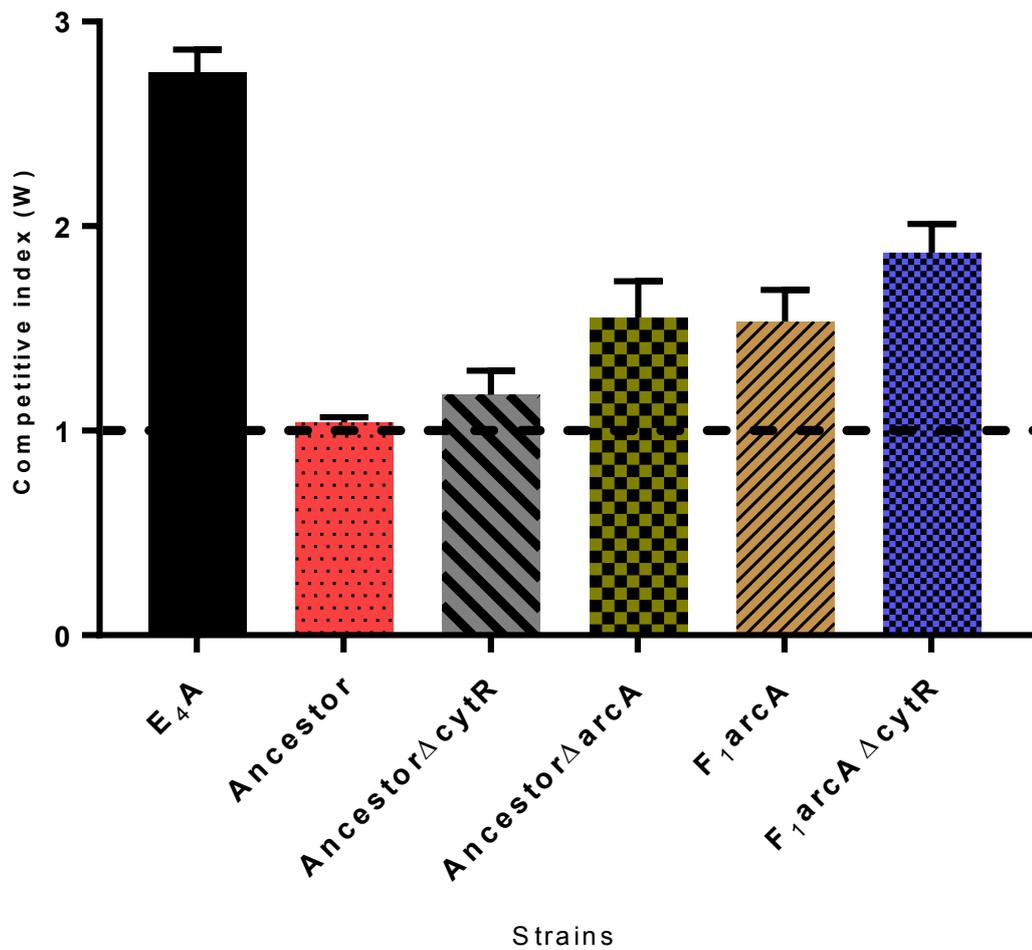
Figure 4.12: Competitive indices of the E4A, ancestor, ancestorΔcytR, ancestorΔarcA, F₁ arcA and F₁ arcAΔcytR at pH 4.5. Values are an average of three independent biological replicates and error bars are the standard error of the mean. The dashed line represents a competitive index of 1 which denotes competitive neutrality. Statistical analysis was done using an unpaired Students t-test, by comparing the mean of each sample to 1.

4.10 Discussion for chapter 4.

This chapter focused on understanding the genetic basis of adaptation of the six evolved strains. This chapter detailed the determination of the timeline of emergence of the mutations in each evolved strain based on the date of first detection using MAMA. It also described the genetic basis of adaptation of the evolved strains. The chapter concluded with a key result which showed the importance of *arcA* and its role in conferring the phenotype of the six evolved strains.

The whole genome sequencing analysis of the of the evolved strains showed different types mutations in their genomes. All the evolved strains had mutations in *arcA*. Five out of the six evolved strains except E$_4$A had an identical point mutation in *rpoA* and five out of the six evolved strains except E$_1$A had point mutations in *cytR*. The *rpoA* mutation could have spread through the populations through a cross contamination during the evolution experiment. To rule out the likelihood of a cross contamination a PCR based method (MAMA) was used to probe the fossil record of each strain. The MAMA results clearly showed that the *rpoA* mutations occurred after the emergence of the *arcA* mutations in all the evolved strains. The results clearly ruled out the possibility of a cross contamination in the experiment.

It is not uncommon for the same mutation to have been repeatedly selected in a lab-based evolution experiment. Another study showed that clones isolated from independently evolved populations of *E. coli* evolved under elevated NADPH stress had the identical mutations in the gene nuoF. This mutation independently altered the enzymatic properties of NuoF to metabolise NADPH and conferred a six -fold increase in the growth capacity of the evolved strains against NADPH stress (Sales *et al.,* 2010). Existence of adaptive parallelism in evolution experiments is not uncommon (Schneider et al., 2006). The *rpoA* mutation may confer a significant fitness advantage to the evolved strains under the conditions in which the evolution experiment was done. As further chapters, will illustrate this mutation has key combinatorial role with the loss of function of arcA to confer the fitness advantage at pH 4.5.

The MAMA results also showed that the *arcA* mutations were the first to emerge across all the evolved strains. The fixation of mutations in *arcA* across all the evolved strains early in the evolution experiment implicated that there was significant selective pressure on *arcA* and its key role to play in the overall phenotype of the evolved strains. A review of the literature on lab-based evolution showed that *arcA* mutations had been reported in other evolution experiments as well. Many of these studies report mutations in *arcA* but haven't characterised

the nature of these mutations. Our study showed that the mutations in *arcA* caused loss of function of the gene. Surprisingly another study, Saxer *et al.,* 2014 reported a mutation in *arcA* (D98A) which has been reported in this study too. The authors propose that the mutation caused a gain of function or conferred a novel property to *arcA* under the conditions of the evolution experiment in their study. But our study has clearly showed that this mutation caused a loss of function of *arcA*. Mutations in *arcA* can lead to a global reprogramming of the gene expression and metabolism (Attey *et al.,* 2014). *arcA* mutants showed increased expression of genes of the involved in the TCA cycle, energy and amino acid metabolism. A recent study showed that *arcA* mutants showed enhanced fitness compared to their wild type counterparts in mouse gut colonization studies (Sousa *et al.,* 2016). Another study suggested that mutations in *arcA* provided *E. coli* with a quicker way to adapt to environmental stresses otherwise which would require the accumulation of several beneficial mutations (Aguilar *et al.,* 2012). A catalogue of lab-based evolution experiments which had reported mutations in *arcA* is shown in table 4.1

Table 4.2: List of evolution experiments which showed mutations in *arcA*

Study reference	Strain	Evolution environment	Outcome
Lee *et al.,* 2010	*E. coli* MG1655	Adaptation to non-native carbon source	12 independent point mutations in *arcA*
Aguilar *et al.,* 2012	*E. coli* MG1655	Adaptation to elevated phosphate levels	2 point mutations in arcA and 1 insertion upstream of *arcA*
Palsson *et al.,* 2009	*E. coli* MG1655	Adaptation to 7%(v/v) ethanol	7 point mutations in *arcA*
Neilsen *et al.,* 2016	*E. coli* MG1655	Adaptation to levilinic acid	5 point mutations in *arcA*
Saxer *et al.,* 2014	*E. coli* RU1	Adaptation to LB (luria-bertuni) medium	11 point mutations in *arcA*
Saxer *et al.,* 2014	*E. coli* RU1	Adaptation to BHI (brain-heart infusion) medium	9 point mutations in *arcA*

Genetic analysis showed that the mutations in *arcA* caused a loss of function of the gene. The loss of function genotype was confirmed by the inability of the evolved strains to repress IcdA expression. The loss of function of *arcA* conferred at intermediate gain in fitness of the ancestor. ArcA is a global regulator of gene expression and primarily acts as a repressor of transcription. Many of the genes and pathways it controls are involved in regulating global cellular metabolism and stress resistance. Loss of function of *arcA* can lead to the derepression of these pathways leading to an altered metabolic state of the cell and elevated stress resistance. As the next chapter, will show the transcriptional landscape shows a global change which is driven by the contribution of ArcA. Another key result from this chapter showed that the loss of function of *arcA* lead to a trade-off to grow anaerobically.

One of the many advantages of knowing the order of emergence of the mutation was the ability to isolate a strain with only the *arcA* mutation from the fossil record of the evolved strains. This strain which was named F_1arcA showed an intermediate fitness phenotype at pH 4.5 like the *arcA* deletion strain of the ancestor. It verified the result that the point mutations in *arcA* caused loss of function of the gene and this strain was used as a proxy for a $\Delta arcA$ strain for further experiments in the study. Moreover, this strain showed an intermediate fitness phenotype. It was clear that *arcA* played a key role in modulating the fitness of these evolved strains at pH 4.5 but it also raised an important question.

Why would *arcA* be active when the evolution experiment was conducted under aerobic growth conditions? This possibly was due to cultures experiencing microaerophilic conditions when growing in a 25ml universal bottle. Mutations which caused a trade-off to grow anaerobically but increased fitness at pH 4.5 would have been selected in course of the evolution experiment. The *arcA* mutations arose early in the evolution experiment suggesting these mutations conferred an intermediate increase in fitness which was further enhanced by mutations arising later in the evolution experiment.

Genetic analysis of the mutations in *cytR* showed that the mutations caused a loss of function of the gene. The loss of function of *cytR* enhanced the fitness of E_1A which was the only evolved strain without a point mutation in *cytR*. This was contrary to what the RNA seq data suggested (next chapter), the transcriptional landscape of E_1A was very similar to that of remaining five evolved strains, suggesting that even E_1A had a mutation in *cytR*. But our genome sequencing analysis showed otherwise. To solve this issue, we decided to re run the

genome sequencing analysis on the genomes of the evolved strains using the Breseq platform (Chapter 3, section 4.3.1).

The Breseq analysis showed that E_1A had an IS5 insertion upstream of *cytR*, this insertion was also seen three other strains. It was surprising as some point mutations and an IS5 insertions occur in the *cytR* locus. Hence, the role of *cytR* needs further investigation and work is underway to understand the role of the IS5 insertion and we are in process of validating the results of the genome sequencing analysis.

The mutations in *cytR* too caused the loss of function of the gene but deletion of *cytR* in the ancestor showed no significant enhancement of its fitness. This meant that *cytR* autonomously couldn't contribute to the overall fitness of the ancestor. This was clearly showed by the results of the genetic analysis. The loss of function of *cytR* was contingent on the loss of function of *arcA*. But the combination of loss of function of *arcA* and *cytR* was insufficient to recover the overall fitness phenotype. This suggested that there might be an important role for *rpoA* to play in the overall fitness enhancements of the evolved strains at pH 4.5

The results of these chapter brought forth some important conclusions which were critical in driving the hypothesis development of the further experimental investigations. The results clearly showed the role of *arcA* in conferring the fitness advantage of the evolved strains at pH 4.5 and its contingent interaction with cytR. The next chapter will focus on understanding the link between these aspects of the evolved strains and their transcriptional signatures.

4.11 Summary of results for chapter 4

1) The order of emergence of mutations in the evolved strains showed a consistent pattern. The first mutation to be detected by MAMA were the *arcA* mutations followed by the rpoA mutation and the mutations in *cytR*.

2) The *rpoA* mutation evolved independently across all the evolved strains it was detected in and this mutation did not spread amongst them because of cross contamination.

3) Mutations in *arcA* and *cytR* caused loss of function of the two genes. The loss of function of arcA enhanced the fitness of the ancestor at pH 4.5

4) The fitness influence of the loss of function of *cytR* was contingent on the presence of *arcA* in the background.

Evolution of acid resistance to pH 4.5 in *E. coli* K12 MG1655

Chapter 5

Identifying the genetic basis of adaptation to pH 4.5- Part 2
Analysing the transcriptional landscape

Chapter 4 described the order of emergence of different mutations in the six evolved strains followed by the experimental evidence which highlights the nature of mutations in *arcA* and *cytR*. A key conclusion from the previous chapter was that the role of *arcA* is very central to the phenotype of the six acid evolved strains. Loss of function of *arcA* showed an improvement in the fitness of the ancestor and the fitness effect of loss of function of *cytR* was contingent on the loss of function of *arcA*. This raised some important questions on the effect about loss of function of *arcA* and *cytR* on the transcriptional landscape of the cell. What were the changes in gene regulation because of the mutations in *arcA* and *cytR*? How were these changes affecting the increase in the general stress resistance of the six evolved strains? What was the role of the *rpoA* mutation? This chapter will focus on understanding some of these questions. The chapter will detail results from RNA-sequencing analysis of five of the six evolved strains and experimental evidence to link the changes in the transcriptional landscape to their stress resistant phenotype.

5.1 RNA sequencing of the evolved strains.

As stated in the previous chapters, one of the objectives of the evolution experiment was to understand the genotype-phenotype relationship of the acid evolved strains. This relationship was the key to deciphering how *E. coli* climbs the fitness peak at pH 4.5. The six evolved strains had mutations in key global regulators which affect transcription of large gene networks. Hence, one of the ways to understand the genotype-phenotype relationship is by probing the transcriptional landscape. Transcription profiling would throw light on the genes and pathways which were significantly differentially expressed in the acid evolved strains.

To probe the transcriptional landscape of the evolved strains, we performed RNA sequencing analysis on E_1A. E_2A, E_3A, E_4A, E_5A and the ancestor. Due to various limiting factors, we could perform RNA-seq only on five evolved strains at that point but RNA seq on the sixth evolved strain is in process. The RNA isolation was done by me under the supervision of Dr Thippesh S. Post quantification and quality analysis RNA samples were shipped to the Centre for Genomic Research (CGR), University of Liverpool. RNA sequencing was carried out at CGR and analysis was done by Dr. John Herbert from the University of Liverpool.

RNA isolation and quantification was done as described in Materials and Methods. Briefly, one biological replicate of each evolved strain was grown in unbuffered LB medium pH 4.5. Samples were retrieved from these cultures when they attained the OD_{600} of 0.8 $\pm$0.03. This

was done in accordance to the first time point of the competition experiments and the optical density of the cells at that time point.

The general stress resistance pathways of *E. coli* are composed of several gene networks which work in tandem and co-regulate each other. The regulatory dynamics of such gene networks are complex and often difficult to disentangle. The six evolved strains have mutations in genes which are key components of these gene networks. To link the genetic changes in these evolved strains to their general stress resistant phenotype we wanted to study their transcriptional landscape. The objective of the RNA sequencing experiment was to systematically look for answers to the following questions. Since the evolved strains had mutations in the same target genes, were they also transcriptionally similar? Are there specific classes of genes which get upregulated and downregulated? Could the differential expression of these genes lead to a specific pathway or regulon responsible for the phenotype? If so could we establish a connection between such changes in the transcriptional profile and general stress resistance.

The RNA sequencing analysis of the evolved strains focused on understanding some of these questions and the results helped us in the hypothesis development of the next phase experimental investigations in the evolved strains. This section will outline and highlight a few of the important results from the RNA sequencing analysis. The RNA sequencing analysis was carried out using a single biological replicate. The absolute gene expression data was recorded in RPKMs (reads per kilobase per million reads). The absolute levels of expression of the ancestor (wild type) was compared to absolute levels of expression of the evolved strains to calculate the fold change in relative levels of expression of genes. The relative levels of expression were normalized and the log2 fold change of the relative levels of expression were calculated (Materials and Methods). The RNA sequencing results are based on the relative levels of expression and log2 fold changes calculated based on comparing the ancestor to the evolved strains. Positive enrichment cores denote a gene or a pathway was upregulated in the ancestor but downregulated in the evolved strains whereas a negative enrichment score denotes a gene or a pathway was upregulated in the ancestor (wild type) but down regulated in the evolved strains. I will present an overall picture of the transcriptional landscape of the five evolved strains and then address the link between their transcriptional response to pH 4.5 and the mutations in their genomes

5.2 Global analysis of gene expression in the evolved strains.

Differential gene expression of a total of 4191 genes was catalogued in the RNA sequencing experiment. The differential gene expression was calculated by comparing the absolute levels of expression (RPKMs) of the ancestor (wild type) with that of the evolved strains. On a global scale majority of the genes showed negligible change in relative gene expression compared to the ancestor (figure 5.1). Approximately 200 genes showed a greater than 2 fold downregulation in the evolved strains and around 150 genes showed a greater than 2 fold upregulation in the evolved strains.

Since the evolved strains had similar mutations in their genomes we predicted that their transcriptional response to pH 4.5 would be similar. To study the overall transcriptional landscape of the strains clustering analysis in the absolute levels of genes expression was carried out using an unweighted arithmetic mean mathematical model. A dendogram was generated to represent the overall gene clustering of the ancestor (wild type) and the evolved strains (fig. 5.2).

As can be seen from figure 5.2, the transcriptional response of the evolved strains to pH 4.5, is significantly different to that of the ancestor (wild type). The evolved strains are fairly similar to each other in their overall transcriptional landscape. E_1A and E_2A form a group whereas E_3A and E_5A form another with E4A having a more divergent transcriptional response.

5.3 KEGG Pathway analysis.

The clustering analysis showed the transcriptional response of evolved strains to pH 4.5 was fairly similar. This result could be correlated to the overall similarities in the fitness and cross protection profile of the evolved strains. Taking these factors into account we wanted to study the overall changes in the different regulatory pathways of the evolved strains. Since their transcriptional response to pH 4.5 was similar we predicted that the overall changes in relative expression of different pathways too would be fairly similar. To study the changes in expression of different pathways a KEGG pathway analysis (Materials and Methods) was performed on the all evolved strains. This analysis was done by creating 15 ranked lists of gene expression data based on the log2 fold change in gene expression. These ranked lists were then compared to the data available from the *E. coli* KEGG database. A total of 115 pathways were analysed and an enrichment score was generated using different false-discovery rates (FDR) (0%, 1%, 5% and 10%). A positive enrichment score meant the pathway was upregulated in

the ancestor (WT) but downregulated in the evolved strain and vice versa. I will present results 5% FDR (figure 5.3).

The KEGG pathways analysis showed a uniform upregulation of pathways which are involved in the catabolic metabolism of different substrate compounds. Three key pathways which were upregulated across all the evolved strains are

a) Oxidative phosphorylation

b) Citric acid cycle

c) Glyoxylate and dicarboxylate metabolism

KEGG pathway analysis also showed uniform downregulation of three pathways across all the evolved

a) Flagella biosynthesis

b) Chemotaxis

c) Two-component systems

5.3.1 Transcription Factor analysis (Regulon DB analysis)

The KEGG pathway analysis showed a global upregulation of pathways involved in cellular metabolism and energy transfer and downregulation of motility and sensory systems. But this global change didn't explain or direct us to a specific route which could explain the phenotype of the evolved strains. A higher level analysis was necessary to understand the link between these global changes in central metabolism and elevated stress resistance. Hence, a transcription factor analysis was performed on the RNA sequencing data of the evolved strains. The analysis was achieved by generating ranked lists of genes regulated by various transcription factors from the RNA sequencing data. The data from these lists was compared to the *E. coli* transcription factor database Regulon DB. The objective of this analysis was to understand the pathways regulated by specific transcription factors which showed maximal change in the evolved strains.

The transcription factor analysis showed that the pathways regulated by *arcA* and *cytR* were the most upregulated whereas the pathways regulated by *flhD* and *flhC* were the most downregulated across all the evolved strains. The evolved strains also showed upregulation of the pathways controlled by key transcription factors such as *IHFα*, *IHFß* and *CRP*.

Interestingly, many of the pathways controlled by known regulators of acid resistance systems such as *gadX, gadW* and *RcsB* were downregulated across all the evolved strains (Fig. 5.4)

5.4 Differential gene expression of the evolved strains

The last three sections gave an overview of the transcriptional landscape of the evolved strains and their differences relative to the ancestor (wild type). This section will focus on understanding some of the key aspects of the differential gene expression profile of the evolved strains. The idea behind this section was to highlight expression of some of the important genes which could be responsible for the changes in the overall transcriptional landscape of these strains and key differences in the transcriptional signatures of the evolved strains. Figure 5.4 and figure 5.5 catalogues the list of genes which showed maximum upregulation and downregulation in the evolved strains relative to the ancestor (wild type).

The gene which is most upregulated across all the evolved strains is *cdd* (fig. 5.6). Cdd or cytidine deaminase is one of the key components of the cytidine metabolism pathway. It is repressed by CytR and activated by CRP. This results correlates with the results of the previous chapter which showed that mutations in *cytR* caused a loss of function of the gene. *cytR* acts as a transcriptional repressor and loss of function of this gene would lead to the upregulation of genes which are repressed by *cytR*. This was clearly seen in case of the genes which constituted the deoxyribonuceloside metabolism pathway. This pathway comprises genes which are involved in pyrimidine transport and biosynthesis. This pathway is repressed by CytR and the genes of this pathway *deoA, deoB, deoC, deoD, nupG* are all upregulated across all the evolved strains. The KEGG pathway analysis showed an upregulation of this pathway (fig. 5.3). The whole change in the transcriptional regulation of this pathway can be explained based on the effect of the loss of function of *cytR*.

Another key observation from the differential expression data was the upregulation of the fimbriae genes (*fimC, fimD, fimG, fimH*) (fig. 5.6). These genes are involved in the biosynthesis and localisation of fimbriae and are positively regulated by *IHFα* and *IHFβ*. *cytR* is the transcriptional repressor of the two genes. Loss of function of *cytR* would lead to an upregulation of *IHFα* and *IHFβ* and in turn increase the expression of genes positively regulated by them. The transcription factor analysis showed an upregulation in the pathways regulated by *IHFα* and *IHFβ* and the fimbriae genes are components of this pathway. This showed a clear link between the transcriptional response of the strains and the genetic changes in them. The upregulation of genes involved in synthesis of fimbriae at pH 4.5 may explain the

enhanced biofilm forming capacities of evolved strains at this pH (results described in chapter 3).

A cluster of genes which were upregulated across all the evolved strains are *sdhA, sdhB, sdhC* (fug. 5.6). These three genes are protein components of the succinate dehydrogenase enzyme. They regulate the flux of succinate to the citric acid cycle and are repressed by *arcA*. The previous chapter detailed the analysis which showed that the mutations in *arcA* caused loss of function of the gene. Since *arcA* is directly involved in repression of these genes loss of function of *arcA* would have led to an increase in expression of these genes in the evolved strains. The KEGG pathway analysis showed that this pathway was uniformly upregulated across all the evolved strains.

The transcription factor analysis showed that the two regulons with highest differential expression across all the evolved strains were the *arcA* and *cytR* regulons. Both these genes have been repeatedly selected across all the evolved strains in the evolution experiment. The RNA sequencing analysis showed changes which correlated with the genetic changes in the evolved strains. The transcription factor analysis also showed that regulons regulated by *galR, galS* and *pdhR* were upregulated across all the evolved strains. These regulons are repressed by *arcA* and loss of function of *arcA* in the evolved strains would explain the upregulation of these pathways. Two regulons repressed by *cytR* which showed upregulation across all the evolved strains were the *CRP* and *Cra* regulons. *CRP* is also involved in regulating several pathways which contribute to the central metabolism of *E. coli* and the upregulation of *CRP* directly impacts the overexpression of some of these pathways which had been highlighted by the KEGG pathway analysis.

But there were some important differences amongst the genes which were upregulated across all the evolved strains. All the evolved strains showed increased expression of the lysine decarboxylases *cadA* and *cadB* except E$_4$A. The two genes are key components of the lysine cadaverine decarboxylase system and are their expression is regulated by the cadaverine antiporter *cadC* (Foster, 2004). E$_4$A has a nonsense mutation (chapter 3, section 3.10) in the regulator *cadC* which would led to a loss of function of the gene and in turn led to the downregulation of *cadA* and *cadB* in this strain.

The differential gene expression analysis showed that a uniform downregulation of the genes involved in the synthesis and assembly of flagella (fig 5.5). The downregulation of these genes was more prominent in case of E$_4$A. E$_4$A has a large genomic deletion which contained the

genes *flhD, flhC, motA, motB, cheW, cheA, tap and tar*. The deletion caused the complete loss of the class I flagella genes which are responsible for the systematic expression of the class II and class III flagella genes, many of which were significantly downregulated in E$_4$A. But the other strains also showed down regulation of the flagella genes and one of the genes which was uniformly downregulated across all the evolved strains was *fliA* (fig. 5.5). *fliA* is the flagellar sigma factor and is responsible for the transcription initiation of the genes which are involved in flagella synthesis and flagella. The downregulation of *fliA* could be responsible for the downregulation of the flagella genes in the other evolved strains.

The overall transcriptional landscape of these strains showed an upregulation of genes and pathways which regulate global metabolism and energy balance of the cell. This shift in the transcriptional landscape can be explained through the role of *arcA* and *cytR*. Both these genes are key regulators of global transcription and mutations in these genes would have led to an altered global transcriptional signature. A large part of the global change in these evolved strains is driven by *arcA* and the transcription factor analysis clearly showed that the regulon most upregulated across all the strains was the *arcA* regulon. This posed an important question regarding what could link the upregulation of the *arcA* regulon and the elevation in the general stress resistance of the evolved strains?

One of the key factors that *arcA* regulates is *rpoS*. RpoS or σ^{38} is the stationary phase sigma factor and controls the expression of a host of genes during stationary phase growth. It also regulates the expression of genes which alter the global stress resistance of the cell. Elevated levels of rpoS is known to increase the overall stress resistance of the cell. The evolved strains showed enhanced cross protection to a variety of stresses. Moreover, the evolved strains from the previous evolution experiment only showed resistance to extreme acid shock (pH 2.5) but no cross protective abilities. The phenotype of the extreme acid resistant strains was *rpoS* independent and did not show any general stress resistance whereas the strains evolved at pH 4.5 showed an elevation in the general stress resistance. Taking all these factors into account it was important that we investigated the role of RpoS in the evolved strains.
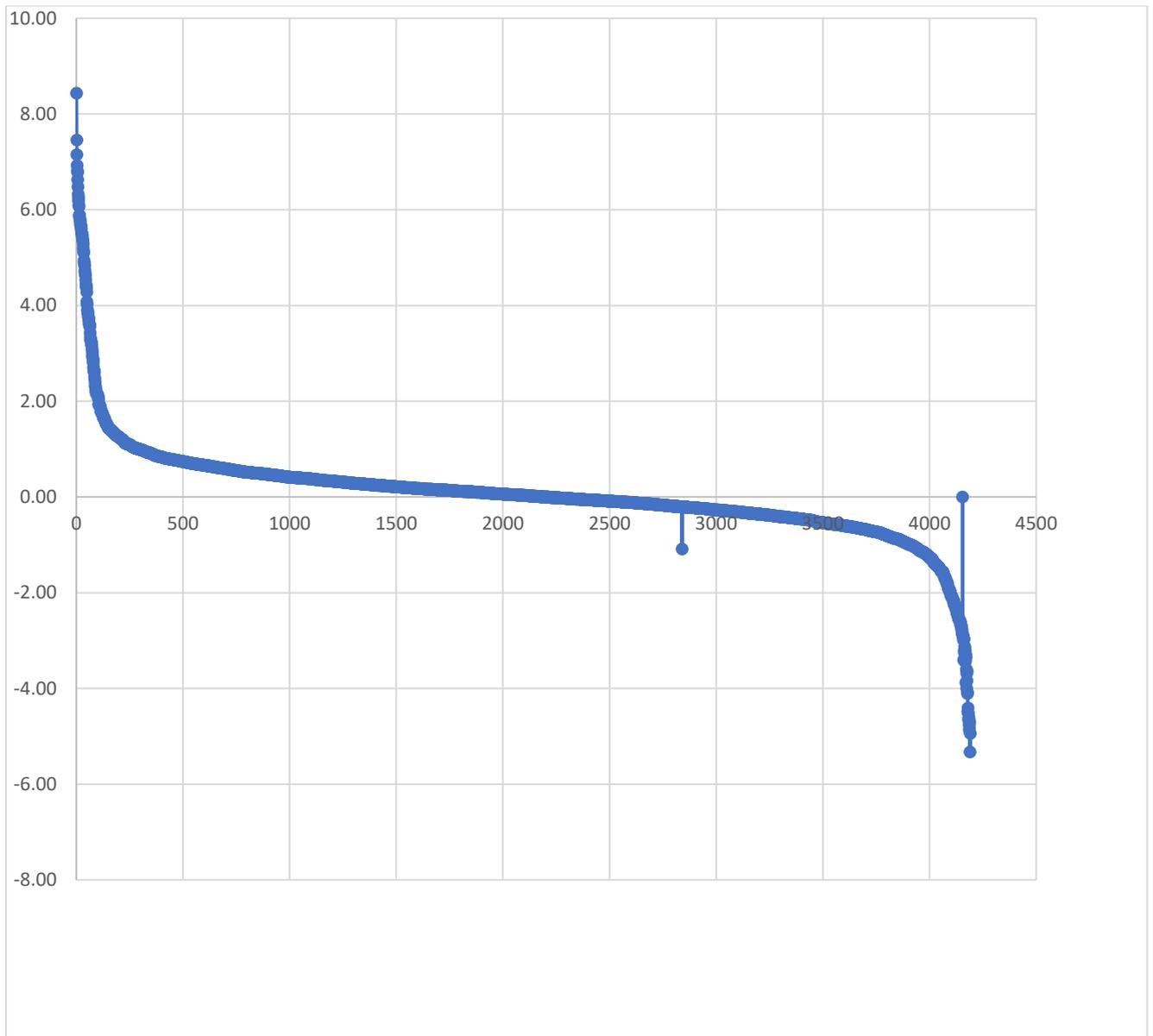
Figure 5.1: Fold change in relative gene expression versus the total number of genes analysed in the RNA sequencing experiment. A positive enrichment score denotes downregulated in the relative levels of expression of the genes in the evolved strains whereas a negative enrichment score denotes upregulation in the relative levels of gene expression in the evolved strains.
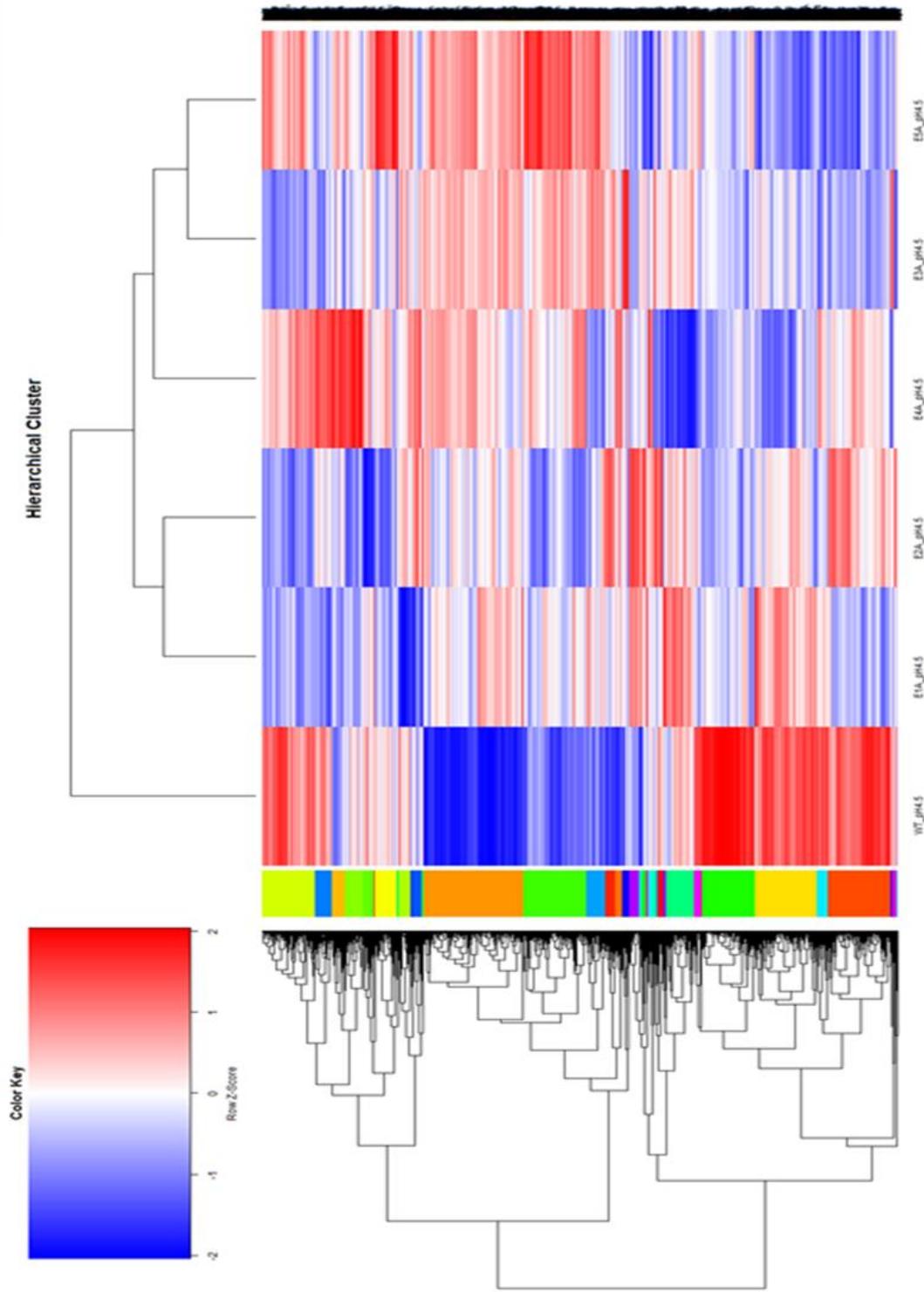
Figure 5.2: Hierarchical clustering of gene expression of the evolved strains and ancestor (WT). Clustering of gene expression in the dendogram is based on the absolute levels of expression in the ancestor and the evolved strains.

