Ph.D. Thesis

Understanding bacterial adaptation to aerobic and anaerobic environments through experimental evolution and whole genome analysis

Thomas Finn

December 2014

A thesis presented in fulfilment of the requirements for the degree of Doctor of Philosophy in Genetics at Massey University, Palmerston North, New Zealand

Abstract

Facultative anaerobic organisms have the metabolic versatility to grow in both aerobic and anaerobic environments. However, molecular mechanisms that underpin adaptation to anaerobic environments are not well understood. This study aims to understand how the facultative anaerobe, *Escherichia coli*, adapts to environments that vary in oxygen content. An experimental evolution experiment was conducted in which replicate lineages were established from a pre-evolved clonal culture of *E. coli* REL4536. Lineages were serially sub-cultured for 4,000 generations within strict aerobic and strict anaerobic environments, and a treatment that fluctuated between the two environments. Significant increases in the relative fitness of lineages exposed to anaerobic conditions were observed, whereas the relative fitness of lineages in aerobic conditions did not increase, likely as the ancestor had been pre-adapted to aerobic growth.

Mutations that arose during evolution were identified by genome sequencing randomly-selected clones from each lineage at 2,000 and 4,000 generations. Traits that contributed to adaptation were predicted via the occurrence of independent mutations affecting common traits among lineages. Adaptation to the anaerobic environment was facilitated by modifications to anaerobic fermentation and the inactivation of virulence genes, whereas in the aerobic environment, mutations predicted to confer a growth advantage in stationary phase were observed. The evolution of generalists involved traits that were similar to those found in both aerobic and anaerobically evolved lineages, as well as the deletion of cryptic prophages from the genome and modifications to amino acid transport.

Phenotypically distinct small colony morphotypes (SCM) arose within anaerobic lineages and two separate adaptive pathways are hypothesised for this divergence. SCM1 were capable of stable co-existence with co-evolved cells of typical colony morphotype, most likely through an acetate cross-feeding mechanism. In contrast, SCM2 was able to out-compete the ancestor within 14 days, despite exhibiting a lower growth rate than the ancestor. SCM2 likely evolved the ability to inhibit the ancestral strain through a contact dependent inhibition mechanism, as evidenced by a mutation in *glgC*. This thesis demonstrates the complex nature of adaptation to anaerobic environments, as revealed by experimental evolution and whole genome sequencing.

Acknowledgements

First and foremost, I would like to thank my supervisors, Dr Christina Moon, Dr Sinead Leahy, Dr Rees Kassen and Prof Paul Rainey. Your advice and insights into this project were invaluable. In particular Christina and Sinead, your guidance, support and constant encouragement over the years have been appreciated every second of the way. Giving me the opportunity to partake in this project brought me to New Zealand, and represents the beginning of my own adult life, and for that I am forever grateful to you both.

I would like to acknowledge AgResearch Ltd and Massey University for providing financial assistance and an excellent working environment during the course of my studies. I would also like to thank the Marsden Fund for providing funding for this project.

I would like to thank a great bunch of people, collectively known as the Rumen Microbiology team at AgResearch, Grasslands. Working amongst you all has been a great pleasure. In particular thanks to Dr Bill Kelly, Dr Suzanne Lambie, Dr Gemma Henderson for assistance during my first year. Thanks also to Dr Peter Janssen and Dr Ron Ronimus for helpful discussions, and Dr Dragana Gagic for her sense of humour. For technical assistance in the lab, thank you Carrie Sang, Dong Li and Dr Sam Noel and especially Rechelle Perry, for her friendship and showing me the ropes with what would become my beloved anaerobic chamber. Outside of the Rumen Microbiology team, thank you also to Dr Mark McCann, Paul Mclean, Paul Lovejoy and Denise Martin. To all those in the AgResearch student room, both past and present, the daily laughs and therapy sessions we enjoyed while sitting around the student table in the AgReseach café kept me hanging in there.

To Dr Milica Ciric, your friendship has meant the world to me. To Sonal Shewaramani and Renee Atua, I cannot think of anyone else who would have made the last few years as much fun as they were, thank you both for that, and here's to many more! To Jamie Lange, thank you for consistently being so dependable, understanding and supportive throughout this process. Without encouragement from you four, I would never have reached this point.

Lastly, I would like to dedicate this thesis to my parents Tom and Collette Finn, and brother, Jim. Being so far away from you has been hard on us all. I hope you can see it as a necessary step towards the only goal that matters to me; to make you guys proud.

Table of contents

List of figures

List of tables

Non-standard abbreviations

Abbreviation	Meaning
Acetyl-CoA	Acetyl coenzyme A
AE	Aerobic
AN	Anaerobic
Anc	Ancestor
AP	Antagonistic pleiotropy
ATP	Adenosine triphosphate
bp	Base pairs
ca	Circa
cAMP	Cyclic adenosine monophosphate
CFU	Colony forming units
d	Day
DNA	Deoxyribonucleic acid
FL	Fluctuating
FSW	Fast switcher
FS	Fuzzy spreader
g	Gravity
GASP	Growth advantage in stationary phase
Gb	Gigabase pairs
GCR	Gross chromosomal rearrangements
hrs	Hours
IS	Insertion element
K	Thousand
kb	Kilo bases
L	Litre
LG	Large
LTEE	Long-term experimental evolution
M	Molar
MA	Mutation accumulation
Mb	Megabase pairs
MP	Mate pair
NADH	Nicotinamide adenine dinucleotide
NFDS	Negative frequency dependent selection
NGS	Next generation sequencing
nm	Nanometers
OD	Optical density
PCR	Polymerase chain reaction
PE	Paired end
PEP	Phosphoenolypyruvate
ppGpp	Guanosine pentaphosphate
Rif^r	Rifampicin resitent reference strain
RNA	Ribonucleoic acid
ROS	Reactive oxygen species
rpm	Rotations per minute
SCDI	Stationary phase contact-dependent inhibition
SCM	Small colony morphotype

SCV	Single colony variant
SL	Small
SM	Smooth
SNP	Single nucleotide polymorphisms
SSW	Slow switcher
TA	Toxin-antitoxin
TAra	Tetrazolium and arabinose
TCA	Tricarboxylic acid
TCM	Typical colony morphotype
U	Units
UPEC	Urethropathogenic
UV	Ultra violet
WGS	Whole genome sequencing
WS	Wrinkly spreader
wt/vol	Weight/volume

Chapter One : Literature review

1.1. Introduction

Historically, life on earth evolved in anaerobic conditions, that is, in the absence of oxygen (1). Photosynthetic organisms living under strict anaerobic conditions, such as Cyanobacteria, played a major role in the the rise of oxygen levels in the atmosphere (2). Today**,** many micro-organisms are still distributed among numerous anaerobic habitats such as in the soil, deep sea thermal vents and the gastro-intestinal tracts of animals. Among these microbes, many are important contributors to human health. The majority of the indigenous symbiotic gastrointestinal microflora consists of obligate anaerobic bacteria, important in promoting digestion of certain foods and stimulating the host immune system (3). Conversely, many obligately anaerobic organisms can be harmful to human health, , such as certain pathogenic *Clostridium* species including *C. botulinum*, *C. difficile*, and *C. tetani* (4). In contrast, facultative anaerobes are have evolved the extraordinary metabolic diversity to grow and reproduce in both aerobic (oxygen rich) and anaerobic environments. These also include human pathogens, such as *Staphylococcus aureus* and *Escherichia coli* that are frequently associated with hospital-acquired infections, and urinary tract infections respectively (5). However, many human commensals that are important for gut health are also facultative anaerobes. Little is known about how such organisms adapt to diverse natural environments that vary with regard to oxygen content.

The broad aim of this thesis is to understand how evolution occurs within strictly aerobic and anaerobic environments. Evolution is the process of descent with modification, where gene frequencies within a population change over time. There are a number of processes that can contribute to changes in gene frequency. Changes in frequency can be random, by genetic drift, or based on reproductive advantage, by natural selection. Natural selection is the evolutionary process whereby organisms that possess traits that allow them to better grow and reproduce in their environment are selected for, resulting in improved survivability in that environment. Such selected organisms are said to have undergone adaptation and the more advantageous the trait, the more common it is likely to become within the population. For traits that are strongly selected

for, they may reach fixation within the population, where all individuals within the population possess the trait, and the underlying mutations that confers it.

Extensive work has been done to understand the biochemical pathways and genetic innovations at the molecular level that are associated with adaptation to aerobic environments (6), but little is known about those which are important for adaptation to anaerobic environments.*Escherichia coli* is a Gram-negative facultative anaerobic bacterium. The normal habitat of *E. coli* is in the lower gastrointestinal tract of mammals, a predominantly anaerobic environment. However, when contents of the lower intestine are excreted, *E. coli* is exposed to an aerobic environment. In order to survive such a varied life cycle, *E. coli* has evolved the capability of efficient energy production and growth in both aerobic and anaerobic environments. In this thesis, I use the facultative anaerobic organism *E. coli*, to understand the dynamics of evolution under aerobic and anaerobic conditions. Furthermore, I am also interested in how the organism adapts to an environment that fluctuates with regard to oxygen content. I will identify the genes and mutations that are most important for adaptation to the two environments and gain an insight into the evolutionary pathways involved in adaptation to the aerobic and anaerobic environments.

1.1.1. Growth in an aerobic environment

In the presence of high concentrations of oxygen, aerobic and facultative anaerobic organisms such as *E. coli* produce energy through aerobic respiration (**Figure 1.1**). There are three main components to aerobic growth on simple sugars such as glucose: glycolysis, the tricarboxylic acid (TCA) cycle and the electron transport chain (7). During glycolysis, the sugar source, such as glucose, is broken down to phosphoenolpyruvate (PEP) and then to pyruvate. Pyruvate is then converted to acetyl-coenzyme A (acetyl-CoA) by the enzyme pyruvate dehydrogenase (8). The second stage occurs when acetyl-CoA becomes fully oxidised by eight enzymatic steps, occurring in a cyclical sequence, known as the TCA cycle (9). For every one pyruvate molecule, four nicotinamide adenine dinucleotide (NADH) molecules are produced by the TCA cycle. NADH is the molecular mediator between the second and final step of respiration; the electron transport chain. NADH acts as a reducing agent by donating electrons, obtained from the TCA cycle, to the electron transport. When oxygen is present, energy is produced by aerobic respiration (10). The end point of respiration is the reduction of the terminal electron acceptor, oxygen, which is reduced to water. Due to the complete breakdown of glucose, aerobic respiration yields a sum total of 36 adenosine triphosphate (ATP) molecules per unit glucose (11).

1.1.2. Anaerobic growth

However, respiration is not solely dependent on the presence of oxygen. In faculatative anaerobes, such as *E. coli,* respiration can also occur in an oxygen free environment, in a process known as anaerobic respiration. In the absence of oxygen, alternative electron acceptors may be used, such as nitrate (NO_3^-), sulphate (SO_4^{2-}), sulphur (S^0) and fumarate ($C_4H_4O_4$). In these conditions, the respiratory electron transport chain is used, however, electrons are donated from NADH to these compounds instead of oxygen. These alternative electron acceptors have less affinity for electrons than oxygen, resulting in less energy per oxidised moleucle. For this reason, energy yields from anaerobic respiration are limited in comparison to those of aerobic respiration.

In the absence of oxygen or the other four alternative electron acceptors, the TCA cycle and oxidative phosphorylation pathways are inoperable (12). Therefore, facultative anaerobic organisms and obligate anaerobic organisms rely upon fermentation for energy production (**Figure 1.1**). Fermentation consists of two parts: glycolysis (similar to the first step of respiration), followed by substrate-level phosphorylation (13). Once again, glucose is oxidised to pyruvate via glycolysis. However, pyruvate cannot be oxidised via the TCA cycle without oxygen or the alternative terminal electron acceptors discussed in **Section 1.1.1**, thus,, the cell must divert to fermentative pathways. Glycolysis is the major energy generating process of anaerobic growth (9). Rather than entering the TCA cycle, substrate level phosphorylation occurs. Pyruvate can be catabolised by three separate reactions through the action of i) lactate dehydrogenase resulting in the production of lactate, ii) pyruvate formate lyase resulting in the production of formate or iii) pyruvate dehydrogenase resulting in the formation of acetyl-CoA (**Figure 1.2**). Acetyl-CoA can be further broken down to acetate by phosphate acetyltransferase and acetate kinase or to ethanol by acetaldehyde dehydrogenase and alcohol dehydrogenase. In addition, substrate level phosphorylation can happen directly at the phosphoenolpyruvate stage with the action of phosphenolpyruvate carboxykinase. A further three enzymatic conversion steps leads to the production of succinate. Succinate, acetate and lactate are each converted to acids and excreted from the cell as waste products. Formate is broken down to hydrogen and carbon dioxide, while ethanol is excreted directly. In all cases, glucose is only partially broken down. Thus, anaerobic fermentation is a much less productive form of energy generation, yielding only two molecules of ATP per unit glucose. However, fermentation proceeds at high rates due to the strong expression of fermentation enzymes under anaerobic conditions (14).

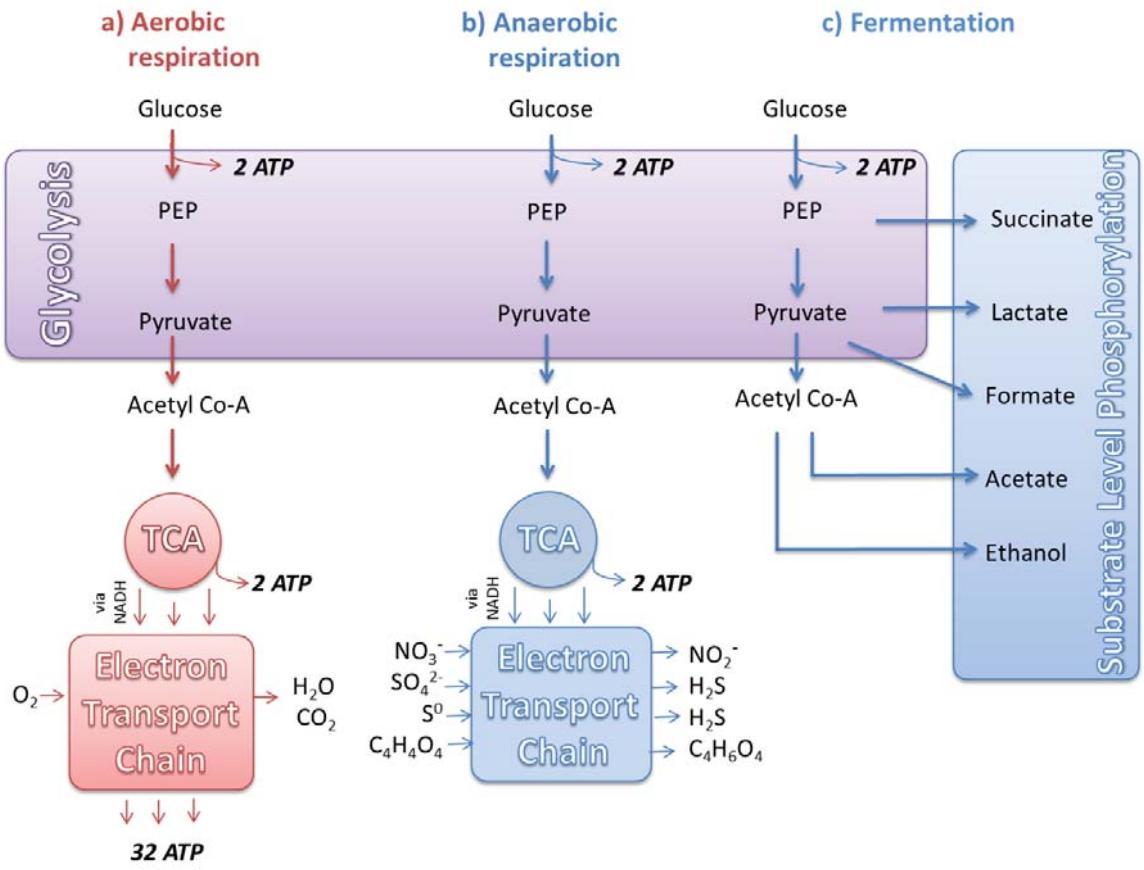

Figure 1.1: Comparison of respiration in aerobic environments and fermentation in anaerobic environments in *E. coli*. For a) aerobic respiration, the three stages are shown: glycolysis; the TCA cycle and the electron transport phosphorylation chain. For b) anaerobic fermentation the two stages are shown; glycolysis and substrate level phosphorylation. The net sum of ATP generated at each step during respiration and fermentation are indicated. PEP denotes phosphoenolpyruvate.

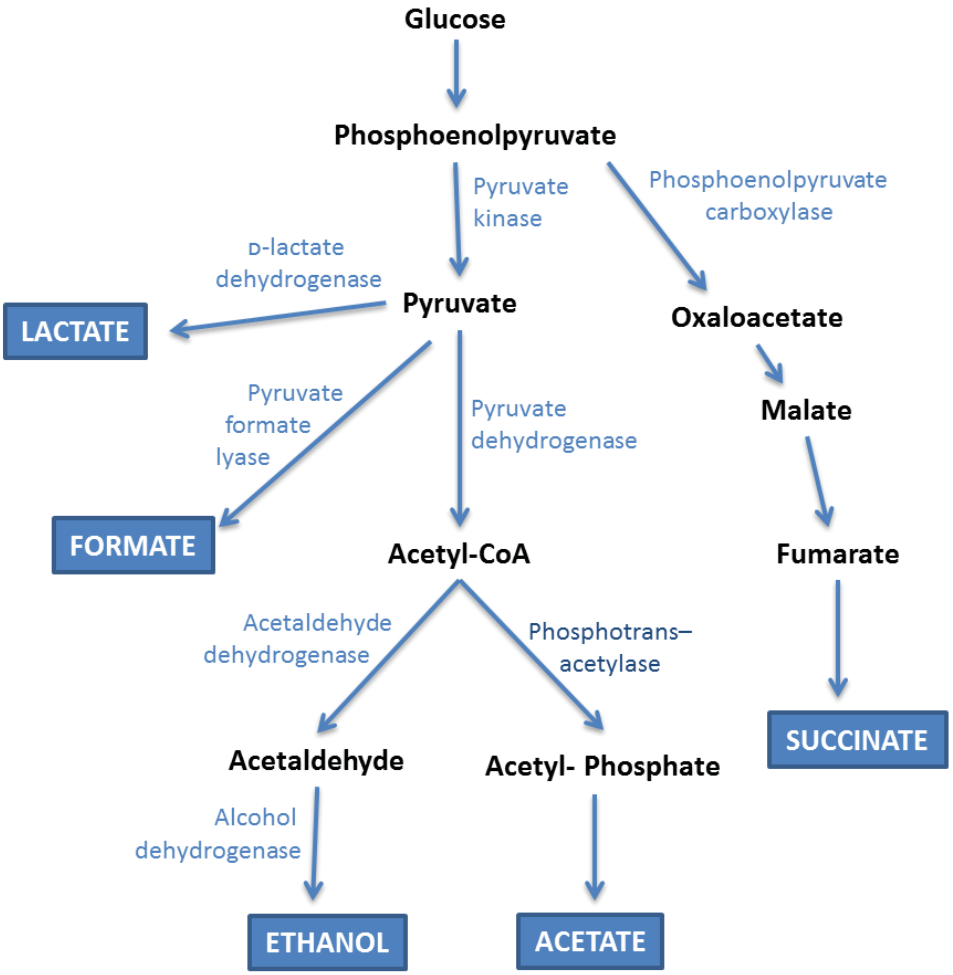

Figure 1.2: Enzymes involved in anaerobic fermentation. The various biochemical reactions involved in the anaerobic fermentation of glucose are indicated. The enzymes involved in selected pathways are shown in blue. Also indicated are the five end-products in blue boxes. See text for details. Figure adapted from Gunsalus *et al.* 1994 (14).

1.1.3. Fluctuating between aerobic and anaerobic growth in *E. coli*

A facultative anaerobic organism must be capable of efficiently switching between respiration and fermentation. Rapid transition between the two energy generating systems is essential for successful growth of facultative anaerobic organisms (9). Physiological alterations allow the cell to switch between respiration and fermentation and ensure that the transition is made readily and smoothly (10). Environmental signals such as the presence or absence of oxygen are fundamental for this response, and result in the biosynthesis of enzymes involved in either respiration or fermentation as appropriate to the environment. Metabolic re-wiring is achieved principally by

transcriptional control and enzyme activation/inhibition (15). At the molecular level, this is mediated through a combination of two global regulatory proteins; the two-component ArcBA (aerobic respiration control) regulatory network and the transcription factor Fnr (fumarate and nitrate reductase). The regulatory elements of these two pathways act in combination to coordinate the flow of carbon into respiration or fermentation in response to the environment.

The ArcBA system is a two component system consisting of the membrane-bound ArcB which senses the environmental signal and the receiver ArcA (**Figure 1.3**). Under aerobic conditions, the ArcB receiver remains unphosphorylated (16), while under anaerobic conditions the ArcBA regulatory network becomes active (16-22). A phosphorylation signal transduction occurs, resulting in the phosphorylation of ArcA via its N-terminus domain. As well as the receiver N-terminus domain, the ArcA protein contains a helix-turn-helix DNA binding domain at its C-terminus and a transcriptional regulator. Once stimulated by ArcB phosphorylation, this transcription factor negatively regulates genes involved in respiration such as *gltA* encoding isocitrate synthase and *icd* encoding citrate dehydrogenase and other genes encoding enzymes for the TCA cycle (14). In total, the ArcBA network represses 17 operons (15).

Co-ordinately, the Fnr transcription factor is also activated under anaerobic conditions (**Figure 1.3**). However, unlike ArcBA, activated Fnr functions as both an activator and a repressor (23). Under aerobic conditions, Fnr exists in the cell as non-functional monomers. Each monomer of Fnr can sense dioxygen directly via redox sensitive $[2Fe\text{-}2S]^{2+}$ clusters located on the N-terminus of the protein. Upon exposure of oxygen to the iron sulphur cluster, monomers become active, and form $[4Fe\text{-}2S]^{2+}$ clusters which causes the monomers to dimerise. Activated Fnr dimers can bind DNA, recognising a highly conserved TTGAT consensus sequence in the promoters of many genes. Fnr stimulates or represses transcription of different genes involved in respiration or fermentation. Fermentation proceeds when Fnr dimers bind to DNA at -39 and -49 bps upstream of the transcriptional start site and act as transcriptional activators of 22 fermentation genes such as pyruvate formate lyase and fumarase B (15). Simultaneously, respiration is blocked via negative regulation of respiration genes**.** Fnr dimers act as repressors of transcription of respiration genes by binding to a region which either overlaps or is adjacent to the sigma 70 recognition sequence, thus blocking binding of RNA polymerase and thereby blocking the transcription of eight respiration genes such as iso-citrate dehydrogenase (24).

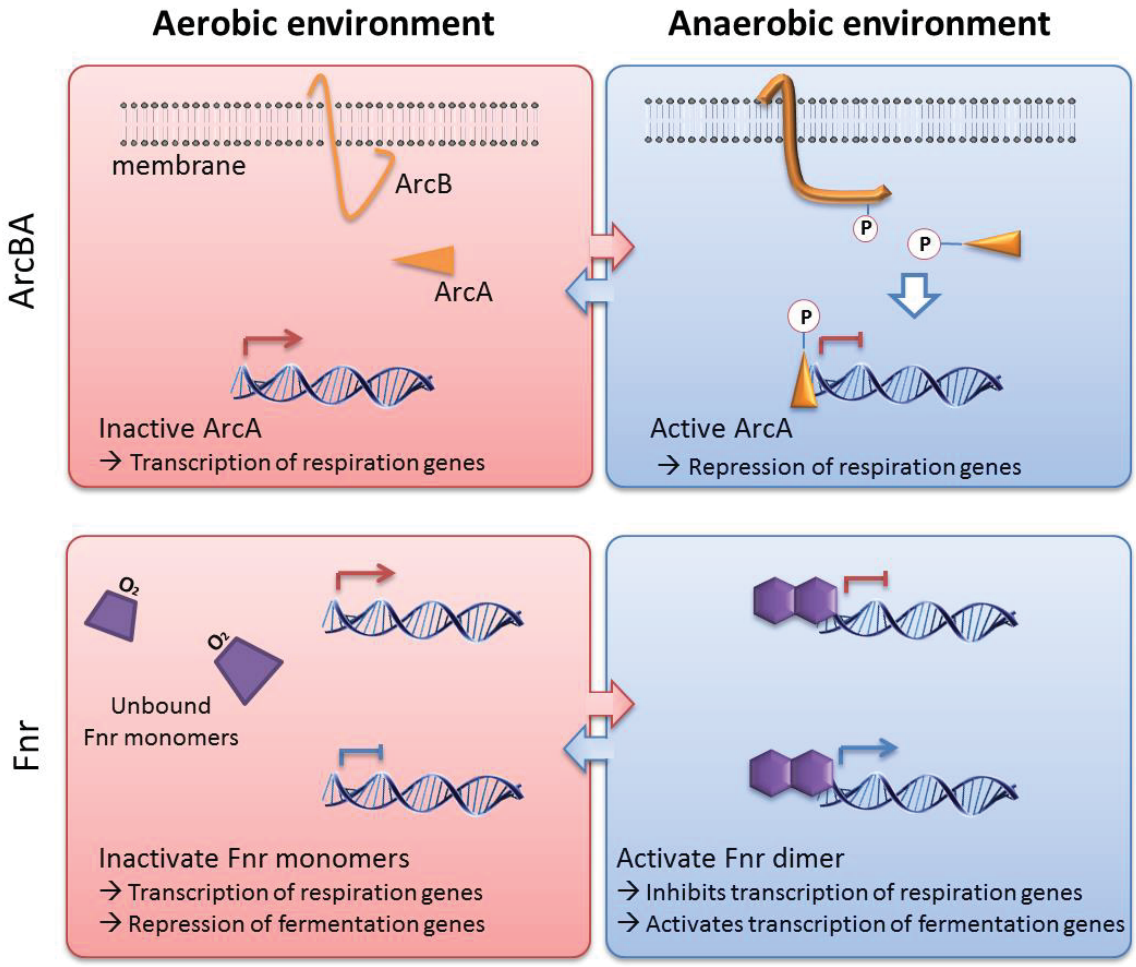

Figure 1.3: The ArcBA and Fnr aerobic/anaerobic response reactions. Both the ArcBA and Fnr global control mechanisms used by *E. coli* during transitions from aerobic to anaerobic and anaerobic to aerobic environments are shown. See text for details. Figure adapted from Gunsalus *et al.* 1994 (14).