Kegg pathway	WT_vs_E1A	WT_vs_E2A	WT_vs_E3A	WT_vs_E4A	WT_vs_E5A
Flagellar assembly	2.62	2.70	2.78	2.78	2.91
Sulfur metabolism	2.18	2.30	2.29	2.38	2.33
Bacterial chemotaxis	2.01	2.10	1.98	2.16	1.93
Fatty acid biosynthesis	0.00	0.00	0.00	0.00	0.00
Mismatch repair	0.00	0.00	0.00	0.00	0.00
Biosynthesis of secondary me	0.00	0.00	0.00	0.00	-1.72
Pyruvate metabolism	0.00	0.00	0.00	-1.73	-2.12
Tryptophan metabolism	0.00	0.00	0.00	-1.67	0.00
Two-component system	0.00	0.00	0.00	0.00	0.00
Glycerolipid metabolism	0.00	0.00	0.00	-1.62	0.00
Lysine degradation	0.00	0.00	0.00	-1.65	-1.60
Fatty acid degradation	0.00	0.00	0.00	-2.17	-2.32
Histidine metabolism	0.00	0.00	-1.84	-1.72	-1.81
Geraniol degradation	0.00	0.00	0.00	0.00	-1.62
Ribosome	0.00	0.00	0.00	-1.67	-1.62
Ascorbate and aldarate metal	0.00	0.00	0.00	-1.78	-1.82
ABC transporters	0.00	0.00	0.00	0.00	2.17
Fructose and mannose metab	0.00	0.00	0.00	0.00	-1.65
Metabolic pathways	0.00	0.00	0.00	0.00	0.00
C5-Branched dibasic acid met	0.00	0.00	0.00	0.00	0.00
Benzoate degradation	0.00	0.00	0.00	0.00	0.00
Pyrimidine metabolism	0.00	0.00	-1.88	-2.08	-1.89
Aminoacyl-tRNA biosynthesis	0.00	-1.72	0.00	-1.81	-1.75
Fatty acid metabolism	0.00	0.00	0.00	0.00	-1.89
Monobactam biosynthesis	0.00	0.00	0.00	-1.89	-1.71
Pentose and glucuronate inte	0.00	0.00	0.00	-1.75	0.00
Glycine, serine and threonine	0.00	0.00	0.00	0.00	-1.59
Phosphotransferase system (	0.00	0.00	0.00	0.00	0.00
Biosynthesis of antibiotics	0.00	0.00	0.00	0.00	0.00
Valine, leucine and isoleucin	-1.73	-1.80	0.00	-1.94	-1.68
Propanoate metabolism	-1.73	0.00	0.00	-1.87	-1.86
Carbon metabolism	-1.75	0.00	0.00	0.00	0.00
Phenylalanine metabolism	-1.80	0.00	0.00	0.00	0.00
Galactose metabolism	-1.80	0.00	0.00	-1.70	-1.84
Amino sugar and nucleotide s	-1.80	-1.70	0.00	-1.76	-1.89
Polyketide sugar unit biosynt	-1.89	-1.82	-1.95	-1.88	-2.21
Glyoxylate and dicarboxylate	-1.90	-2.01	-1.93	-2.00	-1.91
Porphyrin and chlorophyll me	-1.91	-2.03	-1.84	-1.56	-1.72
Oxidative phosphorylation	-1.96	-1.83	-1.81	-2.14	-2.22
Microbial metabolism in dive	-2.01	-1.98	-2.03	-1.88	-2.23
Citrate cycle (TCA cycle)	-2.21	-2.16	-2.31	-2.36	-2.44

Figure 5.3: KEGG pathway analysis of the evolved strains at 5% false discovery rate. Positive enrichment score denotes down regulation in the evolved strains and negative enrichment score denotes upregulation in the evolved strain.

T. Factor	WT_vs_E1A	WT_vs_E2A	WT_vs_E3A	WT_vs_E4A	WT_vs_E5A
FLHD	2.64	2.85	2.76	2.89	2.73
FLHC	2.64	2.85	2.75	2.88	2.74
CSGD	2.14	2.28	2.40	2.37	2.42
GADW	2.03	1.78	2.20	2.13	2.32
APPY	1.99	1.91	2.06	2.07	2.12
ECPR	1.78	1.77	1.85	1.93	1.81
GADX	1.90	1.67	2.11	2.30	2.27
SUTR	1.64	1.70	1.77	1.73	1.72
FLIZ	1.89	2.11	2.10	2.14	2.21
TORR	1.91	1.83	2.08	2.11	2.18
RCSB	1.65	1.48	1.86	2.33	2.12
TDCA	1.86	1.86	1.89	1.94	1.50
TDCR	1.86	1.86	1.88	1.94	1.77
LSRR	1.86	1.88	1.81	2.02	1.75
YDEO	1.72	1.66	1.78	1.66	1.82
ISCR	1.80	1.53	1.82	2.08	1.75
GADE	1.56	1.82	1.64	1.65	1.91
CRP	-1.87	-1.99	-2.03	-2.11	-2.20
NANR	-2.07	-2.89	-2.19	-2.24	-2.24
SRLM	-1.69	-1.87	-1.66	-1.89	-1.74
SRLR	-1.70	-1.89	-1.66	-1.89	-1.74
GALR	-1.79	-1.76	-1.84	-1.90	-1.89
GALS	-1.80	-1.77	-1.84	-1.92	-1.92
DEOR	-1.90	-1.95	-1.93	-1.86	-1.92
GLNG	-1.98	-2.13	-1.84	-2.11	-1.91
GLCC	-2.00	-2.03	-2.01	-1.97	-2.02
CRA	-2.07	-1.90	-2.03	-2.02	-2.17
IHFA	-2.01	-2.17	-2.16	-1.95	-2.24
IHFB	-2.02	-2.18	-2.16	-1.96	-2.24
PDHR	-2.29	-2.25	-2.29	-1.85	-2.26
CYTR	-2.20	-2.28	-2.29	-2.20	-2.30
ARCA	-2.42	-2.52	-2.44	-2.50	-2.55

Figure 5.4: Transcription factor or Regulon analysis of the evolved strains. Positive enrichment score denotes the downregulation of the regulon in the evolved strains and negative enrichment score denotes then upregulation of the regulon in the evolved strains.

	WT_vs_E1A	WT_vs_E2A	WT_vs_E3A	WT_vs_E4A	WT_vs_E5A
fliA	7.81	9.23	7.89	9.08	8.16
flgB	5.73	9.12	5.94	10.42	6.07
flgC	5.69	9.17	5.80	9.29	5.81
fliZ	6.77	6.75	6.93	7.46	6.71
flgE	5.42	8.24	5.74	9.14	5.48
flgD	5.59	8.33	5.58	8.86	5.54
fliE	6.05	7.53	6.55	6.68	6.32
flgF	5.55	7.93	5.47	8.37	5.08
flgG	5.66	7.74	5.34	7.77	5.11
fliL	5.19	7.48	5.23	8.15	5.22
fliC	6.30	5.91	6.20	6.15	6.40
fliD	5.84	6.38	6.18	6.00	6.09
fliF	5.17	7.13	5.32	7.43	5.30
csgC	5.69	5.70	6.59	5.61	5.84
flgH	4.94	6.72	5.06	7.56	5.15
tar	5.89	5.88	5.00	6.92	5.52
fliO	4.73	6.99	4.60	7.26	5.43
fliM	4.59	6.89	5.22	7.48	4.74
flgI	4.95	6.30	5.37	7.22	4.83
yhjH	5.68	5.45	5.83	5.93	5.58
flgA	4.67	6.73	5.25	7.02	4.69
fliN	4.68	6.84	4.84	7.25	4.71
flhB	5.14	6.02	5.45	6.04	5.43
fliG	4.65	6.72	4.60	7.15	4.80
fliS	5.26	5.78	5.39	5.81	5.31
fliH	4.68	6.19	4.80	7.08	4.75
tap	5.09	5.55	5.16	6.31	5.34
flgJ	4.81	6.61	4.80	6.15	4.95
fliK	4.65	6.35	4.76	7.11	4.20
motA	5.02	4.90	5.24	6.39	5.29

Figure 5.5: Differential expression of genes at pH 4.5. Genes differentially downregulated in the evolved strains relative to the wild type. Values represent relative log2 fold change in gene expression. A positive enrichment score meant the gene was upregulated in the ancestor (WT) but downregulated in the evolved strain.

	WT_vs_E1A	WT_vs_E2A	WT_vs_E3A	WT_vs_E4A	WT_vs_E5A
cyoE	-2.95	-1.90	-3.09	-3.28	-3.23
sdhB	-2.81	-2.57	-3.08	-3.03	-2.98
sdhA	-2.81	-2.51	-3.10	-3.05	-3.13
glcE	-3.24	-3.31	-2.83	-1.91	-3.63
nanA	-3.20	-2.03	-3.68	-4.11	-4.01
mhpR	-2.69	-1.95	-3.66	-2.74	-3.80
argT	-2.92	-2.56	-3.56	-2.43	-3.36
sdhC	-3.30	-2.64	-3.55	-2.80	-3.80
deoA	-3.22	-3.20	-3.44	-2.32	-4.10
glcD	-3.31	-3.51	-3.13	-2.19	-3.68
sodA	-1.93	-3.41	-3.97	-3.26	-3.05
putA	-3.59	-2.72	-3.77	-3.26	-3.86
lldP	-3.72	-1.76	-3.91	-3.25	-3.47
tsx	-3.77	-3.01	-3.60	-3.00	-3.73
cyoD	-3.58	-2.40	-3.61	-3.81	-3.13
tsgA	-4.12	-3.48	-3.43	-3.19	-5.16
nanC	-3.36	0.87	-4.45	-5.62	-4.14
ydcI	-3.82	-2.54	-4.57	-2.89	-4.18
phoH	-3.57	-3.27	-4.71	-3.02	-3.84
cyoC	-4.00	-2.88	-4.10	-3.98	-5.04
cadB	-4.47	-4.54	-4.78	-0.29	-5.10
cadA	-4.44	-4.72	-4.86	-0.16	-4.06
cyoB	-4.48	-3.43	-4.60	-4.01	-3.87
fimI	-3.28	-3.91	-4.96	-4.87	-3.51
fimC	-3.58	-4.48	-5.21	-4.80	-4.34
cyoA	-4.93	-3.88	-4.96	-3.94	-4.33
fimH	-2.99	-5.26	-5.05	-4.70	-4.57
udp	-4.60	-4.58	-4.79	-4.03	-5.30
deoC	-4.75	-4.44	-5.06	-3.65	-5.22
nupG	-4.78	-4.10	-4.99	-4.39	-4.75
fimG	-3.20	-5.18	-5.43	-5.16	-4.82
fimF	-3.19	-5.36	-5.35	-5.14	-5.16
ycdZ	-4.85	-4.88	-4.95	-4.20	-5.51
fimA	-3.77	-3.68	-5.73	-5.68	-4.67
fimD	-3.62	-6.22	-5.51	-5.42	-5.87
cdd	-5.30	-4.71	-5.28	-4.74	-4.68

Figure 5.6: Differential expression of genes at pH 4.5. Genes differentially upregulated in the evolved strains relative to the wild type. Values represent relative log2 fold change in gene expression. A negative enrichment score meant the gene was downregulated in the ancestor (WT) but upregulated in the evolved strain.

5.5 Measuring the RpoS or σ³⁸ activity of the evolved strains.

The previous few sections detailed the transcriptional landscape of the evolved strains. A key inference from the RNA sequencing analysis was the upregulation of the *arcA* regulon and the global change in the transcriptional profile of the evolved strains. To understand to genotype-phenotype relationship of the evolved strains it was important to establish a link between the change in the transcriptional landscape of the strains and their phenotype. As described in section 4.4, one of the obvious targets to investigate was RpoS. *arcA* is the transcriptional repressor of *rpoS* and loss of function of *arcA* could directly impact the regulation of *rpoS*. *arcA* regulates *rpoS* through different functional mechanisms which have been briefly described in the previous chapter. *rpoS* is regulated by *arcA* at the level of transcription, post-transcription, translation and post translation. But the RNA sequencing data showed no significant change in the expression of RpoS. This could be possibly due to fact that a large part of the regulation of RpoS by ArcA involved mechanisms which come into effect post-transcriptionally. Hence, any method to be used to consider the role of RpoS in these strains would need to account for these factors. We proposed the use of a promoter probe which would specifically respond to RpoS and allow us to measure the amount of active RpoS in the cell in contrast to just measuring the level of RpoS through western blots.

We sourced a fluorometric promoter probe from Dr Regine Hengge's laboratory at the Institute of Biology, Humboldt University. The plasmid probe had been constructed as described in Carrier and Keasling, 1992. The probe consisted of a synthetically designed *rpoS* responsive promoter fused with a green fluorescent protein reporter. We could use the absolute fluorescence read out as an indirect measure to analyse RpoS activity. The plasmid is denoted pSyn-GFP 21 or p21 (Typas et al., 2007).

To measure the RpoS activity of these strains, the plasmid probe was transformed into the evolved strains and the ancestor (Materials and Methods). A Δ*rpoS* version of the ancestor was used as the negative control for the experiments. Briefly three independent replicates of each strain were grown at pH 4.5 and pH 7. The fluorescence (515nm) and optical density (600nm) of was recorded every 1 hour for 480 minutes. The absolute fluorescence or absolute rpoS activity was calculated by the fluorescence measurement by the optical density measurement (Flu 515nm/OD 600nm) at each time point.

As can be seen from figure 5.7 and 5.8 the ancestor showed a classical RpoS activity profile showing an elevation in activity 300 minutes into the experiment. That time point correlated with early stationary phase growth of the ancestor.

The evolved strains on the other hand showed an enhanced RpoS activity profile. This elevation in RpoS activity was observed at pH 4.5 and pH 7 (figure 5.7 and 5.8). The elevation in *rpoS* activity was observed early in the experiment when compared to the ancestor. Figure 5.9 depicts the final RpoS activities of the evolved strains and the ancestor at pH 4.5 and pH 7.

Hengge *et al.,* 2008 showed that elevation in RpoS activity would lead to an overall increase in the transcription of RpoS specific promoters thereby increasing the expression of genes which elevate the general stress resistance of the evolved strains. If the elevation in RpoS activity was the primary factor contributing to the elevated levels of general stress resistance, then *rpoS* would have a significant role to play in the overall fitness of the evolved strains.
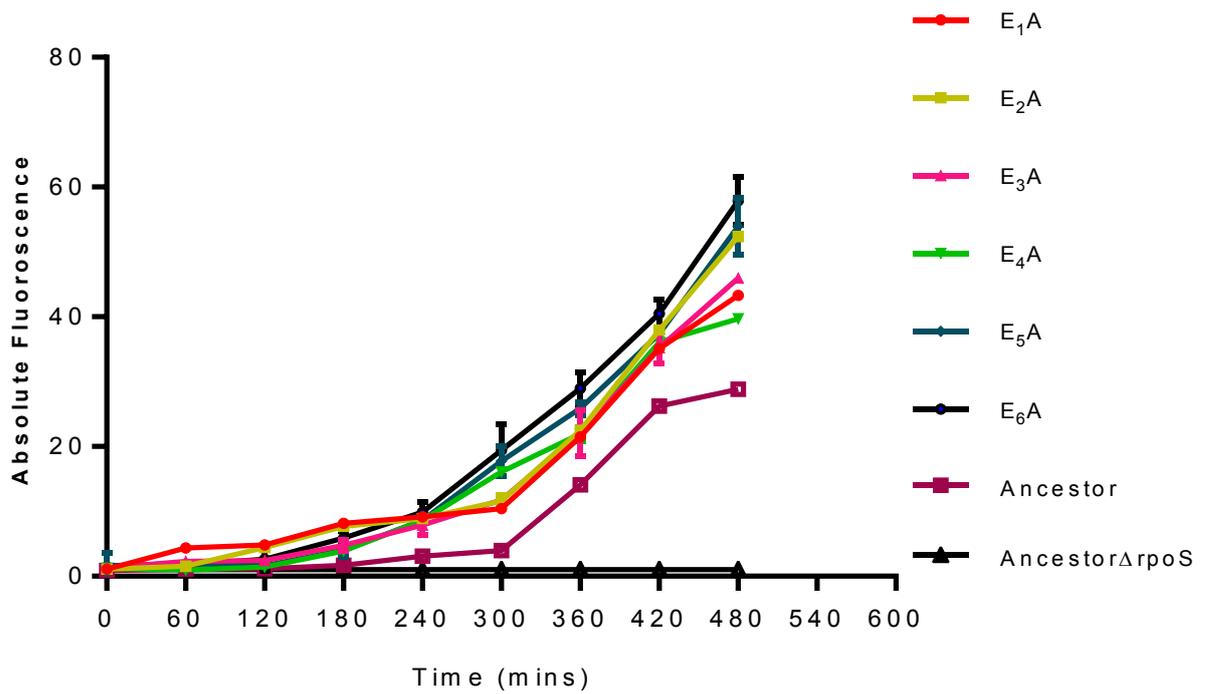
Figure 5.7: Absolute RpoS activity or fluorescence of the evolved strains, ancestor and ancestorΔrpoS at pH 4.5. All the strains are carrying the rpoS responsive p21 fluorometric plasmid probe. The absolute activity is calculated by divining the fluorescence measurement by the optical density measurement (Flu 515nm/OD 600nm) at each time point. The values are an average of three biological replicates and error bars are the standard error of the mean.
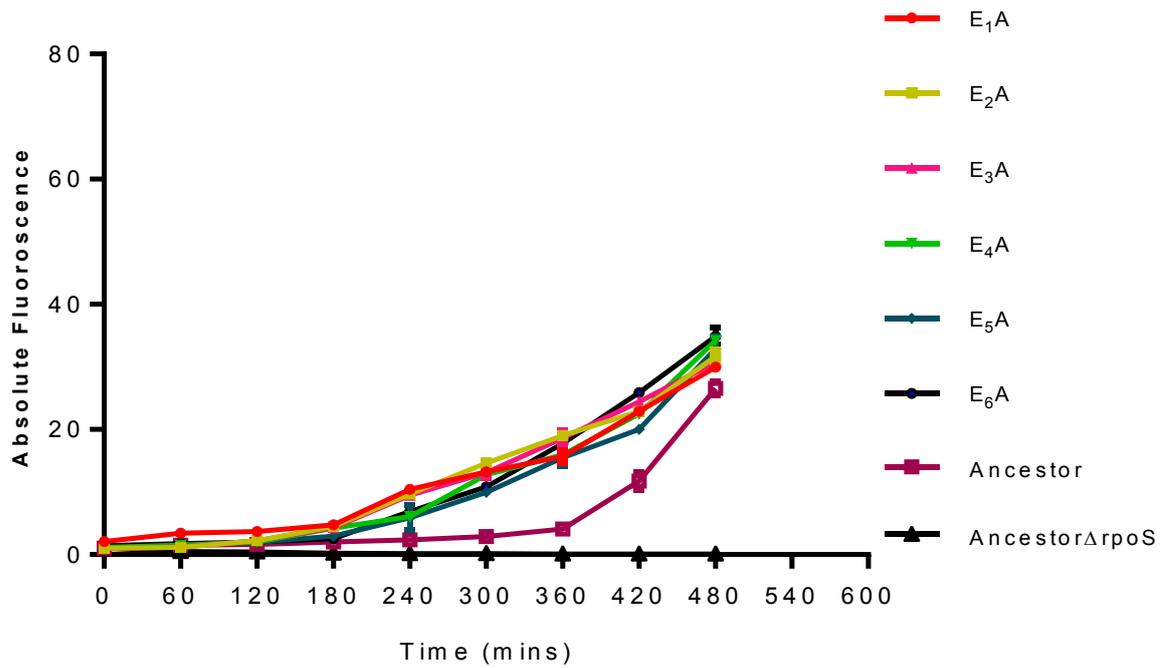
Figure 5.8: Absolute RpoS activity or fluorescence of the evolved strains, ancestor and ancestorΔrpoS at pH 7. All the strains are carrying the rpoS responsive p21 fluorometric plasmid probe. The absolute activity is calculated by divining the fluorescence measurement by the optical density measurement (Flu 515nm/OD 600nm) at each time point. The values are an average of three biological replicates and error bars are the standard error of the mean.
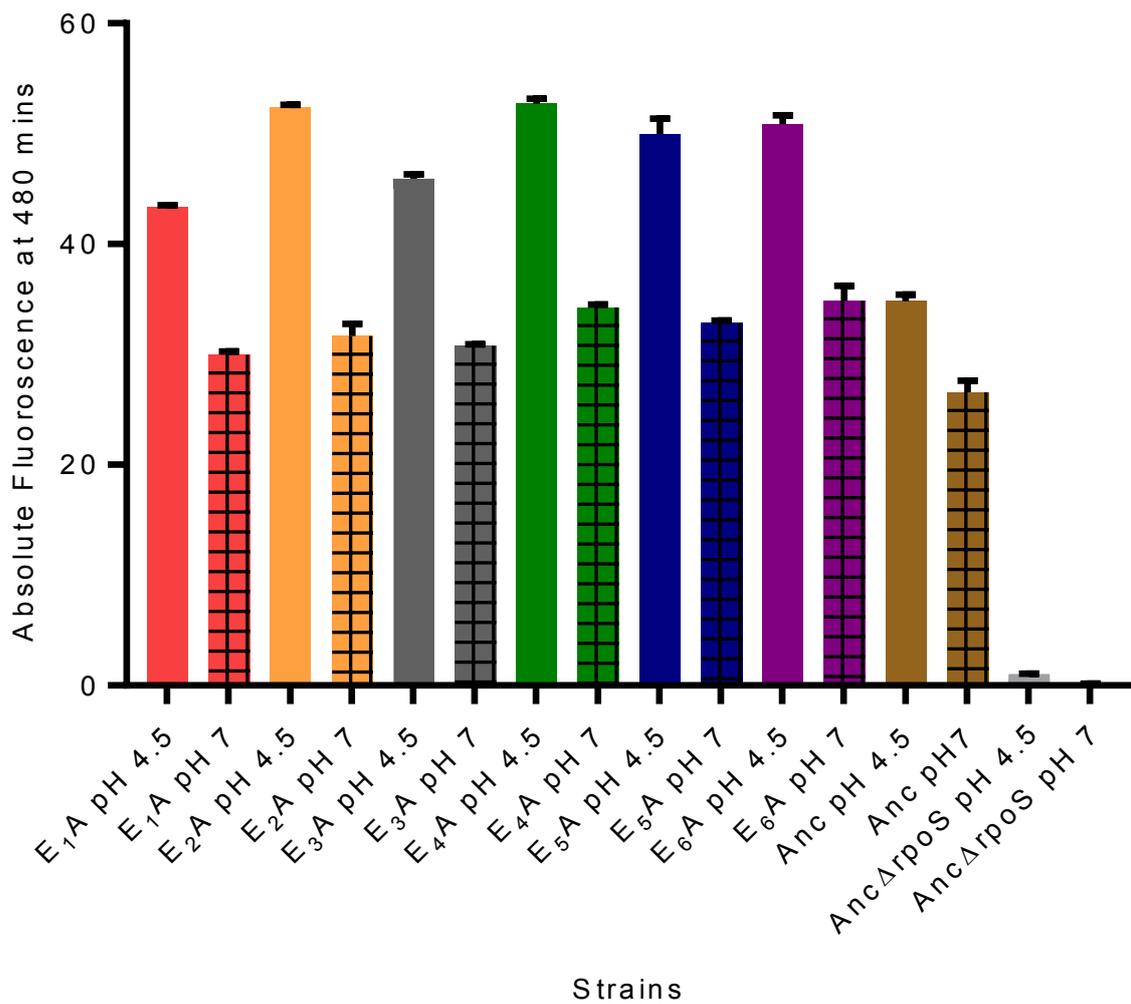
Figure 5.9: Absolute RpoS activity or fluorescence of the evolved strains, ancestor and ancestorΔrpoS at pH 4.5 and pH 7, recorded at 480 minutes. All the strains are carrying the rpoS responsive p21 fluorometric plasmid probe. The absolute activity or fluorescence is calculated by divining the fluorescence measurement by the optical density measurement (Flu 515nm/OD 600nm) at each time point. The values are an average of three biological replicates and error bars are the standard error of the mean. Anc denotes the ancestor.

5.6 RpoS or σ³⁸ plays a crucial role in phenotype at pH 4.5.

The results of the previous section showed that the evolved strains had elevated RpoS activity. Since *rpoS* is responsible for regulating the general stress resistance of *E. coli* it was highly likely that the loss of function of *arcA* resulted in the elevation of the RpoS activity of the evolved strains. If this hypothesis was to be true *rpoS* would be play an important role in conferring the increased cross protection abilities of the evolved strains and their fitness at pH 4.5

To test this the chromosomal copy of *rpoS* was deleted in two of the evolved strains E₄A and E₂A and the ancestor via P1-mediated transduction of the kanamycin marker from the donor strain BW25113Δ*rpoS*::kan of the KEIO library. These *rpoS* deletion strains were competed against KHOO1 and their competitive index was calculated at 24 hours of the competition experiment. We proposed the hypothesis that if *rpoS* was crucial to the fitness of the evolved strains, deletion of *rpoS* would perturb the fitness phenotype of the evolved strains at pH 4.5. Moreover, we hypothesised that if *rpoS* was necessary to combat mild pH shock deletion of *rpoS* in the ancestor would attenuate its ability to resist pH 4.5

As can be seen from figure 5.11 deletion of *rpoS* from the evolved strains completely perturbs the fitness advantage of the strains at pH 4.5. Deletion of *rpoS* also weakened the fitness of the ancestor at pH 4.5. The evolved strains showed a severe reduction in fitness on deletion of *rpoS*.

The outcome of this experiment indicated that the phenotype of the evolved strains was *rpoS* dependent. Another study where *E. coli* populations were evolved under similar conditions to this evolution experiment did not report such a *rpoS* dependent phenotype (Harden et al., 2015). The RpoS activity measurements and the fitness experiment suggests that these evolved strains display a broad range of resistance to a variety of stresses due to higher *rpoS* activity. One of the key factors which might be contributing to this elevation in activity is the loss of function of *arcA*. Loss of function of *arcA* may result in increased transcriptional and translational stability of RpoS. One of the mechanisms by which *arcA* attenuates the level of RpoS in the cell is by increasing its susceptibility to proteolytic degradation. Loss of function of *arcA* may relieve the proteolytic degradation of RpoS and increase the amount of available active RpoS in the cell.
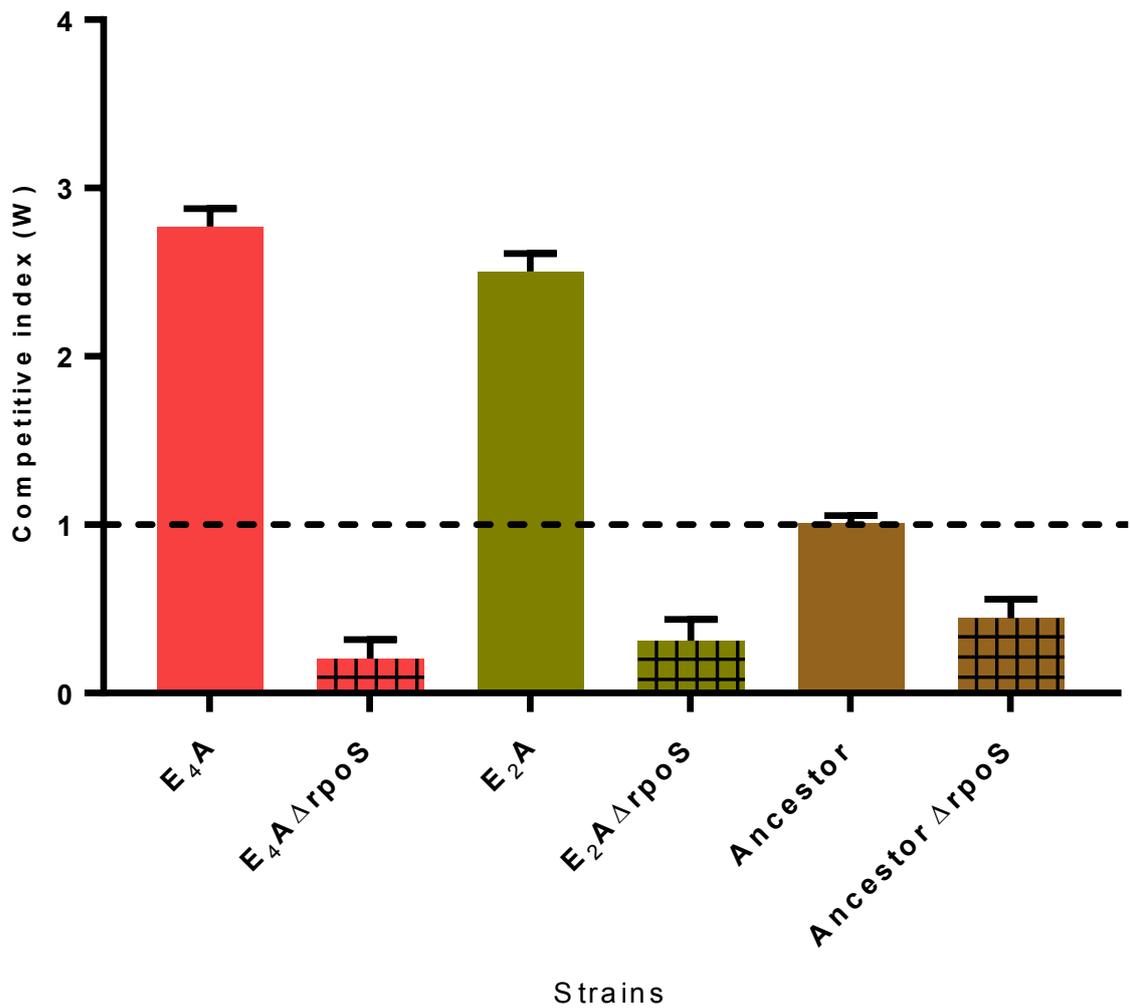
Figure 5.10: Competitive index of E_4A, E_2A, ancestor, $E_4A\Delta rpoS$, $E_2A\Delta rpoS$ and ancestor$\Delta rpoS$ at pH 4.5. The dashed line denotes a competitive index of 1 and represents competitive neutrality. Column values are an average of three independent biological replicates and error bars are standard error of the mean.

5.7 F$_1$*arcA* and F$_1$*arcA*Δ*cytR* show intermediate RpoS activity at pH 4.5

In the previous chapter I have detailed the isolation and characterisation of the F$_1$*arcA* strain. This strain was isolated from the fossil record of E$_1$A and has the *arcA* mutation (M39I) in the genome. It displayed an intermediate fitness phenotype at pH 4.5. The F$_1$*arcA*Δ*cytR* strain showed a minimal increase in fitness compared to the F$_1$*arcA* strain. Since these strains showed intermediate fitness to that of the fully evolved strains it was necessary that the RpoS activity of the two strains was measured and analysed. If the elevation in RpoS activity has a correlation to the fitness of the evolved strains it was important the RpoS activity profile of the two strains at pH 4.5 was understood.

To test the RpoS activities of the two strains the fluorometric RpoS responsive plasmid promoter probe p21 was transformed into the evolved strains and their RpoS activities were measured at different time points in a growth curve. We proposed the hypothesis that if the enhancement in fitness and elevation in RpoS activity were correlated then the F$_1$*arcA strain* would show an intermediate increase in RpoS activity. Since the combination of *arcA* and *cytR* in the F$_1$*arcA*Δ*cytR* strain showed a minimal increase in the fitness we proposed that the RpoS activity of this strain too would show an insignificant increase in RpoS activity.