1.1.4. Impact of aerobic and anaerobic metabolism on the genome

While the aerobic environment is favourable for growth when compared to the anaerobic,environment, due to higher ATP yields in the former, the use of oxygen as a terminal electron acceptor may come at a cost (25, 26). Molecular oxygen (O_2) can be readily modified to toxic by-products in the cell known as reactive oxygen species (ROS). These are normal by-products of aerobic respiration (25-27). These by-products arise as O_2 becomes highly reactive upon accepting electrons. Acceptance of one, two or three electrons results in the formation of superoxide radicals (O_2^-), hydrogen peroxide (H_2O_2) and hydroxyl radicals (OH•) respectively (26). These compounds are frequently implicated in mutagenesis (15, 28-30). Further description of ROS-mediated mutations is provided in **Section 1.2.1.1.1**. As the majority of DNA sequence

modifications are harmful to the cells in which they arise, aerobic organisms have evolved complex prevention and repair mechanisms (31) to mitigate the damaging effects of ROS, such as the SOS response and MutM/MutY in response to oxidative damage (26). During anaerobic growth, the impact of ROS is much less, thus it is expected that specific ROS-mediated mutations are less prevalent. However, it has also been shown that the mutation rate (per generation) during anaerobic growth is higher than that during aerobic growth, though the mutational stresses in this environment are not fully understood (32). The current knowledge in this area is discussed in **Section 1.2.1.1.1**. Given the differences in cell physiology imparted by growth in aerobic and anaerobic environments, I hypothesise that these differences are likely to have implications for the evolution of the organism.

1.2. Overview of evolutionary processes

Evolution is commonly defined as a change in gene frequency over time. In prokaryotes, there are four key mechanisms by which gene frequencies change; genetic variation, genetic drift, genetic draft and natural selection. The four mechanisms do not function in isolation, as the evolutionary process may be the result of a combination of these factors. A commonly held view of evolution is that it is adaptive, whereby natural selection acts on individuals with beneficial mutations to increase their frequency within the population. However, evolution may also be non-adaptive, where random processes such as genetic drift or genetic draft (**Section 1.2.3**) may act to change gene frequencies within a population without positively or negatively affecting the ability of the population to survive and reproduce in its environment. The four general mechanisms of evolution are described below.

1.2.1. Origins of genetic variation

The main building block of evolution is genetic variation, which refers to heritable changes that can result in phenotypic variation. Mutations are changes to the nucleotide sequence of the genome of an organism and are the ultimate source of new genetic variation within biological systems. Mutations, as well as horizontal gene transfer, recombination and gene flow are important sources of variation. Horizontal gene transfer refers to the transfer of genes between organisms in a manner other than that of reproduction (33). Recombination allows for the reordering of the genome and can be a source of much genetic variation (34, 35). Gene flow occurs when immigration or emigration takes place between populations resulting in the loss or gain of genetic variants within a population (36).. . Mutations occur through a variety of endogenous and exogenous processes (37). Endogenous mutations are those which arise

spontaneously as a consequence of the error-prone nature of DNA replication and repair (37). Exogenous mutations are those which arise due to environmental factors such as UV radiation or the prevalence of chemical mutagens such as ethidium bromide (38). In both cases, a change in nucleotide sequence results, which can again be further classified based on the changes that occurred at the sequence level.

1.2.1.1. Types of mutations

There are many different types of mutations. The smallest mutations are point mutations, in which a single nucleotide base is substituted by another base. Such mutations are also called single nucleotide polymorphisms (SNPs). There are two common classifications of point mutations. One classification defines mutations based on the nucleotide alterations at the DNA sequence level. SNP mutations can be divided into transitions or transversions based on whether or not there has been a change in the form of the nitrogenous base. There are two different forms, where purines contain two carbon nitrogen rings, while pyrimidines contain one carbon nitrogen ring. Transition mutations are nucleotide changes that do not result in a change in form, i.e. pyrimidine to pyrimidine or purine to purine. Transversion mutations occur when a nucleotide is replaced by a different form, i.e. from a pyrimidine to a purine (A ↔ C or A ↔ T) or a purine to a pyrimidine (G ↔ C or G ↔ T). Alternatively, if a SNP mutation occurs in coding DNA, the mutation can be classified depending on if the amino acid codon changes. If no change occurs at the protein level, the SNP is most likely to be synonymous. However, if a change does occur, it is said to be a non-synonymous SNP.

Mutations that result in the insertion or deletion of base(s) are collectively termed “indels”. These types of mutations can range in size, from single bases, to whole groups of genes and operons are affected. Large scale insertions involving duplications can arise from replication errors such as DNA polymerase slippage events. Large deletions are often associated with recombination between regions of high homology. Indels typically have large effects on the cells in which they arise as they may cause shifts in the reading frame of genes often leading to a premature stop codon, thereby inactivating many proteins.

A further class of large scale polymorphism is that of insertion (IS) elements, which are mobile genetic elements capable of mediating their own transposition and are prevalent in the genomes of many organisms (39). They are typically less than 2.5 kb in length and broadly categorised into more than 20 families based on their mode of transposition (40). Bacterial IS elements are major

contributors to overall mutagenesis as a result of their transposition activity. They are capable of causing a wide range of mutations, such as gene inactivation as a result of IS element insertion events into transcriptionally active genes (41). They are also known to activate the expression of adjacent genes owing to the presence of constitutive promoters located on the edges of various IS elements (42-44). Their activity has been shown to be a response to a wide variety of conditions such as nutrient limitation (45), stress (46), and non-constant environments (47, 48). In addition to gene activation/inactivation, they have been shown to mediate large scale mutations such as large deletions and rearrangements depending on their orientation (49-51). Large scale rearrangements such as chromosomal inversions can involve as much as a third of the whole bacterial chromosome (52). Larger mobile elements are generally sequences of DNA bearing high homology with bacterial viruses known as prophages, and may range in size from 5 to 40 kb. These mobile genetic elements frequently have genes or whole operons coding for traits crucial to the host genome, and have been acquired by horizontal gene transfer (53).

1.2.1.1.1. Prevalence of mutations in aerobic and anaerobic environments

As discussed in **Section 1.1.4**, growth in the aerobic environment is associated with the production of ROS, which is known to be a source of spontaneous mutagenesis in the aerobic environment (20). Both the base and sugar components of DNA are vulnerable to ROS mediated modification (26). Hydroxyl radicals, one of the most potent ROS, can mutate guanine residues specifically to form 8-oxo-guanine. 8-oxo-guanine is the most widely studied of the mutation products resulting from ROS attack on DNA. This leads to a high occurrence of G:C → T:A transversions in cells grown in the aerobic environment (32). Another transversion mutation A:T → C:G is also known to occur at a high rate due to 8-oxodGTP (32). Thymine residues can be oxidised to form thymine glycol, which results in an increased rate of T:A → C:G transitions (54).

While stresses in the anaerobic environment are not fully understood, it has been shown that the mutation rate – how frequently certain types of mutations arise – of anaerobically grown cells is higher than that of the aerobic environment. Furthermore, the types of mutations that prevail in the anaerobic environment are different(32). Specifically, the frequency of large chromosomal rearrangements including large deletions, duplications, and IS element movement were increased in anaerobically grown cells. A potential reason cited for this observation was due to longer incubation times for growth under the anaerobic environment.

1.2.1.2. Ecological effects of mutations

Mutations are random with respect to the needs of the organism (55). They can be classified according to their effects on the reproductive success of the organism: deleterious, neutral and beneficial. The majority of mutations that occur are harmful to the cell in which they arise and are thus deleterious. Neutral mutations have no effect, while deleterious mutations have negative effects on an organism's reproductive ability in a given environment. Examples of neutral mutations include mutations that occur in intergenic space or synonymous mutations in genes. Conversely, beneficial mutations are those that increase an organism's reproductive rate, and in general occur much less frequently than deleterious or neutral mutations (56). These classifications are relative and dependant on the biotic and abiotic factors within the organism's environment, such that a mutation that is beneficial in one environment may be deleterious in another.

1.2.1.2.1. Classification of beneficial mutations

Beneficial mutations can be classified based on the type of mutation that occurred. However, this tells us little about the underlying mechanisms that allow for the reproductive advantage the mutation confers. Thus, three alternative descriptions are frequently used in the literature. First, adaptive mutations can be classified based on functional disruption caused by the mutation. There are three broad categories; gain of function, modification of function and loss of function (57). Secondly, classification can be based on the type of gene mutated. Specifically, this classification distinguishes between genes that code for proteins that modify the expression of other genes (regulatory genes) or genes which code for functional proteins (structural genes). Currently there is debate as to whether structural or regulatory mutations are more important to the evolutionary process (58-60). Finally, beneficial mutations can be based on the phenotype which they confer to the cell. The phenotypic effects are broadly divided into two categories; optimisation or innovation (61). Optimisation phenotypes improve already existing genetic, metabolic and developmental networks. They do so by fine-tuning gene expression levels, altering regulatory interactions or modifying metabolic fluxes. Frequently, they can reduce the expression of genes which are transcribed at a basal level. In doing so, they conserve energetic resources. Innovation phenotypes allow the organism to occupy a new ecological niche. Innovation phenotypes are not as common as optimisation phenotypes but have been shown to lead to dramatic increases in population numbers in cells which develop these mutations (62).

1.2.2. Genetic drift

Genetic drift is the change in gene frequency in a population due to random sampling (63). Genetic drift changes genotypic frequencies of a population without regard for the phenotypic effect of the genotype. The effects of genetic drift are greater in small populations, as when populations are small, the effects of sampling error can alter gene frequencies significantly. The effects of genetic drift also play an important role in large populations, in which new beneficial mutations have arisen. When new beneficial mutations arise, the proportion of the population with the beneficial mutation is initially small, and the mutation may become extinct due to drift. Conversely, genetic drift may increase the frequency of a mutation that is deleterious. This commonly occurs in populations of small sizes, or in populations that undergo bottleneck events such that the population is temporarily restricted in size. At the molecular level, most of the genetic variation observed within populations is likely to be the product of random genetic drift of neutral alleles that do not affect the organism's ability to survive and reproduce. This neutral theory of molecular evolution was originally proposed by Motoo Kimura in 1968 (64).

1.2.3. Genetic draft

The change in gene frequencies due to linkage is known as genetic draft (65). Neutral and deleterious mutations occur at much higher rates than beneficial mutations. These mutations can reach fixation in asexually reproducing organisms through a form of linkage known as hitchhiking (66). Hitchhiking occurs in genomes where a beneficial (driver) mutation and a neutral or deleterious (passenger) mutation occur in the same genome. Selection will act on the genome based on the net effect of the beneficial and deleterious mutations which, if advantageous, can allow both mutations to reach fixation (65, 66). The phenomenon is especially prevalent in non-recombining asexual bacterial genomes, as the whole genome is essentially one linkage group. Genetic draft is thought to be the main mechanism by which deleterious mutations can spread among populations.

1.2.4. Natural selection

Spontaneous mutation, genetic drift and genetic draft account for the stochastic processes that underlie evolution. In contrast, natural selection is a non-random process by which gene frequencies change over time. Changes in genotypic frequencies of a population are based on differences in reproductive success, such that those that better survive and have a reproductive advantage contribute more to the next generation. In this way, populations become better suited to their environment. For evolution to occur by natural selection, three conditions within a

population must be fulfilled: reproduction, heredity, and variation (67, 68). Organisms must be capable of generating offspring to which their heritable traits are passed on. In addition, variation among these traits within the populations is important as this variation at the genotypic and phenotypic level affects the reproductive success of the organisms. In this way natural selection can transform a population over time.

Depending on the selective pressure, natural selection can be categorised into many different types. For example, directional selection occurs when one extreme of a continuum of variation is selected for; this favoured extreme will become more common. Stabilising selection selects for intermediates in a continuum of variation and generally works to maintain phenotypic status quo, while diversifying selection selects for the extremes of a continuum and may eventually lead to speciation. Another type of selection is frequency dependent selection, described below.

1.2.4.1. Frequency dependent selection

Frequency dependent selection is where the frequency of a variant affects its ability to survive and reproduce. Frequency dependent selection may be positive, where the fitness of the genotype increases as its frequency within the population increases. In contrast, for negative frequency dependent selection (NFDS), the fitness of the genotype increases as its frequency decreases. NFDS maintains genetic diversity over a long-term by favouring a genotype that is rare, rather than a genotype that is abundant (69). Thus, rare genotypes are protected from loss. With regard to competition for resources, when two genotypes share a common environment, but do not compete for the same resource, NFDS may come into play. When a genotype is rare, its resource is relatively abundant in comparison to the other genotype, whose resource is likely to be rare. This allows for the co-existence of genetically distinct forms in sympatry in a density dependent relationship. Thus, biodiversity among two different genotypes in sympatry can be maintained due to the operation of NFDS (69, 70). Examples of studies in which NFDS plays a role are widespread among the literature (71-75) and some are discussed in **Section 1.4.3**.

1.3. Outcomes of evolution

Combinations of the evolutionary processes outlined in **Section 1.2** lead to a variety of potential outcomes due to the random and non-random nature of the processes involved. Thus, evolutionary outcomes may range from extinction to diversification/speciation, co-evolution and adaptive radiation. At the phenotypic level, evolution occurs mainly through natural selection and, as a result, evolution is frequently thought of as an optimising process. The most relevant

outcomes, as they apply to this thesis, include adaptation, specialisation, generalisation and diversification.

1.3.1. Adaptation

By definition, beneficial mutations confer upon their hosts greater survivability and reproductive success relative to others within the population (55). Adaptation is the evolutionary outcome by which an organism becomes better suited to its environment. It is a two-step process (63). The first step is the introduction of genetic variation, whereby mutations with a range of effects on fitness arise spontaneously within a population. The second step is natural selection. As a result of natural selection, beneficial mutations that confer a fitness advantage become more highly represented in subsequent generations (76). Fitness is described as a measure of survival and reproductive success. It is also the phenotype which natural selection acts upon. From this, beneficial mutations that increase the fitness of the organism tend to increase in frequency within the population over time (77).

1.3.1.1. The adaptive landscape

There are many different routes available to an organism when adapting to its environment. A well-documented example of this is that of the many different mechanisms by which bacteria become resistant to antibiotics. There are three broad mechanisms by which, for example, the Gram-positive organism *Bacillus subtilis* becomes resistant to streptomycin: i) the development of efflux pumps to excrete the antibiotic; ii) the modification of the antibiotic active site and iii) the protection of the bacterial target or bacterial property under attack from the antibiotic (78, 79). Sewell Wright's concept of the adaptive landscape is an apt metaphor of this (80). It is a graphical representation of a population's mean fitness as a function of its genotype in a given environment (**Figure 1.4**). The landscape is displayed in three dimensions. On the x and y axes (the floor axes) is the total combination of alleles available to the genome while the third axis (z axis) represents the fitness at each point.

The shape of a landscape varies depending on the environment, with each environment having its own characteristic shape. Landscapes can be smooth, with a single peak, or rugged with many peaks of unequal height. Each peak represents the variety of adaptive solutions available to the cell or population in adapting to its environment (76, 81-83). For rugged landscapes, often there are many local peaks with one global peak. The tops of these peaks represent a genotype of high mean fitness, or individual adaptive mechanisms, each of which may, in the example given, confer

resistance to an antibiotic. The global peak represents the fittest genotype available to the organism within the environment. Random evolutionary processes such as mutation or genetic drift determine which peak the population will climb towards, while natural selection will push the organism or population up the peak (82). Identical populations often diverge in their adaptive traits as populations find alternative adaptive solutions to their specific environment. Additionally, the fixation of mutations of deleterious fitness effects are thought to provide a mechanism by which cells move across valleys, between different fitness peaks within the adaptive landscape (84-86).

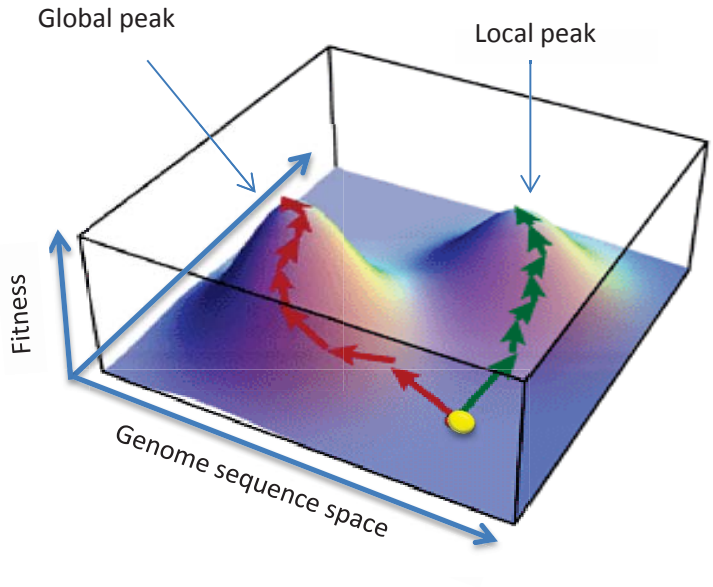

Figure 1.4: Wright's adaptive landscape. A three dimensional adaptive landscape with genome sequence space as the floor (x and y axes) and fitness as the height (z axis) with two fitness peaks is displayed. Within this landscape, a cell or population is located based on its genotype, as indicated by the yellow dot. Red and green arrows indicate two different pathways towards a global (highest) or a local (alternative) peak. See text for details. Figure adapted from Elena *et al.* (2003) (82).

1.3.2. The evolution of specialists

The niche of an organism describes how the organisms mean fitness varies across a range of environments. Furthermore, the niche width describes the range of environments under which an organism can grow and reproduce. Under constant prolonged exposure to any one of these environments**,** selection will lead to the evolution of a specialist (87). Specialists are narrowly adapted types that have the optimal phenotype in one environment. A consequence of this is that

the organism may be relatively less fit across alternative environments. Specialists typically have the highest fitness in the environment of selection than elsewhere. This is a consequence of the specificity of adaptation to a constant environment and because adaptation to one set of conditions is underlain by a cost of adaptation to alternative conditions. Loss of fitness in alternative environments occurs due to trade-offs that arise due to a lack of selection pressure maintaining those characteristics that are important in other alternative environments (82, 88). An example of this phenomenon is the loss of vision and pigmentation in organisms which have become specialised to cave dwelling habitats (89).

There are two mechanisms by which trade-offs arise. In the first case, antagonistic pleiotropy (AP) occurs when a beneficial mutation arises in the environment of selection, but has a negative effect in alternative environments. In an experiment with *E. coli* adapting to a lactose-limited environment, 80% of clonal isolates had deleterious mutations in the *ompF* gene, a gene determined to be responsible for increased fitness in the lactose limited environment. The *ompF* gene is a member of the general bacterial porin family which allows the passage of solutes such as sugars, ions and amino acids into the cell. However, those mutated strains with beneficial (relative to the lactose-limited environment) mutations in *ompF* were shown to have a negative effect in an alternative environment i.e. one with antibiotics. In this environment, strains with mutations in *ompF* displayed increased sensitivity to cloxacillin and penicillin (90).

The second example of trade-offs leading to specialisation is the process of mutation accumulation (MA). MA arises when an organism acquires mutations that are neutral, or have no effect on fitness in the environment of selection, but these same mutations have a negative effect on the reproductive rate of the organism in alternative environments. These findings are frequently observed among MA experiments (91) that minimise the role of natural selection and maximise the role of genetic drift. When this is the case, experimental lineages under the MA regime decline in fitness due to decay of genes throughout the genome at random.

There are many cases of specialisation through AP and MA in the literature (87, 88, 92-96). The two processes do not happen in isolation, but rather a mixture of the two mechanisms contributes to the evolution of niche specialists. However, the sum effect is the specialisation of organisms due to the narrowing of their niche breadth.

1.3.3. The evolution of generalists

While selection in a constant environment leads to the evolution of specialists, selection in a non-constant or heterogeneous environment leads to the evolution of generalists. Specialists are known to have enhanced performance in a constant environment; in contrast, generalists often have intermediary fitness across a range of environments. A generalist is adapted to a wider ecological niche and performs best across the range of environments to which it is exposed (88, 92).

Generalists forego high mean performance in one environment for breadth of adaptation across many environments. This was found in experimental *E. coli* lines adapting to different thermal niches, (32, 37 and 42^{o}C) while also including a heterogeneous sample that transitioned on a daily basis from 32 to 42^{o}C (97). The fitness responses of the experimental lines were investigated after 2,000 generations. In all cases, the fitness response was highest among specialists that were subjected to constant treatments. The generalist line had intermediary fitness as compared to each of the specialists when competed at 32 and 42^{o}C. Furthermore, the temperature at which the generalists displayed the highest increase in fitness was the intermediate temperature of 37^{o}C, the average temperature of the range from which it was evolved despite not growing at this temperature for over 2,000 generations.

Adapting to a range of environments also incurs constraints (88). These constraints occur because specific biological functions in a non-constant environment are rarely performed or maintained as well outside the environment in which they have arisen (82). Thus, there may be fitness costs associated with the maintenance of certain biological functions adapted specifically for a single environment, when applied to a variety of different environments (96). Thus an alternative strategy that functions adequately in both environments is favoured. The second constraint is the rate of adaptation. Specialists evolve in constant environments at a faster rate than generalists evolve in non-constant environments. This is because constant exposure to one environment allows for the fixation of alleles with environment-specific effects more rapidly than when in a non-constant environment over time.

1.3.4. The evolution of biodiversity

When populations are subject to natural selection, the competitive exclusion principle applies, whereby for competing genotypes, only one can occupy the same ecological niche (98). The outcompeted genotype may become extinct. Alternatively, it may evolve to occupy a different

ecological niche, thereby avoiding direct competition. Generally, the outcome of natural selection is sequential clonal replacement within a population. In contrast, biodiversity may be maintained in populations under the force of NFDS. This is explained further in **Section 1.4.3** and **1.4.4**. Studies have also shown that individuals may diversify into distinct sub-populations, particularly in environments where more than one nutrient resource or niche is present (69, 71, 99), whereby initially uniform monocultures become polymorphic (58, 73).

1.4. Empirical studies on adaptive evolution

An indispensable tool in studying evolutionary principles such as those explained in **Section 1.2** is that of experimental evolution. Many valuable insights have been gained from microorganisms subjected to long term evolution experiments (LTEEs) (82). LTEEs allow for direct observation of the adaptation over time. Pure cultures of bacterial cells are divided into replicates, and propagated for many generations. During the course of LTEE, adaptation is typically measured by monitoring the fitness of evolved populations through competitive fitness assays. Combining LTEE and competitive fitness assays together allows us to understand how an organism adapts to its environment.

Bacteria are particularly well suited to LTEEs for a number of reasons. They reproduce asexually, meaning any differences from one generation to the next can be attributed solely to mutation events occurring during the experiment (100). They have rapid generation times, which allow experiments to achieve high generation numbers in relatively short time periods. Bacterial cultures can attain high population numbers in short times and this can be done in small volumes. This also allows for the use of many replicate populations to test the repeatability of outcomes. Their environment can be strictly controlled, which is essential for drawing inferences on the genetic underpinnings of evolution. Bacteria in their evolved states can be frozen in a state of suspended animation, and examined later for further analysis such as direct comparison between ancestral and evolved types.