The results showed significant conformity to our hypothesis. Figure 5.11 illustrates the RpoS activities of F$_1$*arcA* and F$_1$*arcA*Δ*cytR* compared to the ancestor and fully evolved strain E$_4$A. As can be seen from the figure the evolved strain showed a significantly higher RpoS activity profile compared to the ancestor. The F$_1$*arcA* strain showed an intermediate RpoS activity profile compared to E$_4$A and the ancestor (fig. 5.11). The final activity attained by this strain at 480 minutes was intermediate to that of E$_4$A and ancestor (fig. 5.12). The F$_1$*arcA*Δ*cytR* showed a negligible increase in the RpoS activity profile compared to the *arcA* only strain F$_1$*arcA*. The final activity attained by this strain at 480 minutes showed minimal difference to that of F$_1$*arcA* (fig. 5.12).

The results showed that loss of function of *arcA* led to an initial increase in the RpoS activity but this increase was not further complemented by the combination of loss of function of *arcA* and *cytR*. This results suggested that although the fitness effect of *cytR* was contingent on the loss of function of *arcA,* (chapter 4) the combination of *arcA* and *cytR* could neither recover the overall fitness profile of the evolved strains nor the RpoS activity profile compared to the fully evolved strains.
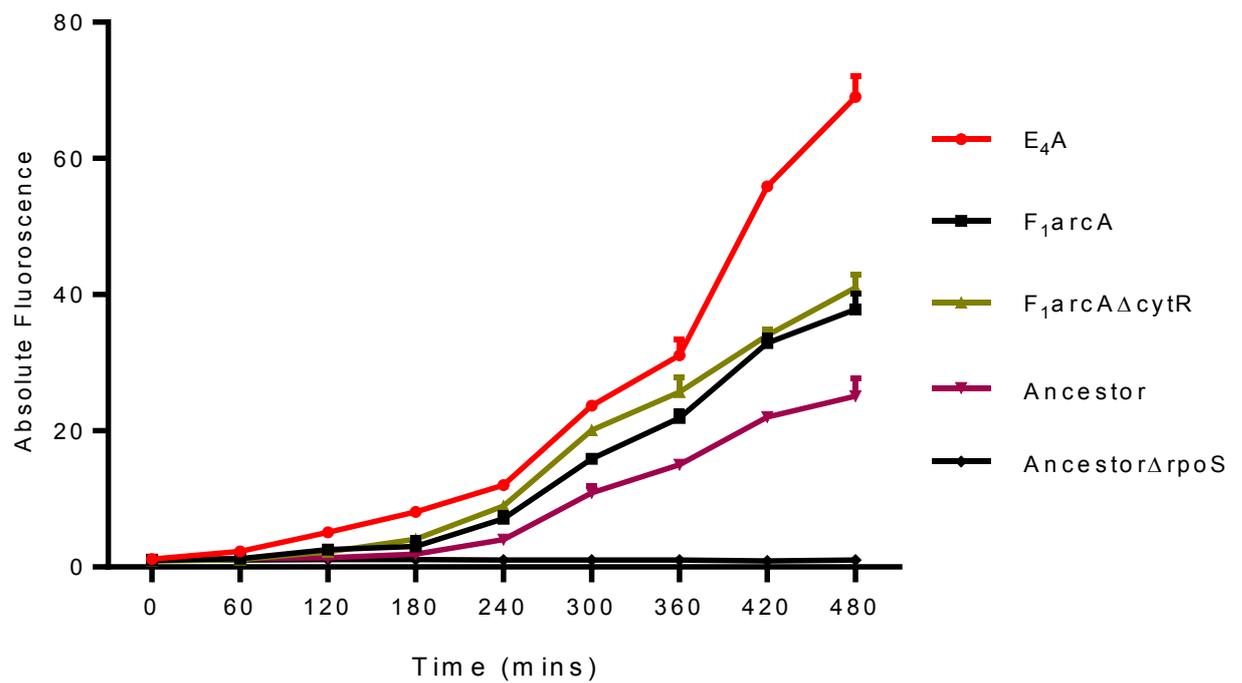
Figure 5.11: Absolute RpoS activity or fluorescence of E4A, F₁arcA, F₁arcAΔcytR, ancestor and ancestorΔrpoS at pH 4.5. All the strains are carrying the rpoS responsive p21 fluorometric plasmid probe. The absolute activity or fluorescence is calculated by divining the fluorescence measurement by the optical density measurement (Flu 515nm/OD 600nm) at each time point. The values are an average of three biological replicates and error bars are the standard error of the mean.
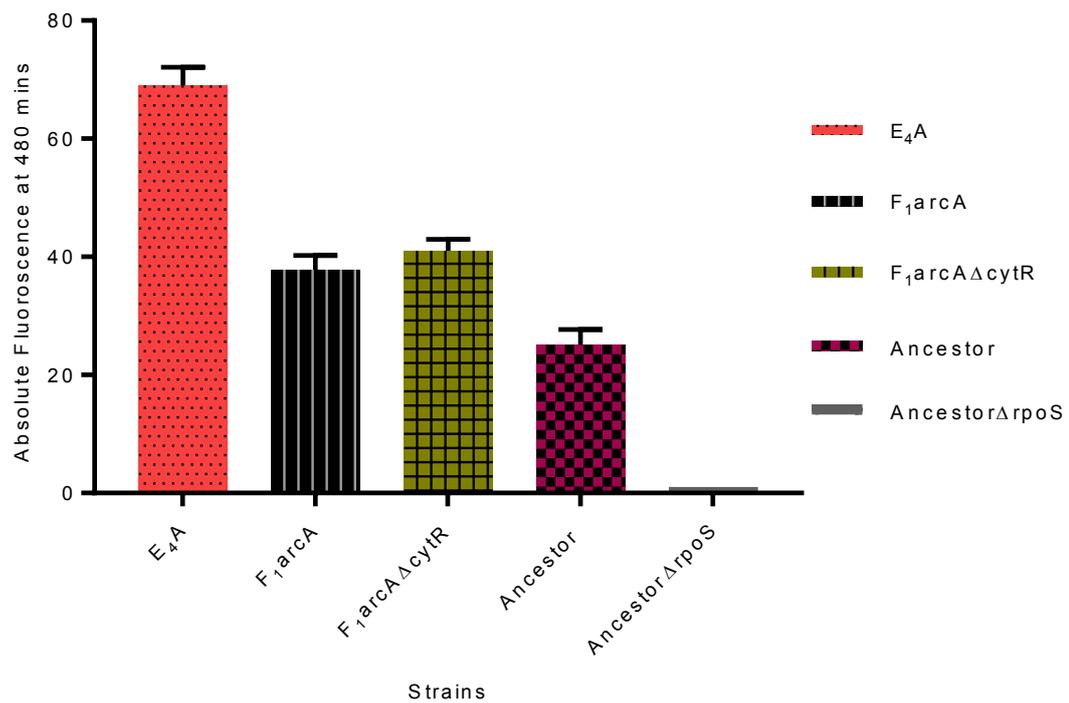
*Figure 5.12: Absolute RpoS activity or fluorescence of E4A, F₁arcA, F₁arcA**Δ**cytR, ancestor and ancestor**Δ**rpoS at pH 4.5, recorded at 480 minutes. All the strains are carrying the rpoS responsive p21 fluorometric plasmid probe. The absolute activity or fluorescence is calculated by divining the fluorescence measurement by the optical density measurement (Flu 515nm/OD 600nm) at each time point. The values are an average of three biological replicates and error bars are the standard error of the mean.*

5.8 The evolved strains show a pH dependent downregulation of HNS (histone-like nucleoid structuring factor) activity

The RNA sequencing analysis of the evolved strains showed that the flagella synthesis and assembly pathway was massively downregulated in all the strains. The transcription factor analysis or regulon analysis showed that the regulon which showed maximum downregulation across all the evolved strains was the *flhDC* regulon. The downregulation in E$_4$A could be attributed to the large genomic deletion which causes the loss of the genes *flhD* and *flhC*. Both these genes are master regulators of flagella synthesis and motility. Since the other evolved strains also showed this transcriptional signature it was important to investigate the factors which could have led to the downregulation of the flagella gene. HNS is a key transcriptional regulator of *flhD* and *flhC* (Shin *et al.,* 1995)

Downregulation of HNS might affect the transcription of *flhD* and *flhC* which in turn would lead to the down regulation of the flagella genes. The evolved strains showed an elevation in RpoS activity and the functional roles of *rpoS* and *HNS* are antagonistic to each other in *E. coli*. Moreover, *arcA* and *HNS* exert similar regulatory influences on the operons they co-regulate by acting as transcriptional repressors. These three arguments necessitated that the analysis of HNS expression in the evolved strains relative to the ancestor. Although the RNA sequencing analysis did not flag a transcriptional change mediated by changes in differential expression of HNS or the HNS regulon it was important that we analysed the role of HNS in the evolved strains.

To do so we used a promoter probe fusion which could report for *HNS* activity. The promoter probe was sourced from David Grainger's laboratory at the IMI, University of Birmingham and the construction of the promoter is described in Haycocks et al., 2015. The heat stable toxin (estA2) gene of enterotoxigenic *E. coli* H10407 has two promoters denoted as *estA1* and *estA2*. The *estA2* promoter is repressed by HNS and the regulatory region of this promoter to which HNS binds was cloned into pRW50 to create a promoter lacZ fusion (P*estA2*::lacZ) (Haycocks *et al.,* 2015).

This promoter lacZ fusion can be transcribed only when HNS is downregulated as HNS represses the transcription of the P*estA2* promoter. Hence, this promoter probe could be used to report for upregulation or down regulation of HNS. The evolved strains, ancestor and the Δ*hns* version of the ancestor were transformed with this promoter probe. Briefly, three independent replicates of the evolved strains, ancestor and Δ*hns* version of the ancestor were grown overnight in LB medium. The overnight cultures were diluted to a starting OD600 of 0.05 in LB medium at pH 4.5 and pH7. They were grown to an OD600 of 0.3-0.4 and their beta-galactosidase activities were recorded (Materials and Methods).

Figure 5.13 shows the beta-galactosidase activities of the different strains assayed with the P*estA2*::lacZ promoter probe. The Δ*hns* strain showed an increased expression of the *estA2* promoter as both pH 7 and pH 4.5. We would expect this as the promoter is *HNS* repressed and the deletion of HNS would

relieve HNS dependent repression. A *Δhns* strain shows a two-fold higher induction of this promoter probe compared to the wild type as can be seen from figure 5.13. The reduction in the activity at pH 4.5 could be attributed to the mild acid shock. The ancestor showed a classic activity profile comparable to published data on *estA2* repression by HNS. Since the cells were assayed at mid-log phase expression of HNS is abundant and it represses the transcription of the *estA2* promoter. The induction profile of the ancestor and its Δ*hns* version were comparable to published data on *estA2* repression by HNS (Haycocks *et al.,* 2015).

The evolved strains showed a pH dependent downregulation of HNS. As can be seen from figure 5.13 all the evolved strains showed higher induction of the promoter probe at pH 4.5 than pH 7. The highest induction was observed in case of E$_4$A. The induction values of the evolved strains at pH 4.5 were like that of the Δ*hns* strain suggesting that HNS was unable to repress the *estA2* promoter. This was not seen at pH 7 where the promoter probe showed reduced induction suggesting HNS could repress the transcription of the *estA2* promoter. The evolved strains showed downregulation of HNS at pH 4.5 but this was not seen in case of the ancestor suggesting HNS was not directly impacted because of mild acid shock. The downregulation of HNS at pH 4.5 could be due to specific changes in the transcriptional landscape of the evolved strains. This result needs further investigation to establish a link between the changes in the transcriptional signature of the evolved strains and downregulation of HNS. This is important as the downregulation of HNS may play a role in attenuated expression of the flagella synthesis pathways in the evolved strains.
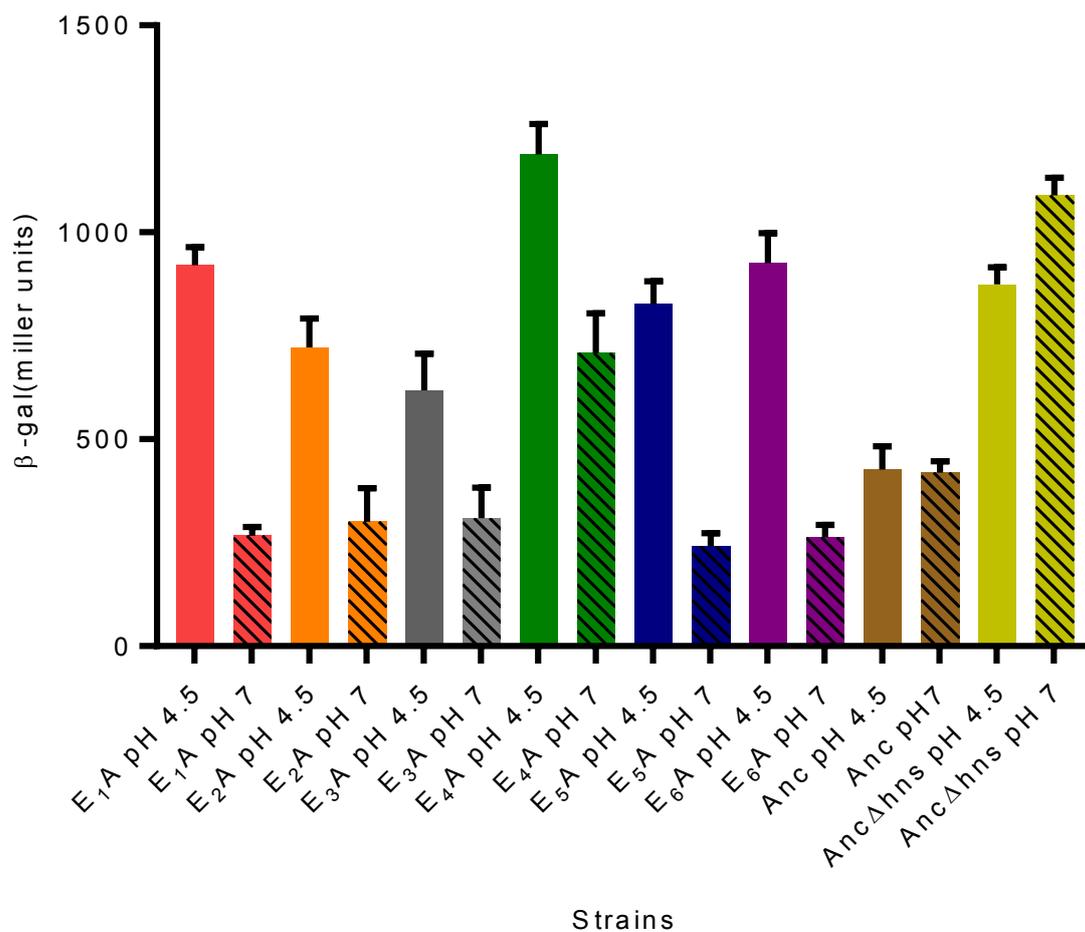
Figure 5.13: Beta-galactosidase activities of the evolved strains, ancestor and ancestorΔhns at pH 4.5 and pH 7 carrying the PestA2::lacZ fusion cloned into the pRW50 plasmid. The background lacZ activity of the empty vector was subtracted from all individual activity measurements. The values are an average of three biological replicates and error bars are the standard error of the mean.

5.9 The evolved strains have elevated levels of CRP (cAMP receptor protein)

Transcription factor analysis of the RNA sequencing data showed that one of the regulons which was upregulated across all the evolved strains was the *CRP* regulon. As discussed in section 4.3, *cytR* and *CRP* co-regulate each other and *cytR* acts as a transcriptional inactivator of CRP under certain conditions. The last chapter detailed the analysis which showed that mutations in cytR caused loss of function of the gene. Loss of function of *cytR* could have led to an upregulation of the CRP regulon. CRP is a global regulator of gene expression and activates the expression of many pathways involved in the catabolic metabolism of various substrates and formation ATP in the cell. Many of the pathways which showed upregulation in the KEGG pathway analysis are co-regulated by CRP. For example, the expression of genes which are components of the citric acid cycle are repressed by *arcA* and activated by CRP. The evolved strains have mutations which cause loss of function of *arcA* which would relive *arcA* dependent repression and upregulation of CRP would increase the expression of this pathway.

To test this, we wanted to measure and compare the levels of CRP in the evolved strains and the ancestor. To do this a similar approach detailed in the previous section was taken. The stable toxin gene of enterotoxigenic *E. coli* H10407 has two promoters denoted as *estA1* and *estA2*. The *estA2* promoter is repressed by *HNS* while the *estA1* promoter is activated by CRP. *HNS* occludes activation of the *estA1* promoter by blocking transcription by binding to the *estA2* promoter. The regulatory region of the *estA1* promoter was cloned into the pRW50 vector to create a *estA1*::lacZ fusion (Haycocks *et al.,* 2015). This beta-galactosidase activity of this promoter lacZ fusion was used an indirect method to report the level of CRP activity (Materials and Methods). The evolved strains, ancestor and ΔCRP version of the ancestor were transformed with this promoter fusion probe and their beta-galactosidase activities the strains were recorded following the same experimental protocol as described in the previous section except that the CRP activities were recorded only at pH 4.5 (Materials and Methods).

As can be seen from figure 5.13 the evolved strains E_4A and E_2A showed higher induction or lacZ activity of the promoter probe compared to the ancestor. The induction was approximately 1-fold higher than the wild type indicating the evolved strains had higher levels of CRP binding to the *estA1* promoter. Hence, we concluded the evolved strains expressed higher levels of CRP than the ancestor.
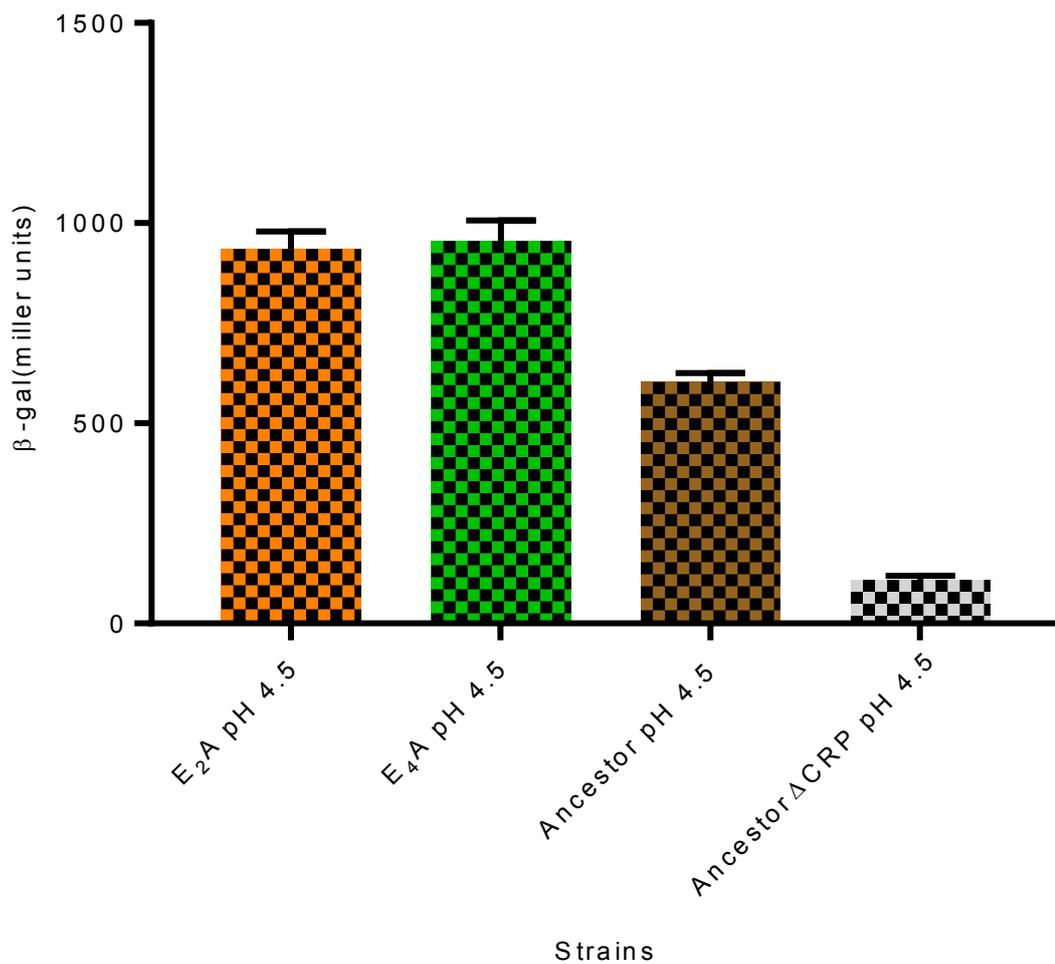
*Figure 5.14: Beta-galactosidase activities of the E₂A, E₄A, ancestor and ancestor**Δ**CRP at pH 4.5 and pH 7 carrying the PestA1::lacZ fusion cloned into the pRW50 plasmid. The background lacZ activity of the empty vector was subtracted from all individual activity measurements. The values are an average of three biological replicates and error bars are the standard error of the mean. The mean activities of E₂A PestA1::lacZ and E₄A PestA1::lacZ were compared to the mean activity of ancestor PestA1::lacZ and ancestor**Δ**CRP PestA1::lacZ using a non-parametric one way ANOVA.*

5.10 The role of *flhD* and *flhC* in the fitness profiles of the evolved strains

The RNA sequencing analysis highlighted the uninform downregulation of the flagella synthesis pathway. The downregulation was observed at the differential gene expression level and at the transcription factor analysis. The two regulons which showed maximum regulation across all the evolved strains were regulated by *flhD* and *flhC*. The downregulation of the flagella genes in E_4A could be attributed to the loss of the two key regulators in the large genomic deletion. But the downregulation of the flagella genes in the other evolved strains was an interesting aspect of the transcriptional signature of these strains. The regulatory network which controls the systematic expression of flagella and motility is complex, especially due to the regulatory influence of different transcriptional activators at the *flhD* and *flhC* locus (Fitzgerald *et al.,* 2009). Nonetheless, we wanted to investigate the contribution of *flhD* and *flhC* to the overall fitness of the evolved strains at pH 4.5. To do so we deleted the chromosomal copy of *flhD* and *flhC* from the evolved strain E_2A and the ancestor via P1 mediated transduction of the kanamycin marker from BW25113Δ*flhD* and BW25113Δ*flhC* of the KEIO library. This work was done by me and Assel Nurmagambetova, a MSc student working under my supervision.

We proposed the hypothesis that deletion of *flhD* and *flhC* would not create a significant difference in the overall fitness of E2A as these genes were already downregulated at the transcriptional level. Since the downregulation of the flagella genes were such a characteristic feature of the evolved strains. We also proposed that deletion of *flhD* and *flhC* would enhance the fitness of the ancestor at pH 4.5. To test this, we competed the deletion strains against KHOO1 at pH 4.5 and generated the competitive index for 24 hours of the competition experiment.

As can be seen from figure 5.14, confirming to our hypothesis the deletion of *flhD* and *flhC* from E_2A showed no difference in the overall fitness at pH 4.5. Surprisingly contrary to our hypothesis the enhancement in fitness upon deletion of *flhD* and *flhC* from the ancestor was minimal when compared to the F_1arcA strain. This showed that inactivation of the master regulators of the flagella operon was not directly responsible for an enhancement of fitness at pH 4.5.
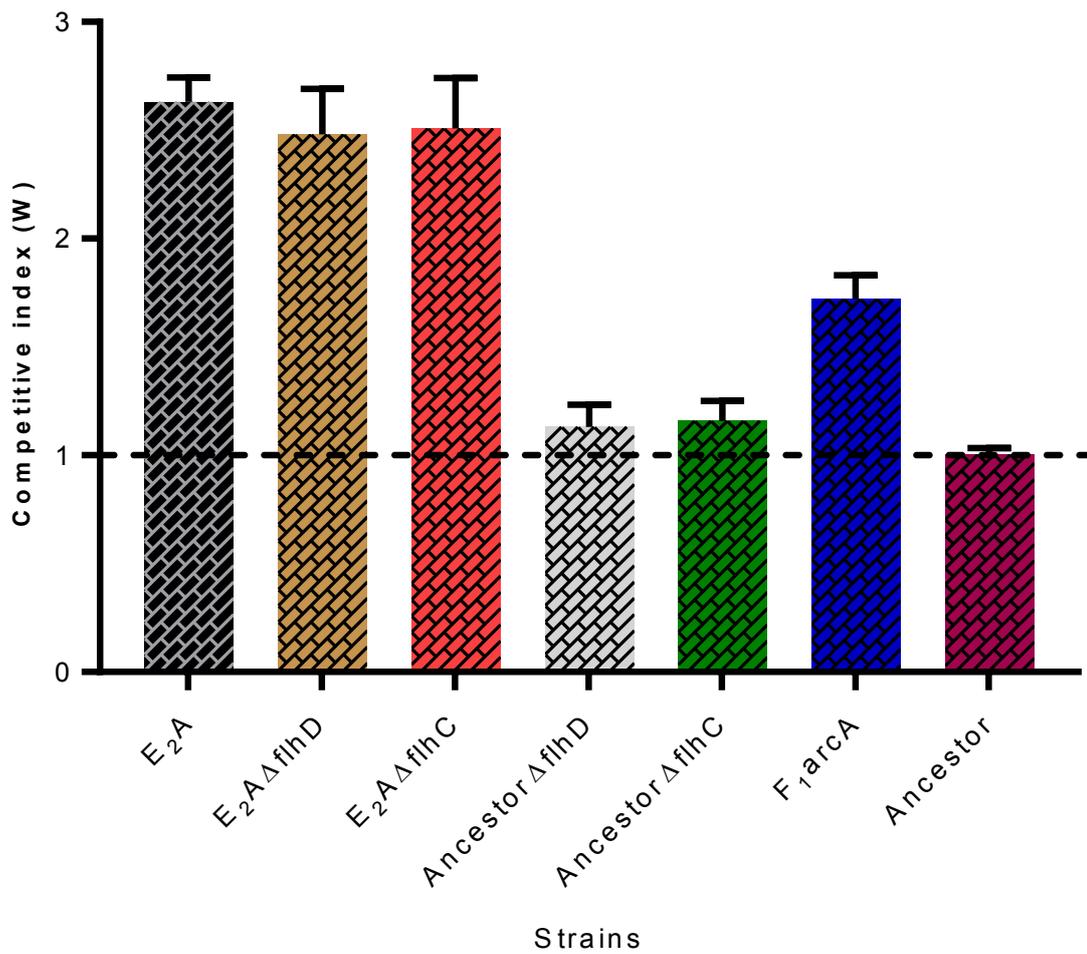
Figure 5.15: Competitive index of E₂A, E₂AΔflhD, E₂AΔflhC, ancestorΔflhD, ancestorΔflhC, F₁arcA and the ancestor at pH 4.5. The dashed line denotes a competitive index of 1 and represents competitive neutrality. Column values are an average of three independent biological replicates and error bars are standard error of the mean. P-values were calculated using a non-parametric one way ANOVA.

5.11 Fitness contribution of *flhD* and *flhC* was contingent on *arcA*.

The results of the previous section showed that the loss of *flhD* and *flhC* did not alter or enhance the fitness of the ancestor at pH 4.5. It was possible that the downregulation of flagella synthesis genes was because of the loss of function of *arcA* or the loss of function of *cytR*. Flagella synthesis requires the systematic expression of genes from three operons known as Type I, Type II and Type III. *flhDC* are the master regulators of flagella synthesis but in addition these two regulators there are other genes which feed into the flagella regulatory network. Hirofumi et al., 2006 had shown that a Δ*arcA* strain had a non-motile phenotype and reduced expression of *flhDC*. This phenotype was independent of *arcB* and dependent on *arcA*. Although the regulatory interactions between *arcA* and *flhDC* is not well understood. Hence, we decided to investigate the effect of *flhD* and *flhC* in a *arcA* background. Downregulation of *flhDC* could result from the loss of function of *arcA* or from the combinatorial effect of loss of function of *arcA* and *cytR*.

To do this we deleted the chromosomal copy of the *flhD* and *flhC* from the F$_1$*arcA* and F$_1$*arcAΔcytR* strains via P1 mediated transduction (Materials and Methods). The kanamycin marker of the Δ*cytR* deletion in F$_1$*arcAΔcytR* was removed using pcp20 dependent excision of the resistance marker. Then these deletion strains were competed against KHOO1 at pH 4.5 and their competitive indices were calculated for 24 hours of the competition experiment.

Coupling the loss of function of *arcA* with the deletion of *flhD* and *flhC* (fig. 5.16) showed a small increase in the overall fitness at pH 4.5. But this improvement in fitness was marginal when compared to the fitness enhancement conferred by the loss of function of *arcA* alone. This can be seen clearly by comparing the fitness index of the F$_1$*arcA* strain with that of F$_1$*arcAΔflhD and* F$_1$*arcA* (fig. 5.16). This showed that the contribution of the loss of *flhD* and *flhC* to the overall fitness of the evolved strains at pH 4.5 was contingent on *arcA*. The exact regulatory effect of the loss of function of *arcA* which led to the downregulation of the flagella genes would need further investigation. Combinations of the loss of function of *arcA* and *cytR* with *flhD* and *flhC* showed a diminished effect on fitness (fig 5.16). This result indicated that the phenotype of the evolved strains could not be reconstructed based on the contribution of these genes. One of the key contributors of the phenotype of the evolved strains could be *rpoA*. The transcriptional downregulation of the flagella genes could be an indirect consequence of the mutations in *arcA* and *rpoA*.
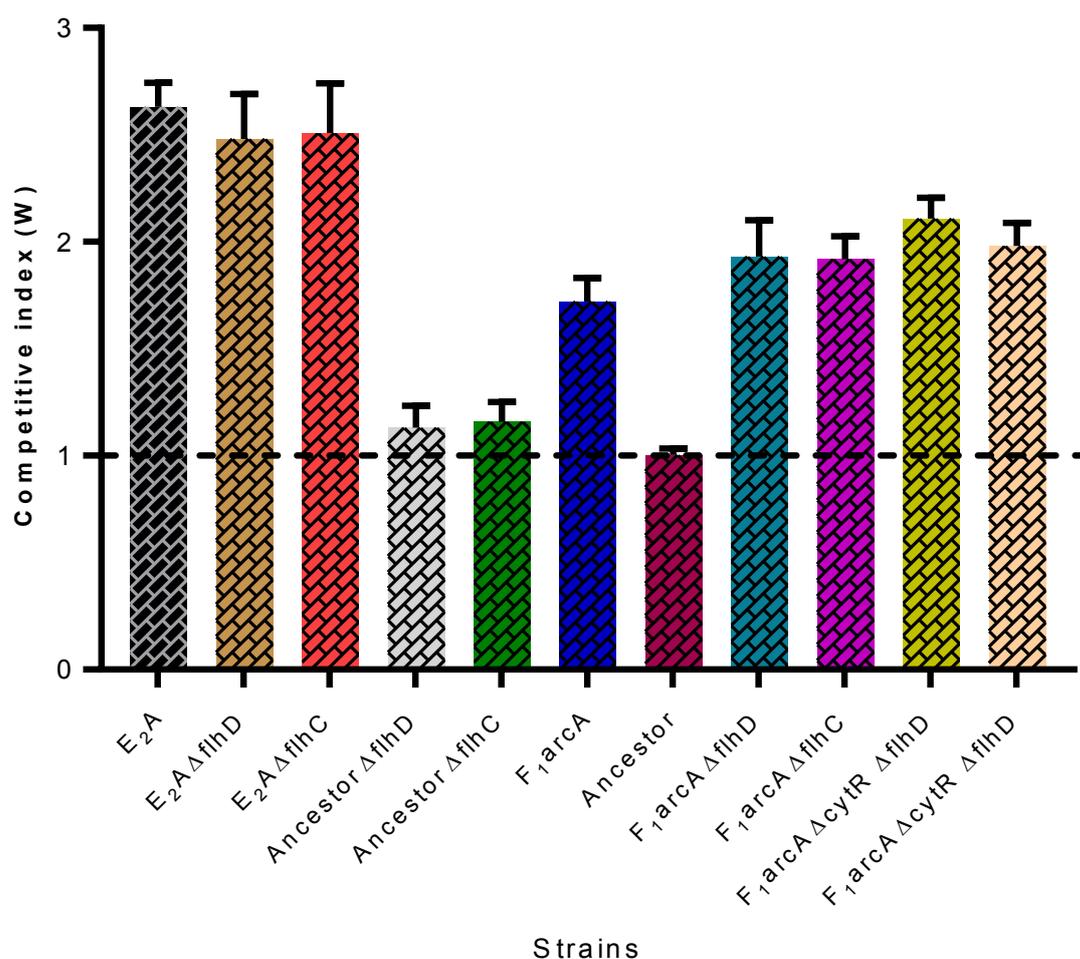
Figure 5.16: Competitive index of $F_1arcA\Delta flhD$, $F_1arcA\Delta flhC$, $F_1arcA\Delta cytR\Delta flhD$ *and* $F_1arcA\Delta cytR\Delta flhC$ *at pH 4.5. The dashed line denotes a competitive index of 1 and represents competitive neutrality. Column values are an average of three independent biological replicates and error bars are standard error of the mean. P-values were calculated using a non-parametric one way ANOVA.*

5.12 Discussion for chapter 5

Chapter 3 and 4 for this study detailed the key results which explain the genetic basis of adaption of the six evolved strains to pH 4.5. The experimental objectives of this chapter were to enable us to create a link between the genetic basis of adaptation and the phenotype of the evolved strains. One of the methods employed by us to understand this link was to investigate the transcriptional landscape of the evolved strains. The RNA sequencing of the evolved strains highlighted some important changes in their transcriptional signatures which were instrumental in driving the development of hypothesis for the experiments that followed in the later sections of this chapter.