Two LTEEs are particularly relevant to this study. The first is that by Prof. Richard Lenski, who in 1988 established a serial batch sub-culture LTEE with the model organism *E. coli* B, known as REL606. This organism was subjected to a LTEE in DM25 minimal glucose media for 2,000 generations initially. The evolution of REL606 in this environment continues to the present day. At the time of writing, the experiment is in its 25th year, achieving over 60,000 bacterial generations

of evolution (http://myxo.css.msu.edu/ecoli/index.html). For this thesis, this experimental system was used as a model, with similar strains and media being used. The second study is one published during the course of this thesis in 2013 (101). The LTEE study of Puentes-Téllez focused on how bacteria adapt to environments varying in oxygen content. This study was also a serial batch sub-culture experiments, in which an *E. coli* K-12 strain (MC1000) was subjected to complex LB media in three different treatments with regard to oxygen content; oxygen-rich, oxygen-limited and a treatment that fluctuated from oxygen-rich and oxygen-limited conditions. The key outcomes from these studies which are relevant to this thesis are discussed in the following sections.

1.4.1. Rate of adaptive evolution from LTEE studies

A major outcome of LTEE studies has been the finding that the rate of adaptation declines over time. The rate of adaptation to a new environment has two distinct phases; it is initially fast but decelerates over time (102). Barrick *et al.* (2009) investigated the rates of fitness improvement and genome evolution during the first 40,000 generations of Prof. Lenski's LTEE study described previously. Experimentally evolving *E. coli* populations initially increased in fitness by 60% during the first 5,000 generations. During the following 15,000 generations, a fitness increase of only 25% was observed (52). In contrast, the rate of genomic change, or the rate at which new mutations arose, was linear throughout. Thus, there is discordance between the rate at which mutations arise and the rate at which populations adapt to the environment. This finding was attributed to three reasons. Firstly, mutations with the highest impact on fitness reach fixation within the population earlier than mutations with lower fitness increases. (102). Secondly, beneficial mutations with high increase in fitness are more likely to evade loss by random genetic drift than less beneficial mutations. Finally, clonal interference, whereby many beneficial mutations may arise in a single population and must compete for fixation, may delay the time taken for the eventual winning mutation to reach fixation (56, 99, 103).

1.4.2. Identification of adaptive mutations from LTEE studies

Adaptive mutations are fitness enhancing mutations that have been acted on by natural selection. However, not all mutations that reach fixation are adaptive mutations**.** Thus, further investigation into the mutated genes that achieve fixation is required to identify those that were most important to the adaptive process and acted on by natural selection (104). Adaptive mutations can be identified based on the extent of parallelism among independent replicate lines. Parallel evolution (or parallelism) occurs when traits, within shared biochemical pathway or network, are mutated independently in multiple evolving lines. In such instances, the resulting phenotypic

change is likely to have a reproductive advantage for the cells in which it has arisen and have been acted upon by natural selection. Thus such mutations would be deemed beneficial for adaptation within the organisms current environment, including media (105). The gold standard for confirming if a mutated gene is adaptive is to recreate the putative adaptive mutation in the ancestral genetic background (usually by allelic replacement), and compare the fitness of the mutant to that of the ancestor. Direct competitive fitness experiments between two strains which are isogenic apart from the mutation of interest allows for a direct measurement of the fitness enhancement due to the introduced mutation.

An important aspect when considering the effects on fitness by potential adaptive mutations is the role that epistasis may play within the genome. Epistasis occurs when the fitness effect of a mutation is masked or modified by its interaction with other genes or mutations in the genome. There can be three types of epistatic interactions. Firstly, synergistic (or positive) epistasis, where the fitness effect of the mutations in combination is greater than expected based on the sum of their individual effects (106). Secondly, sign epistasis occurs when the sign of the fitness effect of a mutation, beneficial (+) or deleterious (-), is altered by epistatic interactions (107). Finally, antagonistic (or negative) epistasis occurs when the contribution of mutations produce fitness benefits that are less than the cumulative sum of their individual fitness effects (108).

1.4.2.1. Examples of adaptive mutations from LTEE studies

Many studies have identified adaptive mutations from the evolved lineages of Prof. Lenski's LTEE study (62, 104, 109-112). In most cases, mutations that were identified as potentially adaptive, via parallelism among replicate lineages, were experimentally verified via allelic replacement and subsequent competition assays (**Table 1.1).** These adaptive mutations are directly relevant to this thesis for three reasons. i) Adaptive mutations in **Table 1.1** were proven to be adaptive in the same minimal media used in this thesis under the aerobic environment and, collectively, these mutations demonstrate the wide variety of adaptive pathways that may be taken during adaptation to an aerobic environment. ii) The same method of identification of adaptive mutations via parallel mutations occurring in replicate lineages was used in this thesis. iii) Finally, many of these adaptive mutations are present in REL4536, the ancestor of this LTEE.

Table 1.1: Summary of competitive fitness data from LTEE studies with *E. coli* B reporting adaptive mutations.

Gene	Relative fitness increase[a]	Mutation in REL4536	Phenotype[b]	Functional disruption[c]	Gene type[d]	Underlying mechanism[e]	Reference[f]
topA	13.30%	Present	Optimisation	Modification	Structural	DNA supercoiling	(109)
fis	5.00%	Present	Optimisation	Modification	Regulatory	DNA supercoiling	(109)
spoT	9.40%	Present	Optimisation	Modification	Regulatory	Interaction Modification of stringent response	(104)
glmU	5.00%	Present	Optimisation	Loss of function	Regulatory	Cell wall biosynthesis	(110)
rbs	1.40%[g]	Present	Optimisation	Loss of function	Structural	Ribose catabolism	(111)
malT	0.015%[g]	Present	Optimisation	Loss of function	Regulatory	Sugar metabolism	(112)
cit	NR[h]	Absent	Innovation	Gain of function	Structural	Citrate metabolism	(62)
pykF	11.10%	Present	-	-	-	-	Unpublished data
nadR	8.10%	Present	-	-	-	-	Unpublished data

[a] Increase in relative fitness reported after introduction of mutation into the ancestral background and competition with the ancestral strain, *E. coli* REL 606 in all cases.

[b] The phenotypic response as a result of the mutation; optimisation or innovation, as classified in **Section 1.2.1.2.1.**

[c] The functional effect of the mutation on the protein, loss, modification or gain of function, as classified in **Section 1.2.1.2.1.**

[d] The type of gene mutated; regulatory or structural, as classified in **Section 1.2.1.2.1.**

[e] Underlying mechanism resulting in the increase in relative fitness.

[f] Reference

[g] Relative fitness increase was significant at the $P < 0.05$ level

[h] Not reported (NR)

1.4.2.1.1. Enhanced DNA supercoiling

DNA supercoiling is an essential mechanism by which DNA is over or under wound in order to compact DNA. Topoisomerases are enzymes that regulate the over- and under-winding of DNA (113). Specifically topoisomerase 1 is responsible for the relaxing of negative supercoils in DNA. The *fis* gene (factor for inversion stimulation) is also involved in supercoiling activity as it encodes a small protein which controls DNA structure through its interaction with the production of

topoisomerase 1 and enzymes that aid in the DNA unwinding processes known as DNA gyrases (114). Mutations in both *fis* and *topA* were sufficient to cause an increased degree of supercoiling activity amongst populations examined. They represent examples of mutations of structural genes that are often have large fitness effects when mutated. The extent of DNA supercoiling is thought to increase fitness as it allows for a global mechanism for the control of many genes simultaneously. These mutations lead to strong pleiotropic effects as changes in DNA topology can influence the expression of many genes (109). At the phenotypic level, an increased degree of supercoiling was found in 10 of the 12 populations by 20,000 generations (109).

1.4.2.1.2. Modification of the stringent response

Mutations in the *spoT* gene were reported in eight of the 12 evolving populations. In *E. coli* the SpoT enzyme is a key enzyme involved in the stringent response, a stress response which causes inhibition of RNA synthesis (115) in response to amino acid starvation, fatty acid limitation and heat shock. The *spoT* gene is known to affect 59 genes that, in turn, are regulated by specific effectors such as guanosine pentaphosphate (ppGpp) and the cAMP-cAMP receptor protein (CRP). Analysis from allelic replacement studies indicated that the competitive advantage was due to a reduced lag phase, which increased the maximum growth rate of the modified ancestor. While the precise mechanism for the effect of the *spoT* mutation on fitness is unknown, two hypothesised mechanisms were proposed. First, fitness increases could be due to a reduction in the expression of genes that are not required, such as the flagella encoding *flg* operon. Secondly, a reduced expression of ppGpp has been shown to increase the rate of transcription of tRNA and rRNA promoters, which may increase the rate of transcription and potentially growth rate in affected cells. This is an example of how point mutations within a regulatory protein can have dramatic effects on fitness of the whole cell (104).

1.4.2.1.3. Decreased cell wall biosynthesis

During the evolution of the bacterial populations, it was observed that cells increased in volume in line with increased relative fitness (83, 116). When investigating the underlying genetics of this, a loss of function mutation in the *glmUS* operon was found. The *glmUS* operon encodes two proteins GlmU and GlmS, both of which are involved in the synthesis of peptidoglycan and lipopolysaccharides of the Gram negative cell wall in *E. coli* (117). The authors hypothesised that it may have been beneficial for cells to produce less cell-wall constituents as the population began to increase its maximum growth rate within the media. Interestingly, this mutation was only found in one (Ara-1) of the 12 populations (110) and it was hypothesised that this was due to strong clonal

interference in other populations which increased in cell volume. Other studies have reported mutations in or upstream of *pbpA*, an operon also involved in cell wall elongation (105) thus implicating many different genetic pathways to attain a large cell volume phenotype.

1.4.2.1.4. Glucose specificity

Within the first 2,000 generations of Prof. Lenski's LTEE experiment, all 12 of the experimental lineages possessed a mutation that resulted in the loss (either partial or complete) of the *rbs* operon, which controls ribose catabolism. In all cases, insertion of an IS150 element upstream of the operon led to the excision of downstream genes. Fitness increase is thought to be due to the optimisation of the glucose specificity (111) and is an example of a beneficial mutation that can arise due to the movement of IS elements.

Another example of a loss of function mutation in an alternative sugar catabolism operon is that of the *malT* gene mutation (112), a transcriptional regulator that activates the transcription of a collection of genes and operons involved in maltose catabolism and transport. Mutations in the *malT* gene were found in eight out of the 12 evolving populations. The beneficial effect of these mutations was shown to be due to mutations which reduce MalT activity, unnecessary in an environment in which glucose is the sole energy source. Like the *rbs* deletion mutation, fitness increases were modest when this mutation was introduced into the ancestral background (0.0146%), however increases in fitness were significant. In addition, the eight events also took place after the initial rapid fitness increase during the first 2,000 generations, therefore contributing to the slower fitness rate increases that took place after 2,000 generations (110).

1.4.2.1.5. Resource switching

An inability to utilise citrate is a key diagnostic feature of *E. coli* during aerobic metabolism due to the lack of a functional transporter protein (62). However, among the 12 lineages of Prof. Lenski's LTEE, one population (the Ara-3 population) evolved the ability to metabolise citrate, which is in abundance in the DM25 media used in the LTEE. This beneficial mutation is an example of an innovation phenotype. The cit+ phenotype is described as having occurred over three steps. The first steps involved the potentiation phase in which the genetic background evolved, such that the cit+ function became accessible through mutation. This potentiating change increased the mutation rate of the nascent cit+ genotype, but did not cause generalised hypermutability (118). During the actualisation process a very weak cit+ phenotypes emerged (**Figure 1.5**). This arose by a duplication mutation that occurred downstream of the *citG* gene and encompassed a portion of

the 3’ portion of the *citG* gene, the *citT* gene, the *rna* gene with its constitutively active promoter and finally the 5’ portion of the *rnk* gene, also including its constitutive promoter. This put the *citT* gene under the regulation of the constitutive *rnk* promoter. The *citT* gene is important for citrate import into the cell.

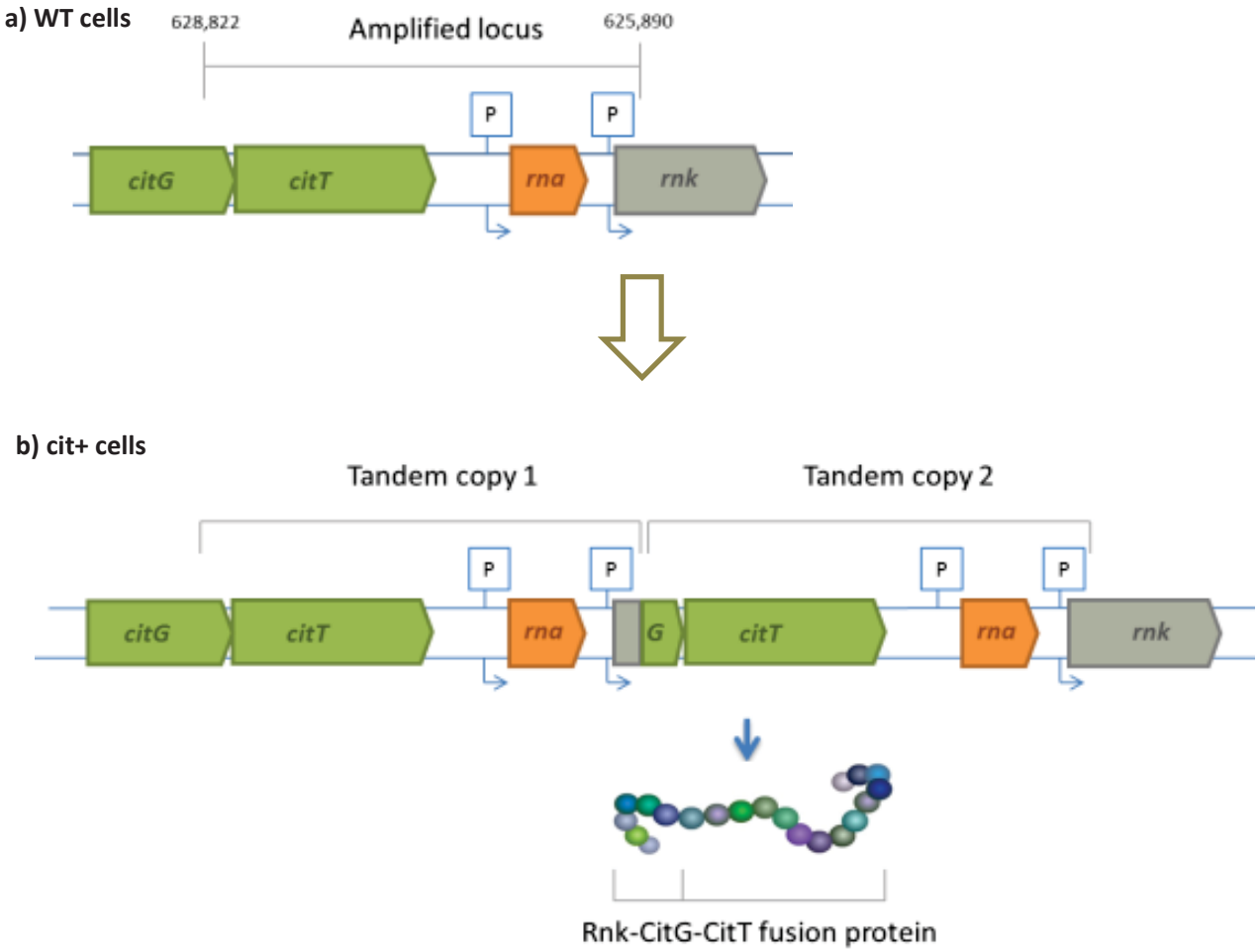

Figure 1.5: The actualisation step of the cit+ phenotype in the Ara-3 population. a) Indicated are the *citG*, *citT*, *rna*, *rnk* genes, constituent promoters of *rna* and *rnk* genes, and the amplified locus in WT cells. b) The tandem copy as a result of duplication and the resulting Rnk-CitG-CitT fusion protein present in cit+ cells is shown. See text for details. Figure adapted from (118).

However, the mutation did not reach fixation until after the third stage. The third stage involved further refinement of the *rnk-citG-citT* modules. This occurred by the tandem duplication of the *rnk-citG-citT* module (**Figure 1.5**) nine separate times that occurred via homologous recombination from 31,000 and 33,000 generations (118). Under the regulation of a constitutively active

promoter, the *citT* gene and the remainder of the operon are transcribed, conferring the ability to import and to metabolise the previously unutilised citrate in the environment. Switching metabolism from glucose to citrate (which is available in the media at a 12-fold higher concentration than glucose in DM25) led to a dramatic population expansion for the cit+ mutant sub-population (119). No allelic replacement studies have been performed on the citrate utilising strain, however population numbers increased impressively leading to the complete replacement of the strain in only ~200 generations (62). The evolution of resource switching from glucose to citrate among *E. coli* in Prof. Lenski's LTEE was only observed in one line and arose relatively late for a mutation with such a large benefit. This is likely as a result of the unknown epistatic mutations that occurred in the Ara-3 population during the potentiation phase, and as such, are thought to have been crucial in allowing the cit+ phenotype to develop in this population alone (118).

1.4.2.1.5.1. Growth advantage in stationary phase

Another example of an innovation mutation regime leading to nutrient resource switching is that of the mutations conferring the growth advantage in stationary phase (GASP) phenotype (120-124). Although these mutations have not been experimentally validated via allelic replacement studies in the strains of Prof. Lenski, they are still considered to be highly beneficial to the cell in which they arise (123). Although not reported in **Table 1.1,** mutations leading to the GASP phenotype have been reported in Prof. Lenski's LTEE strains (124). GASP mutations confer the ability to catabolise one or more amino acids as an alternative source of carbon and energy during prolonged starvation (125). In stationary phase cultures where cell death is equal to the rate of cell growth, strains with the GASP mutations switch from sugar metabolism which may have been depleted and instead catabolise the amino acids that are in abundance in the media as a result of lysed cells.

1.4.2.1.6. Findings from LTEE studies varying oxygen exposure

A limited number of studies have examined the impact of various oxygen levels on evolution. Previously, adaptation was investigated by PCR sequencing of candidate genes in *E. coli* K-12 grown in chemostat cultures under aerobic and oxygen-limited environments by Manché *et al.* 1999 (126). A more comprehensive LTEE approach was undertaken to understand the evolutionary processes that occur under environments with varying oxygen exposure in this thesis and by Puentes-Téllez *et al.* 2013. However, this thesis and the study of Puentes-Téllez differ in two important aspects. Firstly, the aim of the Puentes-Téllez study was to assess bacterial evolution in

LB complex growth medium, while in this thesis, minimal glucose media was used in line with that of Prof Lenski's LTEE study. To this end, a similar model organism was used; a descendent of ancestral REL606 known as *E. coli* REL4536. Furthermore, this thesis looked to examine the evolution of cultures maintained in strict aerobic and strict anaerobic conditions, while the Puentes-Téllez study investigated evolution under oxygen-rich and oxygen-limited conditions. Nevertheless, similar set up and design allow for comparisons between the two LTEE studies, with three of their findings of particular relevance to this thesis.

Firstly, in the Puentes-Téllez paper, the creation of specialist and generalist evolved lineages under oxygen-rich and oxygen-limited conditions were reported, with increased relative fitness in all replicates. Secondly, mutations in *galE* were reported and hypothesized to contribute to the increased relative fitness, based on population level competition assays. This gene is involved in the conversion of UDP-galactose and UDP-glucose as part of galactose metabolism. The ancestral strain of the Puentes-Téllez study was unable to utilise galactose as a sole carbon source due to a mutation in *galE*. However, a transition SNP in *galR* induced expression of the *gal* operon three times higher than in the wild type. This mutation allowed for the innovation of derepression of galactose catabolism and was implicated as being responsible for the fitness increase in evolved forms. Finally, morphological divergence was reported among lineages that had evolved under all treatments. Polymorphic colonies were isolated after 1,000 generations of evolution. The colony polymorphism resulted in the identification of three forms classified as i) small colonies, ii) large/rough/irregular colonies or iii) large/smooth/regular colonies. In the Puentes-Tellez study, all three forms were observed in oxygen-rich and fluctuating treatments, however under oxygen-limited environments, only large/rough/irregular and large/smooth/regular colonies were found. The evolution of different morphotypes within a spatially homogeneous environment is of particular interest within this thesis, and further discussion of sympatric diversification is given below.

1.5. Sympatric diversification

Central to many evolutionary studies is the origin and maintenance of diversity (58, 73, 74, 127, 128). Geographical isolation followed by divergent selection is often cited as a major factor in speciation. In recent studies, diversification arising from a single population has been investigated. This occurs in a two-step process. Initally, naturally arising variants within the population are subject to selection for resources within the environmentthe the resulting genotypes may co-exist

and be maintained by NFDS. This may be seen within spatially heterogeneous and homogeneous environments, and examples are provided below

1.5.1. Diversification in response to spatial heterogeneity

The diversification of the Gram-negative plant-colonising bacterium *Pseudomonas fluorescens* as reported by Rainey and Travisano (127). Isogenic ancestral populations of *P. fluorescens* were cultivated under a spatially heterogeneous regime and after seven days cultures became phenotypically diverse (**Figure 1.6**). Three distinct types were identified based on the morphology of colonies when isolated on agar plates, the ancestral smooth (SM), divergent wrinkly spreaders (WS) and fuzzy spreaders (FS). Only SM types arose in spatially homogeneous (mixed microcosms of nutrient media) environments while in spatially heterogeneous (static microcosms of nutrient media) environments the WS and FS phenotypes arose. This divergence was found to be in response to oxygen concentration gradients. In unshaken cultures, the WS phenotype evolved as an oxygen-rich specialist in the upper layer, with the oxygen-limited specialist FS occupying the lower layer. Different morphotypes of initially homogeneous types were created with regard to different concentrations of dissolved oxygen in non-shaking cultures (75). At the genetic level, four adaptive pathways have been implicated as responsible for the WS, each involved in a range of functions: i) *wss* (wrinkly spreader structural operon) involved in the production of an acetylated cellulosic polymer for biofilm formation (72) and the ii) *wsp* (wrinkly spreader) operon involved in a chemosensory pathway (128); iii) *aws* (alternate wrinkly spreader) operon involved in overproduction of the acetylated cellulosic polymer (ACP) (129) and iv) *mws* (mike's wrinkly spreader) involved in the regulation of the cellular cyclic-di-GMP levels that regulate the level of *wss* expression, and hence polymer formation. The pathway to FS evolution is much simpler than the many pathways involved in WS evolution, with mutations in *fuzY* shown to be responsible where the FS arise in the air-liquid interface and drop to the floor microcosm (130).

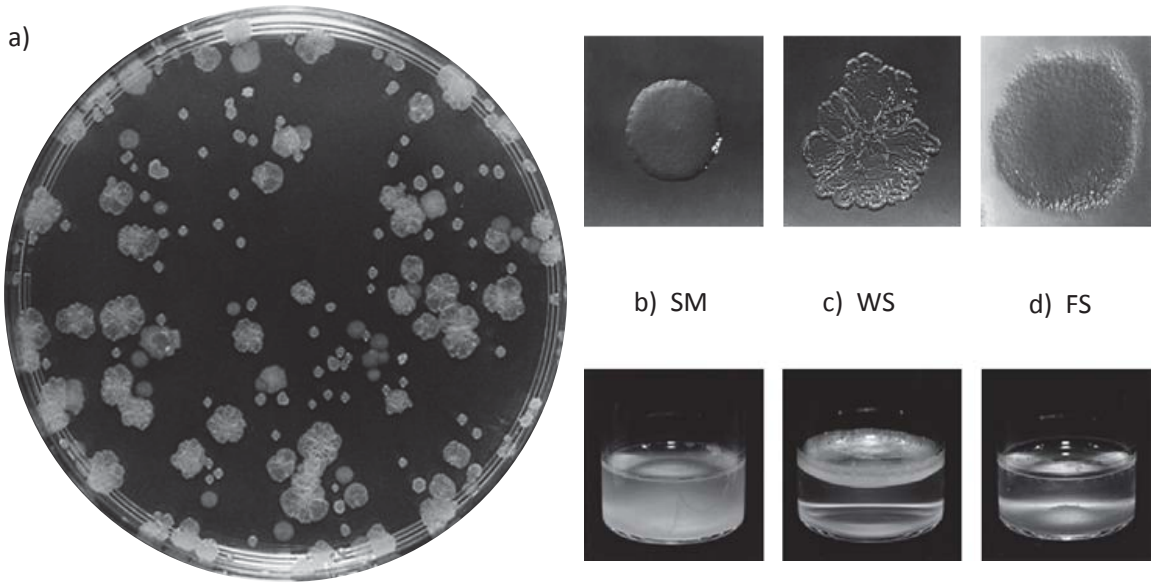

Figure 1.6: Phenotypic diversity among *P. fluorescens* SBW25 populations. a) Range of morphs after plating on King's medium B agar plates. b) Smooth morphs (SM) (ancestral type) in homogenous environments c) Wrinkly spreaders (WS) in the air liquid phase of heterogeneous environment and d) Fuzzy spreader (FS) in anaerobic phase of King's medium B contained in 25 mL broth culture at 28^{o}C. Figure adapted from Rainey *et al*. 1998 (127).