The transcriptional landscape of the evolved strains showed a global change in gene expression with a characteristic upregulation of pathways involved in cellular metabolism and energy dispensation. This contrasted with the changes in gene expression seen in the previous evolution experiment (Johnson et al., 2014). Mutants which were evolved to resist extreme acid stress did not show an overhaul in the global transcriptional profile. Mutations in the sensor kinase in EvgS which conferred resistance to pH 2.5 led to a constitutive activation of the genes in the glutamate decarboxylase system. The Gad system is dedicated to combating and alleviating intracytoplasmic acid stress in *E. coli*. The differences in the transcriptional landscapes of strains evolved to resist different pH values highlighted the specificity of adaptation and selection of different mutations which conferred resistance to low pH stress.

The overall transcriptional landscapes of the evolved strains were very similar to each other but a finer analysis of the individual strains showed that there were a total 230 genes which showed outlier behaviour amongst them. These differences in patterns of genes expression could be attributed to the other mutations in their genomes in addition to the repeatedly selected genes which were *arcA*, *rpoA* and *cytR*. The evolved strains have small deletions in their genomes which were tabulated in chapter 3. These deletions need to be verified and their effect on gene expression and fitness of the evolved strains needs to be analysed in the future. The total number of genes which showed maximal change in differential expression constituted only a small percentage (8%) of the total number of genes analysed in the RNA sequencing experiment.

The evolved strains have mutations in genes which are key regulators of large gene networks. Mutations in global regulators such as *arcA* and *cytR* would cause a global change in gene expression and have pleiotropic effects. A significant number of these pathways which showed

upregulation in the KEGG pathway analysis were either regulated by *arcA* or *cytR*. To determine the contribution of each individual gene to the changes in the transcriptional landscape future experiments would need to be aimed at studying the effect of these on the transcriptional regulation in isolating from each other. The transcription factor analysis gave a clear indication that the upregulation of key regulons such as the CRP regulon were a result of the upregulation of *arcA* and *cytR* regulons in combination. ArcA and CytR are known to interact with CRP, in many cases as transcriptional inactivators or repressors. Loss of function of *arcA* and *cytR* could lead to an elevation in the levels of *CRP* in the cell. Elevated levels of CRP would directly lead to changes in the levels of cyclic-AMP. Cyclic-AMP controls the expression of proton pumps involved in response to low pH stress. Further analysis using promoter probes showed that the evolved strains express higher levels of CRP. The role of CRP and acid resistance is not clearly understood but this study gives an indication of a route through which such interactions can be activated in response to mild stress.

The transcription factor analysis helped us understand the global change in the transcriptional landscape of the evolved strains with a perspective directed towards linking this change to the phenotype of the evolved strains. There were striking similarities between the transcriptional signature of these strains to that of a Δ*arcA* strain studied in Park *et al.,* 2013. It is not surprising that many genes which showed changes in gene expression were *arcA* and *cytR* regulated. The RNA sequencing profiles of the strains complemented their genomic background.

The obvious next question was the link between this global *arcA* dependent change and enhances general stress resistance. *arcA* is a key regulator of RpoS and changes in RpoS activity and RpoS levels in the cell would lead to increased resistance to a variety of stresses. The elevation in rpoS activity in the evolved strains could have altered the SPANC (self-preservation and nutritional competence) balance. A phenomenon commonly associated with rpoS, whereby cells redirect energy sources towards enabling elevated stress resistance and restricted reliance on limited nutritional sources (Ferenci *et al.,* 2001). The upregulation of the glyoxylate pathway and pathways which control catabolism specific amino acids were a characteristic feature of *rpoS* controlled changes in the SPANC balance of the cell. The experimental evidence showed that the contribution of *rpoS* was crucial in the overall fitness of the evolved strains. The loss of function of *arcA* complements the elevation in rpoS activity and enhancement in fitness. But we don't clearly understand the route through which this regulatory rewiring might have taken place in the evolved strains. RNA sequencing of a strain with the *arcA* mutation only might throw light on this direction.

An important observation from the RNA sequencing analysis was the downregulation of the Gad genes, *gadX* and *gadW*. These genes regulate the expression of the glutamate decarboxylases which respond to extreme acid stress. This downregulation was also reported in Penix A *et al.,* 2017 where *E. coli* strains were evolved under similar conditions to our study. The downregulation of these genes could be a direct consequence of elevated RpoS activity and CRP levels. Yamshino *et al.,* 2009 had showed that RpoS and CRP were both involved in downregulation of the gad system. This too may be a result of SPANC where the cell downregulates systems which respond to unbuffered proton stress and redirects energy towards increasing general stress resistance. A key comparison which would help understand the extent to which genes regulated *rpoS* change in the evolved strains would be to perform a regulon or transcription factor analysis of the genes differential expressed from the *rpoS* regulon in the evolved strains. This would also help in investigating the contribution of the rpoA mutation to the phenotype of the evolved strains.

Summary of results for chapter 5

1) RNA sequencing of the evolved strains showed a global change in the transcriptional landscape. The *arcA* and *cytR* regulon were maximally upregulated across all the evolved strains compared to the ancestor.

2) The evolved strains showed a significant downregulation of the genes involved in the synthesis of flagella and regulation of motility

3) The evolved strains have elevated RpoS activity. The elevation in activity was contingent on the loss of function of *arcA*.

4) The evolved strains showed a pH dependent downregulation of HNS and elevated levels of CRP at pH 4.5

5) The fitness contribution of downregulation of *flhD* and *flhC* was contingent on *arcA*.

6) The RpoS activity profile of the evolved strains could not be recapitulated by making combinations of deletions between *arcA, cytR, flhD* and *flhC*

Evolution of acid resistance to pH 4.5 in *E. coli* K12 MG1655

Chapter 6

Identifying the genetic basis of adaptation to pH 4.5- Part 3

Role of *rpoA* in the phenotype of the evolved strains

6.1 *rpoA* and the N294H mutation.

Lab-based evolution is a fundamental technique to understand and study bacterial stress responses. It helps us understand how bacteria employ pre-existing pathways to generate appropriate responses to a variety of stresses. Chapter 3 described the evolution experiment to study *E. coli*'s ability to adapt to a moderately low pH value of pH 4.5. The evolved strains from the evolution experiment showed an enhanced resistance to pH 4.5 and cross protection against other forms of stresses. Chapter 4 showed that a significant part of the enhancement in fitness at pH 4.5 resulted from the loss of function of *arcA* and this was also reflected in the transcriptional landscape which showed upregulation of the *arcA* regulon (Chapter 5).

One of the contributing factors or effects of loss of function of *arcA* was the elevation of RpoS activity in the evolved strains. The elevation in the RpoS activity was crucial to the phenotype of the evolved strains. Investigating the contribution of various target genes in combination with the loss of function of *arcA* showed diminished returns in fitness and RpoS activity. The summative assessment of the results of the previous chapter indicated that the *rpoA* mutation must have an important role to play in linking the genotype of the evolved strains to their phenotype.

I have already introduced the role of *rpoA* in *E. coli* and its importance in the transcriptional machinery of the cell in chapter 2. In the introduction to this chapter I will briefly discuss the role of 294th locus of *rpoA* and the results to follow will show the relationship between *arcA*, *rpoA* and fitness of the evolved strains at pH 4.5

As described in chapter 2, *rpoA* or the RNA polymerase subunit (α) is a critical component of the RNA polymerase holoenzyme (Zhou *et al.,* 1989). Five out of the six evolved strains had mutations in this gene. The mutation caused a histidine to asparagine substitution at the 294th position in the protein. *rpoA* is an essential gene and it is unlikely that the N294H mutation is loss of function mutation. The mutation must have a very specific and niche role to play in the overall phenotype of the evolved strains. The locus of the mutation has been studied in detail in several studies such as Savery *et al.,* 1992, Zhou e*t al.,* 1995 and Ma *et al.,* 1998. It is known to be important in facilitating contact between the RNA polymerase core complex and the DNA binding regions (Gourse et al., 1993). Busby *et al.,* 1998 showed that mutating the 294th residue caused an upregulation of Class I CRP transcribed genes. The evolved strains showed an upregulation of the CRP regulon and many of the key pathways which are Class I CRP transcribed in nature. Figure 5.1 taken from Savery *et al.,* 1992 highlights the residues of *rpoA*

which are critical in the overall RNA polymerase complex assembly and their interactions with DNA binding regions. Another study showed that this residue was critical in facilitating recruitment and contact between the house keeping sigma factor σ^{70} and core RNA polymerase. Alanine substitution of this residue reduced cooperativity between σ^{70} and core RNA polymerase. In vitro transcription assays showed that mutations in and around the 294th locus increased selective interaction between RpoS or σ^{38} directed via occlusion of σ^{70} (Lloyd *et al.,* 2002). The elevation in RpoS activity of the evolved strains could be a result of loss of function of *arcA* which increases RpoS stability and the *rpoA* mutation which enhances the cooperativity between RpoS and RNA polymerase. This could lead to the preferential transcription of genes which caused an alteration in the global transcriptional landscape of the cell finally leading to overall increment in the general stress resistance of the cell.
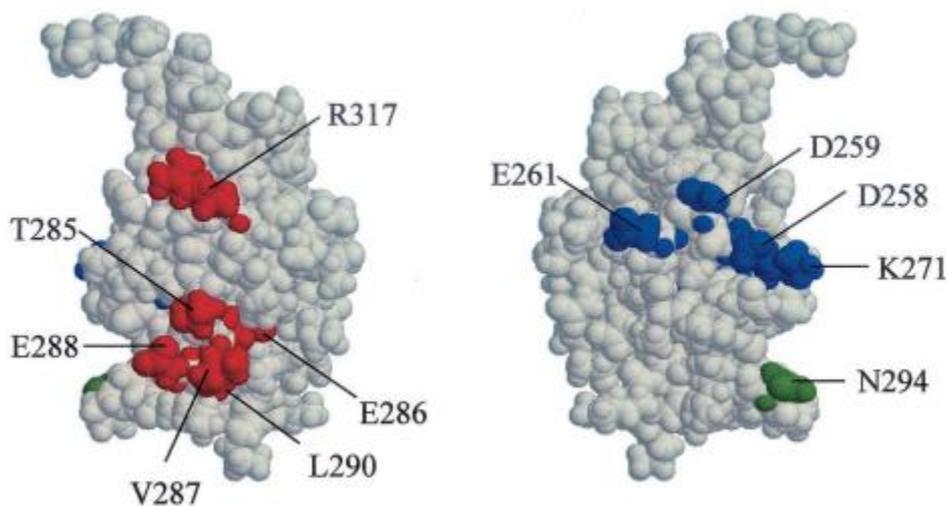
Figure 6.1: Important residues in the alpha subunit of RNA polymerase. This figure has been taken from Savery et al., 1992. The figure highlights residues which are critical in the interactions between RNA polymerase holoenzyme and rpoA. The residues highlighted in red and blue facilitate the interaction between rpoA and the RNA polymerase core complex. The 294th residue in green is critical in driving the interaction between the RNA polymerase complex, sigma factors and DNA binding regions for transcriptional activation.

The above cited arguments suggested that the mutation in *rpoA* must have an important role to play in guiding the phenotype of the evolved strains. This chapter will detail the results which would explain the combinatorial role of *arcA* and *rpoA* in conferring the phenotype of the evolved strains.

6.2 RT-PCR analysis of the F_1arcA.

The F_1arcA strain was isolated from the fossil record of E_1A (chapter 3) and has the M39I mutation its genome. This strain showed an intermediate fitness phenotype and RpoS activity (chapter 4 and 5). Since RNA sequencing of this strain was not possible at that point in the project due to various limitations, we decided to probe the transcriptional signature of this strain using quantitative reverse transcriptase polymerase chain reaction (QRT-PCR). The experiment in this section were conducted by Maria Massoura, a PhD student in the lab and me under the supervision of Dr Thippesh S.

The objective of the experiment was to determine the expression of two genes (upregulated and downregulated) selected from the differential gene expression data collected from the RNA sequencing analysis. The expression profile of the two genes in F_1arcA would be compared to the expression of the genes in the fully evolved strain E_2A. The two genes chosen for the experiment were *flgM* and *fimC*. *flgM* is an anti-*fliA* factor which was significantly downregulated across all the evolved strains. *fimC* is a regulator of the type 1 fimbriae operon and was upregulated in all the evolved strains relative to the ancestor. *pepA*, an aminopeptidase which showed no change in gene expression in any of the evolved strains and ancestor was used as internal reference control.

Briefly, three independent replicates of E_2A, F_1arcA amd ancestor were grown in LB medium at pH 4.5. Total RNA samples were isolated following the same experimental methods used for RNA sequencing (Materials and Methods). cDNA synthesis was done using the Tetra-cDNA synthesis kit and QRT-PCR was performed using Taqman probes as described in materials and methods. Fold expression for each gene was calculated and relative expression of the genes was calculated by comparing the ΔCq values of E_2A, F_1arcA with that of the ancestor. Log_{10} of the fold expression was calculated to represent the results.

Figure 6.2 shows the log_{10} of the fold expression of *flgM* and *fimC* in E_2A and F_1arcA. The fold expression values have been calculated relative to the ancestor. As can be seen from the figure *flgM* is significantly downregulated in E_2A whereas *fimC* is massively upregulated. The data compares correctly with the relative fold expressions of the genes attained from the RNA sequencing analysis. On the other hand, there is a significant difference in the downregulation of *flgM* and upregulation of *fimC* between E_2A and F_1arcA. We compared the mean fold expression values of E_2A against F_1arcA for each gene using an unpaired student's *t*-test. There was significant difference in expression between E_2A and F_1arcA.

The increased upregulation of *fimC* and downregulation of *flgM* in E2A could be because of the *rpoA* mutation. The F$_1$*arcA* strain didn't have the *rpoA* mutation and previous results had already shown that this strain and intermediate fitness and RpoS activity at pH 4.5. The enhanced fitness, RpoS activity and differences in gene expression of the fully evolved strains and F$_1$*arcA* could be the effect of the *rpoA* mutation.
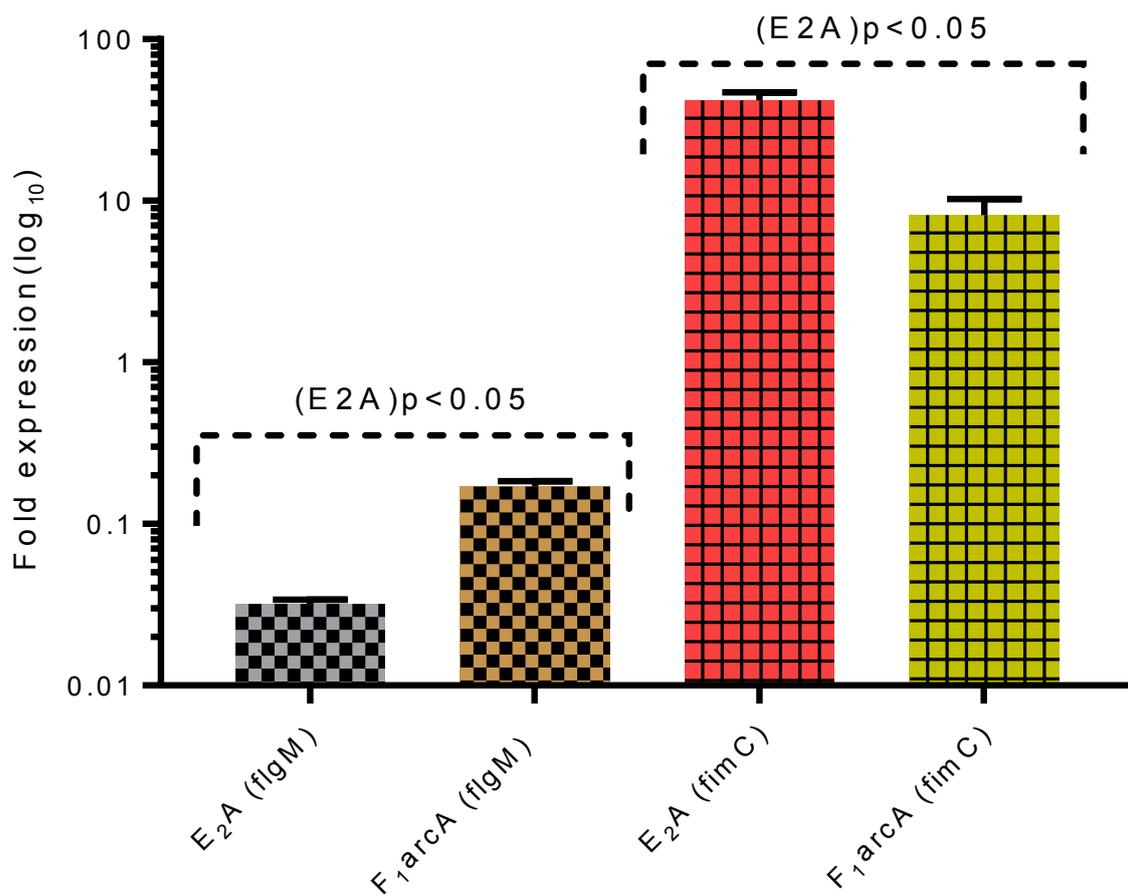
Figure 6.2: Log_{10} fold expression of flgM and fimC in E_2A and F_1arcA in LB medium at pH 4.5. The values are an average of the three independent replicates and error bars are standard error of the mean. P-values were calculated using **an unpaired student's** t-test.

6.3 *arcA* and *rpoA* work together to confer enhanced fitness at pH 4.5

To determine the effect of the *rpoA* N294H mutation on fitness of the evolved strains, we moved the mutation to the genome of the ancestor and F_1*arcA*. The objective of the experiment was aimed at understanding three important question. What is the effect of the *rpoA* mutation on fitness at pH 4.5? Is the effect of the mutation contingent on the presence of loss of function of *arcA* in the background? What is the effect of the *rpoA* mutation on RpoS activity? Before I show the results, I will explain the experimental methodology used to transduce the mutation in brief.

To transduce the N294H mutation we used a recombination based approach. The method can be used to make single base pair replacements in *rpoA* on the genome. This method used an intermediate strain known as WAM123 or phs. This strain was first studied in Rownland *et al.,* 1985. The phs strain has a K271E mutation in *rpoA* which causes a defect in Na^+ transport and renders it incapable of growing on melobiose and methionine supplemented minimal medium. The *rpoA* locus in this strain is marked with a Tn::10 which confers tetracycline resistance.

Zhou *et al.,* 1995 showed that to achieve a methionine and melobiose positive phenotype, this strain could be transformed with a plasmid containing the wild type *rpoA* gene. The transformants could be selected on methionine and melobiose plates for revertants which had undergone recombination with the wild type *rpoA* gene on the plasmid. The recombination event would relieve the K271E mutation and the melobiose and methionine sensitive phenotype.

We used the phs strain as an intermediary vehicle to move the N294H mutation across different genetic backgrounds. To do this we used a plasmid known as pLAW2 *rpoA* which has the wild type *rpoA* gene cloned into it. The construction of the plasmid is explained in Savery *et al.,* 1992. We site directed mutagenized the plasmid to incorporate the N294H mutation.

The resulting plasmid pLAW2 N294H was transformed into the phs or WAM123 strain and successful transformants were selected by ampicillin resistance after overnight incubation at 30 degrees. Successful transformants were replica plated on M9-melabiose and M9-methionine plates Transformants which gave a methionine (met$^+$) and melobiose (mel$^+$) positive phenotype were PCR sequences to verify the N294H mutation and the recombination of the K271E mutation to the wild type. The N294H mutation on the chromosome was transduced to the ancestor and F_1*arcA* via P1-mediated transduction of the tetracycline marker. A schematic diagram of the entire process is shown in the next page.

Transduction of *rpoA* N294H from *phs* strain to the ancestor and F₁*arcA*
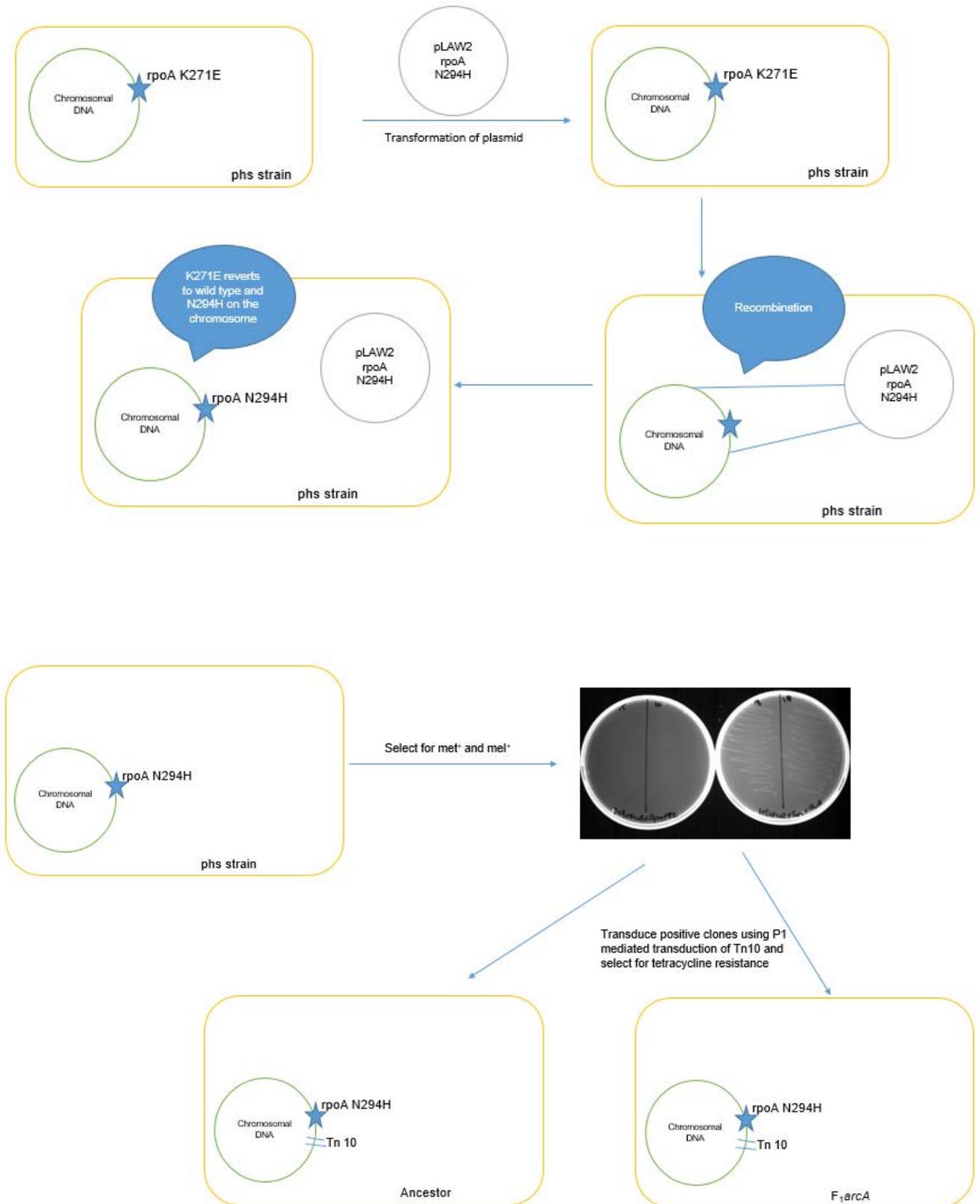
After the N294H mutation was successfully transduced to the ancestor and F$_1$arcA, the resultant strains were competed against KHOO1 and their competitive indices were calculated for 24 hours of the competition experiment. I hypothesised that if the rpoA mutation provided a fitness advantage at pH 4.5, then this would enhance the fitness of the ancestor. But if the advantage was contingent om arcA, then the advantage would manifest only in the presence of a loss of function of arcA in the background. The results are shown in figure 6.2.

The rpoA N294H mutation on its own showed on improvement in fitness at pH 4.5 (fig 6.3). The difference between the ancestral strain and ancestor::N294H strain was insignificant. This result established that the N294H mutation did not confer a fitness advantage at pH 4.5 by itself. But confirming to the hypothesis, the F$_1$arcA N294H strain showed a very similar competitive index to that of the fully evolved strains. The combination of arcA and rpoA was sufficient to recover a considerable amount of the fitness profile. The mean competitive index of F$_1$arcA N294H was compared to all the other strains showed in the figure using an unpaired student's t-test. The results have been tabulated in the table below.

Comparison	p-value	Statistical significance
E$_1$A vs F$_1$arcA N294H	0.0891	Insignificant
E$_2$A vs F$_1$arcA N294H	0.7753	Insignificant
E$_3$A vs F$_1$arcA N294H	0.4629	Insignificant
E$_4$A vs F$_1$arcA N294H	0.8273	Insignificant
E$_5$A vs F$_1$arcA N294H	0.8755	Insignificant
E$_6$A vs F$_1$arcA N294H	0.5612	Insignificant
Ancestor vs F$_1$arcA N294H	0.0005	Significant
F$_1$arcA vs F$_1$arcA N294H	0.0040	Significant
Ancestor::N294H vs F$_1$arcA N294H	0.0009	Significant

The results clearly showed that there was no significant difference between the competitive indices of the evolved strains and F$_1$arcA N294H. But there was a significant difference between F$_1$arcA and F$_1$arcA N294H. The result showed that the combination of loss of function of arcA and the rpoA mutation had an additive effect on fitness at pH 4.5 and were sufficient to confer a significant portion of the phenotype at pH 4.5
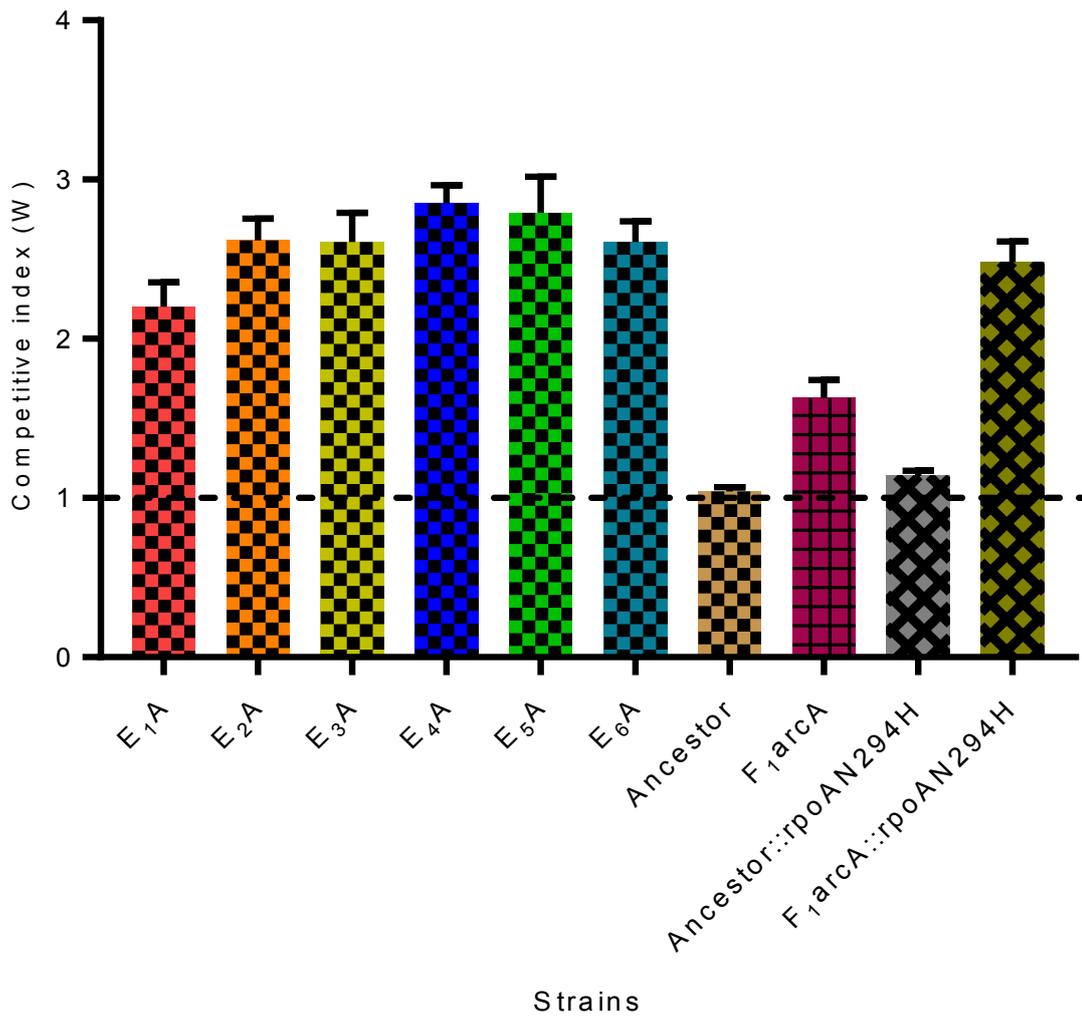
Figure 6.3: Competitive indices of the evolved strains (E_1A-E_6A), ancestor, F_1arcA, ancestor::rpoAN294H and F_1arcA rpoA::N294H at pH 4.5. The values are an average of three independent replicates and the error bars are standard error of the mean. The dashed line represents a competitive index of 1 and denotes competitive neutrality.

6.4 Measuring the RpoS activity of F$_1$*arcA* N294H

The results of the previous section showed that loss of function of *arcA* and the *rpoA* N294H mutation have an additive effect on fitness. To test whether this was true for the RpoS activity of these strains, we transformed the strains with the p21 fluorometric RpoS responsive promoter probe (Materials and Methods) and measured their RpoS activity over 480 minutes in LB medium at pH 4.5(fig. 6.4)

Section 4.7 in chapter 5 detailed the analysis which highlighted that F$_1$*arcA* and F$_1$*arcAΔcytR* showed intermediate RpoS activity. The fitness measurements in the previous section showed that the *arcA* and *rpoA* combination was sufficient to recover significant part of the phenotype of the evolved strains at pH 4.5.

The RpoS activity measurements correlated with that of the fitness measurements. The F$_1$*arcA* N294H strain showed a very similar RpoS activity profile to that of the evolved strains E$_4$A at pH 4.5 (fig 6.4). The ancestor::rpoA N294H showed no significant elevation in its activity profile when compared to the ancestor (fig. 6.4).