1.5.2. Diversification within chemostat cultures on single substrate media

In an LTEE study with *E. coli* K-12 cultures grown in glucose limited media in chemostats, pure cultures were found to develop small colonies when the population was plated onto TAra agar plates (73). After 773 generations, three clones with substantial differences in maximum growth rate and glucose uptake kinetics were identified. Subsequent analysis showed that these two colony morphs represented co-existing genotypes. Further investigation found the existence of two strains CV101 and CV103 each with different mutations in the *acs* gene which encodes acetyl-CoA synthetase.

In the case of CV103 displayed quicker glucose uptake kinetics and conferred the CV103 strain as a glucose specialist which developed as small colonies. However, this came at a cost to catabolism efficiency, which resulted in the incomplete breakdown of glucose under conditions where aerobic respiration should take place. This led to the secretion of acetate as an incomplete glucose breakdown waste product. A build up in acetate concentrations within the chemostat culture resulted. This excreted metabolite presented a new ecological opportunity for the other subset of the population which were distinguishable as small colonies when grown on TAra agar plates (45).

In the CV101 clone, an insertion of an IS30 transposable element inserted into the upstream regulatory region of the *acs* gene. IS30 contains an outward facing promoter-like sequence at its N-terminus. The orientation of the IS element placed *acs* under the semi-constitutive promoter of the IS30 IS element. This led to the up-regulation of *acs*, and created the acetate-scavenging phenotype (45). The generation of glucose and acetate specialists in initially homogeneous populations is an example of population diversity as mediated through niche construction, with diversity maintained by NFDS (131).

1.5.3. Diversification within batch cultures with mixed acetate and glucose media

Divergence and maintenance is also reported with the *E. coli* B strain grown in many serial batch regimes with a mixture of acetate and glucose as carbon sources (58, 74, 132). After 1,000 generations, populations diversify into metabolically distinct strains in relation to their use of carbon sources. When *E. coli* is supplied with glucose and other sugars, glucose is used preferentially. Once glucose is removed a diauxic switch allows the cells to rewire their metabolism and catabolise the secondary sugar sources. It was reported by Spencer *et al.* 2007 (74) that when glucose was grown in DM50:50 (glucose:acetate), populations diverged in their diauxic switch stage. Two colony types, varying at the phenotypic level by the size and growth rate on TAra plates after 48 hours of growth, appeared in all 10 replicates. Growth rate assays identified that the larger colonies achieved higher growth on glucose, impaired growth on acetate and had extended lag phases on switching from glucose to acetate and became known as Slow Switchers (SSW). In contrast, smaller colonies had slow growth on glucose but faster growth on acetate and a very quick lag phase when compared to the SS colony types, and thus became known as Fast Switchers (FSW).

Investigation at the genetic level found that the expression of the genes of the *aceBAK* operon were substantially up-regulated in FSW strains and substantially decreased in ancestral and SSW strains (74). This operon is one of two pathways involved in acetate catabolism. In the FSW strains an insertion of an IS1 into *iclR* (isocitrate lyase regulator) gene was found, which was not present in either the ancestral or SSW strains. The operon is controlled downstream by the *IclR*, a negative regulator which enhances *aceBAK* expression in eight out of nine clones isolated. However, allelic replacement experiments showed that this mutation alone was not sufficient to be responsible for the origin and maintenance of the FSW and SSW variants. The mutated *iclR* gene in the ancestral

strain did not reduce lag switching times, thus highlighting the role that epistasis must play in the origin of this diversity.

1.5.4. Diversification within the glucose utilising population of *E. coli*

Divergent selection has also been found among the evolved lines established by Prof. Lenski. In the Ara-2 populations, polymorphic colonies arose after 6,500 generations of evolution on serial glucose limited batch cultures (71). These colonies differed in size, with large colonies (designated "L", but "LG" for this thesis) and small colonies (designated "S", but "SL" for this thesis) when plated onto solid TAra agar and incubated for 16 hours. In comparison with the FSW and SSW reports from Spencer *et al.* 2007 (**Section 1.6.2.2**), polymorphism in the Lenski lines has two fundamental differences. First, the Lenski lines were serially propagated in media with limited glucose as the only carbon source, with no acetate added to the media. Second, where Spencer *et al.* (74) report polymorphisms in all 12 experimental lineages examined, only one of 12 experimental lineages in Prof. Lenski's LTEE reported this polymorphism. Nevertheless, NFDS has maintained this polymorphic population for over 30,000 generations (124).

Despite extensive investigation (71), the molecular mechanism behind the co-existence of the LG and SL cells in the Ara-2 population remains unknown. This adaptation may create a sub-population if the two clones become specialists for different resources. Based on phenotypic assays, it was hypothesised that a form of GASP-mediated cannibalism could have allowed for the co-existence between the LG and SL cells. While it is rare that GASP mutations would promote diversity rather than purge it with periodic selection, the authors suggest that the famine and feast regime experienced during serial propagation in batch cultures may have allowed for this co-existence (124).

1.6. Combining LTEE with whole genome sequencing to explore the adaptive landscape

There are a wide variety of genetic routes to adaptation. Revealing the adaptive pathways taken by a lineage is challenging in light of processes such as epistasis and clonal interference. Furthermore, complex phenotypes, such as fitness in aerobic and anaerobic environments, are likely to be attributed to mutations in a wide range of structural and regulatory genes functioning in a variety of biological networks. Therefore, a detailed knowledge of the variety of mutations that can occur over the entire genome is fundamental to understanding the process of evolution at the genomic level (133). Previous studies have used traditional phenotype-based assays or

genetic analysis of a few small loci to assess mutations (32, 134). Candidate gene based approaches require *a priori* assumptions on potential genes that could be affected, and are not extensive enough to monitor large scale metabolic changes, potentially overlooking adaptive traits in distant areas of affected networks (135). Even if a mutation in a single candidate gene was found, this may not be the only mutation that contributed to the adaptive process. DNA microarray studies have been used for whole genome mutation discovery (136-138). The system relies on hybridisation of fragmented DNA from evolved strains to microarrays, with mutated DNA hybridising less well than non-mutated DNA. However, microarray based studies, while on the genome-wide scale, are subject to very high false positive rates and require additional downstream sequencing analysis for completion.

The advent of next-generation sequencing (NGS) has heralded a new era in the study of genomics. Three platforms are now commonly used, 454 Life Science's FLX (139), Illumina's HiSeq 2000 (140) and Applied Biosystems' SOLiD™ System (141). Each platform uses unique sequencing technologies such as pyrosequencing for FLX, sequencing-by-synthesis for HiSeq 2000, and sequencing by ligation for SOLiD™ to allow for massively parallel sequencing. Each instrument has inherent advantages and disadvantages, and the ideal platform depends on the application for which it is to be used. Recently, Illumina's HiSeq 2000 genome analyser has become the market leader due to very competitive pricing and high accuracy base-calling ability (142). Each of these NGS technologies have made whole genome sequencing (WGS) a reality for many laboratories, as an organism's entire genome sequence can be obtained quickly and affordably. WGS by next generation sequencers generates millions of reads, which can be assembled in one of two ways. Assembly may be either without a reference genome sequence via *de novo* assembly or, more typically, assembled with the aid of a fully annotated reference genome sequence of a closely related organism. For the identification of mutations in a genome sequence several mutation detection pipelines have recently been developed (143-147). The mutation detection pipeline,breseq was designed for the study of *E. coli* genome sequences and has been used to compare the assembled genome sequences of evolved *E. coli* strains to that of the ancestor to identify regions of sequence variation between the two genomes (52). In the publication by Barrick *et al.* (52), it was shown that the full suite of mutations in genome sequences of evolved clones could be identified via a whole genome sequencing manner during the course of evolution.

However, some difficulties remain in the detection of certain types of mutations using a NGS approach. These difficulties are mostly associated with the short length of reads that are generated by the majority of NGS technologies, such as that provided by the Illumina HiSeq 2000 system, which generates reads of 90 bp in length (141, 148). Short read lengths have limitations in the detection of certain types of mutations such as gross chromosomal rearrangements (GCRs), which are often mediated by short repetitive sequences (149). These mutations arise largely due to the movement of transposable elements and may have implications for the fitness of the genomes in which they arise. Use of mate-pair and paired-end read sequencing strategies have improved accurate detection of these mutation events. They do so by retaining distance information between reads via the use of inserts either between reads (for paired-end reads) or extending from reads (for mate-pair reads) (150). More recently, Pacific Biosciences PacBio's SMRT™ Sequencing produces read lengths of greater than 10 kb (151). This greater read length allows for more complete assemblies and hence facilitate better detection of GCRs in genome sequencing studies.

Since the emergence of next-generation sequencing technologies, and their affordability, WGS using NGS technologies can now be applied as a tool for LTEE studies. The combination of WGS with mutation detection in LTEE studies allows for many advantages. i) Before the advent of NGS, sequencing of entire genomes for LTEE studies was only feasible with smaller, less complex viral genomes, constrained by the technological limitations of Sanger sequencing methods (140). Now NGS technologies are routinely applied to obtain entire prokaryotic genomes, as well as more complex eukaryotic genomes (152-154). ii) Reduced run time and sequencing costs allow for more replicates to be sequenced, a key advantage in the identification of parallelism among traits from replicate populations. iii) Massive read coverage provided by NGS allows mutations to be reported with greater accuracy. iv) Genome re-sequencing strategies of descendent organisms can be employed, where WGS is obtained many times over the course of the LTEE study. In this way, a comprehensive picture allowing for the quantification of evolutionary dynamics of populations through time is captured, important to identify processes such as clonal interference (52). v) The identification of the many mutations that underlie the adaptive process can be identified, creating a comprehensive picture of the role epistasis may play in evolution. This is possible as the full suite of mutations responsible for key phenotypes are identified together with the use of NGS and mutation detection software. vi) WGS, in combination with LTEE studies under different

environmental conditions, allows for the observation of how natural selection shapes genomes in different environments under different selection pressures. Resulting phenotypes that may arise in the different environments can be readily matched to genomes within affected populations (74, 155). Thus, the combination of WGS and LTEE enables insights into the molecular genetics of evolution on an unprecedented scale and provides knowledge of the different adaptive pathways a genotype may take in the adaptive landscape.

1.7. Thesis Outline

My research questions are as follows:

- What are the dynamics of adaptation under aerobic and anaerobic conditions?
- What genes and mutations are more important for adaptation to aerobic, anaerobic and a temporally heterogeneous aerobic to anaerobic treatment?
- How does niche width affect adaptive pathways?
- Do multiple pathways exist for bacterial adaptation to aerobic and anaerobic conditions, and if so how do these pathways differ?

To answer these questions, I used the facultative anaerobic model organism *E. coli* REL4536, a descendent of the REL606 strain, a well-established model organism used in LTEE studies. An overview of my strategy is depicted in **Figure 1.7**. Firstly, replicate lineages were set up and maintained for 4,000 generations in the three conditions of study: strict aerobic, strict anaerobic and a treatment that fluctuated between the two environments. Next, relative fitness of the evolving lineages was monitored with the use of frequent competition assays, performed every 1,000 generations. Finally, the mutations that arose during the 4,000 generations of adaptation were identified using whole genome sequence data of evolved lineages at two time points (at 2,000 and 4,000 generations). I hypothesised that the combination of experimentally evolved lineages of the facultative anaerobic organism REL4536 in aerobic and anaerobic conditions and whole genome sequencing, would allow me to identify the evolutionary pathways undertaken during bacterial adaptation to these environments.

The thesis described will be structured into the following results chapters:

Chapter 3: Adaptation to aerobic and anaerobic environments

In this chapter, long term experimental replicate lineages were established in the three treatments under study: strict aerobic; strict anaerobic and a temporally heterogeneous aerobic to anaerobic treatment. Adaptation of the evolving populations was monitored with relative fitness trajectories from competition assays, performed every 1,000 generations over the 4,000 generations of this experiment. Competitive relative fitness assays were performed in both the environment of selection and the alternative environment to examine the occurrence of tradeoffs within the evolved lines.

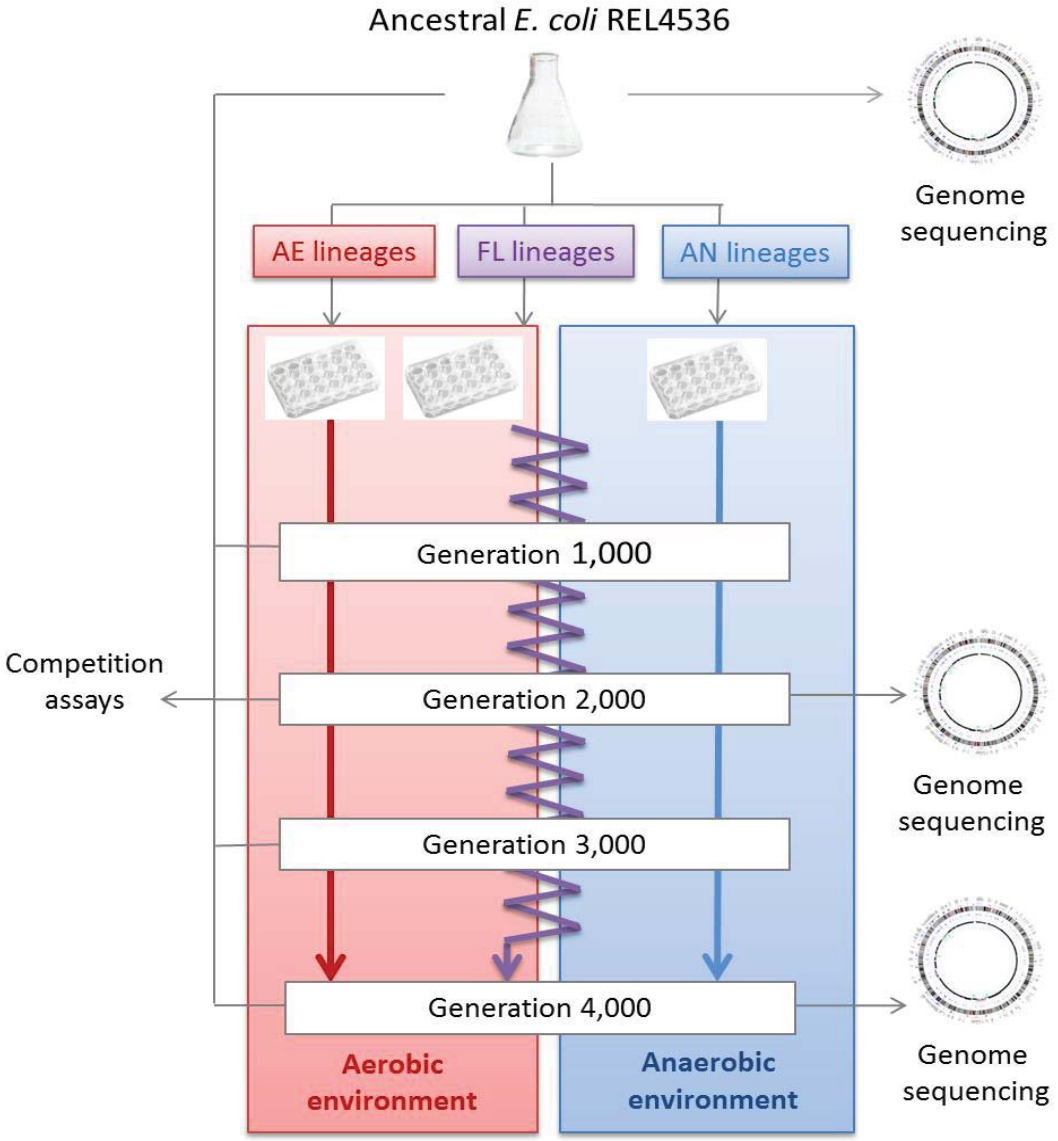

Figure 1.7: Diagrammatic overview of this thesis. The establishment of replicate lineages in the three different environments is indicated, as is the maintenance for the 4,000 generations. Competition assays

performed with the ancestral strain and evolved strains at 1,000, 2,000, 3,000 and 4,000 generations are shown. Genome sequencing performed on the ancestral clone as well as clones isolated at 2,000 and 4,000 generations is also displayed. See text for details.

Chapter 4: Genotypic analysis of lineages adapting to aerobic and anaerobic environments

This chapter examined the mutations that have arisen among clones isolated from evolved populations. Genomes were sequenced at 2,000 and 4,000 generations and mutations were identified using whole genome sequence data and mutation detection pipelines.

Chapter 5: Investigating the origin and maintenance of diversity in the anaerobic lineages

The focus of this chapter was to investigate the genetic basis of the polymorphic colonies that had arisen in the anaerobic environment. Furthermore, the genetic basis for some of the complexities that were encountered among lineages evolving in the anaerobic environments were explored.

Chapter Two : Materials and methods

2.1. Materials

2.1.1. Lab equipment

Conventional glass bottles and DURAN pressure plus bottles, ranging from 50 to 2,000 mL volumes, were originally sourced from Schott Duran® (Mainz, Germany). Solution and buffer sterilisation took place either by filter sterilisation using 0.22 µm Millex syringe-driven filters (Millipore, Co. Cork, Ireland) or in a high pressure steam steriliser SX-700E autoclave (TOMY TECH, Fremont, CA, USA). Sterile syringes ranging from 1 mL to 30 mL were supplied by Becton Dickinson and Co. (Becton, Dickinson and Co., Sparks, MD, USA). Work with high risk chemicals was performed in a fume hood (Thermo Plastic Engineering Ltd. Wellington, NZ). A PHM62 pH meter (Radiometer, Copenhagen, Denmark) was used during media and buffer preparation. Sterile aerobic work was performed in a Class II Type a/B3 biohazard cabinet (Nuaire Biological Safety Cab, Plymouth, MN, USA). Sterile 24-well non-treated polystyrene flat-bottom tissue culture plates with low evaporation lid (referred to as 24-well plates) were supplied by Becton Dickinson and Co..

Sterile anaerobic work was performed in an anaerobic chamber (Coy Laboratory Products Inc., Grass Lake, MI, USA), maintained at a 92% carbon dioxide (CO_2) and 8% hydrogen (H_2) atmosphere. CO_2, H_2 and nitrogen (N_2) gases of industrial grade quality were supplied by BOC Gas (Auckland, New Zealand). AnaeroJar™ 2.5 L with rack plate canisters (Oxoid, Hampshire, UK) were used for incubating anaerobic agar plates. Anaerobic resazurin indicator strips were from Oxoid. Traces of oxygen (O_2) were removed from the gases by passing through a column packed with reduced copper filings that were heated to 400°C in an inline furnace. AnaeroPack® 2.5 L rectangular anaerobic gas boxes (Mitsubishi Gas Company Inc., Tokyo Japan) were used for incubation of anaerobic cultures.

A warm room was maintained for culture incubation at 37°C in which Infors-HT Labtron shakers (Infors USA Inc., Laurel, MD, USA) were placed and used as required. For incubation, an AccuBlock™Digital Dry Bath heating block (Labnet International Inc., Edison, NJ, USA) or a Grant GB

series re-circulating water-bath (Grant, Cambridge, UK) was used. An Infors HT Ecotron incubator (Infors-HT Inc.) or THZ-300 Shaking Incubator (Labtron) were used for the long-term growth experiment and general incubation steps maintained at 37°C with shaking at 150 rpm. Growth of cultures was monitored spectrophotometrically by optical density at 600 nm on an Ultraspec 1100 Pro (Amersham Biosciences, GE Healthcare, Uppsala, Sweden). For microscopic work a DM2500 microscope (Leica Microsystems, Wetzlar, Germany) with Leica Application Suite digital software package (Leica Microsystems) was used.

Conventional polymerase chain reactions (PCRs) were performed in Mastercycler proS machines (Eppendorf, Hamburg, Germany). Horizon gel tanks for running agarose gels were Biometra Analytik Jena (Gottingen, Germany). Photographs of DNA gels were taken using a Gel Logic 200 Imaging System (Eastman Kodak, Rochester NY, USA) with a D700 camera (Nikon Corporation, Tokyo, Japan) with a 60mm macro lens; Camera Control Pro software (Nikon) and Lightroom 2 (Adobe, San Jose, CA, USA) were used for development of the raw format data.

2.1.2. Lab chemicals and enzymes

2.1.2.1. DNA-free water

Distilled water (dH_2O) was collected from a Millipore distillation apparatus and autoclaved. Water was then irradiated with UV light (254 nm, 6 W) for 8 hours.

2.1.2.2. PCR reagents

PCR reagents included 10× Taq buffer minus Mg^{2+} (200mM Tris-HCl, pH 8.4, 500 mM KCl), 50 mM $MgCl_2$ and 10 mM dNTP PCR grade mix were from Roche Diagnostics (Basel, Switzerland). Platinum® Taq or Platinum® High Fidelity Taq (Invitrogen, Carlsbad, CA, USA) were used at a concentration of 5 U/µL polymerase.

2.1.2.3. Ribonuclease A

Ribonuclease A (RNaseA) stock solution was made from Ribonuclease A from bovine pancreas (Sigma-Aldrich Corp., St. Louis, MO, USA) at a concentration of 10.0 mg/mL by dissolving 50.0 mg of Ribonuclease A in 5 mL 10 mM Tris-HCl buffer,filter sterilised, and stored at -20°C.

2.1.2.4. Proteinase K

Proteinase K from *Aspergillus melleus* (Sigma-Aldrich) was made to a stock solution of 5.0 mg/mL by dissolving 25.0 mg of Proteinase K in 5 mL of TE (10:1) buffer, filter sterilised and stored protected from light at -20°C.

2.1.3. Bacterial strains

The bacterial strains used in this study are listed in **Table 2.1**.

Table 2.1: Bacterial strains used in this study

Species	Strain	Application	Source
Escherichia coli	REL4536	Ancestral strain for this study	Richard Lenski, MSU[1]
Escherichia coli	REL606	Ancestral strain for Lenski's adaptation experiment (100)	Richard Lenski, MSU[1]
Escherichia coli	DH5α	Biological control for phage contamination tests	Paul Rainey, MU[2]
Escherichia coli	B113	Biological control for phage contamination tests	AgResearch Ltd[3]

[1] Michigan State University, East Lansing, USA.

[2]Massey University, Albany, Auckland, New Zealand.

[3]AgResearch Ltd, Grasslands Research Center, Palmerston North, New Zealand.

2.1.4. Solutions and buffers

2.1.4.1. Tris-HCl

To make 1 M Tris-HCl, 121.1 g Tris base was dissolved in 800 mL dH_2O. The solution was adjusted to pH 7.6 with concentrated HCl, made up to 1 L, then autoclaved.

2.1.4.2. TE buffer

TE buffer was made by mixing 10 mM Tris with 1 mM EDTA, then dissolved in dH_2O and adjusted to pH 8.0 and autoclaved.

2.1.4.3. 50× TAE buffer

TAE (50×) (2 M Tris, 1 M acetic acid, 50 mM EDTA pH 8.0) was made by adding 242.0 g Tris base, 57.1 mL glacial acetic acid and 100 mL of 0.5 M EDTA pH 9.0 to dH_2O to a volume of 1 L. The solution was adjusted to pH 8.0 then autoclaved. A working concentration of 1× TAE buffer was obtained by 50-fold dilution of 50× TAE buffer.

2.1.4.4. 5 M NaCl solution

A 5 M NaCl solution was made by dissolving 292.2 g NaCl in dH_2O, brought to a volume of 1 L with dH_2O, then autoclaved.

2.1.4.5. 5% (wt/vol) Triphenyltetrazolium chloride solution

A 5% (wt/vol) 2,3,5-triphenyltetrazolium chloride (TTC) (wt/vol) solution was made by dissolving 5.0 mg TTC (Sigma-Aldrich) in dH_2O to a volume of 10 mL, then filter sterilised.

2.1.4.6. Crystal violet staining reagent

A solution of 2.0 g of crystal violet was dissolved in 20 mL 95% ethanol. A second solution of 0.8 g ammonium oxalate dissolved in 80 mL dH_2O was made. The two solutions were combined to make 0.02% (wt/vol) crystal violet reagent.

2.1.4.7. Gram's Iodine

A 0.003% (wt/vol) Gram's Iodine was made by dissolving 1.0 g iodine and 2.0 g potassium iodide in 300 mL of dH_2O and stored in an amber bottle.

2.1.4.8. Decolorising agent

Decolorising agent is composed of a 1:1 ratio of 50 mL acetone and 50 mL 95% ethanol.

2.1.4.9. Safranin

A stock solution of 2.5 g of safranin O dissolved in 100 mL 95% ethanol was made and diluted 10-fold for working concentrations of 0.025% (wt/vol) safranin staining solution.

2.1.4.10. Ethanol

Analytical grade ethanol (VWR International Ltd, Radnor, PA, USA) solutions were supplied at 95% (vol/vol) stock or absolute 99.5%, (vol/vol) stock. Ethanol was diluted and used at 70% (vol/vol), using dH_2O.

2.1.4.11. Isopropanol

Isopropanol (propan-2-ol), analytical grade, was from VWR International Ltd .

2.1.4.12. Liquid N_2

Liquid N_2 was supplied by BOC Gas.

2.1.4.13. Lysis buffer

Lysis buffer was prepared by combining 100 mL of 1 M Tris pH 8.0, 20 mL of 5 M NaCl, 100 mL of 0.5 M EDTA pH 8.0 and 200 mL 10% (wt/vol) SDS. This was brought up to 1 L with 480 mL dH_2O, then filter sterilised.

2.1.4.14. EDTA

To make 0.5 M ethylenediaminetetraacetic acid (EDTA), 186.1 g EDTA was dissolved in 800 mL dH_2O, pH was adjusted to 8.0 with the addition of NaOH, then brought to 1 L with dH_2O and autoclaved.

2.1.4.15. SDS solution

To make 20% (wt/vol) sodium dodecyl sulphate (SDS) solution, 200.0 g of SDS was dissolved in dH_2O. The solution was adjusted to pH 7.2, brought to a volume of 1 L and filter sterilised.

2.1.4.16. Phenol:chloroform: isoamyl alcohol solution

Phenol, chloroform and isoamyl alcohol in a 25:24:1 (vol:vol:vol) ratio, respectively, was manufactured by Invitrogen.

2.1.4.17. Chloroform isoamyl alcohol solution

Chloroform:isoamyl alcohol solution in a 24:1 (vol:vol) ratio, analytical grade, was manufactured by Invitrogen.

2.1.4.18. Sodium acetate solution

To make 3 M sodium acetate, 246.0 g of sodium acetate was dissolved in 500 mL dH_2O. The mixture was adjusted to pH 5.2 and using glacial acetic acid (VWR International Ltd.), brought to 1 L with dH_2O and autoclaved.

2.1.4.19. 1× phosphate buffered saline

To make phosphate buffered saline, 8.0 g NaCl, 0.20 g KCl, 1.44 g Na_2HPO_4, 0.24 g of KH_2PO_4 were mixed and adjusted to pH 7.4. The solution was brought to 1 L with dH_2O, then autoclaved.

2.1.4.20. L-cysteine-HCl reducing agent

L-cysteine reducing agent was made by boiling 1 L of dH_2O for 1 min and cooled to room temperature under (O_2-free) N_2 gas. Empty serum bottles were flushed with N_2 gas. After addition of 12.5 g of L-cysteine HCl to dH_2O the solution was adjusted to pH 9.0 with 4 M NaOH. $Na_2S.9H_2O$ crystals were washed with dH_2O and 12.5 g was added to the solution. After the crystals had dissolved, the solution was divided into the N_2-filled serum bottles, and autoclaved for 10 min at 15 psi at 121°C.

2.1.5. Liquid media

2.1.5.1. Lysogenic-Broth (LB) medium

2.1.5.1.1. Aerobic preparation

To make Lysogenic Broth media, 10.0 g bacto-tryptone, 5.0 g bacto-yeast extract, 10.0 g sodium chloride were mixed and brought to 1 L with dH_2O, then sterilised by autoclaving.

2.1.5.1.2. Anaerobic preparation

The composition of anaerobically prepared LB media was as per **Section 2.1.5.1.1**. However, the solution was prepared in DURAN pressure plus bottles and the media boiled for one minute in a microwave oven to remove O_2. The media was allowed to cool to room temperature with a stream of (O_2-free) CO_2 gas bubbled through it, during which 400 µL of 0.1% (wt/vol) resazurin indicator solution was added, turning the solution to a light blue colour. Once cooled, bottles were sealed with butyl rubber stoppers and polybutylene terephthalate screw caps, then autoclaved. After autoclaving, the medium was brought into an anaerobic chamber. Traces of dissolved O_2 were removed from the solution with the addition of 20 mL L-cysteine HCl reducing agent (**Section 2.1.4.20**), after which the media turned colourless from light blue, indicating trace amounts of O_2 had been removed.

2.1.5.2. Davis minimal (DM) medium

2.1.5.2.1. Aerobic salt preparation

The salts component of Davis minimal media was prepared by mixing 7.0 g potassium phosphate (dibasic trihydrate), 2.0 g potassium phosphate (monobasic anhydrous), 1.0 g ammonium sulfate and 0.5 g sodium citrate. The mixture was brought to 1 L with dH_2O, then autoclaved.

2.1.5.2.1.1. Aerobic additive addition

After autoclaving, DM salts media was supplemented with 0.250 mL of filter sterilised 10% (wt/vol) glucose, 1 mL of filter sterilised 10% (wt/vol) magnesium sulfate and 1 mL of filter sterilised 0.2% (wt/vol) thiamine. After the addition of these additives the medium is herein referred to as DM25. For cultures requiring high cell densities, DM salts media was supplemented with 1 mL of filter sterilised 10% (wt/vol) glucose, creating DM1000.

2.1.5.2.2. Anaerobic preparation

To prepare DM25 anaerobically, the salts components of the media were prepared as described in **Section 2.1.5.2.1.** However, the mixture was prepared in DURAN pressure plus bottles and boiled in a microwave oven for one minute. The media was allowed to cool to room temperature with a stream of (O_2-free) CO_2 gas bubbled through it, during which 400 µL of 0.1% (wt/vol) resazurin indicator solution was added, turning the solution to a light blue colour. Once cooled, bottles were sealed with butyl rubber stoppers and polybutylene terephthalate screw caps, and autoclaved.

2.1.5.2.2.1. Anaerobic additive addition

After autoclaving, the medium was brought into an anaerobic chamber. Residual traces of dissolved oxygen were removed from the solution with the addition of 20 mL L-cysteine HCl reducing agent (**Section 2.1.4.20**) turning the media colourless from light blue, indicating trace amounts of oxygen had been removed. Once colourless, sterile stocks of anaerobically prepared 0.250 mL 10% (wt/vol) glucose, 1 mL of 10% (wt/vol) magnesium sulfate and 1 mL of 0.2% (wt/vol) thiamine were prepared. After the addition of these additives the medium is hereafter referred to as anaerobic DM25.

2.1.6. Solid media

2.1.6.1. LB solid medium

2.1.6.1.1. Aerobic preparation

Medium was prepared as per **Section 2.1.5.1**, with the addition of 16.0 g bacto agar (Becton, Dickinson and Co. Ltd.) prior to autoclaving. After autoclaving, the medium was cooled to 55°C, antibiotics added if required (**Section 2.1.7.3**), poured, and allowed to dry in a biohazard cabinet until solidified.

2.1.6.1.2. Anaerobic preparation

The components were prepared in three DURAN pressure plus bottles; one bottle contained components listed in **Section 2.1.5.1** without dH_2O, the second bottle contained 16.0 g bacto agar, and the final bottle contained 1 L dH_2O. The first two bottles were maintained under continuous (O_2-free) N_2 gas. The dH_2O containing bottle was boiled for one minute, and allowed to cool to room temperature under (O_2-free) N_2 gas. Once cooled, it was split evenly between the first two bottles. Both bottles were mixed, then immediately sealed with butyl rubber stoppers and polybutylene terephthalate screw caps and autoclaved. After autoclaving, the contents were allowed to cool to 55°C and transferred into an anaerobic chamber. Medium was poured into petri dishes and allowed to solidify in the anaerobic chamber. The plates were placed inside Oxoid AnaeroJar™ gas canisters, with an Oxoid indicator strip to monitor the anaerobic conditions, sealed, then taken out of the anaerobic chamber until required.

2.1.6.2. DM solid medium

2.1.6.2.1. Aerobic preparation

DM solid media was made by preparing the media components in three bottles simultaneously. One contained components listed in **Section 2.1.5.2.1**, the second contained 16.0 g bacto agar, the

final bottle contained 4.0 g glucose. A 1 L volume of dH_2O was split evenly between the three bottles, which were then autoclaved. After autoclaving, 1 mL each of filter sterilised 10% (wt/vol) magnesium sulfate and filter sterilised 0.2% (wt/vol) thiamine additives were added. The contents from the three bottles were mixed, poured and allowed to solidify in a biohazard cabinet.

2.1.6.2.2. Anaerobic preparation

The components listed in **Section 2.1.5.2.1** (before the addition of dH_2O) were split into four DURAN pressure plus bottles. Each bottle contained; DM salts component powder; 16.0 g bacto agar powder; 4.0 g glucose powder and 1 L dH_2O. The bottle containing dH_2O was boiled for one minute then allowed to cool to room temperature under (O_2-free) N_2 gas. At the same time, the remaining three bottles containing powder were also under continuous (O_2-free) N_2 gas. Once the bottle containing dH_2O had cooled to room temperature, the contents of the three bottles were combined, sealed with butyl rubber stoppers and polybutylene terephthalate screw caps, then autoclaved. After autoclaving, the solutions were taken into an anaerobic chamber and the contents combined. Here, 1 mL of sterile, anaerobically-prepared 10% (wt/vol) magnesium sulfate and 0.2% (wt/vol) thiamine additives were added to the glucose containing bottle. The contents were poured into petri dishes and allowed to solidify inside the anaerobic chamber. Plates were placed inside Oxoid AnaeroJar™ gas canisters, with Oxoid indicator strip to monitor the anaerobic conditions, sealed, then taken out of the anaerobic chamber until required.

2.1.6.3. Minimal arabinose (MA) solid medium

Minimal arabinose (MA) plates were made by preparing components in three bottles. One contained the components listed in **Section 2.1.5.2.1**, a second contained 16.0 g bacto agar, the final bottle contained 10 g L-arabinose. A 1 L volume of dH_2O was split evenly between the three bottles and autoclaved separately. After autoclaving and cooling to 55°C, 1 mL of sterile, anaerobically-prepared 10% (wt/vol) magnesium sulfate and 0.2% (wt/vol) thiamine additives were added to the bottle containing L-arabinose. The contents from the three bottles were mixed, poured into petri dishes and allowed to solidify in a biohazard cabinet.

2.1.6.4. Tetrazolium arabinose (TA) solid medium

TA plates were made by preparing the components in three bottles separately. The first bottle was prepared by adding 10.0 g of bacto-tryptone, 1.0 g of bacto-yeast extract, 10.0 g of sodium chloride together. The second bottle contained 10.0 g of L-arabinose, while the final bottle contained 16.0 g of bacto agar. A 1 L aliqot of dH_2O was split evenly among the three bottles and

the bottles were autoclaved separately. After autoclaving, the bottles were cooled to 55°C, then mixed together with 1 mL of 5% (wt/vol) TTC (**Section 2.1.4.5**). The TAra agar was poured into petri dishes and allowed to solidify in a biosafety cabinet.

2.1.7. Media additives

2.1.7.1. DM25 additives

2.1.7.1.1. Aerobic preparation

Additives were prepared as stock percentages as described below.

2.1.7.1.1.1. Glucose

A 10% (wt/vol) glucose stock was prepared by dissolving 1.0 g of glucose in 10 mL of dH_2O and sterilised by filter sterilisation. The solution was aliquoted in 0.250 mL volumes into sterile microcentrifuge tubes and stored at -20°C.

2.1.7.1.1.2. Magnesium sulphate

A 10% (wt/vol) magnesium sulfate stock was prepared by dissolving 1.0 g of magnesium sulfate in 10 mL of dH_2O and sterilised by filter sterilisation. The solution was aliqoted in 1 mL volumes into sterile microcentrifuge tubes and stored at -20°C.

2.1.7.1.1.3. Thiamine

A 0.2% (wt/vol) thiamine stock was prepared by dissolving 0.1 g of thiamine in 50 mL of dH_2O and filter sterilised. The solution was aliquoted in 1 mL volumes and into sterile microcentrifuge tubes stored at -20°C.

2.1.7.1.2. Anaerobic preparation

Stock solutions of additives were prepared as in **Section 2.1.7.1.1**. Before sterilisation samples were gassed with (O_2-free) CO_2 for 30 minutes. Samples were brought into an anaerobic chamber, filter sterilised and aliquoted as in **Section 2.1.7.1.1**.

2.1.7.2. 70% glycerol saline

2.1.7.2.1. Aerobic preparation

Glycerol saline was made by adding 8.5 g NaCl to 700 mL of glycerol, then brought to 1 L with dH_2O and autoclaved.

2.1.7.2.2. Anaerobic preparation

Components listed in **Section 2.1.7.2.1** were mixed in DURAN pressure plus bottles and boiled. After boiling for one minute, the bottles were allowed to cool to room temperature under (O_2-

free) CO_2 gas. Once cooled, bottles were sealed with butyl rubber stoppers and polybutylene terephthalate screw caps and autoclaved.

2.1.7.3. Antibiotics

Antibiotics (supplied by Sigma-Aldrich) used in culture media are listed in **Table 2.2**. For antibiotics dissolved in dH_2O or 1 M NaOH, the solution was filter sterilised.

Table 2.2: Antibiotics used in this study

Antibiotic	Abbreviation	Stock Concentration	Working concentration	Solvent
Ampicillin	Amp	50 mg/mL	50 µg/mL	dH_2O
Chloramphenicol	Cm	34 mg/mL	34 µg/mL	100% (vol/vol) ethanol
Kanamycin	Kan	50 mg/mL	50 µg/mL	dH_2O
Nalidixic Acid	Nal	30 mg/mL	30 µg/mL	1 M NaOH
Rifampicin*	Rif	50 mg/mL	100 µg/mL	100% (vol/vol) methanol
Tetracycline*	Tet	10 mg/mL	10 µg/mL	70% (vol/vol) ethanol

* Light sensitive antibiotics were protected from light both as stocks, and when used in growth media.

2.2. Methods

2.2.1. Growth conditions

2.2.1.1. Aerobic cultures

All aerobic work with cultures took place inside a biohazard cabinet (**Section 2.1.2**). Cultures in 24-well plates were grown in 1 mL volumes with aerobically prepared media with a low evaporation lid. Cultures of larger volume (10 mL or 30 mL) were grown in sterile 50 mL centrifuge tubes with lids slightly open to allow aeration. All aerobically grown cultures were incubated at 37°C with orbital shaking at 150 rpm (unless otherwise stated).

2.2.1.2. Anaerobic cultures

All anaerobic work with cultures took place inside an anaerobic chamber. Anaerobic cultures in 1 mL volumes were grown with anaerobically prepared media in 24-well plates with a low evaporation lid. The 24-well plates were placed in AnaeroPack® gastight boxes and sealed to maintain an anaerobic environment. Cultures required at higher volumes were grown in sterile 250 mL serum bottles sealed with butyl rubber bungs and metal casings, or in DURAN pressure

plus bottles sealed with butyl rubber stoppers and polybutylene terephthalate screw caps. All anaerobic cultures were incubated at 37°C with orbital shaking at 150 rpm (unless otherwise stated).

2.2.2. Culture resuscitation

To resuscitate cultures for the long term evolutionary experiment (**Section 2.2.5**) and other experiments, cells stored in glycerol saline at -85°C storage were first washed in growth media to remove glycerol. Small pieces of frozen culture stock (approximately 10 µL) were added to 990 µL of DM25. Samples were vortexed and centrifuged for 3 min at 11,000 *g*, then media removed. This was repeated twice. This culture was then used for inoculation. Aerobic culture resuscitation took place in a biohazard cabinet and was performed with aerobic DM25, while anaerobic culture resuscitation took place in an anaerobic chamber with anaerobic DM25.

2.2.3. Growth courses

Growth courses were performed for cultures growing in LB, DM0 and DM25 prepared aerobically and anaerobically. In all cases, growth courses took place in 24-well plates, with total culture volumes of 1 mL. Frozen cultures were resuscitated as per **Section 2.2.2**. The culture was inoculated into two replicate wells containing 990 µL of the media under investigation and incubated overnight as per **Section 2.2.1.1** for aerobic growth or **Section 2.2.1.2** for anaerobic growth. The following day, one of the two replicates was used to measure overnight growth using optical density readings, giving an indirect measure of overnight growth of the remaining replicate, which was used as the inoculum. Following this, 10 µL of the inoculum was inoculated into three replicates of 990 µL of tested media and incubated as before. At three hour intervals (typically 0, 3, 6, 9, 12, 18, 24, 36 and 48 hour time points), cultures were removed from the incubator and 10 µL samples of each of the three biological replicate cultures were taken. Colony forming units (CFU) per mL was used to determine total viable count rather than optical density readings in an attempt to estimate the number of live cells only. Samples were then serially diluted to 10^{-5} in DM Salts and plated on LB solid media with 3-fold technical replication. Plates were then incubated at 37°C overnight and counted the next day to estimate growth dynamics. All culture manipulation was performed in a biohazard cabinet for aerobically grown cultures, and in an anaerobic chamber for anaerobically grown cultures.

2.2.4. Gram stains

A sample of the culture under investigation was added onto a glass microscope slide, air-dried, and fixed by passing the slide over a flame several times. Slides were flooded with crystal violet stain (0.02% wt/vol), left for 1 min, and washed off with H_2O. The samples were covered with Gram's Iodine solution (0.003% wt/vol) for 2 min, then washed off with H_2O. The samples were decolourised with acetone:alcohol (1:1, vol:vol), rinsed with H_2O, counterstained with safranin (0.025% wt/vol) for 1 min, and rinsed with H_2O. The samples were viewed under a Leica DM2500 microscope, with settings manually optimised for Gram stains (exposure time of 573.0 ms, gamma 0.92, gain 9.2×, saturation 1.50).

2.2.5. Long-term adaptation experiment

2.2.5.1. Establishment of long term evolving populations

The ancestral strain for this study, *E. coli* REL4536, stored at -85°C was resuscitated as per **Section 2.2.2** and streaked sequentially three times on DM25 plates and incubated overnight at 37°C. A single random colony was picked and grown aerobically in 9.90 mL DM25 to establish an ancestral REL4536 culture. From this, 21 independent *E. coli* populations were created for long term experimental evolution under the three treatments of interest: aerobic growth, anaerobic growth and growth with daily alternating (fluctuating) exposure to aerobic and anaerobic environments (**Figure 2.1**). Establishment of the long term aerobic and fluctuating cultures (seven independent replicates each) took place aerobically, with inoculation of 10 µL of the ancestral culture into 990 µL of aerobic DM25 in 24-well plates. Establishment of the long term anaerobic cultures were performed similarly, however, within an anaerobic chamber, and using anaerobic DM25 medium. All cultures were incubated aerobically or anaerobically as appropriate, under the conditions described in **Section 2.2.1**. In total, seven replicate populations (lineages) were established per treatment, and an uninoculated media-only control was included in each treatment group to monitor for cross-contamination within the 24-well plates (**Section 2.2.5.4**).

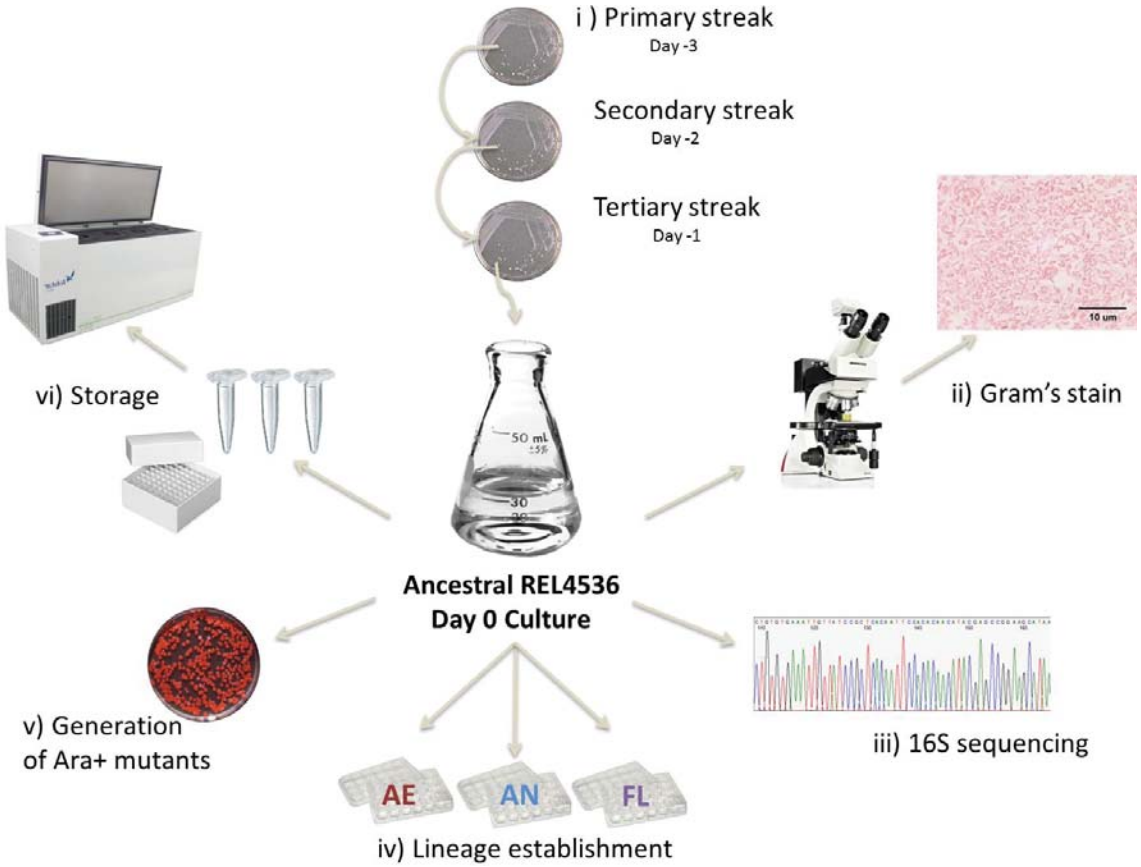

Figure 2.1: Overview of lineage establishment at Day 0. i) *E. coli* REL4536 clones were streaked sequentially three times for a pure clone and added to DM25 culture and grown overnight (**Section 2.2.5.1**). This was the ancestral REL4536 Day 0 culture. ii) Aliquots were Gram stained to ensure purity (**Section 2.2.4**). iii) Sequencing of 16S rRNA genes was used to confirm identity as *E. coli* REL4536. iv) Aliquots of 10 µL were taken from the Day 0 culture to establish the long-term lineages in their respective treatments (**Section 2.2.5.1**). v) Further aliquots were taken for the generation of Ara+ mutants for competition assays (**Section 2.2.6.1**). vi) The remainder of the culture was then stored at -85°C (**Section 2.2.5.5**).

2.2.5.2. Maintenance of evolving populations

The long term evolutionary cultures were propagated by daily sub-culturing. A diagrammatic representation of lineages in the 24-well plate format is shown (**Figure 2.2 a**). The positions of lineages within the plate were moved to the adjacent well, in an anti-clockwise fashion, each time a new plate was used (every fourth day), in order to minimise effects such as minor variation in evaporation rates due to the proximity of wells to the edge of the plate. For the aerobic cultures, sub-culturing took place approximately every 24 hours in a biohazard cabinet, where 10 µL

(1/100th volume) of each culture from the previous day was transferred to 990 µL fresh DM25 medium within the 24-well plate. Sub-culturing of the anaerobic cultures took place similarly, but within the anaerobic chamber and using anaerobically prepared DM25. The fluctuating treatment required culture growth to be alternated between aerobic and anaerobic environments on a daily basis, and bacteria were sub-cultured under each condition accordingly. All cultures were incubated aerobically or anaerobically as appropriate, under the conditions described in **Section 2.2.1**. For each treatment, one 24-well plate allowed for three sub-cultures. Thus, when a new plate was required, the lineage positions were re-organised to minimise potential plate position effects on growth.

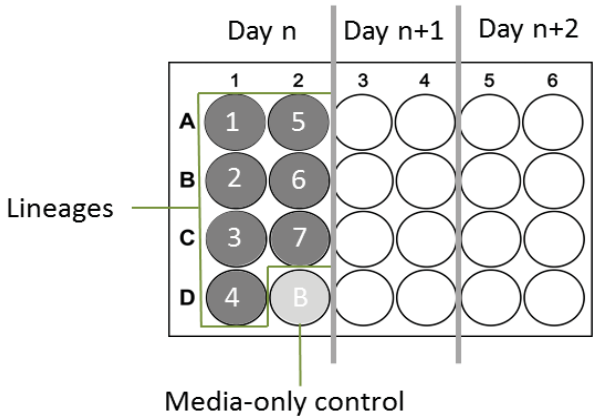

Figure 2.2: Set up of lineages in 24-well plate format. A diagrammatic representation of the lineage set up in the 24-well plate format.

2.2.5.3. Cell density monitoring of evolving populations

Cell densities of the lineages were monitored by measuring total viable counts of cells after 24 hours of growth similar to **Section 2.2.1**.