Figure 6.4 compares the final RpoS activity of the strains calculated at 480 minutes. The F$_1$*arcA* N294H attained similar end-point activity as that of E$_4$A. The combination of *arcA* and *rpoA* showed an additive effect on RpoS activity (fig. 6.5) and sufficient to elevate the RpoS activity to levels similar to that of the fully evolved strains.
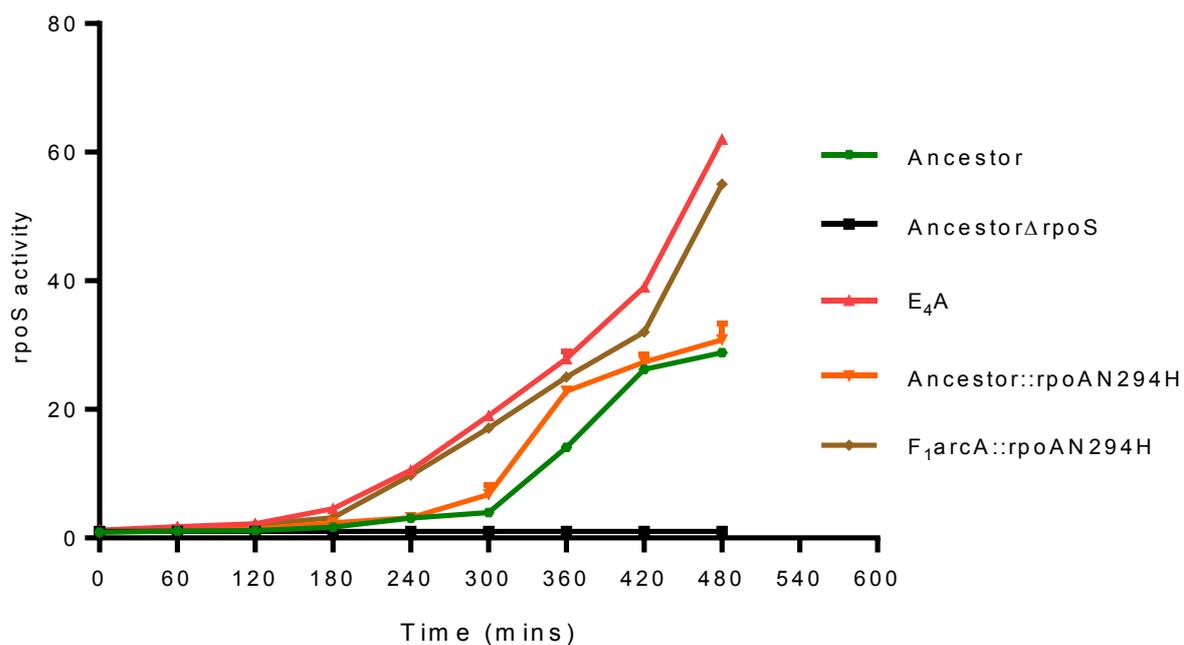
Figure 6.4: Absolute RpoS activity or fluorescence of the evolved strains, ancestor and ancestorΔrpoS at pH 4.5. All the strains are carrying the RpoS responsive p21 fluorometric plasmid probe. The absolute RpoS activity is calculated by divining the fluorescence measurement by the optical density measurement (Flu 515nm/OD 600nm) at each time point. The values are an average of three biological replicates and error bars are the standard error of the mean.
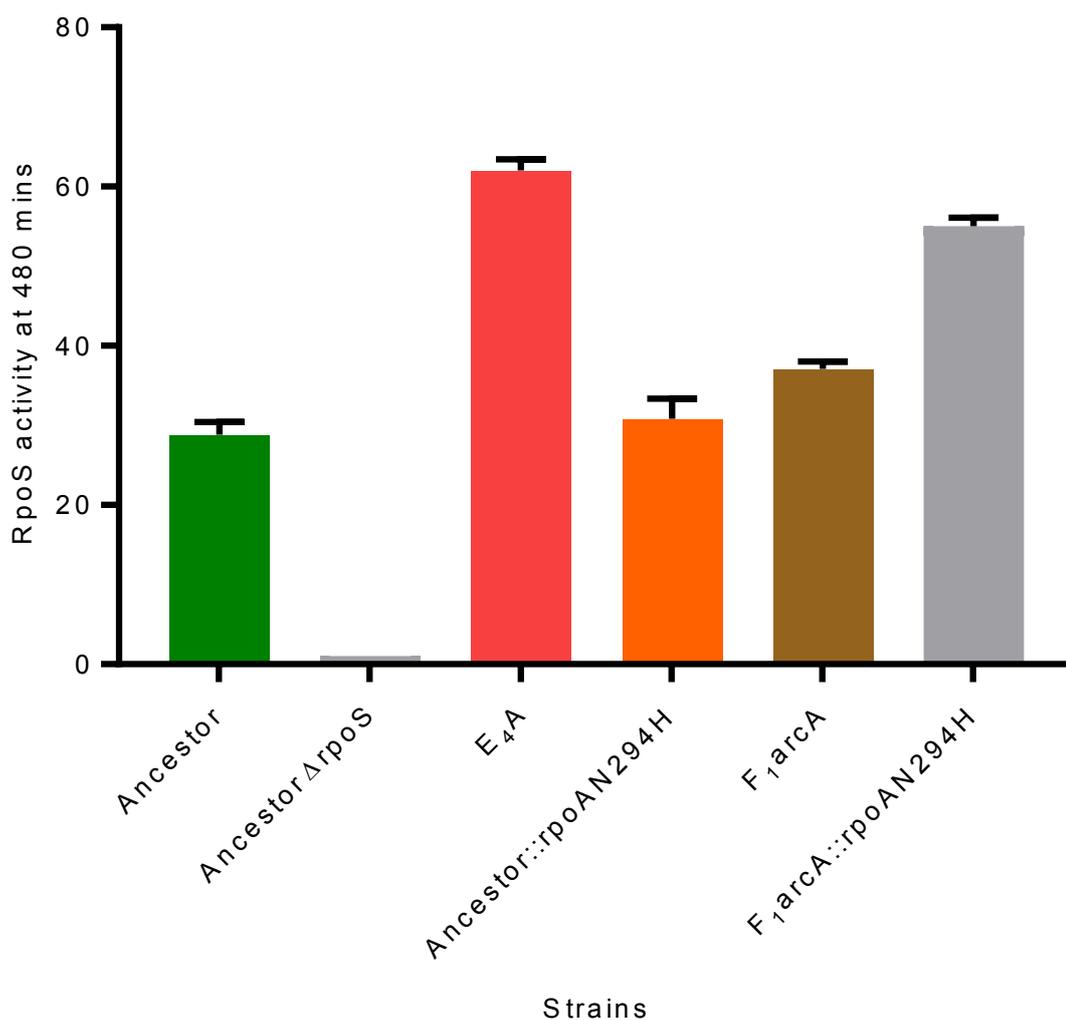
*Figure 6.5: Absolute RpoS activity or fluorescence of E4A, F₁arcA, ancestor::rpoAN294H, F₁arcA::N294H, ancestor and ancestor**Δ**rpoS at pH 4.5 recorded at 480 minutes. All the strains are carrying the RpoS responsive p21 fluorometric plasmid probe. The absolute RpoS activity or fluorescence is calculated by divining the fluorescence measurement by the optical density measurement (Flu 515nm/OD 600nm) at each time point. The values are an average of three biological replicates and error bars are the standard error of the mean.*

6.5 Competitive index and RpoS activity showed a positive correlation.

One of the key experimental conclusions on this study was the relationship between fitness and RpoS activity. The evolved strains were evolved under conditions of mild acid stress and their end-point phenotype showed that they were resistant to acid stress and other forms of stresses. Analysis of the genetic basis of adaptation showed that loss of function of *arcA* was partly responsible for the enhancement in fitness. Fitness contributions of other genes selected in the evolution experiment such as *cytR*, *flhD* and *flhC* were contingent on the loss of function of *arcA*. Investigations of the transcriptional landscape showed that a significant percentage of the changes in the transcriptional landscape of the evolved strains were driven by the upregulation of the *arcA* regulon. This is where the role of rpoS came into the picture.

There was an established connection between *arcA*, *rpoS* and the role of *rpoS* in regulating general stress resistance. Hence, it was logical that we investigated its fitness contribution in the evolved strains. The RpoS analysis brought forth three key conclusions. First the phenotype of the evolved strains was significantly *rpoS* dependent. Secondly, loss of function of *arcA* caused an intermediate elevation in *rpoS* activity and thirdly, the combination of *arcA* and *rpoA* showed an additive effect in elevating the RpoS activity comparable to the fully evolved strains.

The above arguments bring to light the overall role of RpoS in conferring the phenotype of the evolved strains. To quantitively represent this contribution, we wanted to evaluate the correlation between the RpoS activity of the some of the key strains analysed in this study and their competitive indices. This was done by plotting the RpoS activity of these versus their competitive indices (CI). A linear regression was constructed to determine the correlation between the two parameters.

The strains that were used in the comparison have been listed in Table 6.1 with their corresponding RpoS activities and competitive indices.

Table 6.1: List of strains used to construct the RpoS activity versus competitive index correlation

Strains	Competitive index	SEM	RpoS activity	SEM
Ancestor	1.008	0.091	28.836	0.608
E$_1$A	2.21	0.114	43.323	0.217
E$_2$A	2.56	0.201	52.377	0.256
E$_3$A	2.77	0.18	45.945	0.387
E$_4$A	2.83	0.17	52.724	0.445
E$_5$A	2.78	0.11	49.989	1.373
E$_6$A	2.68	0.103	50.892	0.767
AncestorΔarcA	1.62	0.11	35.65	0.672
E$_2$AΔrpoS	0.216	0.104	1.044	0.01
E$_4$AΔrpoS	0.311	0.128	0.108	0.071
F$_1$arcA	1.78	0.156	37.82	2.4406
Ancestor N294H	1.14	0.092	27.67	0.44
F$_1$arcA N294H	2.66	0.109	49.09	1.02

The comparison showed a significant correlation between RpoS activity and competitive index. The slope of the regression line was 0.927 which represent a high degree of correlation between the two parameters. The analysis showed four distinct clusters of strains. The first cluster (highlighted in blue) was that of the ΔrpoS strains, second (highlighted in maroon) was the ancestor, third (highlighted in black) were the ancestorΔarcA and F$_1$arcA and fourth (highlighted in red) were the evolved strains (E$_1$A-E$_6$A) and F1arcA::N294H.

The correlation showed that there was a step wise increase in RpoS activity between the loss of function of arcA and strains which had the arcA and rpoA mutations in combination. This correlation between RpoS activity and fitness has not been reported by other lab-based evolution studies focused on understanding E. coli's ability adapt to mild acid stress.
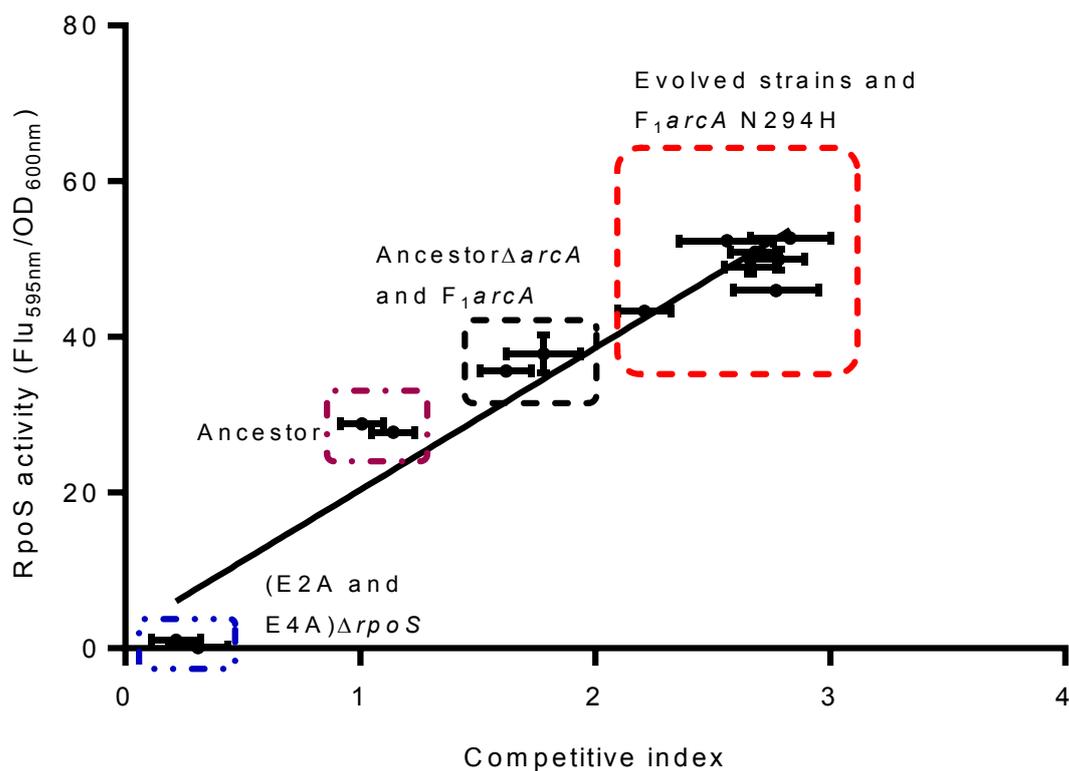
Figure 6.6: Correlation of RpoS activity of the strains listed in Table 6.1 and their competitive indices. RpoS activity was plotted on the Y-axis and the competitive indices on the X-axis. The horizontal and vertical error bars represent the standard error the mean of the two parameters. The values are an average of three independent replicates. Since the competitive indices (CI) are much smaller in value than the RpoS activities, the vertical error bars showing the standard error of the CI was not clearly visible in the plot.

6.6 Discussion for chapter 6

This chapter illustrated the results which showed one the key links between the genotype and phenotype of the evolved strains. Loss of function of *arcA* contributes to an intermediate increase in fitness at pH 4.5. This intermediate gain of fitness was further complemented by the *rpoA*N294H mutation to enhance the fitness to level observed in six fully evolved strains. The combination of loss of function of *arcA* and *rpoA*N294H was sufficient to recover significant portion of the phenotype displayed by the six evolved strains. We now understand this aspect of their genotype to phenotype link with considerable clarity. This combinatorial effect of mutations in *arcA* and *rpoA* which lead to increased *rpoS* activity in the cell has not been reported widely in the available literature on regulation of either of the three genes. Now that we understand there is a clear relationship between these three genes in conferring fitness at pH 4.5, we need to understand the functional basis of this relationship. The key questions to which need to be addressed to complete the overall picture are:

a) How does loss of function of *arcA* cause an elevation in *rpoS* activity?
b) What percentage of the genes differential expressed in the evolved strains are components of the rpoS regulon?
c) What does the transcriptional landscape of the F_1arcA strain look like?
d) Does loss of function of *arcA* cause an overall change in gene expression which is further modulated by the rpoAN294H mutation?
e) What are the differences in the transcriptional signatures of the strains F_1arcA and $F_1arcA::rpoA$N294H?
f) Can we deduce a functional route or regulatory pathway which can link the upregulation of the *arcA* regulon and elevation in *rpoS* activity?

These questions would need to be investigated to further our understanding of lab-based evolution of resistance to pH 4.5 and cross stress protection. We have done some preliminary experimental investigations to start answering some of these questions. One of the questions above is focused on understanding how does the loss of function of *arcA* causes an elevation in *rpoS* activity. The RNA sequencing analysis did not give a clear answer to that aspect of the regulatory changes observed in the evolved strains. Although differential gene expression did show the downregulation of *fliA* or σ^{28}, the flagellar sigma factor. Based on a preliminary review of the literature on *fliA*, we learnt that Kato *et al.,* 2007 had shown that a *arcA* mutant showed reduced expression of FliA. The reduction in expression of FliA was *arcB* independent.

Whether phosphorylated ArcA directly fingerprints fliA was not very clear from the papers considering the *arcA* regulon. Hence, an in-vitro transcription assay of *fliA* in the F_1arcA strain would help us understand if *arcA* had a direct role in the downregulation of *fliA* in the evolved strains. FliA is known to increase the cooperativity of the house keeping sigma factor with core RNA polymerase σ^{70}. FliA and σ^{70} compete for the core RNA polymerase complex during transcriptional activation. Since both these factors are significantly overexpressed compared to other sigma factors in *E. coli* (Antonio et al., 2008) they selectively initiate transcription of genes by occlusion of other transcription factors.

It is possible that the loss of function of arcA causes downregulation of *fliA* in the evolved strains. The downregulation of *fliA* increases competition between σ^{70} and other transcription factors such as *rpoS*. The *rpoA*N294H mutation alters the conformation of the RNA polymerase core complex to a state which favours rpoS binding to the core complex. This would allow *rpoS* to transcribe its target genes due to preferential occlusion of the σ^{70} dependent promoters. Whether the above hypothesis is valid would depend on studying the transcriptional signatures of the F_1*arcA* and F_1*arcA::rpoA*N294H strains.

This would highlight the core set of transcriptional changes which are caused by loss of function of *arcA* alone and the combination of loss of function of *arcA* and the *rpoA*N294H mutation. Preliminary results suggest that deletion of *fliA* in the ancestor marginally improves its fitness at pH 4.5. This increase in fitness is accompanied by a minimal elevation in rpoS activity. It is possible that downregulation of fliA is contingent on the presence of arcA as seen in case of the other target genes such as *cytR*, *flhD* and *flhC*. These aspects need further investigation in the future.

Although there are many key factors which need further investigation, what is very clear from our analysis of the evolved strains is the correlation between fitness and RpoS activity. The comparison between the two parameters showed a step wise increase between the ancestor, loss of function of *arcA* and the combination of *arcA* and *rpoA*. This correlation was not seen in previous evolution experiment conducted in the lab where *E. coli* strains were evolved to resist an extreme pH of 2.5. The extreme acid resistant strains had narrow phenotypic range and their phenotype was completely *rpoS* independent. Whereas the strains evolved at pH 4.5 show a completely different phenotypic profile to the extreme acid resistant strains. This goes to emphasise the power of lab-based evolution as a tool study how bacteria activate different modes of stress responses to varying environmental conditions.

6.7 Summary of results for chapter 6.

1) QRT-PCR analysis showed difference in the differential gene expression of *flgM* and *fimC* between the strains F_1arcA and E_2A.

2) The *rpoA* mutation N294H confers a gain of fitness to the evolved strains at pH 4.5.

3) The effect of the *rpoA* mutation is contingent on the presence of the loss of function of *arcA* in the background

4) RpoS activity and competitive fitness correlate each other in lab-based evolution of resistance to pH 4.5

Evolution of acid resistance to pH 4.5 in *E. coli* K12 MG1655

Chapter 7

Conclusion

7.1 Fitness phenotype at pH 4.5

In this study, we examined the ability of *E. coli* K-12 MG1655 to adapt to a moderately low pH value of 4.5. This pH is relevant with respect to bacterial passage through the intestine and the environment enteric pathogens would get exposed to in the lower intestine (Evans *et al.,* 1988). Following 740 generations of serial passage in unbuffered LB medium at pH 4.5, the six evolved populations displayed a significantly enhanced fitness phenotype at the selection pH. We isolated six clones or acid evolved strains which also displayed a similar phenotype at pH 4.5. The six acid evolved strains were also resistant to other forms of stresses showing a cross stress resistant phenotype. Our study showed that the fitness phenotype observed was much higher that the fitness values reported by other studies, such as Harden *et al.,* 2015, which have studied lab-based evolution of acid resistance. We also showed the fitness advantage observed at the selection pH was seen in exponential and stationary phase. The evolved strains attained significantly higher end-point culture density and log-phase growth rates. This showed that the conditions under which this evolution experiment was done, selected for mutants' fitter in both phases of the growth cycle. The evolved strains also showed enhanced resistance to other forms of stress. Evolution experiments previously conducted in the lab, did not report this cross stress protective behaviour of mutants selected to resist extreme acid stress pH 2.5 (Johnson *et al.,* 2015). This showed that changing the conditions under which evolution experiment was conducted selected for different phenotypic properties. This also suggests that the specificity of adaptation is influenced by the environmental conditions of the lab-based evolution experiment.

7.2 Role of *arcA* in increasing resistance to pH 4.5.

The six acid evolved clones we isolated, had different mutations in their genomes. These included single nucleotide polymorphisms, insertion sequence movements, small and large deletions. One of the target genes which was repeatedly selected across all the six evolved strains was *arcA*. As described in chapter 4, arcA is a global transcriptional regulator and controls the switching of carbon oxidation pathways during exposure to microaerophilic and anaerobic growth conditions. It is involved in regulating a significantly large number of operons in *E. coli* and has been shown to impact the expression of other stress response networks.

Our study showed that all the mutations in *arcA* were loss of function in nature. The loss of function of *arcA* conferred an intermediate gain of fitness at the selection pH. At the protein

level, some of the mutations seen in the evolved strains are in the dimerization interface. The mutations in the *arcA* may be causing loss of function by disrupting the structure-function relationship of the protein. It would be interesting to study some these mutations in isolation and understand their role in maintaining the functional integrity of the protein. Johnson *et al.,* 2014 showed that mutations in the sensor kinase which EvgS which conferred resistance to extreme acid stress were involved in disrupting the functional properties of the protein, One of the important conclusions by comparing the results to other lab-based evolution experiments, was that *arcA* is a common target in such experiments. It is possible that mutations in ArcA enable the cell to bring about global changes which enhances their phenotype against stress which otherwise would require the accumulation many beneficial mutations to climb their fitness peaks.

An important aspect which needs to be investigated is whether loss of function of ArcA would also confer increased resistance to pH 4.5 in other standard lab-strains of *E. coli* such BW25113 and W3110. Both these strains are more sensitive to acid challenge compared to MG1655. It is necessary to understand if loss of function of arcA aids to enhance their fitness against acid challenge. This could be done in parallel to making RpoS activity measurements to understand if there is a clear link between loss of function of ArcA, increased RpoS activity and enhanced fitness at low pH in these strains of *E. coli*.

This was clearly seen when the transcriptional landscape of the evolved strains was analysed. Analysis of the genes differential expressed showed had 125 of the most upregulated genes were part of the *arcA* regulon. These genes were mostly responsible for the energy dispensation and distribution in the cell. For example, the genes which are key components of the citric acid cycle were significantly upregulated in the acid evolved strains. This could cause an overall increase in the ATP pool of the cell. Increments in the cellular pool of ATP would lead to an increased activity of the proton pumps in the cell, which energetically remove protons. This would alleviate intracytoplasmic stress without having the need to express different acid resistance systems which aid in alleviating intracytoplasmic proton stress.

The transcriptional analysis of the evolved strains also showed the massive downregulation of the motility operons. This would be clearly the case for E_4A as it has a large deletion which causes the loss of the type 1 motility operon. But the downregulation of flagella in case of the other evolved strains was peculiar. Although, this is a feature which has been commonly reported in experimental evolution studies and explained as a mechanism to conserve energy.

This aspect will need further investigation, preliminary analysis suggests that *fliA* may be an important intermediary factor in conferring the fitness phenotype of the evolved strains. Some studies have suggested that arcA may play a role in regulating fliA, arcA mutants reported very low levels of FliA expression. Our study has showed this with respect to acid stress response and this aspect has not been reported before. As *arcA* doesn't directly fingerprint fliA, the downregulation of *fliA* could be through an indirect effect of loss of function of *arcA*. For purposes of further experiments, we need to investigate if the downregulation of *fliA* is contingent on loss of function of *arcA*. This could be done by making a *fliA* deletion in the F_1arcA strain and measuring its fitness relative to the ancestor.

Another key experimental hypothesis that was made from reflecting on the transcriptional landscape of the evolved strains was to investigate the role of *rpoS* in the six evolved strains. ArcA is a known repressor of RpoS and controls the stability of rpoS at various stages of RpoS expression. Loss of function of *arcA* showed an intermediate increase in fitness at pH 4.5. Our analysis showed that the evolved strains had enhanced RpoS activity and their phenotype was completely dependent on *rpoS*. We now also know that the evolved strains show increased levels of RpoS in the cell (data not shown in the thesis). Loss of function of arcA would directly impact the level of RpoS in the cell. This impact could be directly through gene regulation and by increasing the post translational stability of RpoS in the cell.

Increased stability of RpoS would allow for the preferential transcription of which are components of the RpoS regulon. This would inherently increase the overall stress resistance of the cell and thereby increase fitness at pH 4.5. One important comparison which needs to be done using the existing data on the transcriptional landscape of the evolved strains, is to deduce the differentially expressed genes that are components of the RpoS regulon. This will help us in understanding what percentage of the genes differentially expressed could be due to increased RpoS activity in the evolved strains. This could also be understood by analysing the transcriptional landscape of the F_1arcA strain. This would throw light on the differential expression of genes which would result from directly due to loss of function of *arcA* and help in connecting the link between loss of function of *arcA* and its effect on RpoS. This could be done by comparing RNA seq data of the F_1arcA with that of the fully evolved strains and extracting the core set of genes which are differentially expressed due to loss of function of *arcA*. Also, this would highlight the role of *arcA* in the overall adjustment of the different regulons which are upregulated and downregulated in the evolved strains. These experiments will allow us to deduce functional routes that connect the genotype and phenotype of the

evolved strains. This work is already underway and results will be incorporated in the overall findings of the study.

7.3 Mutation in RNA polymerase is crucial to the enhanced fitness at pH 4.5.

Five out of the six acid evolved strains had the identical mutation in the RNA polymerase subunit (α). Several lab-based evolution studies on E. coli stress responses have reported mutations in RNA polymerase subunits (Conrad *et al.,* 2011), (Palsson *et al.,* 2009). Our analysis of the evolved strains showed that the *rpoA* mutation was crucial to the fitness phenotype of the evolved strains. Our study highlighted two key principles of evolutionary biology. The first one being diminishing-returns epistasis (Lenski *et al.,* 1994), (Itou *et al.,* 1998), (Chou *et al.,* 2011). Our genetic analysis showed that the loss of *arcA* function conferred intermediate gain in fitness at pH 4.5. This intermediate gain of fitness couldn't be enhanced to levels observed in the fully evolved strains by making combination of mutants. The only route through which the overall phenotype at pH 4.5 could be recovered was by combining the *arcA* and *rpoA* mutation. The second principle which was highlighted was the role of historical contingency. The fitness effect of the *rpoA* mutation was completely contingent on the loss of function of *arcA*. This showed that there is important relationship between the two genes in enabling adaptation and evolution of resistance at pH 4.5.

There could be three possible mechanisms by which we could explain the effect the *rpoA* mutation on the phenotype of the acid evolved strains. (i) The RNA polymerase mutation could increase the stability of the core complex at moderately low pH. This would contain the integrity of the complex for transcription even when the cytoplasm is moderately acidified due to proton accumulation from the external medium. This is supported by the fact that our acid evolved strains also showed resistance permeant acids. (ii) The mutation could increase the specificity of RpoS towards the RNA polymerase holoenzyme. The mutation has been described in the literature to increase specificity of the core complex towards RpoS through occlusion of binding to the house keeping sigma factor, σ^{70}. This would increase preferential expression of genes which are transcribed by RpoS. This is supported by the upregulation of the genes *cadA* and *cadB*, which encode the lysine decarboxylases. The two genes are key acid resistance genes and transcribed by RpoS. (iii) The loss of function of *arcA* could increase the expression of different regulons which are repressed by arcA. The upregulation of these regulons would get modulated and adjusted to the required optimum by the *rpoA* mutation. The mutated locus is known to impact the expression of Class I CRP transcribed genes, most of the

genes which constitute the *arcA* regulon are class I transcribed in nature. This features in combination could lead to an overall increase in the RpoS activity of the evolved strains. These proposed mechanisms could be studied through Transposon transposon-directed insertion site sequencing (Tradis). We intend to make a Tradis library in one of the evolved strains and study its behaviour at pH 4.5. This would allow us to decipher genes which are key to the overall phenotype of the evolved strains.

7.4 Downregulation of the glutamate decarboxylase genes.

One of the key features of the transcriptional signatures of the evolved strains was the uniformly observed downregulations of the Gad genes. Two key regulators *gadX* and *gadW* were significantly downregulated across all the evolved strains. These two genes are key regulators of *gadE*. GadE is the transcriptional activator of GadA and GadB. Previous work done in the lab and other studies have shown that the Gad system is critical during exponential phase acid stress as well as stationary phase. Since, this system is such a key component of the acid response unit of *E. coli*. It was surprising that key regulators of this system would be downregulated during in course of adaptation to mild acid stress. The downregulation of the Gad system was also reported by Penix et al., 2017. The authors investigated the transcriptional landscape of *E. coli* W3110 strains evolved at pH 4.6 and observed a similar downregulation of the Gad genes.

The strains evolved at pH 4.5 in this study, showed an overall increase in the general stress resistance. Transcriptionally, these strains showed significant upregulation of the genes which are involved in the bioenergetics and metabolism of cell. The overall increased expression of this genes would create an energy surplus, which can be employed in combating variety of stress conditions. The expression of the Gad system on the other hand is an energy consuming process as the end goal of the activation of the Gad system is remove protons from the cytoplasm by ATP consumption. It is possible that the evolved strains downregulated this key system as the energy requirement would not be beneficial in the conditions to which they were exposed to. The downregulation of the Gad system can compensate for the energy required to upregulate other essential cellular processes which could be enabling the increment in the general stress resistance of the evolved strains. The downregulation of the Gad genes can be a direct consequence of the increased RpoS activity of the evolved strains. RpoS is a known to downregulate the Gad system and the conditions under which the evolved strains were

propagated may have selected for the selective downregulation of the Gad system. This may serve as a trade-off to gain fitness at pH 4.5 and increased general stress resistance

7.5 Future directions

This study has highlighted and re-emphasized the role of lab-based evolution as a tool to study bacterial stress responses. Understanding the different ways by which *E. coli* responds to acid stress is very important with respect to clinical microbiology. Our study has clearly shown that evolution of *E. coli* under mild acid stress leads to activation of systems which are involved in controlling the global energy balance and metabolism of the cell. This very important with respect to understanding how enteric bacterium survive and respond to stress in the human gut. Our experiments clearly indicated that under the conditions in which this evolution experiment was conducted, cells presumably also experienced anaerobic stress. This possibly lead to ArcA being active and loss of function of *arcA* lead to evolution of fitter mutants. This is physiologically relevant as enteric bacteria would face similar conditions in the human mid-gut, where the conditions would couple acid and anaerobic stress. Taking inferences from this study, the role ArcA needs to be further investigated in enabling survival in the intestine. Two other studies have shown that *arcA* mutants then to be fitter in mouse colonisation studies (Lasaro *et al.,* 2014 and Jones *et al.,* 2011). The results of our study establish this fact from the view of stress response in *E. coli*. Tis aspect needs to be further studies and investigated to understand the role of global regulators in enabling adaptation under conditions of stress.

Structure-function relationship of the acid sensor kinase of *Escherichia coli*, EvgS

Chapter 8

Sen et al., 2017

Structural and Functional Analysis of the *Escherichia coli* Acid-Sensing Histidine Kinase EvgS

Hrishiraj Sen,[a] Nikhil Aggarwal,[a] Chibueze Ishionwu,[a] Nosheen Hussain,[a] Chandni Parmar,[a] Mohammed Jamshad,[a] Vassiliy N. Bavro,[b] Peter A. Lund[a]

School of Biosciences, University of Birmingham, Birmingham, United Kingdom[a]; School of Biological Sciences, University of Essex, Colchester, United Kingdom[b]

ABSTRACT The EvgS/EvgA two-component system of *Escherichia coli* is activated in response to low pH and alkali metals and regulates many genes, including those for the glutamate-dependent acid resistance system and a number of efflux pumps. EvgS, the sensor kinase, is one of five unconventional histidine kinases (HKs) in *E. coli* and has a large periplasmic domain and a cytoplasmic PAS domain in addition to phospho-acceptor, HK and dimerization, internal receiver, and phosphotransfer domains. Mutations that constitutively activate the protein at pH 7 map to the PAS domain. Here, we built a homology model of the periplasmic region of EvgS, based on the structure of the equivalent region of the BvgS homologue, to guide mu- tagenesis of potential key residues in this region. We show that histidine 226 is re- quired for induction and that it is structurally colocated with a proline residue (P522) at the top of the predicted transmembrane helix that is expected to play a key role in passing information to the cytoplasmic domains. We also show that the constitu- tive mutations in the PAS domain can be further activated by low external pH. Ex- pression of the cytoplasmic part of the protein alone also gives constitutive activa- tion, which is lost if the constitutive PAS mutations are present. These findings are consistent with a model in which EvgS senses both external and internal pH and is activated by a shift from a tight inactive to a weak active dimer, and we present an analysis of the purified cytoplasmic portion of EvgS that supports this.

IMPORTANCE One of the ways bacteria sense their environment is through two-component systems, which have one membrane-bound protein to do the sensing and another inside the cell to turn genes on or off in response to what the membrane-bound protein has detected. The membrane-bound protein must thus be able to detect the stress and signal this detection event to the protein inside the cell. To understand this process, we studied a protein that helps *E. coli* to survive ex- posure to low pH, which it must do before taking up residence in the gastrointesti- nal tract. We describe a predicted structure for the main sensing part of the protein and identify some key residues within it that are involved in the sensing and signal- ing processes. We propose a mechanism for how the protein may become activated and present some evidence to support our proposal.

KEYWORDS *Escherichia coli*, acid resistance, histidine kinase, periplasm, signal transduction

Among the many different two-component systems (TCS) that detect signals and regulate gene expression in bacteria, one intriguing group is typified by histidine kinases (HKs) that possess one or more periplasmic Venus flytrap (VFT) domains (1–3). VFT domains are found in proteins across all kingdoms. In bacteria, they are largely a feature of amino acid binding proteins, particularly periplasmic binding proteins (PBPs) that bind solutes for subsequent import into the cell via transporters. Such proteins

have been classified in various ways based on their structure, and such classifications correlate well with the type of solute that is bound (4–6). VFT domains contain two globular subdomains or lobes containing mostly alternating alpha-helices and beta-sheets that flank the binding site of the ligand, and they can exist in open or closed states. Transition in the equilibrium between the two states is caused by the binding of the ligand, which causes "closure" of the trap (7). Although mostly associated with PBPs, VFT domains are also found on many membrane-bound HKs, with the number ranging from one to five. Many of these HKs are members of TCS in pathogens that are associated with virulence, and their ligands are unknown (1, 3).

The best characterized of this large group of HKs is the BvgS protein found in the bordetellae. This protein regulates the expression of virulence factors, including toxins and adhesins, via the response regulator BvgA, and it has a large number of genes in its regulon (8). The natural ligand for BvgS is unknown, and while it is constitutively active in cells grown at 37°C, it is switched off by low temperature or exposure to high concentrations of sulfate ions or nicotinic acid (reviewed in reference 1). The protein has a molecular mass of approximately 135 kDa, of which nearly 60 kDa is in the periplasm, linked to the cytoplasmic part by a single transmembrane helix. It is dimeric, and the structure of the periplasmic domain has been solved (2, 3), revealing two VFT domains in each protomer, wrapped around each other with one in the open and one in the closed conformation. Downstream of the transmembrane helix is a PAS domain, a common feature of many regulatory proteins, including HKs, which in turn is followed by phospho-acceptor (carrying the target histidine residue, H729) and kinase domains. Unusually compared to most TCS kinases, the phosphate is transferred to an aspartate, D1023, in an internal receiver domain before being transferred to a final histidine, H1172, in a histidine phosphotransfer domain (9–11). From here, it is ultimately transferred to the receiver aspartate in BvgA. This pattern of internal phosphotransfer in so-called unorthodox HKs has been proposed to make them more sensitive to the signal within a critical range and more robust against noise (12, 13). A detailed mechanism relating movement of domains in the periplasm to a subsequent alteration in the dynamics of a coiled-coil region immediately after the PAS domain has recently been proposed, which may be generic to all proteins with this domain structure (14).