2.2.5.4. Contamination checks of evolving populations

To monitor for internal contamination between replicate lineages, the media-only control well was was checked for growth by growing aliquots of the blank on LB agar with overnight incubation. To minimise aerosols between replicate lineages a low-evaporation lid was used to seal the plate. To monitor for external contamination, routine phage contamination tests were performed as described by Lenski and colleagues (100). Tests were performed by streaking two aliquots of 20 µL of T5 or T6 phage stocks (1.3×10^6 plaques/mL and 6.4×10^6 plaques/mL respectively) on an LB agar plate and allowed to dry. Aliquots of 3 µL, from the previous days' culture were streaked perpendicularly across phage streaks and plates were incubated at 37°C overnight. Descendents of

E. coli B strains display a T5 sensitive and T6 resistant phenotype. *E. coli* B113 was used as a control for phage stock viability in this study. Contraction of lineage growth along T5 streak, and no contraction along T6 streak indicated no external contamination had occurred. Contamination checks, both internal and external, were performed every 14 days before lineages were stored at -85°C (**Section 2.2.5.5**).

2.2.5.5. Storage of evolving populations

For aerobic and fluctuating lineages, 0.4 mL of 70% (vol/vol) glycerol saline (**Section 2.1.7.2.1**) was mixed with 0.5 mL of overnight culture from aerobic and fluctuating treatment lineages in 1.7 mL Safe-Lock™ tubes in a biohazard cabinet. For anaerobic culture storage, 0.4 mL anaerobic 70% (vol/vol) glycerol saline (**Section 2.1.7.2**) was mixed with 0.5 mL overnight culture from anaerobic treatment lineages in 1.7 mL Safe-Lock™ tubes in the anaerobic chamber. Samples were stored at -85°C every two weeks.

2.2.6. Competitive fitness assays

2.2.6.1. Generation of spontaneous Ara+ mutant strains

To assess relative fitness of the adaptive lineages, neutrally marked reference strains of the ancestor are required. Spontaneous mutants capable of utilising L-arabinose have been shown to be neutral in a variety of conditions (87, 100, 155). The ancestral *E. coli* REL4536 culture (**Section 2.2.5**) was concentrated by splitting the culture into two 5 mL volumes in 15 mL centrifuge tubes and centrifuging at 8,000 *g* for 10 min. The supernatant was removed and each pellet re-suspended in 200 µL DM25, then combined. Aliquots (50 µL) of the cell suspension were plated on to eight MA plates (**Section 2.1.6.3**) and incubated overnight at 37°C. All colonies were single-colony streaked three times on MA plates and with overnight incubation at 37°C after each sub-culture. The colonies were also streaked on TAra plates (**Section 2.1.6.4**), to verify the Ara+ phenotype (discrete white colonies). Single colonies from each of the eight mutants on MA plates were randomly chosen and used to inoculate 9.90 mL DM25 medium and grown at 37°C with 150 rpm shaking overnight. Each culture was stored in aliquots as described (**Section 2.2.5.5**).

2.2.6.2. Generation of antibiotic resistant mutant strains

Frozen replicates of the ancestral *E. coli* REL4536 culture were resuscitated (**Section 2.2.2**) and grown overnight in DM1000 media. The cells were concentrated as described in **Section 2.2.6.1** and plated onto DM plates containing antibiotics: chloramphenicol, kanamycin, nalidixic acid or rifampicin (**Section 2.1.7.3**). Plates were incubated overnight at 37°C. Any colonies that grew were

single-colony streaked three times sequentially on the same medium and antibiotic from which they were isolated. A randomly selected colony was used to inoculate 9.90 mL DM25 medium with the appropriate antibiotic and grown at 37°C with 150 rpm shaking overnight. The culture was stored in aliquots as described (**Section 2.2.5.5**).

2.2.7. Fitness assays with Ara+ marker

Relative fitness was assayed by competing a reference competitor strain against an evolved clone or population in DM25 both aerobically and anaerobically. Cultures were resuscitated from frozen population samples as per **Section 2.2.2**. For fitness assays using the Ara+ competitor clone, 10 µL of both the resuscitated competing and reference Ara+ strains were added individually to 990 µL of DM25 in 24-well plates. Cultures were grown overnight at 37°C with 150 rpm orbital shaking. Each culture was used to initiate four biological replicate cultures by inoculating 10 µL of sample culture into 990 µL DM25 for each replicate. Cultures were grown at 37°C overnight with orbital shaking at 150 rpm. At T = 0, cultures were mixed at an estimated cell ratio of 1:1 of evolved population competitor to Ara+ strain, ca. 5.0×10^5 cells of each in a 1 ml total volume of fresh DM25. A 10 µL aliquot of the combined T = 0 culture was serially diluted to 10^{-2} and 100 µL plated onto two technical replicates of TAra solid media and the total number of each competitor was determined by direct counts based on colony colour (white versus red) on TAra indicator plates following incubation of plates. The competition cultures were incubated at 37°C with orbital shaking at 150 rpm for 24 hours. The following day, 10 µL of the T = 24 culture was serially diluted to 10^{-4}, and 100 µL of dilutions, plated with two fold technical replication on TAra solid media, grown overnight at 37°C. Colonies were counted from each of the plates and recorded.

Anaerobic fitness assays were carried out as above. However, media was anaerobically prepared (**Section 2.1.5.2.2.1**), culture manipulation took place in an anaerobic chamber (**Section 2.1.1**) and cultures were incubated in gas tight anaerobic boxes (**Section 2.1.1**).

2.2.8. Fitness assays with Rif^r marker

Relative fitness of evolved clones or populations was determined similar to **Section 2.2.7.** However, the ancestral reference strain in these assays was that of the Rif^r clone, in place of the Ara+ strain, such that the assay could be performed both aerobically and anaerobically. Cultures were resuscitated from frozen population samples as per **Section 2.2.2**. For competition assays using the Rif^{r}2 competitor clone, 10 µL of sample inoculum was added to 990 µL of DM25, while 10 uL of the Rif^r competitor clone culture was added to 990 µL of DM25 + Rif$^{100\ \mu g/\mu L}$ in 24-well plates.

Cultures were grown overnight at 37°C with 150 rpm orbital shaking. Each culture was used to initiate five biological replicate cultures by inoculating 10 µL of sample culture into 990 µL DM25 for each replicate, while the Rif^{r}2 competitor culture, 10 µL, was inoculated into 990 µL of DM25 + Rif$^{100\ \mu g/\mu L}$ for each replicate. Cultures were grown at 37°C overnight with orbital shaking at 150 rpm. At T = 0, cultures were mixed at an estimated cell ratio of 1:1 of evolved population competitor to Rif^{r}2 strain, ca 5.0 × 10^{5} cells of each in a 1 mL total volume in fresh DM25 media. A 10 µL aliquot of the T = 0 culture was serially diluted to 10^{-2} and 100 µL plated onto two technical replicates of LB solid agar plates. After overnight incubation, total number of mixed cells in the culture, and two replicates of LB + Rif$^{100\ \mu g/\mu L}$ agar plates were examined to determine the number of Rif^{r} cells in the culture following incubation of plates. The competition cultures were incubated at 37°C with orbital shaking at 150 rpm for 24 hours. The following day, 10 µL of the culture was serially diluted to 10^{-4}, and 100 µL of the dilution was plated in duplicate on LB and LB + Rif$^{100\ \mu g/\mu L}$ agar plates and grown overnight at 37°C. Colonies were counted from each of the plates and recorded. At both T = 0 and T = 24, the total number of evolved cells was calculated as in Equation 3.1.

$$n_E = n_{Total} - n_R$$

Equation 3.1: Calculation for the number of evolved cells (n_E). Where n_{Total} is the number of cells on LB plate and n_R is the number of Rif resistant reference cells.

Anaerobic fitness assays were carried out as above. However, media was anaerobically prepared (**Section 2.1.5.2.2.1**), culture manipulation took place in an anaerobic chamber (**Section 2.1.1**) and cultures were incubated in gas tight anaerobic boxes (**Section 2.1.1**). To control for marker inconsistencies in aerobic and anaerobic assays, an ancestor versus Rif^{r}2 clone sample was incorporated with each assay performed.

2.2.8.1. Fitness calculation

Relative fitness is the change of the ratio between two competing strains, when grown together in the same environment (100) and was calculated as presented in **Equations 3.2, 3.3** and **3.4.**

Equation 3.2: Calculation of Malthusian parameter (m_n) for competitors. $n_E(0)$ and $n_R(0)$ are initial densities of evolved and reference strain, respectively. $n_E(1)$ and $n_R(1)$ are final densities after one day of growth of the evolved and reference strains, respectively. t is time (1 day).

$$\mathrm{m}_n = \ln\left[\frac{n_{E,R}(1)}{n_{E,R}(0)}\right] \Big/ t$$

Equation 3.3: Calculation of the doubling time (D_n). where the D_n ratio is a log(2) normalisation of the rate of increase.

$$\mathrm{D}_n = \ln\left[\frac{n_{E,R}(1)}{n_{E,R}(0)}\right] \Big/ \ln(2)$$

Equation 3.4: Calculation of relative fitness ($\boldsymbol{\omega}_{E,R}$). Where D_E is the doubling time of the evolved cells and D_R is the doubling time of the reference strain.

$$\boldsymbol{\omega_{E,R}} = D_E \Big/ D_R$$

2.2.9. Reciprocal invasion assay

In this thesis, polymorphic colonies based on size were observed. Pure cultures of each morphotype, both typical colony morphotypes (TCM) colonies and small colony morphotypes (SCM), were obtained for three populations that had become polymorphic by 2,000 generations (AN-2K-4, AN-2K-6 and AN-2K-7) by streaking sequentially three times on solid LB agar as per **Section 2.2.5.1** and stored as per **Section 2.2.5.5**. No TCM isolates could be identified from the AN-2K-7 population, as such ancestral REL4536 cells represented TCM cells for the AN-2K-7 population for reciprocal invasion and cross-feeding experiments.

To investigate the dynamics of the long-term fitness effects underlying the two morphotypes, reciprocal invasion experiments were used. Briefly, on Day -3, both colony morphotypes were isolated from three sample populations each, AN-2K-4, AN-2K-6 and AN-2K-7 population, and were anaerobically resuscitated from frozen stocks as per **Section 2.2.2** and incubated overnight**.** On Day -2, an additional day of culture acclimatisation was allowed for each culture. On Day -1, CFU/mL counts were performed on these cultures when they had reached stationary phase. On Day 0, CFU/mL counts from the preceding day's cultures were used to estimate the required volume of inoculum to add to the Day 0 cultures at a TCM to SCM cell ratio of 9:1, 1:1 and 1:9 for the three competing populations. Each population was sub-cultured by transferring 10 µL of

growing culture into 990 µL fresh anaerobic DM25 after 24 hours of growth for 14 days. Prior to sub-culturing, the frequencies of each morphotype were obtained by spreading 100 µL of the populations on LB agar media. The morphotypes were distinguished based on colony size after 30 hours of growth at 37^{o}C. Reciprocal invasion experiments were performed with three-fold technical replication.

2.2.10. Cross-feeding assay

2.2.10.1. Media preparation

To investigate if two sub-populations were able to co-exist through a cross-feeding dynamic in the anaerobic environments, cross-feeding experiments were undertaken. Pure cultures of both TCM and SCM from AN-2K-4, AN-2K-6 and AN-2K-7 (**Section 2.2.2**) with REL4536 in place of AN-2K-7 TCM cells once again. To obtain morphotype specific filtrates, revived cultures were inoculated into 12 replicates each of 990 µL anaerobic DM25 and incubated anaerobically for 24 hours. After 24 hours, replicate cultures were then combined and filter sterilised through 0.22 µm filters (Millex). The absence of viable cells in the filtrate was confirmed by plating the filtrate on LB agar and incubating overnight at 37C, and checked for the absence of growth the following day. Anaerobic DM25 additives glucose, magnesium sulphate and thiamine (**Section 2.1.7.1.2**) were added to the media to generate TCM-treated DM25 and SCM-treated DM25 for each population.

2.2.10.2. Assay

Pure ancestral REL4536, and both TCM and SCM morphotype representatives of AN-2K-4, AN-2K-6 and AN-2K-7 were resuscitated in DM25 (**Section 2.2.2**). Each sample was inoculated into untreated DM25 (control), and TCM-treated DM25 and SCM-treated DM25 derived from the lineages of inoculation. For AN-2K-7, SCM cultures were also inoculated into REL4536-treated DM25. All cultures were incubated as per **Section 2.2.3** for 24 hours. Stationary phase cell densities of all samples were obtained via CFU counts (as per **Section 2.2.3**) by plating onto LB agar and incubating for 30 hours at 37^{o}C. Cell densities of each culture were obtained in the corresponding TCM- and SCM-treated DM25, as well as untreated DM25 and compared.

2.2.11. Polymerase chain reaction

2.2.11.1. Primers

Primers were ordered from Integrated DNA Technologies Inc. (Custom Science, Auckland, NZ) or Invitrogen (Life Technologies, Carlsbad, CA, USA), made up to stock solutions of 100 µM with DNA-free water and diluted to working concentrations of 10 µM when required. Stock and working

concentrations aliquots were stored in a -20°C freezer. Primers used in the study are listed in **Appendix Table 7.1.**

2.2.11.2. Reactions

Polymerase chain reactions (PCRs) were performed in 25 µL volumes, unless otherwise specified. All PCRs were performed with primers listed in **Appendix Table 7.1** and with either Platinum® Taq or Platinum® High Fidelity *Taq* (Invitrogen) enzyme. The reaction mixtures are indicated in **Table 2.3**. The following PCR protocol was used for cycles run with Platinum® Taq or Platinum®: initial denaturation at 94°C for 1 minute followed by 30 cycles of denaturation at 94°C, a 30 seconds annealing step at 2°C less than the lowest melting temperature of the primer pair used in each reaction, followed by a 72°C extension step for 1 minute per kb of product length. A final extension step at 72°C for 5 minutes completed the reaction before holding the samples at 16°C. Where Platinum® High Fidelity Taq was required for proofreading ability, the reaction composition was similar to **Table 2.3** except 10× High Fidelity PCR buffer (600 mM Tris-SO_4, pH 8.9, 180 mM $(NH_4)_2SO_4$) and 50 mM $MgSO_4$ was used in the place of Platinum buffer and 50 mM $MgCl_2$, and extension temperatures were dropped to 68°C during cycles.

Table 2.3: PCR reaction composition for PCR using Platinum® Taq

Component	Volume	Final concentration
10× PCR buffer (200 mM Tris-HCl, pH 8.4, 500 mM KCl)	2.5 µL	1×
50 mM $MgCl_2$	0.75 µL	1.5 mM
10 mM dNTP mixture	0.5 µL	0.2 mM
10 µM Primers (each)	0.5 µL	0.25 µM
Platinum ® Taq (5U/µL)	0.1 µL	1 U
MilliQ H_2O	with above, to 24 µL	-
Template DNA	1 µL	100 – 200 ng

2.2.11.3. Agarose gel electrophoresis

Agarose gels, 1% (wt/vol), were made using UltraPure agarose (Oxoid) in 1× TAE buffer (**Section 2.1.4.3**) containing 1× final concentration of SYBR Safe DNA Gel Stain (Life Technologies, Carlsbad,

USA). DNA samples were mixed with 10×BlueJuice™ Gel loading buffer (Invitrogen) (typically 4 µL of product and 1 µL of dye) and loaded into the wells. DNA size standards of 100 bps or 1 kb+ markers (Life Technologies) were loaded into the first lane. Gels were run in 1× TAE buffer in either a Horizon Gel Tank or Liberty Fast Tank (6MGel, Fast Agarose System, BioKeyston Co., CA, USA) at 100 V/cm for 30 min. Bands were visualised using UV trans-illumination and photographed using a Nikon D700 camera with Kodak Gel Logic 200 Imaging System software (Eastman Kodak, NY, USA).

2.2.11.4. PCR purification

PCR products were cleaned with a QIAquick PCR Purification Kit (Qiagen) or exoSAP-IT (Affymetrix Inc., Santa Clara, CA, USA), as per the manufacturers protocol.

2.2.11.5. Colony PCR

Individual colonies were picked from plates using sterile pipette tips and re-suspended in 10 µL of MilliQ water. Cell suspensions were used as templates for PCR reactions (**Section 2.2.11.2**).

2.2.12. DNA extractions

DNA was extracted from *E. coli* cultures using three different methods: A Genomic-tip 100/G kit (Qiagen) following the manufacturer's protocol, or, a NEXTprep™- Bacterial DNA Isolation Kit (Bioo Scientific Corp, Austin, TX, USA) following the manufacturer's manual, or using a phenol-based extraction method (**Section 2.2.12.1**)

2.2.12.1. DNA extractions using phenol:chloroform

DNA was extracted from liquid cultures of *E. coli* using a phenol-based extraction method (156). Stationary phase cultures in 500 mL volumes were centrifuged for 10 min at 8,000 *g*, and the media discarded. The cell pellets were processed immediately, or frozen with liquid N_2 and DNA extracted the following day. Pellets were re-suspended in 5 mL lysis buffer (**Section 2.1.4.13**), centrifuged for 10 min at 8,000 *g* and the supernatant discarded. Pellets were re-suspended in 350 µL of lysis buffer containing 5 µL of 100 mg/mL lysozyme and 5 µL of 10 mg/mL RNAse and incubated for 1 hr at 37°C with 150 rpm shaking. One millilitre of 0.5 M EDTA and 1 mL of 20% (wt/vol) SDS were added to the lysates and incubated for 1 hr at 65°C. Proteinase K (Sigma-Aldrich), 25 µL of a 5 mg/mL stock, was added and incubated overnight at 65°C. The following day, an equal volume of phenol:chloroform:isoamyl alcohol 25:24:1 (vol/vol/vol) was added to each sample, mixed by vortexing and left for 3 min. Samples were centrifuged for 20 min at 8,000 *g* at 4°C. The aqueous layer was transferred to a new tube and the process repeated. An equal volume

of chloroform:isoamyl alcohol 24:1 (vol/vol) was added to each sample, mixed, and left for 3 min. Samples were centrifuged for 20 min at 8,000 *g* at 4°C. The aqueous layer was transferred to a new tube. Three molar sodium acetate (pH 5.5) was added to each sample at 1/10th the volume of aqueous layer, followed by 7 mL of ice-cold isopropanol, and incubated at -20°C for 2 hrs. Samples were centrifuged for 20 min at 8,000 *g* at 4°C, and the supernatant discarded. DNA pellets were washed twice with 70% (vol/vol) ethanol and centrifuged for 5 min at 8,000 *g* at 4°C after each wash. After discarding the supernatant in the final wash step, the DNA pellets were air-dried for 15 min, then re-suspended in 100 µL MilliQ water at 65°C.

2.2.13. DNA quantification

DNA was quantified by using at least one of the following methods: 1) Spectrophotometrically using a NanoDrop ND-1000 UV-Vis spectrophotometer (NanoDrop Technologies, Wilmington, DE, USA). DNA purity was determined by analysis of the A_{260}/A_{280} and A_{260}/A_{230} ratios, which for pure DNA samples, should be ~1.8 and ~2.0, respectively. 2) Fluorometrically, using a Qubit2.0 Fluorometer (Invitrogen) with Quant-iT dsDNA Broad-Range (BR) Assay kit for high DNA concentrations (2 – 1,000 ng), and Quant-iT DNA High-Sensitivity (HS) Assay kit (Life Technologies) for low DNA concentrations (0.2 – 100 ng), as per the manufacturer's instructions. 3) Quantifying bands intensity from an electrophoresis gel image using a High DNA Mass Ladder (Invitrogen), with band densitometry calculated using Kodak ID 1.6 software.

2.2.14. Whole genome sequencing

DNA from the evolved lineages for whole genome re-sequencing was extracted from either whole populations or from clones isolated from the populations. Individual clones were isolated by streaking the population on either aerobic or anaerobic DM plates (consistent with the lineage environment), and randomly selecting an isolated clone for single-colony streaking. Clones were streaked a total of three times, after which a colony was chosen at random to inoculate aerobic or anaerobic DM25 (as appropriate) and incubated to generate a pure culture. Aliquots of the culture were stored (**Section 2.2.5.5**), and genomic DNA was extracted (**Section 2.2.7**). DNA from all clones at 2,000 generations was extracted with a Genomic-tip 100/G kit. DNA from all clones at 4,000 generations of the aerobic and fluctuating lineages was extracted using a NEXTprep™ - Bacterial Isolation Kit. DNA from all anaerobic populations and anaerobic clones at 4,000 generation were extracted using the phenol-based extraction method (**Section 2.2.12.1**).

2.2.15. Illumina sequencing

High quality genomic DNA was commercially sequenced by the Beijing Genomics Institute Co. Ltd. (Shenzhen, China). Sequencing was performed on an Illumina HiSeq 2000 instrument. For the 2,000 generation genomes, the Illumina Paired-End (PE) Library Preparation Kit v2 was used to construct 500 bp paired-end inserts. Paired-end sequencing with 90 bp reads generated 1 Gb of sequencing data for each sample, allowing for over 200× fold coverage of each genome. For 4,000 anaerobic genomes Illumina Mate Pair (MP) Preparation Kit v2 was used to construct 2 kb insert libraries (**Table 2.4**). Mate pair sequencing with 90 bp reads generated 700-900 Mb of clean data for each 4,000 generation sample. Read data for each sample was received in the form of two FASTQ files for each genome, representing forward and reverse reads. FASTQ files are text-based files containing the nucleotide sequence of a read and additional quality scores of each base within the read. The FastQC program, a quality control tool for FASTQ files (157), was used to assess the quality of all reads for each genome.

Table 2.4: Sample and library construction details for the genomes sequenced in this study

Samples	Generation	Library construction type	Insert size	Fold coverage
Ancestral clone	0	90 bp PE libraries	500 bp	200×
7 Aerobic clones	2,000	90 bp PE libraries	500 bp	200×
6 Anaerobic clones	2,000	90 bp PE libraries	500 bp	200×
7 Fluctuating clones	2,000	90 bp PE libraries	500 bp	200×
7 Aerobic clones	4,000	90 bp PE libraries	500 bp	200×
7 Fluctuating clones	4,000	90 bp PE libraries	500 bp	200×
7 Anaerobic clones	4,000	90 bp MP libraries	2,000 bp	100×
3 Anaerobic clones	2,000	90 bp MP libraries	2,000 bp	100×
4 Anaerobic clones	1,000	90 bp MP libraries	2,000 bp	100×
Rif^R clone	0	90 bp MP libraries	2,000 bp	100×

2.2.16. Mutation identification

The genome sequence of *E.coli* REL4536 is not publicly available, however it's ancestral genome sequence, *E.coli* REL606 is available (GenBank accession number NC_012967.1). The GenBank file of REL606 was downloaded from the NCBI database

(ftp://ftp.ncbi.nih.gov/genomes/Bacteria/Escherichia_coli_B_REL606_uid58803) and manually edited using Artemis (158) and the European Molecular Biology Open Source Suite (EMBOSS) application (159) to incorporate the 28 mutational differences known between the genomes of REL606 and REL4536 as reported by Barrick *et al.* (52). The resulting GenBank file was used as the reference REL4536 genome sequence for all downstream analyses.

To identify the mutations that occurred in the evolved strains the mutation detection pipeline breseq v.014 was used (52). Breseq is a computational pipeline used to identify mutations relative to a reference sequence in short-read DNA re-sequencing data. It uses reference-based alignment approaches to predict mutations in a sample relative to a previously sequenced genome. Breseq is ideal for microbial genomes and re-sequenced samples that are only slightly diverged from a reference sequence. Sequencing reads were aligned to the manually created REL4536 genome using breseq and the mutations observed between the reference and query genome sequences were analysed.

Mutated genes were grouped and ordered in Excel v14.0 (Microsoft Corporation, Redmond, WA, USA). To determine the putative function of genes mutated in this study, EcoCyc (160) and DAVID v6.7 (161) were used. EcoCyc is a database containing literature based curation of the entire *E. coli* K-12 MG1655 genome. DAVID or Database for Annotation, Visualisation and Integrated Discovery uses existing biological databases and analytic tools to extract biological meanings from large gene or protein lists. Mutated genes were grouped by similarities in function, to determine which biological processes were affected based on KEGG and Gene Ontology annotation.

2.2.17. Bioinformatic resources and software

Bioinformatic resources and software used in this study are listed in **Table 2.5**

Table 2.5: Bioinformatic resources and software used in this study.