In *Escherichia coli*, the closest homologue to BvgS is the EvgS HK, which shares the same domain organization. The *evgS* gene and the gene for its cognate response regulator, *evgA*, were first cloned as multicopy suppressors of a Δ*envZ* mutation (15). Early studies showed that EvgA binds to the region upstream from the *evgAS* operon and in doing so regulates both the operon and the divergently transcribed *emrKY* genes, which encode a multidrug efflux pump (16, 17). Overproduction of EvgA leads to elevated resistance to multiple drugs and toxins, which was attributed to upregulation of the EmrKY and the YhiUV (subsequently renamed MdtEF) transporters (18–20). EvgA overproduction also leads to the expression of an acid resistance phenotype in exponential phase, a finding that enabled the identification of several new genes implicated in acid resistance in *E. coli*, including the one encoding the central regulator GadE, as well as the elucidation of part of the complex network of interactions governing expression of the glutamate-dependent acid resistance mechanism of *E. coli*, AR2 (also called the GAD system) (21–24). Deletion of either *evgA* or *evgS* leads to the complete loss of acid induction of all the key genes in the AR2 network (25, 26). Evolution experiments with selection for long-term survival at pH 2.5 in exponential phase led repeatedly to the isolation of mutations that rendered EvgS constitutively active at pH 7, and mutants of EvgS selected for their constitutive activation of an *emrK-lacZ* fusion also showed an acid resistance phenotype (17, 25, 27). Thus, the *evgAS* regulon is important both in acid resistance and in drug efflux in *E. coli*.

Comparative analysis of the transcriptomes of strains expressing different constitutively active EvgS proteins shows a considerable overlap of upregulated genes, including the genes in the acid fitness island, genes adjacent to the *evgAS* operon (involved in drug efflux and in oxalate decarboxylation), and components of cytochrome *bd-II* oxidase and periplasmic hydrogenase. Activation of the regulators GadE, GadX, GadW,

and YdeO by phosphorylated EvgA has a complex and central role in activation of the main structural genes of the glutamate-dependent acid resistance system (*gadA*, *gadB*, and *gadC*) plus other genes in the acid fitness island that encode periplasmic chaperones and the outer membrane lipoprotein Slp, as well as a range of unlinked genes, including those for the flagellar protein FliC and the large and small subunits of glutamine synthetase (25, 26, 28, 29). In pathogenic *E. coli* O157:H7, the GadE regulon has also been shown to include the Ler protein, which itself represses many of the genes in the locus of enterocyte effacement (LEE) pathogenicity island, and mild acidification leads to downregulation of LEE gene expression (30). Thus, the EvgAS TCS has the potential, through its function in regulation of GadE and other central acid response regulators, to have important roles in *E. coli* pathogenicity. Indeed, overproduction of EvgA has been found to reduce expression of the components of the type three secretion system encoded by the LEE region (31), while *evgS* mutants are attenuated in an avian lung infection model with an avian-pathogenic strain of *E. coli* (32).

Our previous study on EvgS showed that many different mutations in the PAS domain led to constitutive activation of the protein and that deletion of part of the periplasmic domain led to loss of inducibility by low pH but no loss of activity of these constitutive mutants (27). We proposed a model based on these data whereby detection of the inducing signal takes place in the periplasmic domain. This results in transmission of an activating signal across the inner membrane, which in turn leads to the weakening of subunit interactions in a tight and inactive dimer of EvgS. By this model, the active form is proposed to be a loose dimer, and the activating mutations in the PAS domain are proposed to also weaken interactions in the dimer, thus mimicking the activation of EvgS that normally occurs only on external mild acidification. Here, we have constructed a model structure of the EvgS periplasmic domain, based on the published BvgS periplasmic domain structure, to guide site-directed mutagenesis of residues in the periplasmic domain. We report the identification of key periplasmic residues needed for low-pH induction of EvgS. We also show that the cytoplasmic domain alone can still mediate some pH-dependent gene expression, and we propose from this finding a more complete model for how EvgS may function, with some supporting biochemical evidence.

RESULTS

Prediction of the structure of the EvgS periplasmic domain: do amino acids influence induction of EvgS activity? The BvgS and EvgS proteins share a common domain organization and exhibit a high degree of sequence similarity (28.8% identity). Both are examples of unconventional HKs which have internal receiver domains to which the phosphate is initially transferred after autokinase activity has phosphorylated a histidine residue in the HK domain, rather than the more common direct transfer to a response regulator. The phosphate is subsequently passed to a second histidine in the histidine phosphotransfer domain for eventual transfer to an aspartate in the cognate response regulator protein (33, 34). Both proteins have unusually large periplasmic domains, each containing two VFT domains. The recent availability of high-resolution structures of the BvgS periplasmic domains (the isolated VFT2 domain [PDB 3MPK] [2] and the dimeric complete periplasmic portion of the protein [PDB 4Q0C] [3]) allowed us to construct high-fidelity homology models of the EvgS periplasmic domain. The available structure of BvgS covers residues 33 to 544 of the mature protein, corresponding to residues 25 to 538 of *E. coli* EvgS (NCBI reference sequence WP_072282315), after which the protein is predicted to cross the inner membrane via a single transmembrane helix. These periplasmic regions have 25% identity (122/497 residues) and 46% similarity, with only three single-residue gaps and one two-residue gap in the alignment produced using MAFFT (see Fig. S1 at http://epapers.bham.ac .uk/3021/), allowing unequivocal assignment of all the secondary structure elements and the building of reliable homology models. A range of models was created using the I-TASSER modeling suite, yielding models with a C-score value between +0.75 and
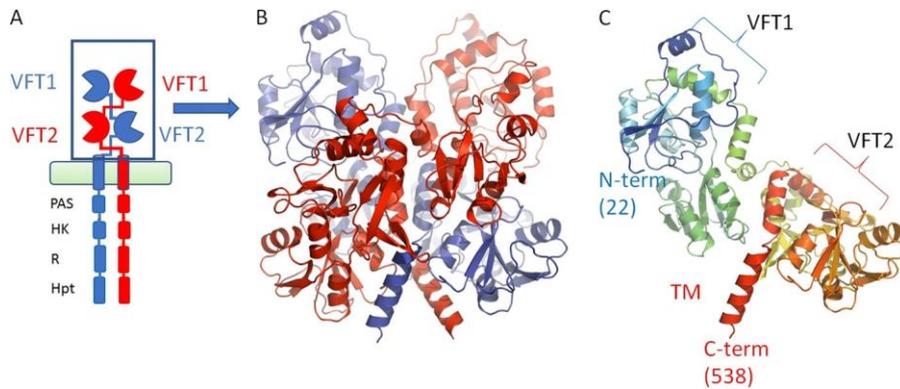
FIG 1 Schematic of EvgS showing the domain organization and the modeled section discussed in the text. (A) Domain organization of EvgS, showing the membrane and intracellular kinase domains. (B) The periplasmic domain is composed of two Venus flytrap domains, VFT1 and VFT2, which dimerize. The different chains are colored red and blue, respectively. (C) A single protomer showing the domain organization and colored in rainbow from N terminus (blue) to C terminus (red). The homology model presented covers residues 22 to 538 of the full-length protein.

+0.95, indicative of high confidence and strong predictive power. The highest overall scoring model was used for analyzing the structural properties of EvgS and making selections for the residues to be targeted by point mutagenesis as described below.

As we used the BvgS structure as the primary model template, the predicted EvgS periplasmic domain is compatible with the dimeric organization of BvgS, having two VFT domains in each protomer, with a 2-fold symmetry axis parallel to the long axis of the dimer. The peptide chain shows broadly the same path in each protomer and crosses twice between the sides of the dimer before passing into the alpha-helix, which is predicted to cross the membrane (Fig. 1A and B). The conformation of the EvgS model is also consistent with the state of the BvgS in the crystal structure, and correspondingly each monomer in the model structure contains the two distinct VFT domains, with VFT1 in an open conformation and VFT2 in a closed conformation. While the conformation presented is highly probable and is compatible with the EvgS sequence based on its close homology to BvgS, one needs to keep in mind that this conformation is chosen due to model bias and is impossible to validate directly at this point. Figure 1C shows a single protomer to make the arrangement of the two domains in the complex clearer. A detailed description of the characteristics of the full structure can be found in reference 3.

To see what biological insights we could get from homologous structures, we searched the Protein Data Bank (PDB) database for the closest structural matches to the EvgS structure other than the template BvgS structure itself. From this search we identified PDB 4PRS and 4PSH, which are structures of an arginine binding protein that works with an ABC transporter from the extreme thermophile Gram-negative bacterium *Thermotoga maritima* (35) without and with a bound arginine, respectively. The holo- structure reveals a well-defined arginine moiety coordinated by a number of residues, including R82 (as numbered in the PDB file), and closely resembles the VFT2 domains of both BvgS and EvgS (see Fig. S2 at http://epapers.bham.ac.uk/3021/). The same type of coordination to bound amino acids is seen in other examples, which all show significant structural homology with the VFT2 domain, including 3VVE (a protein involved in lysine transport in *Thermus thermophilus*) (36), 4H5F (an arginine binding ABC transporter from *Streptococcus pneumoniae*) (P. J. Stogios et al., unpublished reference from the PDB entry), and 3LSL (glutamate binding domain of GluA2, human AMPA receptor) (37). The coordinating arginine residue common to all these structures is conserved in both BvgS (R380) and EvgS (R375), although its removal in BvgS has been shown to have no effect on BvgS activity (2).

The binding of glutamate (shown in Fig. S3 at http://epapers.bham.ac.uk/3021/) was of interest, as EvgS activates the AR2 system in *E. coli*, which requires the import of
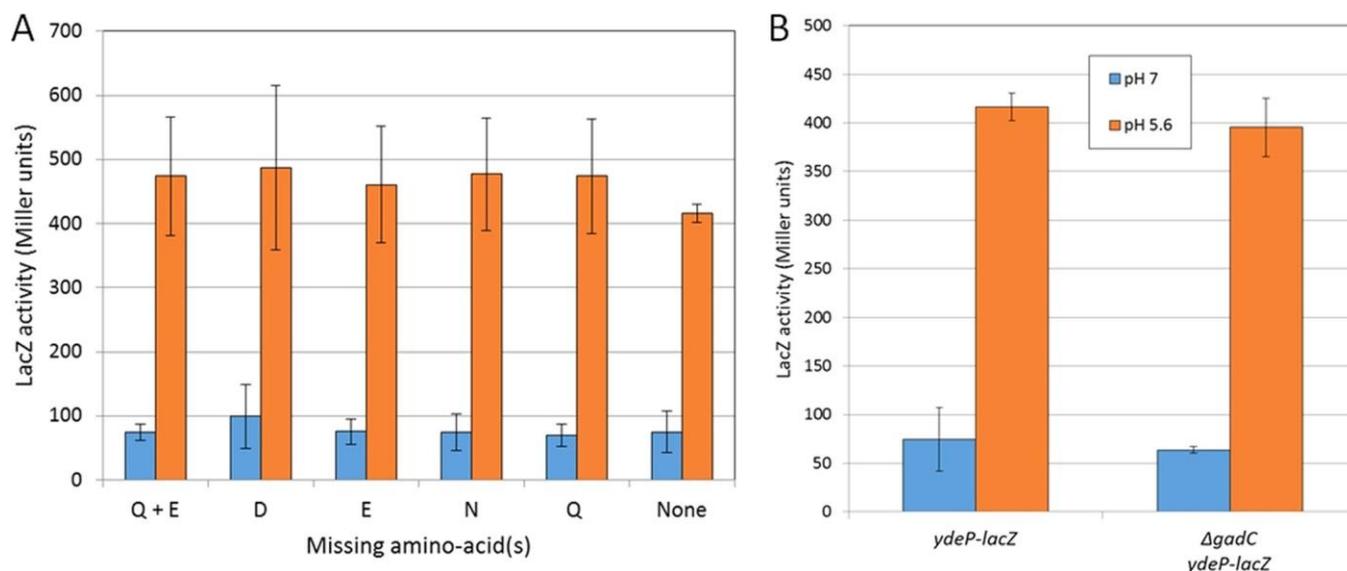
FIG 2 EvgS induction does not require amino acid binding or activity of the GadC glutamate transporter. (A) Activity of the *ydeP-lacZ* reporter at pH 7 and pH 5.6 in the absence of selected amino acids (the induction ratios for all amino acids are shown in Fig. S4 at http://epapers.bham.ac.uk/3021/). (B) Activity of the *ydeP-lacZ* reporter at pH 7 and pH 5.6 in the presence and absence of the GadC antiporter.

external glutamate as a substrate via the GadC antiporter (28, 29), and this led us to speculate that binding of glutamate might be required for EvgS activation. We tested this possibility by assaying EvgS-mediated induction of a *ydeP-lacZ* promoter fusion as described in Materials and Methods but with different amino acids or groups of amino acids omitted from the growth medium. No effects on induction were seen when different combinations of glutamate, glutamine, aspartate, or asparagine were omitted from the growth medium (Fig. 2A). Data for other groups of amino acids are summa- rized in Fig. S4 at http://epapers.bham.ac.uk/3021/ and also show no effect. We also checked to see whether glutamate transport might be necessary for EvgS activation, as there are precedents for TCS being activated by the transport of their ligand (for a recent review, see reference 38). To do this, we deleted the gene for the GadC antiporter from the reporter strain by P1 transduction. As shown in Fig. 2B, this deletion also had no effect on EvgS-mediated induction of *ydeP-lacZ*. Thus, despite the close structural link of EvgS to the amino acid binding proteins, we have detected no amino acid influence by either binding or transport over the activation of EvgS, although we cannot rule out the possibility that one or more amino acids bind to VFT2.

Potential roles of periplasmic histidine residues in EvgS activity. The VFT domain closer to the N terminus of the protein, VFT1, was also investigated by initially looking for structural clues. We superimposed the arginine binding region of ArgBP as defined for 4PSH over this domain. Structural similarity could be seen, but the root mean square deviation (RMSD) in this case was over 4 Å, as opposed to 2.3 Å for the overlay with VFT2, showing that VFT1 is more divergent structurally from ArgBP. When we analyzed VFT1 from the model and ArgBP more closely, two points were noted. The first was that there is no conserved arginine corresponding to R82 (the centrally important residue which coordinates the carboxyl ring of the ligand in all the amino acid binding proteins) in this domain. Instead, EvgS has T127 and S128 in the corre- sponding positions, which would make it impossible to coordinate a canonical amino acid in VFT1. The second is a striking cluster of three histidine residues which protrude into the binding cleft of VFT1 (H63, H106, and H124). Both of these features are shown in Fig. S5 at http://epapers.bham.ac.uk/3021/. Although the histidines do not sterically clash with the position that a bound amino acid would take, the VFT1 domain is in its open conformation in this structure, whereas if a ligand were to bind, it would change to the closed conformation, potentially increasing the impact of this histidine triad on the VFT1 binding domain. Taken together, and allowing for the uncertainty associated
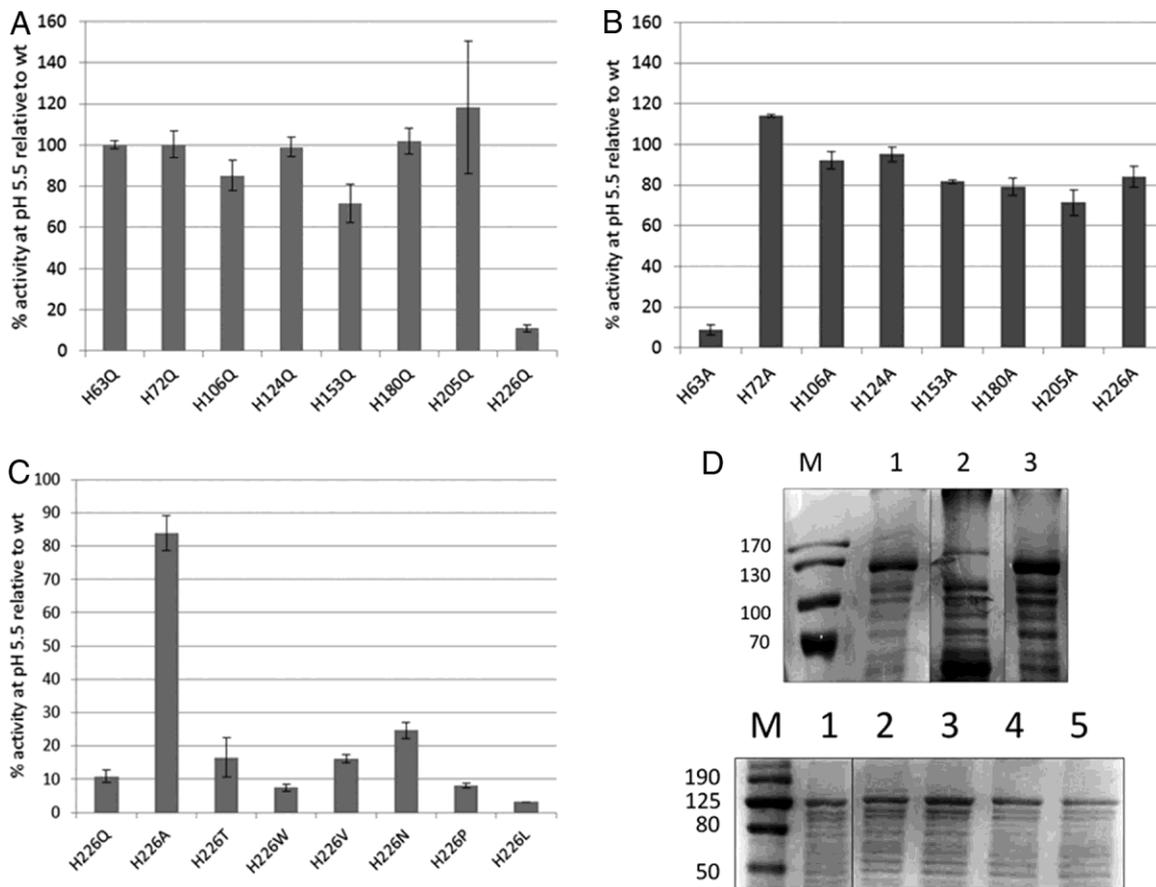
FIG 3 (A to C) Induction at pH 5.6 of EvgS proteins mutated at different histidine residues in the periplasmic domain, expressed as a percentage of the induction seen with wild-type EvgS. (A) Histidine to glutamine. (B) Histidine to alanine. (C) H226 to multiple different amino acids. (D) Top gel, membrane fractions from cells expressing wild-type EvgS (lane 1), EvgS H63A (lane 2), or EvgS H226Q (lane 3). Bottom gel, membrane fractions from cells expressing wild-type EvgS (lane 1), EvgS H226A (lane 2), EvgS H226V (lane 3), EvgS H226W (lane 4), or EvgS H226P (lane 5). Markers (lane M) are labeled in kilodaltons. Vertical lines on the gel images show where additional lanes run on the original gels have been spliced out for clarity.

with modeling bias, these two observations make it less likely that an amino acid-like ligand might stably bind to the VTF1 domain.

The histidine triad presents an interesting possibility as a potential pH sensor. Histidine residues are good candidates for detecting changes of pH due to their protonation under mild acid conditions. The optimum pH for induction of EvgS activity has been shown to be around 5.5 to 5.7 (26, 39), and the pK_a of histidine residues can often fall in this range (40). To test this possibility, particularly for the histidine triad but also for all the other histidine residues in VFT1, we constructed mutations where we individually replaced all the histidines in the VFT1 domain with alanine or glutamine and measured their impact on EvgS-induced *ydeP-lacZ* activity. The data (Fig. 3A and B) show that all histidines except H63 could be changed to alanines without loss of induction and that all except H226 could be changed to glutamines without loss of induction.

One plausible reason for loss of induction is that the mutation leads to a failure to correctly fold the protein or to loss of targeting of the protein to the membrane. To check whether this possibility was an explanation for our data, we prepared crude membrane fractions of strains expressing wild-type EvgS and the two noninducible mutants EvgS H63A and EvgS H226Q and compared them on SDS-PAGE. As can be seen in Fig. 3D (top gel), no band corresponding to EvgS H63A could be detected, whereas EvgS H226Q was clearly visible in membrane fractions, as was the wild-type protein. This result shows that loss of induction of H63A is almost certainly a trivial consequence
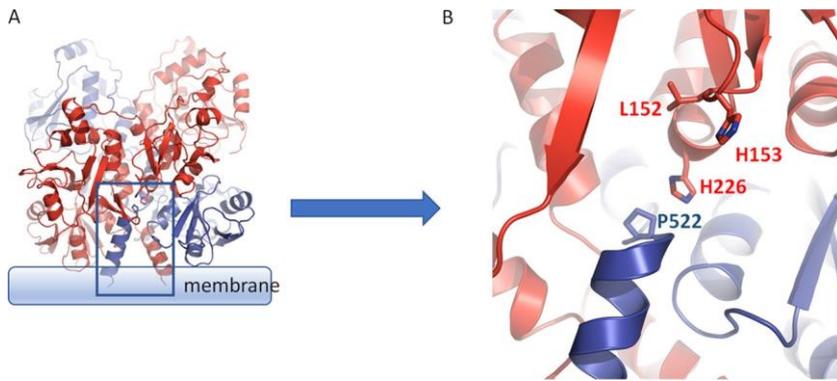
FIG 4 Residues in the interdomain of the VFT1 and VFT2 interface are crucial for EvgS activation and may be involved in signal transduction involving Pro522. (A) A general view of the region discussed. (B) Image rotated and zoomed in to show individual amino acid side chains.

of the absence of the protein, but loss of induction for H226Q must have a different explanation.

The fact that EvgS H226Q cannot be induced at low pH but EvgS H226A can suggests that while this residue is likely to play a critical role in some aspect of signal detection or transduction, protonation is not involved in this process. To investigate this further, we made another series of mutations at this position (to T, L, V, W, N, and P) and confirmed that all of the mutants except H226A showed a significant decrease in inducibility at pH 5.6 relative to that of the wild-type protein (Fig. 3C). Several of these mutants were tested for EvgS protein levels in membrane fractions, and all still produced a visible protein band (Fig. 3D, bottom gel), confirming that this residue is crucial for EvgS activity.

None of the mutations in any of the histidine residues in the potential triad led to loss of inducibility of EvgS, with the trivial exception of H63A, which led to loss of the protein. To test whether functional redundancy might be built into this triad, we constructed all possible pairwise combinations of histidine to glutamine and the triple H63Q H106Q H124Q mutant. As shown in Fig. S6 at http://epapers.bham.ac.uk/3021/, all of these mutants still showed the same level of induction as the wild-type EvgS, ruling out any role of this triad in detection of the inducing signal.

Identification of a potential interaction between H226 and P522 at the top of the predicted transmembrane helices. We examined the position of the critical H226 residue in the modeled structure of EvgS. H226 is one of two histidines (the other being H153) that protrude in an antenna-like fashion from each monomer when those are considered in isolation. In the context of the dimer, however, H226 is deeply embedded in the structure and makes several possible intraprotomer and interprotomer contacts which may relate to its function. In particular, we have noted that H226 is located in close proximity to a proline residue (P522) which we predict is likely to cap the predicted transmembrane helix of the opposite protomer (Fig. 4), though it should be noted that the confidence of the model in this region (positions 521 to 525) is somewhat lower, with a gap in the template (BvgS) alignment due to an extra residue in the EvgS sequence after position 523. P532 in BvgS clearly caps alpha-helix 17, which leads to the transmembrane domain. If we extend the alignment from the transmembrane domain backwards, the residue P522 in EvgS is found before an extra predicted helical turn, to cap the corresponding helix 17. The proximity of P522 to a number of other residues suggests that it could play a role as the transducer of conformational switches both from intraprotomer transitions (e.g., opening and closing of VFT domains) and interprotomer transitions. Based on these predictions and to test whether P522 is indeed also needed for induction of activity of EvgS, we constructed a P522A mutant and analyzed its inducibility. We also noted that another residue that maps to the same locality and that protrudes from each monomer toward the neighboring
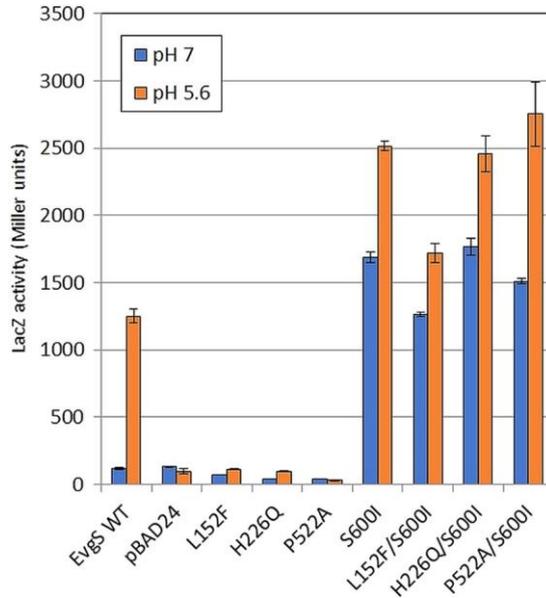
FIG 5 Impact of the periplasmic L152F, H226Q, and P522A mutations on EvgS-mediated induction of the *ydeP-lacZ* reporter at pH 5.6, alone and when combined with the constitutively activating S600I mutation.

protomer is L152, adjacent to H153, which has previously been mutated to F and identified as being required for activation of EvgS (39). We reconstructed this mutant and tested it in the same analysis as a positive control. The results (Fig. 5) confirmed that none of these mutants showed any detectable activity above background at pH 7 or pH 5.6. Due to difficulties in producing reproducible results when trying to detect these mutated proteins in membrane fractions, we added the activating mutation S600I to the cytoplasmic PAS domain of each mutant to see if they regained activity. S600I is one of the mutations whose presence makes EvgS constitutively active at pH 7, and we have previously shown that if S600I is introduced into an EvgS mutant that has been rendered noninducible at pH 5.6 by deletion of part of the periplasmic domain, the protein shows constitutive activity (27). Thus, by introducing the S600I mutation into EvgS proteins that carry single-amino-acid mutations in the periplasmic domain, we can distinguish mutated proteins that may be misdirected, misfolded, or degraded (and which hence will have no activity) from those that do reach the membrane but are not activated as normal by low external pH. In each of the three cases tested (L152F, H226Q, and P522A), the addition of the S600I mutation to EvgS that contained these mutations in the periplasmic domain led to levels of activity at pH 7 which were very similar to that of EvgS S600I. We also noted that in EvgS S600I and in each of the three double mutants, further induction (an average of 1.51-fold [standard deviation, 0.21]) was seen when cells were incubated at pH 5.6 (Fig. 5). The periplasmic mutations thus do not compromise the potential for EvgS to be active, but they do render it noninducible by low pH. However, they do not alter the constitutive activity at pH 7 or the inducibility at pH 5.6 of EvgS S600I when they are introduced into it, a point that is considered in the next section. Possible roles of the H226 and P522 residues are considered further in Discussion.

Cytoplasmic EvgS shows pH-dependent dimer formation *in vitro*. The inducibil- ity at pH 5.6 of EvgS S600I, whether the mutation is present on its own or combined with a mutation in the periplasmic domain that blocks induction, could result from the detection of a signal associated with low pH taking place at the membrane or in the cytoplasm. The first explanation is rendered less likely by the fact that a membrane-anchored form of EvgS lacking the whole periplasmic domain has already been studied in the same assay system as used here and shows no activity at either pH 7 or pH 5.6 (39). We therefore decided to express the entire cytoplasmic region of EvgS with no
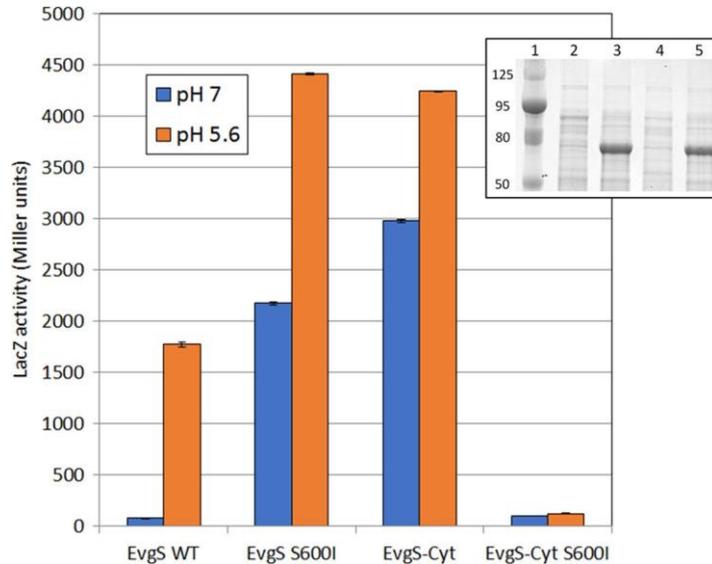
FIG 6 Comparison of activities of wild-type EvgS and EvgS-Cyt without and with the activating S600I mutation in the cytoplasmic PAS domain. The inset shows a 12.5% SDS-polyacrylamide gel of crude extracts from the reporter strain expressing EvgS-Cyt (lanes 2 and 3) or EvgS S600I-Cyt (lanes 4 and 5), without and with induction, respectively; lane 1 shows molecular mass markers in kilodaltons.

membrane attachment to see whether it still showed any pH-dependent increase in activity. The construction and expression of this protein, which we call EvgS-Cyt, is described in Materials and Methods.

It is quite common for the cytoplasmic portions of TCS HKs to show constitutive activity when expressed without the transmembrane and periplasmic regions (see, e.g., references 41 and 42), and this proved to be the case with EvgS-Cyt, in that the *ydeP-lacZ* reporter showed high activity at pH 7 in the presence of EvgS-Cyt and the absence of any full-length EvgS. Moreover, as observed with the constitutive mutant EvgS S600I, we saw a further induction of activity when cells expressing EvgS-Cyt were incubated at pH 5.6, supporting the hypothesis that EvgS can sense a cytoplasmic signal that changes when the external pH drops (Fig. 6). Whether this signal is pH itself or something else remains to be determined.

We previously proposed a model to explain the high frequency with which muta- tions that activate EvgS and map in the PAS domain are found. In this model, we suggested that the nonactive state of EvgS could be a tight inactive dimer. Activation is caused when a detection signal is received which moves the equilibrium of EvgS from a tight, inactive dimer to a weak, active dimer. According to this model, the PAS domain mutants cause constitutive activation at pH 7 by weakening the dimer interactions in the absence of an inducing signal, and consistent with this, many of these mutations map at or close to the predicted dimer interface (27). The model can explain the activation of EvgS signaling when the cytoplasmic domain alone is expressed, as the domain will now have lost those interactions in the periplasmic and transmembrane domains that normally contribute to the tight, inactive state.

In its simplest form, our hypothesis proposes that EvgS can exist in three states: a tight, inactive dimer; a weak, active dimer; and a weaker inactive dimer or monomer, with this third state being seen only when the periplasmic and transmembrane domains are missing and an activating mutation is present in the PAS domain. If this hypothesis is correct, we can predict that introducing a PAS domain mutation (which weakens subunit interactions) into the non-membrane-anchored cytoplasmic domain (in which interactions have already been weakened) would further reduce the degree of association of the subunits of the dimer, potentially to the point that it no longer shows any activity at all. We tested this prediction and confirmed that if the activating S600I mutation is introduced into EvgS-Cyt, all activity and inducibility is lost (Fig. 6)
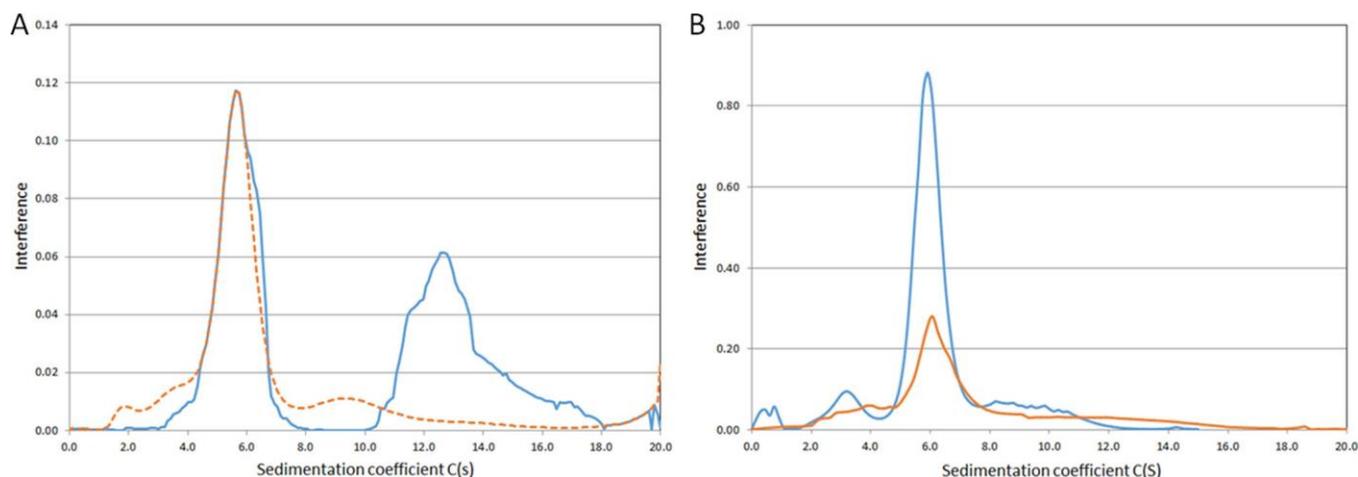
FIG 7 Analytical ultracentrifugation analysis of purified EvgS-Cyt (A) and EvgS-Cyt S600I (B) at pH 7 (blue lines) and at pH 5.6 (orange lines). The pH 5.6 data are plotted as a dashed line in panel A to make the coincidence of the lower-molecular-weight peaks clear. Different peak heights are due to different amounts of protein loaded in AUC cells.

even though the protein is still made and can be clearly seen on an SDS-polyacrylamide gel at the same level as EvgS-Cyt (Fig. 6, inset).

To test our model biochemically, we purified the EvgS-Cyt protein and compared its behavior in an analytical ultracentrifuge when run at either pH 7 or pH 5.6 (Fig. 7A). Strikingly, the protein sediments predominantly as two peaks at pH 7, consistent with a monomer-dimer equilibrium as required by our model. The breadth of the dimer peak and the presence of shoulders suggests that even higher-order structures may form. The mass distribution shows that the molecular mass of the first peak is 81.7 kDa and that of second peak is 163 kDa, quite close to the expected values for monomer and dimer (72.3 kDa and 144.6 kDa, respectively). At pH 5.6, a small amount of the protein aggregates, and the lower-molecular-mass peak becomes somewhat broader, with a clear shoulder at a slightly high molecular mass, although the peak exactly coincides with the peak at pH 7. The higher-molecular-mass peak seen at pH 7 was not visible at pH 5.6. This result is consistent with a weakening of the dimer at the inducing pH, as predicted by our model. When the same analysis was done with purified cytoplasmic domain containing the activating S600I mutation, the analytical ultracentrifugation (AUC) results were practically identical at pH 7 and pH 5.6 (Fig. 7B), with a mass distribution of 76.22 kDa for both peaks, which is very close to the expected value. This result is consistent with our model, where the activating mutations in the PAS domain mimic the effect of low pH by producing a shift in the equilibrium away from the tight, inactive dimer state even at a normally noninducing pH of 7, such that the EvgS protein is now active at neutral pH.

DISCUSSION

Several obvious questions can be asked for any HK which functions as part of a TCS, and answering these questions for any one kinase will contribute to our understanding of TCS in general, as well as improving our understanding of the particular system that it regulates. The first question usually concerns the nature of the signal that is recognized. In the case of EvgS, the actual signal is not known, and although a reasonable primary assumption is that the protein responds directly to changes in pH, there is no experimental proof for this assumption. Indeed, although several other HKs which are activated by low pH have been successfully studied to identify key residues for signal detection, direct demonstration that pH is the actual signal detected would require approaches such as reconstitution in a proteoliposome system where the pH can be manipulated or direct identification of pH-affected residues by deuterium exchange, and such experiments have not been done to date. In living cells, the consequences of changing external pH will be both significant and multifaceted,

particularly given the importance of transmembrane proton gradients in energy metabolism, so proof that pH is the direct signal is hard to obtain. Nonetheless, systematic studies of several other HKs have revealed a range of amino acids that may directly or indirectly detect pH change. In the ArsS kinase of *Helicobacter pylori*, a specific histidine residue (H94) was shown to be required for pH detection, although other residues were also shown to make a contribution (43). The SsrA HK in *Salmonella enterica* was shown to require two clusters of histidine residues in the periplasmic domain to detect low pH, with mutations in individual residues having relatively small effects (44). The *E. coli* CadC protein, which regulates the lysine-dependent AR4 system, is a ToxR-like protein with its C terminus in the periplasm, which has to interact with the lysine transporter LysP at low pH in order to activate gene expression (which as a one-component system it does directly, rather than by interacting with a response regulator protein [45]). Histidines play no role in the pH detection mechanism, which has been shown to rely on a cluster of acidic residues that form a negatively charged patch in the periplasmic region of the protein (46). pH detection is also required by the transporters GadC and AdiC of the *E. coli* acid resistance systems AR2 and AR3 (both of which function only below pH 6), and it has been recently shown that a tyrosine residue (Y74) is critical for pH sensing in AdiC, a finding that has been confirmed in an *in vitro* proteoliposome transport system (47). Thus, a range of different protonatable residues are used by different proteins as pH detectors.

None of the above examples are of proteins with large periplasmic VFT domains, so they cannot be directly used to identify candidate residues in EvgS. We have not yet succeeded in identifying residues in EvgS which detect low pH in EvgS, and given the large size of the periplasmic part of the protein, systematic mutagenesis of all candidate residues will be a considerable undertaking. It is also the case that EvgS may not directly detect pH at all, but detects some other signal. Indeed, earlier studies have shown that soluble fragments of both EvgS and BvgS were strongly inhibited in their *in vitro* phosphorylation by oxidized ubiquinone, suggesting a possible direct link to the electron transport chain and energy metabolism of the cell (48). However, the work reported here does rule out several candidates for signal detection. Although the triad of histidines in VFT1 was an attractive candidate, mutagenesis rules out a role in signal detection for this cluster, as it does for most of the other histidine residues in VFT1. It remains possible that key histidine residues in VFT2 will yet be found, although nothing in the structure points to any strong candidates. We also ruled out the possibility that amino acid binding in VFT2 modulates EvgS activity, and we showed on a structural basis that amino acid binding to VFT1 is very unlikely. We examined the potential role of GadC activity in the light of systems such as the AR4 system, which requires lysine transport by the LysP protein for activation, but we showed that GadC was not needed for EvgS activation. However, we did identify one histidine (H226) as being important for activation of the kinase. The mechanism for its role is unknown, but the structural model showed it to be close to L152, which already has been identified as being required for activation (39), and to P522, which caps the predicted transmembrane helix. The location of the H226 and L152 residues close to the dimer interface in the modeled structure provides a plausible mechanism whereby a structural change in the periplasmic domain could lead to a movement being transmitted via the transmembrane helix to the cytoplasm, with such movements being intraprotomer or interprotomer. Communication of structural changes from extracellular and periplasmic detector domains to transmembrane helices that rely on the presence of a conserved proline residue at the top of the helix or in the loop joining two consecutive transmembrane helices is seen in numerous receptor structures, such as in the family of pentameric ligand-gated ion channels (e.g., the acetylcholine and gamma-aminobutyric acid [GABA] receptors in humans), where such changes are key parts of the gating mechanism (reviewed in reference 49). The resulting small displacements in the transmembrane helices of the EvgS dimer could activate the cytoplasmic HK activity of EvgS in a way analogous to those reported in a related family of sensory kinases that includes EnvZ (50, 51).

Once the signal has been transmitted across the membrane, it has to produce a structural change somewhere in the cytoplasmic domains of the kinase in order for the kinase activity to be activated (or in the case of inhibitors, for it to be repressed, or for an auto-phosphatase activity, for it to be activated). Significant evidence reported by us and others previously had already pointed to the PAS domain just after the membrane as having a key role in the activation of both EvgS and BvgS. Mutations in this domain that activate EvgS at pH 7 or block the negative effects of the modulators of BvgS are easy to find, which implies that they represent a loss of some aspect of the protein's normal structure rather than a novel functional structure (27). The simplest explanation for this is that they cause weakening of an inactive tight dimer. Histidine kinases are generally found to be active as dimers, and there are precedents for a weakening of the dimer being the mechanism of activation in some cases. For example, the DctB HK found in *Sinorhizobium meliloti*, which detects and responds to the presence of C_4 dicarboxylic acids, is believed to proceed to an activated state by weakening of interactions at a dimer interface (52, 53), and the DcuS protein, which carries out a similar function in *E. coli*, has been proposed to act by a similar mechanism, in part via dimer interactions in the PAS domain (54).

Our proposal that EvgS may exist in three states (an inactive tight dimer, an active weak dimer, and an inactive weaker dimer or monomer, with the last being seen only in mutated forms of the protein) is supported by several lines of evidence. First, the large number of mutations found in the PAS domain that activate EvgS suggest that they are loss-of-function mutations at the structural level, and model building places many of them close to or at a putative dimerization interface in the PAS domain (27). In support of the latter point, the BvgS PAS domain has been shown to form a dimer when expressed as a soluble protein in *E. coli* (55). Second, we show here that releasing the cytoplasmic portion of the protein from its membrane anchor is sufficient to activate it. This suggests that some of the interactions forming the tight dimer are in the periplasmic domain (as shown by the structure) and the transmembrane domain and that without these, the cytoplasmic part of the protein forms mainly a weak, active dimer. However, expression of the cytoplasmic portion of the protein together with an S600I mutation in the PAS domain (which activates the full-length protein) leads to complete loss of activity even though the protein is still expressed, which is consistent with a further weakening of the weak dimer to an inactive state. AUC data on the cytoplasmic portion of the protein demonstrate the existence of a pH-regulated monomer-dimer equilibrium. As predicted by our model, there is a shift away from the dimer present at pH 7 when the protein is incubated pH 5.6, and this is also is seen if the protein contains the S600I mutation in the PAS domain.

We have not ruled out the possibility that a similar shift in a tight-weak dimer equilibrium in the periplasmic domain is also involved in enabling EvgS to respond to low pH in the periplasm, but two observations make this explanation less likely. First, no periplasmic mutations that give constitutive activation of EvgS have been found, despite nonbiased searches using three different selection methods (17, 27). Second, deletion of the periplasmic domain alone leads to an inactive protein (47) implying that the interactions responsible for the inactive state are predominantly in the transmembrane and cytoplasmic domains.

A recent detailed study on BvgS has generated data that support a model where a two-helix coiled-coil region immediately downstream of the BvgS PAS domain is crucial in mediating the activation of the kinase activity. According to this model, this well-conserved region may be affected by the state of the PAS domain itself, which could act as a "toggle switch," changing the dynamics of the coiled coil from fairly rigid (in the phosphatase mode) to highly dynamic (in the kinase mode) (14). The authors of that study also mention the possibility that the PAS domain senses internal conditions in the cell, which hence may contribute to BvgS activity. The work we report here supports this view. The isolation of noninducible mutations in the periplasmic part of the protein has enabled us to clearly distinguish periplasmic and cytoplasmic detection of signal by EvgS, and by studying these mutations in the presence of the constitutive mutation

S600I (which is necessary to give a high enough level of activity above background), together with the activity of the cytoplasmic part of EvgS alone, we have shown that induction of EvgS activity can be caused directly or indirectly by signals in the cytoplasm. This is an intriguing finding in the light of the recent demonstrations that *Salmonella* can detect a decreased cytoplasmic pH via the PhoP HKs, even when the external pH is neutral (56, 57). However, it poses a further functional puzzle, in that it has been shown that the decrease in the pH of the bacterial cytoplasm of *E. coli* when the external pH drops is only transitory, with a rapid restoration of neutrality (58, 59). Determination of how this transient drop might lead to sustained activation of EvgS-dependent gene expression will require detailed kinetic studies; the availability of the luciferase promoter probes which were used to obtain highly time-resolved data on induction of the members of the AR2 system should be useful here (26). These data also raise the intriguing possibility that EvgS has the ability to detect the pH gradient itself, which perhaps links back to the observations referred to above that it is affected strongly by oxidized ubiquinone.

In conclusion, our data are consistent with a model in which detection by EvgS of a ligand or ligands in the periplasm leads to structural rearrangements that, via H226 and P522, cause transduction of the inducing signal across the membrane. This signal, together with further information from the state of the cytoplasm, leads to a weakening of the dimer interface in the PAS domain, which in turn, probably through the mechanism described in reference 14, activates the autokinase activity of EvgS and strongly induces expression of the acid resistance AR2 genes together with several drug efflux pumps. It is not hard to see how such a mechanism may have a role in aiding the colonization of the gut by *E. coli*, but the nature of the ligand(s) remains unclear for the present.

MATERIALS AND METHODS

Bacterial strains and plasmids. The bacterial strains and plasmids used in this study are listed in Table S1 at http://epapers.bham.ac.uk/3021/. HST08 supercompetent cells supplied with the In-Fusion hybrid cloning kit (TaKaRa-Bio) were used as the host for gene cloning. *E. coli* MG1655 Δ*evgS*::*cat ydeP-lacZ* Kan^r is the reporter strain for the acid induction assay (60). This strain was transformed either with the control plasmid pBAD24-his or with plasmids expressing wild-type or mutated versions of EvgS to determine the activity of the EvgS protein.

Bacterial cells were grown aerobically at 37°C in lysogeny broth (LB) (1% [wt/vol] NaCl, 1% [wt/vol] tryptone [BD Biosciences], and 0.5% [wt/vol] yeast extract) with added antibiotics (carbenicillin [100 μg/ml], kanamycin [100 μg/ml], or chloramphenicol [35 μg/ml]) as appropriate. For the acid induction assay, cells were grown in M9 medium (pH 5.6 or pH 7) supplemented with 100 mM KCl, 0.2% Casamino Acids (BD Biosciences), and 0.4% glucose. L-Arabinose (0.2%) was added to the culture at an optical density at 600 nm (OD$_{600}$) of 0.1 to 0.2 to induce protein expression from the pBAD promoter.

Plasmid construction. All site-directed mutagenesis experiments were conducted using the QuikChange Lightning kit (Agilent) according to the manufacturer's instructions. To construct pBADEvgS-Cyt, codons for the cytoplasmic domains of EvgS (comprising amino acid residues 553 to 1193) plus a His$_6$ C-terminal tag were PCR amplified from the plasmid pBADEvgS using primers 5=-GAGGAATTCACC ATGGCGCTCAGTTCGTCGTCGT-3= (forward) and 5=-CAAAACAGCCAAGCTTGTCATTTTTCTGACAGAAAAC-3= (reverse). The amplified fragment was gel purified using the In-Fusion NucleoSpin (TaKaRa-Bio) kit and cloned into NcoI-HindIII-cut pBAD24 using the In-Fusion (TaKaRa-Bio) restriction-free cloning kit following the manufacturer's guidelines. The NcoI site was modified to ensure that the insert was in frame. To construct pET41c-EvgS-Cyt, the codons for the same fragment of EvgS were PCR amplified from the plasmid pBADEvgS using primers 5=-AAGGAGATATACATATGCGCTCAGTTCGTCGTCGT-3= (forward) and 5=-CGCGTGGCACAAGCTTGTCATTTTTCTGACAGAAAAC-3= (reverse). The amplified fragment was gel purified as described above and restriction cloned (NdeI-HindIII) into pET41c using the In-Fusion (TaKaRa-Bio) restriction-free cloning kit as described above. The plasmid pET41c used in our study is a modified version of its parent where the glutathione *S*-transferase (GST) tag has been replaced with C-terminal His$_8$ tag. The resultant plasmid, pET41c-EvgS-Cyt, was transformed into BL21(DE3*) for overexpression and purification. The plasmid pET41c-EvgS-Cyt was also used as a template for mutagenesis to introduce the S600I mutation, using primers 5=-GTTCAAAAGCACTATTATGAATAATGACATTACCTTGCCAGTTTACAA-3= and 5=-TTGTAAACTGGCAAGGTAATGTCATTATTCATAATAGTGCTTTTGAAC-3=, to enable purification of the cytoplasmic domain containing this activating mutation.

Acid induction assay. A single colony of each of the strains of interest was inoculated into 5 ml LB medium with the requisite antibiotics and grown overnight (10 to 12 h) aerobically at 37°C. The following day, cultures were diluted into supplemented M9 medium (pH 5.6 or pH 7) to an initial OD$_{600}$ of 0.05, grown with shaking at 37°C to an OD$_{600}$ of 0.1 to 0.2, and assayed for beta-galactosidase activity (61). All assays were done at least as biological triplicates.

Overexpression and purification of the EvgS cytoplasmic domains. BL21(DE3*)/pET41c-EvgS-Cyt was grown overnight aerobically in 50 ml of LB medium, pH 7. The following day, the culture was diluted into 800 ml of LB medium containing 100 μg/ml of kanamycin, to a starting OD_{600} of 0.05. The cultures were grown aerobically at 37°C to an OD_{600} of 0.3, and then the temperature was reduced to 24°C and the culture grown further to an OD_{600} of 0.4. Isopropyl-[3-D-thiogalactopyranoside (IPTG) (0.1 mM) was then added to the cultures to induce expression of the protein, and the cultures were incubated overnight. They were then centrifuged at 6,000 rpm in JLA 8.1000 rotor (Beckman Coulter), and the harvested cells were resuspended in 100 ml of HEPES buffer (pH 7.4) (200 mM HEPES, 500 mM NaCl) containing a tablet of EDTA-free protease inhibitor (Sigma). The resuspended cells were lysed using an Emulsiflex-C3 homogenizer as per the manufacturer's guidelines. The lysate was centrifuged at 20,000 rpm using a JA 20 rotor (Beckman Coulter) and filtered through a 0.22-μm filter. The lysate was bound overnight to a 5-ml His-trap HP (GE Healthcare) column. The bound protein was eluted in HEPES buffer containing 500 mM imidazole (Sigma) and analyzed by SDS-PAGE. Fractions containing a band of the correct size were concentrated using a Vivaspin 20-ml concentrator (50,000 molecular weight cutoff [MWCO]) (GE Healthcare) and gel purified using an Akta Pure 25 (GE Healthcare LS) with a prepacked Hi-Load 16/600 Superdex 200 PG column. The cytoplasmic domain containing S600I was expressed and purified in the same way, except that cells were resuspended after harvest in 150 ml 150 mM HEPES–50 mM MOPS (morpholinepropanesulfonic acid)–200 mM NaCl (pH 7.4) and were eluted from the His-trap column in 150 mM MOPS–50 mM KOH plus 500 mM imidazole.

AUC. For analytical ultracentrifugation (AUC), sedimentation equilibrium runs were done using a Beckman XLA Optima after samples had been dialyzed extensively against 50 mM Tris-HCl buffer (pH 7 or pH 5.6). The dialysis buffer contained EDTA-free protease inhibitor and 2 mM phenylmethylsulfonyl fluoride (PMSF) (Sigma). After dialysis, protein solutions at an OD_{280} of 1 (corresponding to a protein concentration of 0.1 μM) were analyzed at 40,000 rpm and 20°C. A total of 200 scans were collected to capture the complete extent of the sedimentation of the protein samples at both pH values. The sedimentation distribution was analyzed using Sedfit (62, 63), and data were fitted using the following parameters: partial specific volume, 0.73000; buffer viscosity (poise), 0.01002. The sedimentation coefficient and mass distributions were extracted from the data to generate plots.

Structural modeling. To identify the best templates for the creation of the models, we used protein BLAST (BLASTp) against the PDB database. The dimeric structure of the periplasmic domain (PDB 4Q0C) of *Bordetella pertussis* BvgS (UniProtKB P16575) was identified as the top match to the EvgS query sequence. Structural alignments were performed using MAFFT (64) and visualized using ESPript 3 (65). EvgS homology models were built using I-TASSER (66). Output models had a C-score spread between 0.75 and 0.95. The C-score is a confidence score, the value of which varies from −5.0 to 2.0, with higher scores being indicative of high confidence in model building. Additional model optimization was performed manually using Coot as part of the CCP4 package (67). SSM superposition was used for structural alignment as implemented in CCP4. Similarity scores were calculated using SIM (http://web.expasy.org/sim/). Visualization of structural data was done with the PyMOL molecular graphics system, version 1.8 (Schrödinger, LLC.).

ACKNOWLEDGMENTS

REFERENCES

1. Jacob-Dubuisson F, Wintjens R, Herrou J, Dupré E, Antoine R. 2012. BvgS of pathogenic Bordetellae: a paradigm for sensor kinase with Venus Flytrap perception domains, p. 57– 83. *In* Gross R, Beier D (ed), Two component systems in bacteria. Caister Academic Press, Poole, UK.
2. Herrou J, Bompard C, Wintjens R, Dupré E, Willery E, Villeret V, Locht C, Antoine R, Jacob-Dubuisson F. 2010. Periplasmic domain of the sensor-kinase BvgS reveals a new paradigm for the Venus flytrap mechanism. Proc Natl Acad Sci U S A 107:17531–17535. https://doi.org/10.1073/pnas.1006267107.
3. Dupré E, Herrou J, Lensink MF, Wintjens R, Vagin A, Lebedev A, Crosson S, Villeret V, Locht C, Antoine R, Jacob-Dubuisson F. 2015. Virulence regulation with Venus flytrap domains: structure and function of the periplasmic moiety of the sensor-kinase BvgS. PLoS Pathog 11:e1004700. https://doi.org/10.1371/journal.ppat.1004700.
4. Tam R, Saier MH, Jr. 1993. Structural, functional, and evolutionary relationships among extracellular solute-binding receptors of bacteria. Microbiol Rev 57:320 –346.
5. Berntsson RP, Smits SH, Schmitt L, Slotboom DJ, Poolman B. 2010. A structural classification of substrate-binding proteins. FEBS Lett 584:2606 –2617. https://doi.org/10.1016/j.febslet.2010.04.043.
6. Scheepers GH, Lycklama A, Nijeholt JA, Poolman B. 2016. An updated structural classification of substrate-binding proteins. FEBS Lett 590: 4393–4401. https://doi.org/10.1002/1873-3468.12445.
7. Mao B, Pear M, McCammon J, Quiocho F. 1982. Hinge-bending in L-arabinose-binding protein. The Venus's-flytrap model. J Biol Chem 257:1131–1133.
8. Cummings CA, Bootsma HJ, Relman DA, Miller JF. 2006. Species- and strain-specific control of a complex, flexible regulon by Bordetella BvgAS. J Bacteriol 188:1775–1785. https://doi.org/10.1128/JB.188.5.1775-1785.2006.
9. Uhl MA, Miller JF. 1994. Autophosphorylation and phosphotransfer in the Bordetella pertussis BvgAS signal transduction cascade. Proc Natl Acad Sci U S A 91:1163–1167. https://doi.org/10.1073/pnas.91.3.1163.
10. Uhl MA, Miller JF. 1996. Central role of the BvgS receiver as a phosphorylated intermediate in a complex two-component phosphorelay. J Biol Chem 271:33176 –33180. https://doi.org/10.1074/jbc.271.52.33176.
11. Uhl MA, Miller JF. 1996. Integration of multiple domains in a two-

component sensor protein: the *Bordetella pertussis* BvgAS phosphorelay. EMBO J 15:1028 –1036.

12. Kim JR, Cho KH. 2006. The multi-step phosphorelay mechanism of unorthodox two-component systems in *E. coli* realizes ultrasensitivity to stimuli while maintaining robustness to noises. Comput Biol Chem 30:439 – 444. https://doi.org/10.1016/j.compbiolchem.2006.09.004.

13. Csikász-Nagy A, Cardelli L, Soyer OS. 2011. Response dynamics of phosphorelays suggest their potential utility in cell signalling. J R Soc Inter- face 8:480 – 488. https://doi.org/10.1098/rsif.2010.0336.

14. Lesne E, Krammer EM, Dupre E, Locht C, Lensink MF, Antoine R, Jacob-Dubuisson F. 2016. Balance between coiled-coil stability and dynamics regulates activity of BvgS sensor kinase in Bordetella. mBio 7:e02089. https://doi.org/10.1128/mBio.02089-15.

15. Utsumi R, Katayama S, Taniguchi M, Horie T, Ikeda M, Igaki S, Nakagawa H, Miwa A, Tanabe H, Noda M. 1994. Newly identified genes involved in the signal transduction of *Escherichia coli* K-12. Gene 140:73–77. https://doi.org/10.1016/0378-1119(94)90733-1.

16. Tanabe H, Yamasaki K, Kato A, Yoshioka S, Utsumi R. 1998. Identification of the promoter region and the transcriptional regulatory sequence of the *evgAS* operon of *Escherichia coli*. Biosci Biotechnol Biochem 62: 286 – 290. https://doi.org/10.1271/bbb.62.286.

17. Kato A, Ohnishi H, Yamamoto K, Furuta E, Tanabe H, Utsumi R. 2000. Transcription of *emrKY* is regulated by the EvgA-EvgS two-component system in *Escherichia coli* K-12. Biosci Biotechnol Biochem 64:1203–1209. https://doi.org/10.1271/bbb.64.1203.

18. Nishino K, Yamaguchi A. 2001. Overexpression of the response regulator evgA of the two-component signal transduction system modulates multidrug resistance conferred by multidrug resistance transporters. J Bacteriol 183:1455–1458. https://doi.org/10.1128/JB.183.4.1455-1458 .2001.

19. Nishino K, Yamaguchi A. 2002. EvgA of the two-component signal transduction system modulates production of the yhiUV multidrug transporter in *Escherichia coli*. J Bacteriol 184:2319 –2323. https://doi.org/ 10.1128/JB.184.8.2319-2323.2002.

20. Eguchi Y, Oshima T, Mori H, Aono R, Yamamoto K, Ishihama A, Utsumi R. 2003. Transcriptional regulation of drug efflux genes by EvgAS, a two-component system in *Escherichia coli*. Microbiology 149:2819 –2828. https://doi.org/10.1099/mic.0.26460-0.

21. Masuda N, Church GM. 2002. *Escherichia coli* gene expression responsive to levels of the response regulator EvgA. J Bacteriol 184:622– 625. https://doi.org/10.1128/JB.184.22.6225-6234.2002.

22. Masuda N, Church GM. 2003. Regulatory network of acid resistance genes in *Escherichia coli*. Mol Microbiol 48:699–712. https://doi.org/10 .1046/j.1365-2958.2003.03477.x.

23. Nishino K, Inazumi Y, Yamaguchi A. 2003. Global analysis of genes regulated by EvgA of the two-component regulatory system in *Esche- richia coli*. J Bacteriol 185:2667–2672. https://doi.org/10.1128/JB.185.8 .2667-2672.2003.

24. Ma Z, Masuda N, Foster JW. 2004. Characterization of EvgAS-YdeO-GadE branched regulatory circuit governing glutamate-dependent acid resis- tance in *Escherichia coli*. J Bacteriol 186:7378 –7389. https://doi.org/10 .1128/JB.186.21.7378-7389.2004.

25. Itou J, Eguchi Y, Utsumi R. 2009. Molecular mechanism of transcriptional cascade initiated by the EvgS/EvgA system in *Escherichia coli* K-12. Biosci Biotechnol Biochem 73:870 – 878. https://doi.org/10.1271/bbb.80795.

26. Burton NA, Johnson MD, Antczak P, Robinson A, Lund PA. 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. https://doi .org/10.1016/j.jmb.2010.06.054.

27. Johnson MD, Bell J, Clarke K, Chandler R, Pathak P, Xia Y, Marshall RL, Weinstock GM, Loman NJ, Winn PJ, Lund PA. 2014. Characterization of mutations in the PAS domain of the EvgS sensor kinase selected by laboratory evolution for acid resistance in *Escherichia coli*. Mol Microbiol 93:911–927. https://doi.org/10.1111/mmi.12704.

28. Foster JW. 2004. *Escherichia coli* acid resistance: tales of an amateur acidophile. Nat Rev Microbiol 2:898 –907. https://doi.org/10.1038/ nrmicro1021.

29. Lund P, Tramonti A, De Biase D. 2014. Coping with low pH: molecular strategies in neutralophilic bacteria. FEMS Microbiol Rev 38:1091–1125. https://doi.org/10.1111/1574-6976.12076.

30. Kailasan Vanaja S, Bergholz TM, Whittam TS. 2009. Characterization of the *Escherichia coli* O157:H7 Sakai GadE regulon. J Bacteriol 191: 1868 – 1877. https://doi.org/10.1128/JB.01481-08.

31. Nadler C, Shifrin Y, Nov S, Kobi S, Rosenshine I. 2006. Characterization of

enteropathogenic *Escherichia coli* mutants that fail to disrupt host cell spreading and attachment to substratum. Infect Immun 74:839 – 849. https://doi.org/10.1128/IAI.74.2.839-849.2006.

32. Dziva F, Hauser H, Connor TR, van Diemen PM, Prescott G, Langridge GC, Eckert S, Chaudhuri RR, Ewers C, Mellata M, Mukhopadhyay S, Curtiss R 3rd, Dougan G, Wieler LH, Thomson NR, Pickard DJ, Stevens MP. 2013. Sequencing and functional annotation of avian pathogenic *Escherichia coli* serogroup O78 strains reveal the evolution of *E. coli* lineages patho- genic for poultry via distinct mechanisms. Infect Immun 81:838 – 849. https://doi.org/10.1128/IAI.00585-12.

33. Stock AM, Robinson VL, Goudreau PN. 2000. Two-component signal transduction. Annu Rev Biochem 69:183–215. https://doi.org/10.1146/ annurev.biochem.69.1.183.

34. Mascher T, Helmann JD, Unden G. 2006. Stimulus perception in bacterial signal-transducing histidine kinases. Microbiol Mol Biol Rev 70:910 –938. https://doi.org/10.1128/MMBR.00020-06.

35. Ruggiero A, Dattelbaum JD, Staiano M, Berisio R, D'Auria S, Vitagliano L. 2014. Structure of apo ArgBP from T maritima. PLoS One 9:e96560. https://doi.org/10.1371/journal.pone.0096560.

36. Kanemaru Y, Hasebe F, Tomita T, Kuzuyama T, Nishiyama M. 2013. Two ATP-binding cassette transporters involved in (S)-2-aminoethyl-cysteine uptake in Thermus thermophilus. J Bacteriol 195:3845–3853. https://doi .org/10.1128/JB.00202-13.

37. Ahmed AH, Oswald RE. 2010. Piracetam defines a new binding site for allosteric modulators of ◌-amino-3-hydroxy-5-methyl-4-isoxazole- propionic acid (AMPA) receptors. J Med Chem 53:2197–2203. https://doi .org/10.1021/jm901905j.

38. Piepenbreier H, Fritz G, Gebhard S. 2017. Transporters as information processors in bacterial signalling pathways. Mol Microbiol 104:1–15. https://doi.org/10.1111/mmi.13633.

39. 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. https://doi.org/10.1128/JB.01742-14.

40. Edgcomb SP, Murphy KP. 2002. Variability in the pKa of histidine side- chains correlates with burial within proteins. Proteins 49:1– 6. https://doi .org/10.1002/prot.10177.

41. Cavicchioli R, Chiang RC, Kalman LV, Gunsalus RP. 1996. Role of the periplas- mic domain of the Escherichia coli NarX sensor-transmitter protein in nitrate- dependent signal transduction and gene regulation. Mol Microbiol 21:901– 911. https://doi.org/10.1046/j.1365-2958.1996.491422.x.

42. Kwon O, Georgellis D, Lynch AS, Boyd D, Lin EC. 2000. The ArcB sensor kinase of *Escherichia coli*: genetic exploration of the transmembrane region. J Bacteriol 182:2960 –2966. https://doi.org/10.1128/JB.182.10 .2960-2966.2000.

43. Müller S, Götz M, Beier D. 2009. Histidine residue 94 is involved in pH sensing by histidine kinase ArsS of *Helicobacter pylori*. PLoS One 4:e6930. https://doi.org/10.1371/journal.pone.0006930.

44. Mulder DT, McPhee JB, Reid-Yu SA, Stogios PJ, Savchenko A, Coombes BK. 2015. Multiple histidines in the periplasmic domain of the *Salmonella enterica* sensor kinase SsrA enhance signaling in response to extracel- lular acidification. Mol Microbiol 95:678 – 691. https://doi.org/10.1111/ mmi.12895.

45. Tetsch L, Koller C, Haneburger I, Jung K. 2008. The membrane-integrated transcriptional activator CadC of *Escherichia coli* senses lysine indirectly via the interaction with the lysine permease LysP. Mol Microbiol 67: 570 – 583. https://doi.org/10.1111/j.1365-2958.2007.06070.x.

46. Haneburger I, Eichinger A, Skerra A, Jung K. 2011. New insights into the signaling mechanism of the pH-responsive, membrane-integrated tran- scriptional activator CadC of *Escherichia coli*. J Biol Chem 286: 10681– 10689. https://doi.org/10.1074/jbc.M110.196923.

47. Wang S, Yan R, Zhang X, Chu Q, Shi Y. 2014. Molecular mechanism of pH- dependent substrate transport by an arginine-agmatine antiporter. Proc Natl Acad Sci U S A 111:12734 –12739. https://doi.org/10.1073/pnas .1414093111.