Resource	Application	Reference
Artemis	Genome sequence viewing and analysis	Rutherford *et al.* 2000 (158)
Basic local alignment search tool (BLAST)	Heuristic alignment of query sequence to sequence database	Altschul *et al.* 1990 (162)
Geneious	DNA alignment, assembly and analysis software	Drummond *et al.* 2011 (163)
FastQC	Quality control of next generation sequencing	Andrews *et al.* 2010 (157)
Open CFU	Colony counting software	http://opencfu.sourceforge.net/
breseq	Bacterial genome reference based assembly	Barrick *et al.* 2009 (52)

2.2.18. Statistical analysis

Student's t-tests of unequal variances were used for the statistical analysis of competitive fitness data (101). When the fitness of strains was expected to be equal to 1, one sample, one-tailed test was used with the null hypothesis stating fitness data was equal to 1. All other tests performed, were two sample tests with the null hypothesis stating that the two sample sets were equal. For the comparison of relative frequencies of TCM and SCM ratios in reciprocal invasion experiments single factor analysis of variance (ANOVA) was used as in (71), with the null hypothesis stating that there was no difference between the relative frequencies. Statistical evaluation of data was performed on Microsoft Office Excel v14.0.

Chapter Three : Adaptation to aerobic and anaerobic environments

3.1. Introduction

Adaptive evolution is the outcome of natural selection – a process by which populations become better suited to the environment. In 1988, Prof. Richard Lenski initiated an experiment in which he propagated *E. coli* via daily transfer in batch culture. The study, which has become known as the "long-term experimental evolution" (LTEE) experiment, has allowed direct observation of the adaptive processes through time. In his landmark study published in 1991, Prof. Lenski used a clone obtained from *E. coli* B, which became known as REL606 (6). Pure cultures of clone REL606 bacterial cells were divided into replicates, and propagated for many generations. This organism was subjected to LTEE in the minimal glucose media known as DM25 for 2,000 generations and its evolution in this environment continues to the present day. During the course of LTEE, adaptation is measured by monitoring the relative fitness of evolved populations through competitive fitness assays. Combining LTEE and competitive fitness assays allows us to understand how an organism adapts to its environment.

The aim of this chapter is to develop a LTEE study to observe adaptation of *E. coli* to aerobic and anaerobic conditions and to monitor relative fitness through competition assays. *E. coli* has the extraordinary metabolic capability to produce energy in the presence or absence of oxygen (14). Thus, from one cell, adaptation to two distinct conditions, each with their unique stresses and benefits, can be compared over time. Other advantages of using *E. coli* as a model organism in LTEE include its short generation times and ability to grow to high densities in small culture volumes. Furthermore, the ability to store cultures in suspended animation and to revive them at a later date for further analysis, and more recently the availability of a fully annotated genome sequence, make *E. coli* a popular model organism for LTEE studies. The present study will use a modification of the experimental system developed by Lenski *et al.* 1991 (6). *E. coli* REL4536 will be used as the model organism for this thesis. REL4536 is a 10,000 generation descendent of *E. coli* REL606, used in Prof. Lenski's studies. This strain was selected to minimise adaptations to DM25 minimal glucose media (62, 164, 165) as opposed to adaptation to the specific treatments of

growth under strict aerobic or strict anaerobic conditions. By treating DM25 via removal of dissolved oxygen and maintaining cultures in strict anaerobic conditions, the experimental design used in this thesis allowed us to determine the adaptation of *E. coli* REL4536 to the anaerobic environment in parallel to evolution in aerobic environment. Furthermore, a treatment that fluctuated between aerobic and anaerobic conditions allowed us to examine how *E. coli* REL4536 may adapt to non-constant exposure to both aerobic and anaerobic environments.

3.2. Objectives

The objectives of this chapter are:

1. To determine the growth dynamics of *E. coli* REL4536 in batch cultures in aerobic and anaerobic environments.
2. To establish long-term adaptive lineages in aerobic, anaerobic and fluctuating treatments.
3. To assess the relative fitness of the lineages as they adapt to their treatments.

3.3. Results and discussion

3.3.1. Growth dynamics of REL4536 in batch cultures in aerobic and anaerobic environments

For the LTEE study described in this thesis, a large number of cultures needed to be maintained simultaneously. However, working within an anaerobic work station posed technical challenges. Access to and from the anaerobic work station was through a small airlock system which limited the number of cultures in large volumes, such as the 10 mL cultures in 50 mL flask format used by Lenski *et al.* 1991 (6), that could be brought in and out of the anaerobic work station together. As such, smaller culture volumes were required to maintain a high number of replicates. Furthermore, smaller culture volumes allowed for the use of compact anaerobic gas canisters during incubation. Therefore, it was deemed practical for the cultures to be maintained in smaller 1 mL volumes in 24-well tissue culture plates (**Section 2.1.1**).

The growth of *E. coli* B strains has been studied extensively in DM25 in aerobic conditions (102, 166, 167). However, the growth rates of *E. coli* B strains in DM25 under anaerobic conditions were unknown. Furthermore, different growth vessels and culture volumes are known to affect the growth dynamics of cultures, and potentially, their subsequent evolution (168, 169). Thus, the growth dynamics of REL4536 under both aerobic and anaerobic conditions in 1 mL total volumes

were investigated (as per **Section 2.2.3**) to determine the time period between sub-culturing events. Estimates of cell growth were initially based on optical density measurements. However, to determine viable cell counts, colony forming units (CFU) per mL of growing cultures were determined, since these would indicate the amount of inoculum that would be sub-cultured in the proposed serial transfer regime. Results are displayed in **Figure 3.1.**

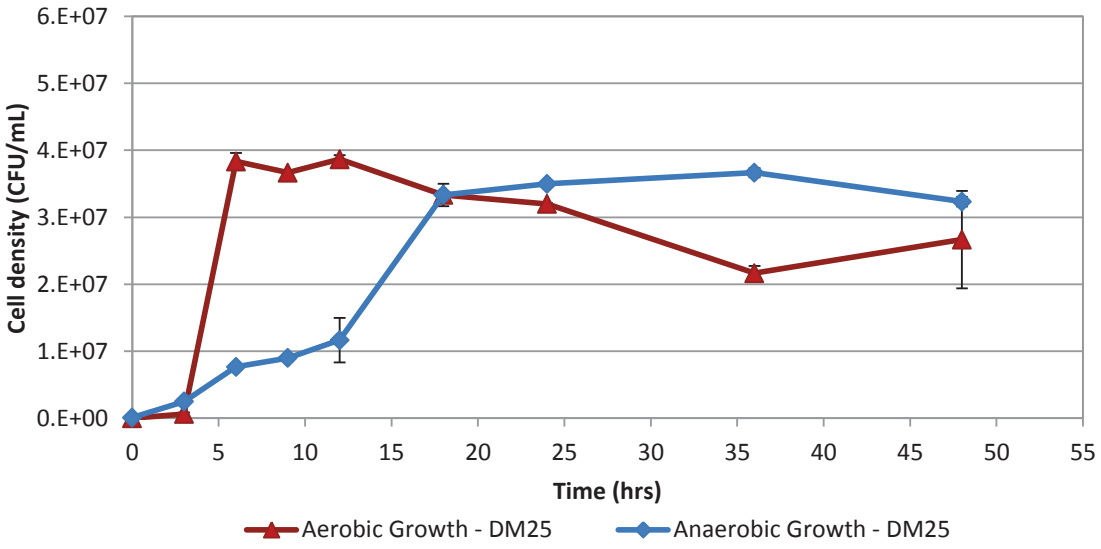

Figure 3.1: Growth curves of *E. coli* REL4536 in aerobic and anaerobic conditions. Stationary phase is reached within 24 hours for both conditions. Data points represent the mean values from three biological replicates, with error bars representing standard error of the mean.

Under aerobic conditions, a peak cell density of 4×10^7 CFU/mL was reached within six hours, presumably the point at which all glucose in the media has been metabolised. After this, there was a steady decline in cell density to 2×10^7 CFU/mL after 36 hours of growth. Growth conditions such as the media, inoculum to volume proportion (1:100), temperature and speed of shaking were all maintained similar to those of Lenski *et al.* 1991, who reported a stationary phase density of 5×10^7 CFU/mL within eight hours of growth (6). Growth dynamics of cultures in this thesis were in line with those reported by Lenski *et al.*, despite reduced culture volumes (1 mL versus 10 mL), different growth vessels (24-well plate versus flask) and different oxygen content as final cell densities reported for both aerobic and anaerobic growth respectively (~3.5×10^7 CFU/mL and 4.0×10^7 CFU/mL).

While the maximum cell densities were similar between aerobic and anaerobic environments, under anaerobic conditions, the growth rate was considerably slower (**Figure 3.1**). During anaerobic growth, a peak cell density of ~3.5 × 10^7 CFU/mL was reached within 20 hr. An increased lag phase and reduced rate of exponential growth delayed the time at which stationary phase was reached within the anaerobic environment. After the maximum cell density was reached, it remained relatively constant until the 48 hr time point. As stationary phase was reached within 24 hours for both aerobic and anaerobically grown treatments in the 24-well plate format, it was determined that sub-culturing of the long-term adaptive lineages would take place by batch culture transfer every 24 hr in each of the growth treatments.

Different growth rates of REL4536 in aerobic and anaerobic environments meant that during the 24 hour growth cycle, the aerobic cultures were predominantly in stationary phase, while anaerobic populations predominantly remained in lag and exponential phase. The different growth rates between aerobic and anaerobic conditions are likely to affect the mechanisms of adaptation that arise in each environment (125, 166, 170). In the aerobic environment, cultures under prolonged periods of stationary phase are known to have a high incidence of growth advantage in stationary phase (GASP) mutations (125, 171). These adaptations are unlikely to be found in cultures exposed to the anaerobic environment due to their limited time in stationary phase during the 24 hour growth cycle. During the sub-culturing regime, the metabolic machinery of the cell is required to adjust from a period of famine to feast (172, 173). This adjustment takes place during the lag phase of growth which is often found to be extended under anaerobic growth conditions (174-176). Thus, in the anaerobic environment, we are likely to see mutations that may allow the bacteria to reduce the duration of lag phase, thereby attaining higher population densities earlier (166).

3.3.1.1. Citrate utilisation of REL4536 in the anaerobic environment

We had expected that the differences in energy production (aerobic respiration as compared to anaerobic fermentation) in the two environments would lead to noticeable differences in the maximum cell densities attained in each environment. As illustrated in **Figure 3.1**, *E. coli* REL4536 reached similar cell densities when grown aerobically and anaerobically in DM25 (~4.0 and ~3.5 × 10^7 CFU/mL, respectively). Different growth rates in aerobic and anaerobic environments can be attributed to the use of different metabolic networks by *E. coli* to obtain energy within each environment. In the aerobic environment, glucose is the sole energy source supplied in DM25. In

this environment, the tricarboxylic acid (TCA) cycle is capable of complete oxidation of glucose, yielding 36 ATP molecules per molecule of glucose (177). Under anaerobic conditions, however, glucose is partially oxidised through the lower energy yielding fermentation network, yielding two ATP molecules per unit glucose. Although the rate of growth between the aerobic and anaerobic environments differed considerably, the similar cell densities achieved in both conditions suggests that during anaerobic growth, REL4536 metabolised an additional component of the media.

Under anaerobic conditions, *E. coli* is able to metabolise citrate (62, 178, 179). Citrate is present in DM25 at a high concentration of 1,700 µM. As such, citrate utilisation by REL4536 in the anaerobic environment was investigated. Aerobic and anaerobic growth courses of REL4536 were performed in DM0 and DM25 media (as per **Section 2.2.3**). DM0 contained no supplementary glucose while DM25 contained 139 µM of glucose, with both DM0 and DM25 each containing 1,700 µM of citrate. Results of the four growth courses are outlined in **Figure 3.2.**

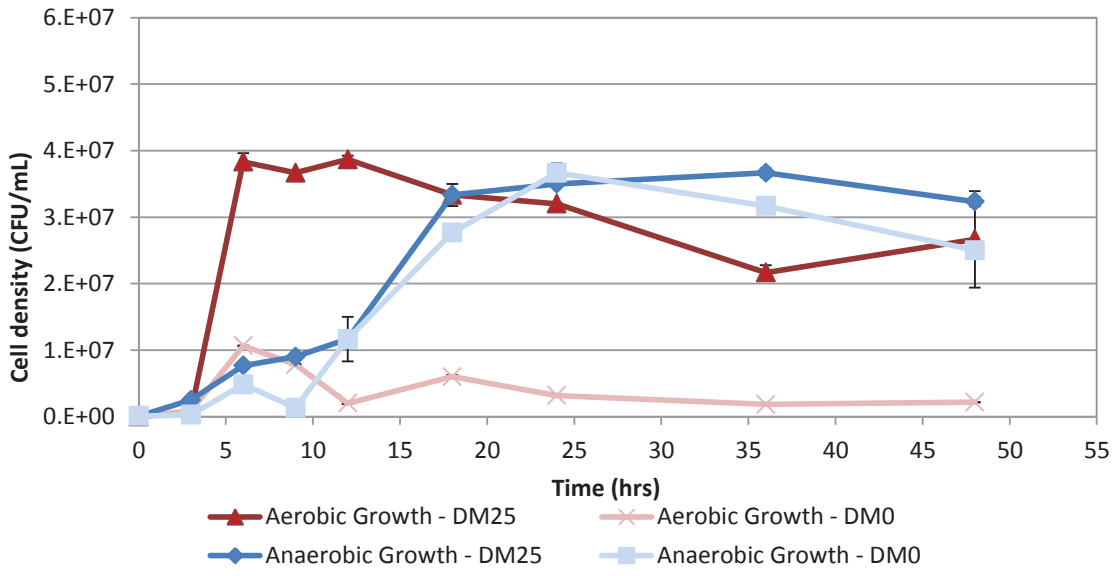

Figure 3.2: Growth of *E. coli* REL4536 in DM0 and DM25 in aerobic and anaerobic conditions. Data points represented the mean values of three biological replicates, and error bars represent the standard errors of the mean.

Under aerobic conditions REL4536 was positive for growth in DM25, reporting a cell density of ~4.0×10^7 CFU/mL which was reached within 6 hours of growth. In contrast, only a slight increase in cell density was observed in the inoculated DM0 culture at the 6 hour time point. It is not certain what this increase is a result of. Limited growth in aerobic DM0 media suggests that

aerobic growth of REL4536 in DM25 is largely dependent on glucose. Under anaerobic conditions however, the growth dynamics of REL4536 in both DM0 and DM25 were similar, with cultures reaching peak densities of ~3.5×10^7 CFU/mL within 24 hours of growth. The absence of glucose in DM0 did not limit growth of REL4536 in the anaerobic environment. The growth experiment described here showed that glucose alone was not solely responsible for anaerobic growth.

E. coli is unlike other enterobacteria such as *Salmonella* and *Klebsiella* in that it cannot metabolise citrate under aerobic conditions (180). The inability to metabolise citrate is a key attribute of *E. coli*, and is due to a non-functioning *citT* gene. The *citT* gene encodes a citrate-succinate antiporter which does not function in aerobic conditions (181-183). However, the non-functioning *citT* gene is overcome in the anaerobic environment, where in the presence of co-substrates such as L-lactate, L-malate, fumarate, pyruvate, lactose or glucose, *E. coli* can metabolise citrate (179). These co-substrates act as the electron donors for the conversion of citrate to acetate and succinate (179, 184) despite the non-functioning *citT* transport system. The only co-substrate that is available in DM25 is glucose, and thus, available citrate is oxidised to acetate and succinate. During anaerobic growth, both citrate and glucose are degraded simultaneously (178, 185). Once glucose has become exhausted, citrate utilisation also stops (179) as the cell requires the reducing power of the co-substrate to reduce available citrate.

The evolution of citrate utilisation among *E. coli* in LTEE was observed in one population (the Ara+3 line) of Prof. Lenski's LTEE study, which enabled the aerobic utilisation of citrate by generation 31,500 (62). The *cit*+ strain evolved by a tandem duplication that captured an aerobically expressed promoter of the downstream *rna* gene for the expression of the previously inactive *citT* gene (118). However, citrate utilisation observed by Blount *et al.* (2008) is distinct from the metabolism of citrate observed in this thesis. In Blount *et al.* (2008), citrate utilisation occurred in the aerobic environment while in this thesis citrate utilisation only occurred in the anaerobic environment. Cells containing the aerobically expressed *citT* gene reported optical density values at 420 nm of ~0.25, as compared to ~0.04 prior to the promoter capture mutation event (118). Aerobically, citrate is metabolised through the TCA cycle, capable of generating much higher yields of ATP than from fermentation in the anaerobic environment.

The similar cell densities reported in this study in the two environments may be accounted for by the ability of *E. coli* to anaerobically utilise the high concentration of citrate within the anaerobic media. Assuming the complete breakdown of glucose, where 1 mol of glucose yields 36 mol ATP,

energy production in the aerobic environment is likely to produce 5,004 µM ATP. In the anaerobic condition, again assuming the complete breakdown of glucose, where 1 mol of glucose yields 2 mol ATP and further assuming 1 mol of citrate produces 1.3 mol ATP (184), the combined energy production in the anaerobic environment is likely to be near 2,488 µM ATP. Thus, the difference in energy generation in the two environments is still unexplained.

3.3.2. Establishment of long-term lineages in aerobic, anaerobic and fluctuating treatments

Long-term lineages of REL4536 were established after confirming the growth characteristics of REL4536 in batch culture. Lineages were seeded from an ancestral aerobic stock culture derived from a single colony of REL4536 reaching a maximum cell density of 4.5×10^7 CFU/mL (as described in **Section 2.2.5**). The identity of the ancestral culture was confirmed as REL4536 by Gram stain (**Section 2.2.4**), phage contamination tests (**Section 2.2.5.4**), and sequencing of the 16S rRNA gene using fD1 and rD1 primers (**Section 2.2.9**). Aliquots of the ancestral stock culture were stored at -80^{o}C for downstream experimental work (**Section 2.2.5.5**). In total, 21 replicate lineages were established in the three treatments of this study: aerobic (AE), anaerobic (AN) and a treatment that fluctuated daily from aerobic and anaerobic environments (FL). These lineages were maintained as per **Section 2.2.5.2**, similar to Lenski *et al.* 1994 (6). Lineage adaption took place during the course of the LTEE (**Section 2.2.5.2**) for 4,000 bacterial generations. During this time, contamination checks (**Section 2.2.5.4**) and storage were performed routinely (**Section 2.2.5.5**) on a bi-weekly basis (approximately every 100 generations). Cell densities were monitored over the course of the 4,000 generations (**Figure 3.3**). It is noted that during the experiment, the blank well of the anaerobic plate had become contaminated three times, ca. generations 2,133, 3,199 and 3,600, as seen by *E. coli* colonies during CFU counts. CFU counts of other lineages on the same plate were negative for non-*E. coli* cross contamination. However, cultures were not restarted from the last frozen stock as during a pilot study conducted before this LTEE, frequent resuscitation of anaerobic cultures were not viable in the longterm for unknown reasons. As such, the independent cultivation of the individual AN lineages could not be guaranteed from ca. 2,000 generations. It is thought that the cross contamination events took place in the anaerobic lineages inside the airlock of the anaerobic chamber during equilibration to the ananerobic atmosphere. During this process, the airlock and it's contents undergo six rounds of gas exchange followed by a vacuum cycle, generating 20 inches of mercury (0.69 kg/square cm) of negative pressure. It is thought that aerosols, generated from cultures during this time, were

likley to have cross contaminated into the replicate wells within the anaerobic 24 well plate. The implications of this result are discussed further in **Section 4.3.3.3**.

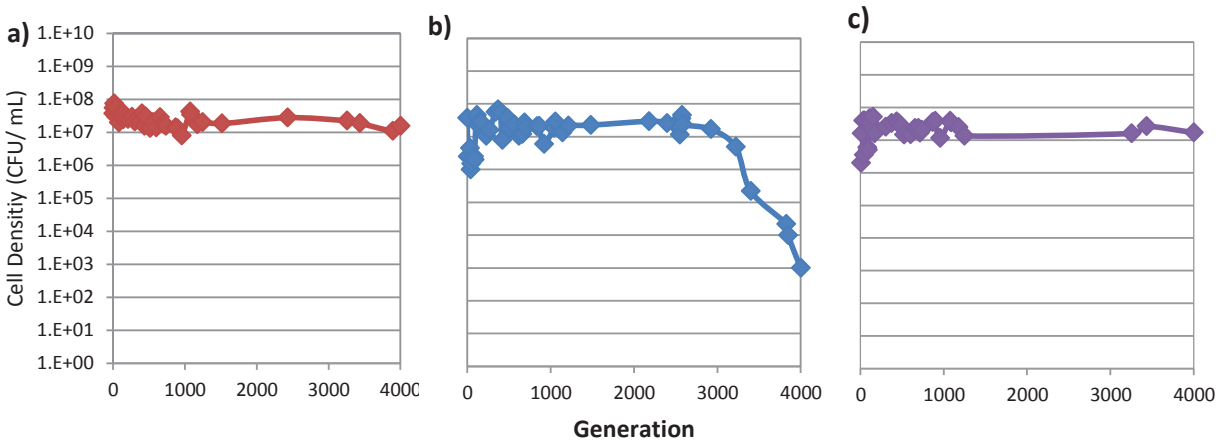

Figure 3.3: Cell densities in all treatments over 4,000 generations. a) Represents aerobic lineage AE2 b) represents anaerobic lineage AN3 and c) Represents fluctuating lineages FL6. The data points display mean CFU/mL of each lineage. Each data-point is the average of counts from two technical replicates.

The cell densities of both the aerobic and fluctuating lineages (see data of representative lineages in **Figure 3.3 a** and **c**) remained relatively constant throughout the 4,000 generations. However, for the anaerobic lineages, there was an initial decrease in the total cell population within the first 50 generations. This lasted for approximately 100 generations. After 150 generations, an increase in cell density of anaerobic lineages to levels more in line with the aerobic and fluctuating lineages was found. This may have been the result of a beneficial mutation fixing early in the course of the experiment. Overall the cell yields in the anaerobic population were more variable than those of the aerobic and fluctuating lineages and an example of data from a single lineage, AN3, is shown in **Figure 3.3 b**. Also of note is the decline in cell density in the anaerobic populations around the 3,000 generation time-point. This was consistent among all anaerobic lineages but with very slight differences in the time at which they started to decline (discussed further in **Section 3.4.3**).

3.3.3. Assessment of fitness of evolving populations

3.3.3.1. Development of neutrally marked strains for competitive fitness assays

3.3.3.1.1. Neutrality of Ara+ marked strains under aerobic and anaerobic conditions

To measure the fitness of evolved populations using competition assays, phenotypic markers are used to distinguish between the competing evolved versus ancestral cell types. The marker most commonly used is for arabinose utilisation (186), which confers the ability to utilise L-arabinose and allows for the identification of mutants based on the colour of colonies on TAra indicator agar plates, Ara+ are white and Ara- are red. REL4536 contains a G → A point mutation in the *araA* gene. This substitutes a glycine for an aspartic acid at position 92 of the AraA polypeptide, which disrupts the protein's ability to catalyse the first step in the degradation of L-arabinose. Spontaneous mutants that possess the Ara+ phenotype were generated from the ancestral culture (**Section 2.2.6.1**). In total, six spontaneous Ara+ mutant colonies were isolated, streaked to single colonies for purification, and their phenotypes confirmed on TAra agar plates as pure white colonies (**Section 2.2.6.1**). Pure cultures were subsequently grown and stored at -85°C.

For the marked reference strains to be representative of the ancestral population, the presence of the marker must not affect its fitness. The marker is said to be neutral if it has no effect on the fitness of the reference strain when compared to the ancestor. To test for this, each of the six Ara+ mutants were competed with the ancestral cell, first in the aerobic environment (as per **Section 2.2.6.1**) with the results displayed in **Table 3.1**. Neutrality of the marker is found if the relative fitness values of mutants have values close to one, indicating that their underlying mutations are selectively neutral under the assay conditions.

Table 3.1: Relative fitness of the six Ara+ mutants compared to REL4536 under aerobic conditions.

Spontaneous mutant	Relative fitness (ω)[a]	Standard error of mean
Ara+1 [b]	0.988	0.067
Ara+2 [b]	1.037	0.024
Ara+3	1.088	0.131
Ara+4 [b]	1.037	0.064
Ara+5	0.950	0.023
Ara+6	1.049	0.106

[a] Relative fitness was calculated from four independent competitions

[b] Spontaneous Ara+ mutants with relative fitness closest to one in the aerobic environment

As fitness was also required to be determined under anaerobic growth conditions, the three Ara+ mutants showing relative fitness values closest to one (Ara+1, Ara+2 and Ara+4) were further tested under anaerobic conditions. As can be seen from **Table 3.2**, none of the three Ara+ mutants that had displayed neutral fitness under the aerobic competitions displayed neutral fitness when competed anaerobically.

Table 3.2: Relative fitness of three Ara$^+$ mutants compared to REL4536 under anaerobic conditions.