48. Bock A, Gross R. 2002. The unorthodox histidine kinases BvgS and EvgS are responsive to the oxidation status of a quinone electron carrier. Eur J Biochem 269:3479 –3484. https://doi.org/10.1046/j.1432-1033.2002 .03029.x.

49. Taly A, Hénin J, Changeux JP, Cecchini M. 2014. Allosteric regulation of pentameric ligand-gated ion channels: an emerging mechanistic per- spective. Channels 8:350 –360. https://doi.org/10.4161/chan.29444.

50. Botelho SC, Enquist K, von Heijne G, Draheim RR. 2015. Differential repositioning of the second transmembrane helices from *E. coli* Tar and

EnvZ upon moving the flanking aromatic residues. Biochim Biophys Acta 1848:615– 621. https://doi.org/10.1016/j.bbamem.2014.11.017.

51. Heininger A, Yusuf R, Lawrence RJ, Draheim RR. 2016. Identification of transmembrane helix 1 (TM1) surfaces important for EnvZ dimerisation and signal output. Biochim Biophys Acta 1858:1868 –1875. https://doi.org/10.1016/j.bbamem.2016.05.002.

52. Nan B, Liu X, Zhou Y, Liu J, Zhang L, Wen J, Zhang X, Su XD, Wang YP. 2010. From signal perception to signal transduction: ligand-induced dimeric switch of DctB sensory domain in solution. Mol Microbiol 75: 1484 –1494. https://doi.org/10.1111/j.1365-2958.2010.07069.x.

53. Liu J, Yang J, Wen J, Yang Y, Wei X, Zhang X, Wang YP. 2014. Mutational analysis of dimeric linkers in peri- and cytoplasmic domains of histidine kinase DctB reveals their functional roles in signal transduction. Open Biol 4:140023. https://doi.org/10.1098/rsob.140023.

54. Etzkorn M, Kneuper H, Dünnwald P, Vijayan V, Krämer J, Griesinger C, Becker S, Unden G, Baldus M. 2008. Plasticity of the PAS domain and a potential role for signal transduction in the histidine kinase DcuS. Nat Struct Biol 15:1031–1039. https://doi.org/10.1038/nsmb.1493.

55. Dupré E, Wohlkonig A, Herrou J, Locht C, Jacob-Dubuisson F, Antoine R. 2013. Characterization of the PAS domain in the sensor-kinase BvgS: mechanical role in signal transmission. BMC Microbiol 13:172. https://doi.org/10.1186/1471-2180-13-172.

56. Chakraborty S, Mizusaki H, Kenney LJ. 2015. A FRET-based DNA biosen- sor tracks OmpR-dependent acidification of Salmonella during macro- phage infection. PLoS Biol 13:e1002116. https://doi.org/10.1371/journal.pbio.1002116.

57. Choi J, Groisman EA. 2016. Acidic pH sensing in the bacterial cytoplasm is required for Salmonella virulence. Mol Microbiol 101:1024 –1038. https://doi.org/10.1111/mmi.13439.

58. Wilks JC, Slonczewski JL. 2007. pH of the cytoplasm and periplasm of Escherichia coli: rapid measurement by green fluorescent protein fluo- rimetry. J Bacteriol 189:5601–5607. https://doi.org/10.1128/JB.00615-07.

59. Martinez KA 2nd, Kitko RD, Mershon JP, Adcox HE, Malek KA, Berkmen MB, Slonczewski JL. 2012. Cytoplasmic pH response to acid stress in individual cells of Escherichia coli and Bacillus subtilis observed by fluo- rescence ratio imaging microscopy. Appl Environ Microbiol 78: 3706 –3714. https://doi.org/10.1128/AEM.00354-12.

60. Eguchi Y, Itou J, Yamane M, Demizu R, Yamato F, Okada A, Mori H, Kato A, Utsumi R. 2007. B1500, a small membrane protein, connects the two-component systems EvgS/EvgA and PhoQ/PhoP in Escherichia coli. Proc Natl Acad Sci U S A 104:18712–18717. https://doi.org/10.1073/pnas.0705768104.

61. Miller JH. 1972. Experiments in molecular genetics. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

62. Schuck B. 2000. Size distribution analysis of macromolecules by sedimen- tation velocity ultracentrifugation and Lamm equation modeling. Biophys J 78:1606 –1619. https://doi.org/10.1016/S0006-3495(00)76713-0.

63. Brown PH, Schuck P. 2008. A new adaptive grid-size algorithm for the simulation of sedimentation velocity profiles in analytical ultracentrifu- gation. Comput Phys Commun 178:105–120. https://doi.org/10.1016/j.cpc.2007.08.012.

64. Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res 30:3059 –3066. https://doi.org/10.1093/nar/gkf436.

65. Robert X, Gouet P. 2014. Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res 42:W320 –W324. https://doi.org/10.1093/nar/gku316.

66. Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. 2015. The I-TASSER Suite: protein structure and function prediction. Nat Methods 12:7– 8. https://doi.org/10.1038/nmeth.3213.

67. Emsley P, Lohkamp B, Scott WG, Cowtan K. 2010. Features and devel- opment of Coot. Acta Crystallogr D Biol Crystallogr 66:486 –501. https://doi.org/10.1107/S0907444910007493.

Bibliography

Adnan, M. *et al.* (2010) 'Contribution of *rpoS* and *bolA* genes in biofilm formation in Escherichia coli K-12 MG1655', *Molecular and Cellular Biochemistry*, 342(1–2), pp. 207–213

Aguilar, C. *et al.* (2012) 'Genetic changes during a laboratory adaptive evolution process that allowed fast growth in glucose to an *Escherichia coli* strain lacking the major glucose transport system', *BMC Genomics*, 13(1), p. 385.

Al-Ajmi, D. *et al.* (2006) 'Evaluation of a PCR detection method for Escherichia coli O157:H7/H- bovine faecal samples', *Letters in Applied Microbiology*, 42(4), pp. 386–391.

Amato, S. M. and Brynildsen, M. P. (2014) 'Nutrient transitions are a source of persisters in Escherichia coli biofilms', *PLoS ONE*, 9(3), pp. 1–9.

Andersson, S. G. E. (2016) 'Stress management strategies in single bacterial cells', *Proceedings of the National Academy of Sciences*, 113(15), pp. 3921–3923.

Andersson, S. G. E. *et al.* (2017) 'Importance of stress-response genes to the survival of airborne *Escherichia coli* under different levels of relative humidity', *AMB Express*. Springer Berlin Heidelberg, 7(15), p. 71.

Arnold, C. N. *et al.* (2001) 'Global Analysis of *Escherichia coli* Gene Expression during the Acetate-Induced Acid Tolerance Response', *Journal of bacteriology*, 183(7), pp. 2178–2186.

Atwood, K. C., Schneider, L. K. and Ryan, F. J. (1951) 'Periodic Selection in Escherichia Coli', *Proceedings of the National Academy of Sciences*, 37(3), pp. 146–155.

Arnoldini, M. *et al.* (2012) 'Evolution of Stress Response in the Face of Unreliable Environmental Signals', *PLoS Computational Biology*, 8(8).

Baik, H. S. *et al.* (2016) 'The acid tolerance response of *Salmonella typhimurium* provides protection against organic acids', *Microbiology society*, 142(1 996), p. 3 195-3200.

Bak, G. *et al.* (2014) 'Roles of *rpoS*-activating small RNAs in pathways leading to acid resistance of Escherichia coli', *Microbiology Open*, 3(1), pp. 15–28.

Baker, D. R. *et al.* (2007) 'Differences in virulence among *Escherichia coli* O157:H7 strains isolated from humans during disease outbreaks and from healthy cattle', *Applied and Environmental Microbiology*, 73(22), pp. 7338–7346.

Barrick, J. E. and Lenski, R. E. (2013) 'Genome dynamics during experimental evolution.', *Nature reviews. Genetics*. Nature Publishing Group, 14(12), pp. 827–39.

Barrick, J. E. *et al.* (2009) 'Genome evolution and adaptation in a long-term experiment with *Escherichia coli.*', *Nature*. Nature Publishing Group, 461(7268), pp. 1243–1247.

Behrends, V. *et al.* (2014) 'A metabolic trade-off between phosphate and glucose utilization in *Escherichia coli* TL - 10', *Molecular Biosystems*. Royal Society of Chemistry, 10 VN-r(11), pp. 2820–2822.

Bennett, A. F. and Lenski, R. E. (2007) 'An experimental test of evolutionary trade-offs during temperature adaptation', *Proc Natl Acad Sci U S A*, 104 Suppl, pp. 8649–8654.

Birdsell, D. N. *et al.* (2012) 'Melt analysis of mismatch amplification mutation assays (melt-MAMA): A functional study of a cost-effective SNP genotyping assay in bacterial models', *PLoS ONE*, 7(3).

Blattner, F. R. *et al.* (2010) 'The Complete Genome Sequence of Escherichia coli K-12', *Science AAAs*, 1453(1997), p. Science. 1998 Mar 20;279(5368):1827.

Bohannon, D. E. *et al.* (1991) 'Stationary-phase-inducible gearbox'' promoters: Differential effects of katF mutations and role of ??70', *Journal of Bacteriology*, 173(14), pp. 4482–4492.

Bougdour, A., Wickner, S. and Gottesman, S. (2006) 'Modulating RssB activity: IraP, a novel regulator of σS stability in Escherichia coli', *Genes and Development*, 20(7), pp. 884–897.

Bradbeer, C. (1993) 'The proton motive force drives the outer membrane transport of cobalamin in *Escherichia coli.*', *Journal of bacteriology*, 175(10), pp. 3146–50.

Bradley S. Hughes*, Alistair J. Cullum, and A. F. B. (2007) 'An Experimental Evolutionary Study on Adaptation to Temporally Fluctuating pH in Escherichia coli', *The University of Chicago Press Journals*, 4(Physiological and Biochemical Zoology 80), pp. 406–421.

Breland, E. J., Eberly, A. R. and Hadjifrangiskou, M. (2017) 'An Overview of Two-Component Signal Transduction Systems Implicated in Extra-Intestinal Pathogenic *E. coli* Infections', *Frontiers in Cellular and Infection Microbiology*, 7(May), pp. 1–14.

Brennan, F. P. *et al.* (2010) 'Characterization of environmentally persistent *Escherichia coli* isolates leached from an Irish soil', *Applied and Environmental Microbiology*, 76(7), pp. 2175–2180.

Brown, D. R. *et al.* (2014) 'Nitrogen stress response and stringent response are coupled in Escherichia coli.', *Nature communications*. Nature Publishing Group, 5(May), p. 4115.

Bueno, D. J., Silva, J. O. and Oliver, G. (2004) *Public Health Microbiology: Methods and protocols*, *Journal of Chemical Information and Modelling*.

Bull, H. J., Lombardo, M. J. and Rosenberg, S. M. (2001) 'Stationary-phase mutation in the bacterial chromosome: recombination protein and DNA polymerase IV dependence.', *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), pp. 8334–8341.

Canino-Koning, R., Wiser, M. J. and Ofria, C. (2016) 'The Evolution of Evolvability: Changing Environments Promote Rapid Adaptation in Digital Organisms', *Journal of bacteriology*, 18(4), pp. 1–8.

Castanié-Cornet, M. P. *et al.* (2010) 'Acid stress response in Escherichia coli: Mechanism of regulation of gadA transcription by RcsB and GadE', *Nucleic Acids Research*, 38(11), pp. 3546–3554.

Chang, D.-E. *et al.* (2004) 'Carbon nutrition of Escherichia coli in the mouse intestine.', *Proceedings of the National Academy of Sciences of the United States of America*, 101(19), pp. 7427–32.

Chattopadhyay, S. *et al.* (2012) 'Convergent molecular evolution of genomic cores in Salmonella enterica and Escherichia coli', *Journal of Bacteriology*, 194(18), pp. 5002–5011.

Chaudhuri, R. R. and Henderson, I. R. (2012) 'The evolution of the Escherichia coli phylogeny', *Infection, Genetics and Evolution*. Elsevier B.V., 12(2), pp. 214–226.

Chen, X. *et al.* (2013) 'Metabolic engineering of Escherichia coli: A sustainable industrial platform for bio-based chemical production', *Biotechnology Advances*. Elsevier Inc., 31(8), pp. 1200–1223.

Christie-Oleza, J. A. *et al.* (2009) 'Conjugative interaction induces transposition of ISPst9 in *Pseudomonas stutzeri* AN10', *Journal of Bacteriology*, 191(4), pp. 1239–1247.

Cornforth, D. M. and Foster, K. R. (2013) 'Competition sensing: the social side of bacterial stress responses', *Nature Reviews Microbiology*. Nature Publishing Group, 11(4), pp. 285–293.

Chung, H. J., Bang, W. and Drake, M. A. (2006) 'Stress response of Escherichia coli', *Comprehensive Reviews in Food Science and Food Safety*, 5(3), pp. 52–64.

Cohen, S. N. *et al.* (1973) 'Construction of biologically functional bacterial plasmids in vitro.', *Proceedings of the National Academy of Sciences of the United States of America*, 70(11), pp. 3240–4.

Colloms, S. D., Alén, C. and Sherratt, D. J. (1998) 'The ArcA/ArcB two-component regulatory system of *Escherichia coli* is essential for Xer site-specific recombination at psi', *Molecular Microbiology*, 28(3), pp. 521–530.

Compan, I. and Touati, D. (1994) 'Anaerobic activation of arcA transcription in *Escherichia coli*: Roles of Fnr and ArcA', *Molecular Microbiology*, 11(5), pp. 955–964.

Conrad, T. M. *et al.* (2009) 'Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations.', *Genome biology*, 10(10), p. R118.

Conrad, T. M., Lewis, N. E. and Palsson, B. Ø. (2011) 'Microbial laboratory evolution in the era of genome-scale science.', *Molecular systems biology*, 7(509), p. 509.

Cunha, B. *et al.* (2017) 'Optimization of a Bioassay to Evaluate Escherichia coli Stress Responses', *Bioengineering (ENBENG), 2017 IEEE 5th Portuguese Meeting on 16-18 Feb. 2017*, (10.1109/ENBENG.2017.7889424).

Davies, S. J. *et al.* (1999) 'Inactivation and Regulation of the Aerobic C 4 -Dicarboxylate Transport (dctA) Gene of Escherichia coli', *Journal of bacteriology*, 181(18), pp. 5624–5635.

De Biase, D. *et al.* (1999) 'The response to stationary-phase stress conditions in Escherichia coli: role and regulation of the glutamic acid decarboxylase system.', *Molecular microbiology*, 32(6), pp. 1198–211.

De Biase, D. and Lund, P. A. (2015) 'The Escherichia coli Acid Stress Response and Its Significance for Pathogenesis', in *Advances in Applied Microbiology*, pp. 49–88.

Deatherage, D. E. *et al.* (2017) 'Specificity of genome evolution in experimental populations of *Escherichia coli* evolved at different temperatures', *Proceedings of the National Academy of Sciences*, 114(10), pp. E1904–E1912.

Detains, J. R. *et al.* (2012) 'Evolutionary insight from whole-genome sequencing of experimentally evolved microbes', *Molecular Ecology*, 21(9), pp. 2058–2077.

Dong, T. and Schellhorn, H. E. (2009) 'Global effect of RpoS on gene expression in pathogenic Escherichia coli O157:H7 strain EDL933', *BMC Genomics*, 10, p. 349.

Dragosits, M. and Mattanovich, D. (2013) 'Adaptive laboratory evolution -- principles and applications for biotechnology.', *Microbial cell factories*. Microbial Cell Factories, 12(1), p. 64.

Dragosits, M. *et al.* (2013) 'Evolutionary potential, cross-stress behaviour and the genetic basis of acquired stress resistance in Escherichia coli.', *Molecular systems biology*. Nature Publishing Group, 9(643), p. 643.

Elena, S. F. and Lenski, R. E. (2003) 'Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation.', *Nature reviews. Genetics*, 4(6), pp. 457–69.

Evans, D. F. *et al.* (1988) 'Measurement of gastrointestinal pH profiles in normal ambulant human subjects', *PLoS Genetics*, 29(48), pp. 1035–1041.

Finn, T. J. *et al.* (2017) 'Dynamics and genetic diversification of *Escherichia coli* during experimental adaptation to an anaerobic environment', *PeerJ*, 5, p. e3244.

Foster, J. W. (2004) 'Escherichia coli acid resistance : tales of an amateur acidophile', *Nature reviews. Microbiology*, 2(November), pp. 898–907.

Fotadar, U., Zaveloff, P. and Terracio, L. (2005) 'Growth of Escherichia coli at elevated temperatures', *Journal of Basic Microbiology*, 45(5), pp. 403–404.

Gonzalez, A. and Bell, G. (2013) 'Evolutionary rescue and adaptation to abrupt environmental change depends upon the history of stress.', *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, 368(1610), p. 20120079.

Harden, M. M. *et al.* (2015) 'Acid-adapted strains of Escherichia coli K-12 obtained by experimental evolution', *Applied and Environmental Microbiology*, 81(6), pp. 1932–1941.

Herring, C. D. *et al.* (2006) 'Comparative genome sequencing of Escherichia coli allows observation of bacterial evolution on a laboratory timescale.', *Nature genetics*, 38(12), pp. 1406–1412.

Hryckowian, A. J. *et al.* (2014) 'IraL is an RssB anti-adaptor that stabilizes RpoS during logarithmic phase growth in Escherichia coli and Shigella', *mBio*, 5(3), pp. 1–8.

Hu, P. *et al.* (2009) 'Global functional atlas of Escherichia coli encompassing previously uncharacterized proteins', *PLoS Biology*, 7(4), pp. 0929–0947.

Ibarra, R. U., Edwards, J. S. and Palsson, B. O. (2002) 'Escherichia coli K-12 undergoes adaptive evolution to achieve in silico predicted optimal growth', *Nature*, 420(6912), pp. 186–189.

Ibekwe, V. C. *et al.* (2008) 'Interplay between intestinal pH, transit time and feed status on the in vivo performance of pH responsive ileo-colonic release systems', *Pharmaceutical Research*, 25(8), pp. 1828–1835.

Iuchi, S. *et al.* (1990) 'The arcB gene of Escherichia coli encodes a sensor-regulator protein for anaerobic repression of the arc modulon', *Molecular Microbiology*, 4(5), pp. 715–727.

Iyer, R., Williams, C. and Miller, C. (2003) 'Arginine-Agmatine Antiporter in Extreme Acid Resistance in Escherichia coli Arginine-Agmatine Antiporter in Extreme Acid Resistance in Escherichia coli', *Journal of Bacteriology*, 185(22), pp. 6556–6561.

Jordan, K. N., Oxford, L. and O'Byrne, C. P. (1999) 'Survival of low-pH stress by Escherichia coli O157:H7: Correlation between alterations in the cell envelope and increased acid tolerance', *Applied and Environmental Microbiology*, 65(7), pp. 3048–3055.

Kaas, R. S. *et al.* (2012) 'Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes', *BMC Genomics*, 13(1), p. 577.

Kram, K. E. *et al.* (2017) 'Adaptation of Escherichia coli to Long-Term Serial Passage in Complex Medium: Evidence of Parallel Evolution', *mSystems*, 2(2), pp. e00192-16.

Kannan, G. *et al.* (2008) 'Rapid acid treatment of Escherichia coli: transcriptomic response and recovery.', *BMC Microbiology*, 8(1), p. 37.

Kivisaar, M. (2010) 'Mechanisms of stationary-phase mutagenesis in bacteria: Mutational processes in pseudomonads', *FEMS Microbiology Letters*, 312(1), pp. 1–14.

Koseki, S. and Yamamoto, K. (2006) 'Recovery of Escherichia coli ATCC 25922 in phosphate buffered saline after treatment with high hydrostatic pressure', *International Journal of Food Microbiology*, 110(1), pp. 108–111.

LaCroix, R. A. *et al.* (2015) 'Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium', *Applied and Environmental Microbiology*, 81(1), pp. 17–30.

LaCroix, R. A. *et al.* (2015) 'Use of adaptive laboratory evolution to discover key mutations enabling rapid growth of Escherichia coli K-12 MG1655 on glucose minimal medium', *Applied and Environmental Microbiology*, 81(1), pp. 17–30.

Larsson, G., Enfors, S. -O and Pham, H. (1990) 'The pH-auxostat as a tool for studying microbial dynamics in continuous fermentation', *Biotechnology and Bioengineering*, 36(3), pp. 224–232.

Lawrence, J. G. and Ochman, H. (1998) 'Molecular archaeology of the Escherichia coli genome.', *Proceedings of the National Academy of Sciences of the United States of America*, 95(16), pp. 9413–9417.

Lech, T. (2016) 'Ancient DNA in historical parchments - Identifying a procedure for extraction and amplification of genetic material', *Genetics and Molecular Research*, 15(2).

Lenski, R. E. *et al.* (2015) 'Sustained fitness gains and variability in fitness trajectories in the long-term evolution experiment with Escherichia coli', *bioRxiv*, (282), p. 20152292.

Leuko, S. and Raivio, T. L. (2012) 'Mutations that impact the enteropathogenic escherichia coli Cpx envelope stress response attenuate virulence in galleria mellonella', *Infection and Immunity*, 80(9), pp. 3077–3085. doi: 10.1128/IAI.00081-12.

Lukjancenko, O., Wassenaar, T. M. and Ussery, D. W. (2010) 'Comparison of 61 Sequenced Escherichia coli Genomes', *Microbial Ecology*, 60(4), pp. 708–720.

Lin, J. *et al.* (1995) 'Comparative analysis of extreme acid survival in Salmonella typhimurium, Shigella flexnari, and Escherichia coli. *Journal of bacteriology*, 177(14), pp. 4097–4104.

Lin, J. *et al.* (1996) 'Mechanisms of acid resistance in enterohemorrhagic Escherichia coli', *Appl Environ Microbiol*, 62(9), pp. 3094–3100.

Łó, J. M. *et al.* (2013) 'Hypothesis and theory article Altruism of Shiga toxin-producing Escherichia coli: recent hypothesis versus experimental results', *Frontiers in Cellular and Infection Microbiology*, 4(January), p. 166.

Lourenço, M. *et al.* (2016) 'A Mutational Hotspot and Strong Selection Contribute to the Order of Mutations Selected for during Escherichia coli Adaptation to the Gut', *PLoS Genetics*, 12(11), pp. 1–23.

Lukacs, G. L., Rotstein, O. D. and Grinstein, S. (1991) 'Determinants of the Phagosornal pH in Macrophages', *The Journal of biological chemistry*, 266(36), pp. 24540–24548.

Márcia Silva Campos Mata, G., Moura Ferreira, G. and Spira, B. (2017) 'RpoS role in virulence and fitness in enteropathogenic Escherichia coli', *PLoS one*, pp. 1–18.

Ma, Z., Richard, H. and Foster, J. W. (2003) 'pH-Dependent Modulation of Cyclic AMP Levels and GadW-Dependent Repression of RpoS Affect Synthesis of the GadX Regulator and Escherichia coli Acid Resistance', *Journal of Bacteriology*, 185(23), pp. 6852–6859.

Ma, Z., Richard, H. and Foster, J. W. (2003) 'pH-Dependent Modulation of Cyclic AMP Levels and GadW-Dependent Repression of RpoS Affect Synthesis of the GadX Regulator and Escherichia coli Acid Resistance', *Journal of Bacteriology*, 185(23), pp. 6852–6859.

Martin, M. *et al.* (2016) 'Laboratory evolution of microbial interactions in bacterial biofilms', *Journal of Bacteriology*, 198(19), pp. 2564–2571.

Matsushika, A. and Mizuno, T. (1998) 'A dual-signalling mechanism mediated by the ArcB hybrid sensor kinase containing the histidine-containing phosphotransfer domain in Escherichia coli', *Journal of Bacteriology*, 180(15), pp. 3973–3977.

Mears, P. J. *et al.* (2014) 'Escherichia coli swimming is robust against variations in flagellar number', *eLife*, 2014(3), pp. 1–18.

Mellata, M. (2013) 'Human and avian extraintestinal pathogenic Escherichia coli: infections, zoonotic risks, and antibiotic resistance trends.', *Foodborne pathogens and disease*, 10(11), pp. 916–32.

Meng, W. *et al.* (2000) 'The Escherichia coli RNA polymerase alpha subunit linker: length requirements for transcription activation at CRP-dependent promoters.', *The EMBO journal*, 19(7), pp. 1555–1566.

Metris, a *et al.* (2014) 'E. coli under Salt Stress: Metabolic Shift in the Presence of Glycine Betaine.', *Applied and environmental microbiology*, 80(15), pp. 4745–4756.

Mika, F. and Hengge, R. (2005) 'A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of ??S (RpoS) in E. coli', *Genes and Development*, 19(22), pp. 2770–2781.

Papadopoulos, D. *et al.* (1999) 'Genomic evolution during a 10,000-generation experiment with bacteria.', *Proceedings of the National Academy of Sciences of the United States of America*, 96(7), pp. 3807–3812.

Papatzitze, O. *et al.* (2011) 'Study of the Diversity of E . Coli Strains Isolated From Aquatic Environments', *Microbial cell factories*, 1551(9), pp. 155–161.

Park, D. M. *et al.* (2013) 'The Bacterial Response Regulator ArcA Uses a Diverse Binding Site Architecture to Regulate Carbon Oxidation Globally', *PLoS Genetics*, 9(10).

Partridge, J. D. and Harshey, R. M. (2013) 'More than motility: Salmonella flagella contribute to overriding friction and facilitating colony hydration during swarming', *Journal of Bacteriology*, 195(5), pp. 919–929.

Pesavento, C. (2011) 'Regulatory mechanisms in the coordination of motility and curli fimbriae-mediated adhesion in Escherichia coli', *Journal of bacteriology*, 7(April), pp. 1919–1934.

Philippe, N. *et al.* (2007) 'Evolution of global regulatory networks during a long-term experiment with Escherichia coli', *BioEssays*, 29(9), pp. 846–860.

Ponciano, J. M. *et al.* (2009) 'Evolution of diversity in spatially structured Escherichia coli populations', *Applied and Environmental Microbiology*, 75(19), pp. 6047–6054.

Portnoy, V. A., Bezdan, D. and Zengler, K. (2011) 'Adaptive laboratory evolution — harnessing the power of biology for metabolic engineering', *Current Opinion in Biotechnology*. Elsevier Ltd, 22(4), pp. 590–594.

Porwollik, S., Wong, R. M.-Y. and McClelland, M. (2002) 'Evolutionary genomics of *Salmonella*: Gene acquisitions revealed by microarray analysis', *Proceedings of the National Academy of Sciences*, 99(13), pp. 8956–8961.

Pup, G. M., Lan, R. and Reeves, P. R. (2000) 'Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics.', *Proceedings of the National Academy of Sciences of the United States of America*, 97(19), pp. 10567–72.

Raeside, C. *et al.* (2014) 'Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli.*', *mBio*, 5(5), pp. e01377-14.

Rangel, J. M. *et al.* (2005) 'Epidemiology of Escherichia coli O157:H7 outbreaks, United States, 1982-2002', *Emerging Infectious Diseases*, 11(4), pp. 603–609.

Rasko, D. A. *et al.* (2008) 'The pangenome structure of Escherichia coli: Comparative genomic analysis of E. coli commensal and pathogenic isolates', *Journal of Bacteriology*, 190(20), pp. 6881–6893.

Riehle, M. M., Bennett, A. F. and Long, A. D. (2001) 'Genetic architecture of thermal adaptation in Escherichia coli', *Proceedings of the National Academy of Sciences*, 98(2), pp. 525–530.

Riehle, M. M. *et al.* (2003) 'Evolutionary changes in heat-inducible gene expression in lines of Escherichia coli adapted to high temperature', *Physiological Genomics*, 14(1), pp. 47–58.

Russo, T. A. and Johnson, J. R. (2003) 'Medical and economic impact of extra intestinal infections due to Escherichia coli: Focus on an increasingly important endemic problem', *Microbes and Infection*, 5(5), pp. 449–456.

Rutherford, S. L. (2000) 'Rutherford SL. From genotype to phenotype: buffering mechanisms and the storage of genetic From genotype to phenotype: buffering mechanisms and the storage of genetic information', *BioEssays*, 22(December 2000), pp. 1095–1105.

Sandberg, T. E. *et al.* (2014) 'Evolution of *Escherichia coli* to 42 °c and subsequent genetic engineering reveals adaptive mechanisms and novel mutations', *Molecular Biology and Evolution*, 31(10), pp. 2647–2662.

Sannasiddappa, T. H. *et al.* (2011) 'The influence of *Staphylococcus aureus* on gut microbial ecology in an in vitro continuous culture human colonic model system', *PLoS ONE*, 6(8), pp. 2–9.

Savery, N. J. *et al.* (2002) 'Determinants of the C-terminal domain of the Escherichia coli RNA polymerase alpha subunit important for transcription at class I cyclic AMP receptor protein-dependent promoters.', *Journal of bacteriology*, 184(8), pp. 2273–80.

Saxer, G. *et al.* (2014) 'Mutations in Global Regulators Lead to Metabolic Selection during Adaptation to Complex Environments', *PLoS Genetics*, 10(12).

Scheutz, F. and Strockbine, N. A. (2005) 'Genus I. Escherichia Castellani and Chalmers 1919, 941T', *Bergey's Manuel of Systematic Bacteriology*, 2, pp. 607–624.

Schluter, J. and Foster, K. R. (2012) 'The Evolution of Mutualism in Gut Microbiota Via Host Epithelial Selection', *PLoS Biology*, 10(11).

Schneider, D. *et al.* (2000) 'Long-term experimental evolution in *Escherichia coli*. IX. Characterization of insertion sequence-mediated mutations and rearrangements', *Genetics*, 156(2), pp. 477–488.

Schneider, D. and Lenski, R. E. (2004) 'Dynamics of insertion sequence elements during experimental evolution of bacteria', *Research in Microbiology*, 155(5), pp. 319–327.

Scott, D. R. *et al.* (2002) 'Mechanisms of acid resistance due to the urease system of Helicobacter pylori', *Gastroenterology*, 123(1), pp. 187–195.

Sen, H. *et al.* (2017) 'Structural and functional analysis of the *Escherichia coli* acid-sensing histidine kinase, EvgS', *Journal of Bacteriology*, (July), p. JB.00310-17.

Sernova, N. V. and Gelfand, M. S. (2012) 'Comparative Genomics of CytR, an Unusual Member of the LacI Family of Transcription Factors', *PLoS ONE*, 7(9), pp. 1–16.

Shah, D. *et al.* (2006) 'Persisters: a distinct physiological state of E. coli.', *BMC microbiology*, 6, p. 53.

Shala, A. A., Restrepo, S. and Gonzalez Barrios, A. F. (2011) 'A network model for biofilm development in Escherichia coli K-12.', *Theoretical biology & medical modelling*. BioMed Central Ltd, 8(1), p. 34.

Sheldon, J. R. *et al.* (2012) 'Role of rpoS in *Escherichia coli* O157:H7 strain H32 biofilm development and survival', *Applied and Environmental Microbiology*, 78(23), pp. 8331–8339.

Sultan, Z. *et al.* (2004) 'Comparison of mismatch amplification mutation assay with DNA sequencing for characterization of fluoroquinolone resistance in Neisseria gonorrhoeae', *Journal of Clinical Microbiology*, 42(2), pp. 591–594.

Touchon, M. *et al.* (2009) 'Organised genome dynamics in the Escherichia coli species results in highly diverse adaptive paths', *PLoS Genetics*, 5(1).

Tyerman, J. G. *et al.* (2013) 'The evolution of antibiotic susceptibility and resistance during the formation of Escherichia coli biofilms in the absence of antibiotics', *BMC Evolutionary Biology*, 13(9), p. 22.

Vogt, R. L. and Dippold, L. (2005) 'Escherichia coli O157:H7 outbreak associated with consumption of ground beef, June-July 2002.', *Public health reports*, 120(2), pp. 174–178.

Wang, S. *et al.* (2009) 'Transcriptomic response of Escherichia coli O157:H7 to oxidative stress', *Applied and Environmental Microbiology*, 75(19), pp. 6110–6123.

Wei, Y., Ocampo, P. and Levin, B. R. (2010) 'An experimental study of the population and evolutionary dynamics of *Vibrio cholerae* O1 and the bacteriophage JSF4.', *Proceedings. Biological sciences / The Royal Society*, 277(1698), pp. 3247–54.

Wu, X. *et al.* (2014) 'Adaptation of *Escherichia coli* to elevated sodium concentrations increases cation tolerance and enables greater lactic acid production', *Applied and Environmental Microbiology*, 80(9), pp. 2880–2888.

Wu, X. *et al.* (2014) 'Adaptation of *Escherichia coli* to elevated sodium concentrations increases cation tolerance and enables greater lactic acid production', *Applied and Environmental Microbiology*, 80(9), pp. 2880–2888.

Zhou, Y. and Gottesman, S. (2006) 'Modes of regulation of RpoS by H-NS', *Journal of Bacteriology*, 188(19), pp. 7022–7025.