Spontaneous mutant	Relative fitness (ω)[a]	Standard error of mean
Ara+1	0.328	0.072
Ara+2	0	0
Ara+4	0.612	0.032

[a] Relative fitness was calculated from four independent competitions

The finding that the arabinose marker is not neutral in the anaerobic environment was unexpected. REL4536 has a mutation in the *araA* gene which renders it unable to metabolise the sugar L-arabinose. The six Ara+ mutants generated on Day 0 of the LTEE are spontaneous revertants of REL4536 (**Section 2.2.7**) which have obtained the ability to utilise L-arabinose. The *araA* gene codes for L-arabinose isomerase which catalyses the isomerisation of L-arabinose to L-ribulose (187). Why the activation of this pathway would lead to such a dramatic decline in competitive fitness as compared to the ara- in the absence of arabinose under anaerobic conditions remains uncertain. Potentially, other mutations, some of which may have been deleterious in the anaerobic environment, may have hitch-hiked with the mutation, conferring arabinose utilisation during the generation of the mutant.

These results indicated that arabinose utilisation is not a suitable marker for assessing fitness in both aerobic and anaerobic environments, and that the development of alternative marked strains was required. During this thesis an attempt was made to engineer strains with a *lacZ* deletion (that could be detected by utilisation of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal), resulting in blue colony formation), and 16S rRNA gene mutations that could be distinguished by fluorescence *in situ* hybridisation (FISH) with detection on a fluorescence activated cell sorter (FACS). FACS-based methods (188-192) to distinguish and enumerate REL4536 cells, using allelic variant *16S* rRNA gene FISH probes, were successful (results not shown). However, as both the *lacZ*

and *16S* rRNA gene strategy required the genetic manipulation of REL4536, genetic constructs were engineered using a variant of the lambda red recombination system (193, 194). Despite numerous attempts, the rate of chromosomal integration of the construct was poor. All resulting transformants grew extremely slowly, and were deemed impractical to be used further (results not shown), thus this approach was abandoned. The use of spontaneous antibiotic resistant mutants to assess fitness was investigated.

3.3.3.1.2. Neutrality of antibiotic resistance markers under aerobic and anaerobic conditions

Spontaneous mutants of REL4536, with resistance to various antibiotics with different modes of action, were screened for neutral fitness against REL4536 in aerobic and anaerobic competitive fitness assays. The following four antibiotics were tested in this study: i) Chloramphenicol, which inhibits protein synthesis by inhibiting the peptidyl transferase activity of ribosomes, preventing peptide bond formation during protein chain elongation (195). ii) Kanamycin, which interacts with 30S ribosomes, inhibiting protein synthesis (196). iii) Nalidixic acid, which prevents DNA from unwinding and therefore duplicating, through inhibition of bacterial DNA gyrase (197). iv) Rifampicin, which inhibits bacterial transcription through inhibition of DNA-dependent RNA polymerases. Spontaneous resistant mutants were isolated on DM agar plates containing one of the following antibiotics; chloramphenicol, kanamycin, nalidixic acid and rifampicin at working concentrations of 34, 50, 30 and 100 µg/mL, respectively (**Section 2.2.6.2**). This generated 30, 45, 21 and seven colonies per plate. Eight colonies from the chloramphenicol, kanamycin and nalidixic acid plates, and all seven of the colonies on rifampicin plates were picked at random and streaked sequentially on appropriate antibiotic plates to purify. Four antibiotic resistant mutants, from both the nalidixic acid and rifampicin antibiotic selection were competed against the ancestor in both aerobic and anaerobic environments (as per **Section 2.2.8**). The results of fitness assays are displayed in **Table 3.3**. Both types of antibiotic resistant marked strains display more neutral fitness values than that of the Ara+ mutants, with mutant Rif^r2 being selected for further study.

Table 3.3: Aerobic and anaerobic competitive fitness assay results for nalidixic acid resistant mutants and rifampicin resistant mutants.

Mutant ID	Resistant Antibiotic	Aerobic Assay		Anaerobic Assay	
		Relative fitness (ω)[a]	SEM[b]	Relative fitness (ω)[a]	SEM[b]
Nal^{r}1	Nalidixic acid	1.1619	0.1553	0.8865	0.0961
Nal^{r}2	Nalidixic acid	0.9034	0.0858	0.8536	0.0981
Nal^{r}3	Nalidixic acid	0.8255	0.0798	0.8338	0.0387
Nal^{r}4	Nalidixic acid	0.8783	0.1116	1.0298	0.1290
Rif^{r}1	Rifampicin	1.0347	0.0512	0.9344	0.0701
Rif^{r}2	Rifampicin	0.9726	0.0596	1.0417	0.0466
Rif^{r}3	Rifampicin	0.9520	0.1046	1.1034	0.0183
Rif^{r}4	Rifampicin	1.0571	0.0323	1.0298	0.1290

[a] Relative fitness (ω) values are calculated as the mean of four biological replicates.

[b] Standard error of the mean

On average, the rifampicin resistant mutants had fitness values closer to one than those of the four nalidixic acid mutants tested (**Table 3.3**). The relative fitness of mutant Rif^{r}2 as compared to REL4536 was closest to neutral in both aerobic and anaerobic assays. To identify the genetic cause of the resistance phenotype and to determine if Rif^{r}2 was isogenic to the ancestral cell, genomic DNA was extracted (**Section 2.2.12.1**) and the complete genome was sequenced (**Section 2.2.12.1**) and compared to the ancestral genome. Bacteria typically evolve resistance to rifampicin through mutations in the *rpoB* gene, which encodes the β subunit of RNA polymerase (198-200). A single A → T transversion in *rpoB* (genome position 4,128,442 bps) was identified, resulting in the substitution of a phenylalanine, a large aromatic amino acid, for an isoleucine, a large hydrophobic residue at amino acid 581 of RpoB. In general, rifampicin resistant mutations map to three distinct clusters in the middle of the *rpoB* gene, which decreases the binding affinity of rifampicin to RNA polymerase (201). The mutation identified in the case of Rif^{r}2 occurred in the second last residue within the conserved RNA polymerase Rpb2 domain 3, just upstream of the DNA-directed RNA polymerase β subunit external 1 domain, previously identified as mutation cluster 2 (198). Resistance to rifampicin has been shown not to confer a loss of fitness in an antibiotic-free environment (202), and this can now be extended to the anaerobic environment.

All competition assays performed in this study were performed with Rif^{r}2 representing the ancestral competitor strain. Furthermore, in the recent LTEE study, Puentes-Téllez *et al.* 2013 concurrently identified rifampicin resistance as a suitable marker for testing fitness in oxygen limited conditions (101). However, the use of the antibiotic resistant marker had its challenges.

The assay developed using the Rif^{r}2 strain was technically demanding, as it required duplicate plating to estimate the number and proportion of antibiotic resistant cells in the population (in contrast to Ara+ based assays where competitors can be directly observed on a single plate). As such, using antibiotic resistance to identify competitor strains resulted in greater statistical variability around fitness estimates. To compensate for this, additional replicate competition assays were undertaken, and a Rif^{r}2 versus ancestor control competition was included with each set of assays.

3.3.3.2. Adaptation to narrow niches

The niche of an organism is the range of conditions under which it can grow or reproduce (88). Exposure of organisms to a narrow niche or constant environment tends to promote the evolution of a specialist, an organism that has developed high fitness in the environment of selection. This generally comes at the expense of fitness in environments other than that to which they have evolved (87, 93, 94, 203). This expense, known as the cost of specialisation, is typically seen when lineages of a high fitness in the environment of selection show a reduction in fitness in alternative environments. In this thesis, lineages were set up such that niche specialisation to both the aerobic and anaerobic environment could be studied. To do this, competitive fitness assays (**Section 2.2.8**) were performed on populations that had evolved under strict aerobic (**Section 3.3.3.2.1**) and strict anaerobic conditions (**Section 3.3.3.2.2**) in both the environment of selection and the alternative environment over the course of the 4,000 generations. Three lineages in each treatment were randomly selected and competition assays were performed at 1,000 generation intervals from generation 1,000 to generation 4,000. To monitor neutrality of the Rif^{r}2 strain, a control competing Rif^{r}2 with the ancestor was included.

3.3.3.2.1. Evolution of the aerobic lineages – general trends

To assess how fitness changed over time in the aerobic lineages, mean relative fitness was averaged over the three aerobic populations tested at random from the seven aerobic lineages; AE2, AE3 and AE7. On average, mean relative fitness of aerobic lineages in the aerobic environment from 1,000 to 4,000 generations is displayed in **Table 3.4 a)**. However, while there is a general trend showing a slight increase in fitness, these increases were not statistically significant. This is likely due to the ancestral REL4536 strain having previously pre-adapted to DM25 under aerobic conditions for 10,000 generations (52, 166).

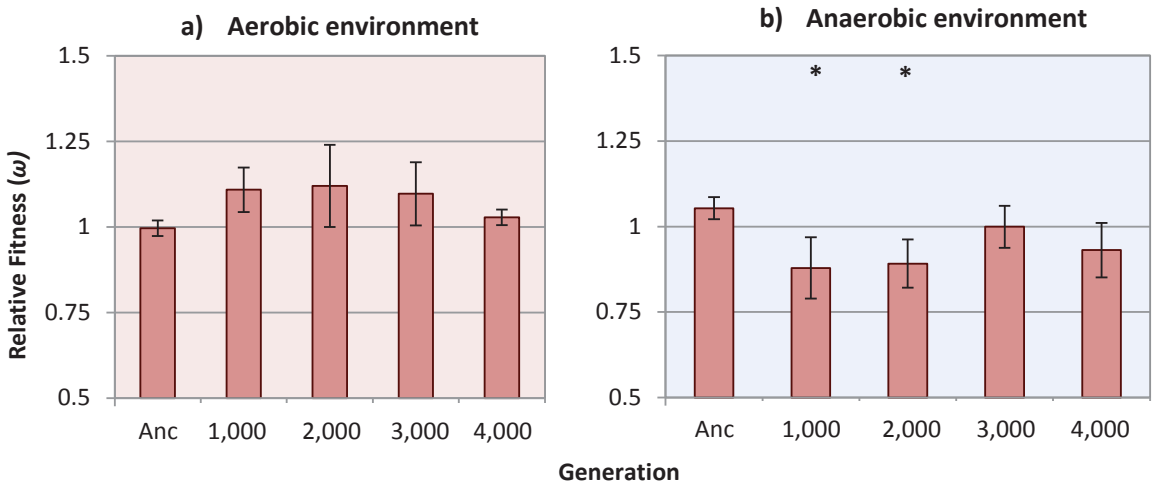

Figure 3.4: Average relative fitness of aerobic lineages over 4,000 generations. Displayed are the mean relative fitness values of three aerobic lineages AE2, AE3 and AE7 as compared to the ancestor performed in a) the aerobic (red background) and b) the anaerobic (blue background) environments. Error bars represent the standard error of the mean.

The ancestral strain used in this thesis was the 10,000 generation descendent in the Ara+1 population from Lenski *et al.,* 1991 (6). Barrick *et al.* (52) studied the adaptation of Prof. Lenski's LTEE strains to minimal glucose media and found that the rate of adaptation declined over time and that the fitness trajectory had a characteristic hyperbolic plus linear curve over the first 20,000 generations. The ancestral strain used in this thesis, REL4536 had previously undergone 10,000 generations of adaptation to aerobic DM25. The 4,000 generations of adaptive evolution in this thesis are thus comparable to the 10,000 to 14,000 generation period of adaptation, which was during the linear phase of adaptation. Mutations that become fixed in the population during the linear phase are likely to be those that have smaller effects on the overall fitness of the aerobic lineages in the aerobic environment. As a result of this, I would likely encounter a linear increase in relative fitness among the aerobically evolved lineages in this thesis. This explains the lack of significance of the relative fitness increase (as seen in **Table 3.4 a**) of aerobic lineages under the aerobic environment from 1,000 to 4,000 generations in this thesis, as compared to the 10% fitness increase reported by Barrick *et al.* (2009) from the 10,000 to 14,000 generations. Furthermore, variability in the successive fitness values of aerobically evolved lineages was observed by Barrick *et al.* (2009) and indicated the extent of sampling error which is found during the measurement of relative fitness, where the authors cited statistical uncertainty inherent in the Ara based competition assay. Variability was also found in this thesis with the use of the Rif based

competition assay, and is a likely explanation for the apparent decline in fitness of the 4,000 generation samples. Lastly, the combined averages of the three different populations examined per treatment (**Figure 3.4**) could mask the potentially varied and unique pathways to adaptation that may have been taken by individual populations.

To examine if there has been a cost of specialisation within aerobic lineages adapting to the aerobic environment, aerobic lineages were competed in the anaerobic (alternative) environment (**Figure 3.4 b**). In general, when the aerobically evolved lineages were competed in the anaerobic environment, there was on average, a trend of decreased fitness throughout the 4,000 generations but it was not statistically significant. A potential reason for this could be because averages are displayed, which represent the general trend in the aerobic lineages but may mask the potentially distinct individual fitness trajectories each lineage could have taken. Additionally, wide error bars in **Figure 3.4 a** and **b** support this view. Thus the evolution of the aerobic lineages was investigated at the individual lineage level.

3.3.3.2.1.1. Evolution of the aerobic lineages – individual lineage trends

From **Section 3.3.3.2.1**, when average fitness increases of three populations (AE2, AE3 and AE7) were considered, there was no statistical support for specialisation in the aerobic condition. To determine if this was a result of variability among individual fitness values or a lack of fitness increases among all three populations, individual fitness trajectories of the three populations were considered (**Figure 3.5**). Examining the fitness dynamics of each population in the aerobic environment **Figure 3.5 a**), AE2 displayed the largest increase in relative mean fitness, which significantly increased to 1.31 ± 0.09 (two sample t-test, $P < 0.05$) relative to the ancestor by 2,000 generations. This rate of increase slowed at later time-points. AE3 appeared to have a more gradual rate of fitness increase, with an apparent decrease in fitness at 4,000 generations. For both AE2 and AE3, it appeared that fitness had declined after the 2,000 generation time-point for AE2 and 3,000 generation time-point for AE3. Curiously, the fitness of AE7 did not appear to increase throughout the 4,000 generations and the narrow standard error of the means suggests that these observations are real. Despite wide error bars around the mean relative fitness at each time point, there was considerable variability in the fitness dynamics of individual aerobic lineages in aerobic conditions, suggesting that different mutational pathways of varying efficacy, particularly between AE2 and AE7, were taken by individual lineages.

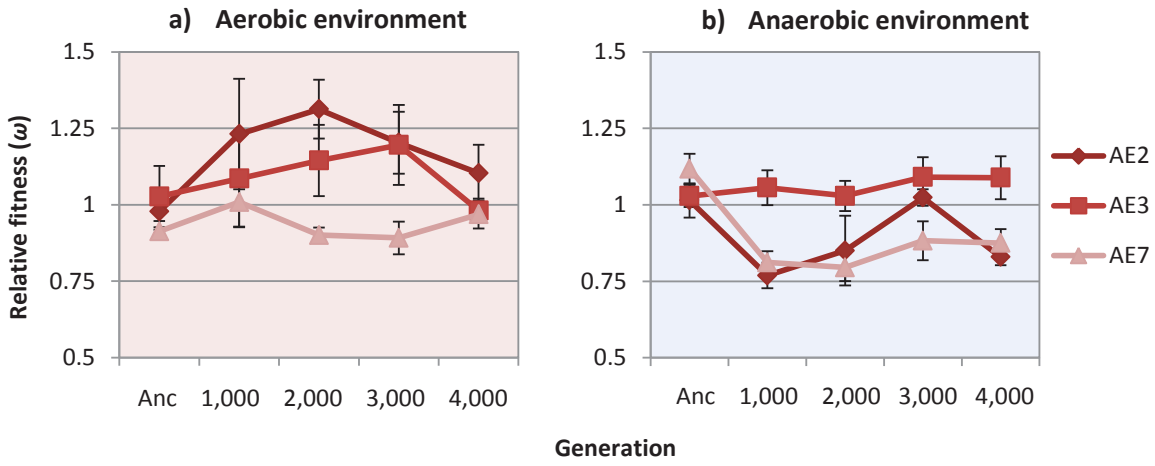

Figure 3.5: Relative fitness trajectories of individual aerobic lineages over 4,000 generations. Displayed are the mean relative fitness values of three individual populations (AE2, AE3 and AE7) relative to the ancestor in a) the aerobic (red background) and b) the anaerobic environments (blue background). Mean values were obtained from average of five biological replicates with error bars representing the standard error of the mean.

As with the aerobically performed competitions, there was variability observed in each lineage's fitness response to the anaerobic environment **(Figure 3.5 b**). Among the three populations tested, there was no change in the fitness of AE3 in the anaerobic competition. However, AE2 and AE7 displayed a marked reduction in fitness. Between 1,000 and 3,000 generations, there appeared to be an increase in fitness in AE2 under the anaerobic conditions. This trend is likely an artifact of high variability of the assay. Fixation of a beneficial mutation in the aerobic condition which confers an advantage to the cell in the anaerobic environment is unlikely as a reduction in fitness of AE2 is again reported at 4,000 generations. The relative fitness of AE7 decreased to ~0.84 which was significant at each generation (two sample t-test, $P < 0.05$ in all cases except at the 3,000 generation time-point). AE2 decreased in fitness significantly at 1,000 and 4,000 generations (two sample t-test, $P < 0.05$), with other time points showing a less pronounced decline, most likely due to statistical variability of the assay.

Comparing the fitness trajectories of aerobic lineages between aerobic and anaerobic conditions **(Figure 3.5 a** and **b**), we see evidence of different modes of specialisation among different populations. The fitness response of AE2 was the highest of the three populations competed in the aerobic environment, but decreased significantly in the anaerobic environment. This is

characteristic of specialisation through antagonistic pleiotropy (AP) (87). AP occurs through the fixation of mutations that are beneficial in the aerobic environment but have negative effects in the anaerobic environment (82, 204). AP is a common mode of specialisation and has been reported in *E. coli* populations during thermal niche adaptation (87, 97), metabolic specialisation (205) and the emergence of wrinkly spreaders in *P. fluorescens* (128). The observation that the AE7 population did not increase in fitness in the aerobic environment but dropped in fitness in the anaerobic environment leads me to assume that AE7 became a specialist through mutation accumulation. This occurs when neutral mutations arise by drift in the adaptive environment. In the alternative environment however, these same mutations have a negative effect on the cell, leading to a decline in fitness (82, 87, 94).

3.3.3.2.2. Evolution of the anaerobic lineages – general trends

To determine the mode of anaerobic lineage specialisation, competition assays of three randomly selected anaerobic lineages (AN1, AN4 and AN6) were performed in both the aerobic and anaerobic environments, with results displayed in **Figure 3.6**. In the first 2,000 generations, fitness increased in all three tested lineages. On average, mean relative fitness of anaerobic lineages in anaerobic environments at 1,000 and 2,000 generations were 1.36 ± 0.03 and 1.38 ± 0.02 respectively. After 2,000 generations however, there was a decrease in mean fitness among all three anaerobic lineages in the anaerobic environment. By 3,000 generations, anaerobic lineages were difficult to grow in both the anaerobic and aerobic environment, reaching maximum cell densities of 10^4 CFU/mL. The lower stationary phase cell densities in these cultures resulted in difficulty obtaining a 1:1 ratio of evolved cells to reference cells, necessary for the competitive fitness assays. Cell densities declined further to a maximum of 10^3 CFU/mL by 4,000 generations. The extent of this poor growth in evolved cultures impaired the ability to reliably determine the number of evolved cells from the number of Rif^r cells by subtraction (as per **Equation 3.1** in **Section 2.2.8.1**). Among replicates, the number of evolved cells was often calculated to be less than zero at Day 0 of the competitions, resulting in unreliable fitness values for 4,000 generation lineages. For this reason, competition assays were deemed not viable at the 4,000 generation time-point. This is indicated on subsequent figures as not determined (ND). Nevertheless, from 3,000 generation mean fitness values, it was apparent that fitness of the anaerobic lineages decreased in the latter stages of this LTEE. Due to the continual decline of population numbers, it is assumed that the relative fitness of 4,000 anaerobic populations is below the relative fitness reported for populations at 3,000 generations. This observation was highly unexpected, and has

been seldom observed (206) and is explored further in **Chapter 5** of this thesis. In the current chapter, the relative fitness data from the 0 to 2,000 generation time points are reported.

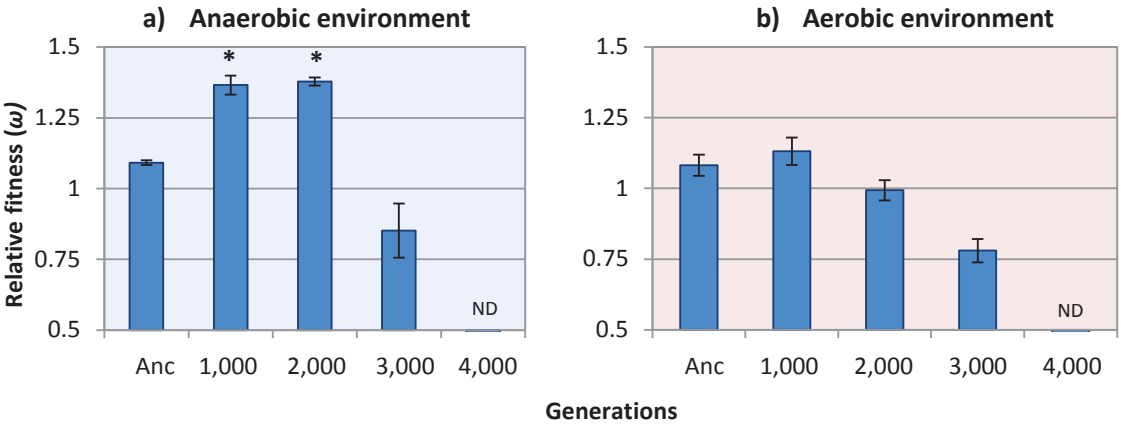

Figure 3.6: Average relative fitness of anaerobic lineages over 4,000 generations. Displayed are the mean relative fitness values of three anaerobic lineages AN1, AN4 and AN6 relative to the ancestor in a) the anaerobic (blue background) and b) the aerobic (red background) environments. Error bars represent the standard error of the mean. ND represents not determined, see text for details.

Within the first 2,000 generations, there was initially a much greater increase in fitness among the anaerobic lineages adapting to the anaerobic environment than the aerobic lineages adapting to the aerobic environment. Within the first 1,000 generations, the mean fitness of the three anaerobic lineages tested increased to a relative mean fitness of 1.36 ± 0.03 (two sample t-test, $P < 0.001$), followed by a reduced rate of fitness increase between 1,000 and 2,000 generations to 1.38 ± 0.02 (two sample t-test, $P < 0.01$). (**Figure 3.6 a**). Up to the 2,000 generation time point, this fitness trajectory was similar to the hyperbolic model which provided the curve of best fit for the fitness trajectory of REL606 over 10,000 generations of evolution (55, 56, 82, 83, 100, 101). This highly significant increase in fitness could be a result of two factors. Firstly, as exposure to the anaerobic environment is novel for REL4536, the genes mutated in the initial stages of adaptation may have been those with the highest improvement to fitness in the anaerobic environment. Furthermore, the mutation rate of REL4536 grown on DM25 was found to be higher in the anaerobic environment (1.88×10^{-3} mutations per genome per generation) than in the aerobic environment (1.14×10^{-3} mutations per genome per generation) (S. Shewaramani, unpublished data). A higher mutation rate could imply that more beneficial mutations could occur, leading to increases in mean fitness in the anaerobic environment (207, 208).

Comparing fitness trends of anaerobic lineages in **Figure 3.6 a** and **b,** evidence of adaptation through trade-offs is seen by an increase in fitness in the anaerobic environment (**Figure 3.7 a**) accompanied with no change in fitness reported when these same lineages are competed in the aerobic environment (**Figure 3.7 b**). The cost of specialisation reported among anaerobic lineages was distinct from those observed in the aerobic populations (**Section 3.3.3.2.1**). During the course of anaerobic specialisation, mutations that occurred in the anaerobic environment were beneficial, as seen by the increases in fitness among the anaerobic populations in the anaerobic environment. However, this did not come at a cost to adaptation to the aerobic environment, as no change in fitness was observed (**Figure 3.7 b**). This type of specialisation, with mutations that are beneficial in the environment of selection and also neutral in the alternative environment (82), suggests that genes mutated in the anaerobically evolved lineages did not adversely affect the cells performance in the aerobic environment.

3.3.3.2.2.1. Evolution of the anaerobic lineages – individual lineage trends

To examine if anaerobic lineage adaptation varied to the extent seen in the aerobic lineages, competitive fitness of AN1, AN4 and AN6 populations was investigated individually under anaerobic (**Figure 3.7 a**) and aerobic (**Figure 3.7 b**) conditions. In both the aerobic and anaerobic environments, fitness trajectories of the anaerobic lineages showed much less variation when compared to those of the aerobic lineages (**Figure 3.5**). Of the three lineages tested, the relative mean fitness of AN1 exhibited the greatest increase to 1.43 ± 0.08 within the first 1,000 generations. By this time the AN4 and AN6 also increased in relative mean fitness significantly, to 1.34 ± 0.07 and 1.32 ± 0.04 respectively (two sample t-test, with $P < 0.001$ in both cases). By 2,000 generations the rate of fitness increase slowed, but relative fitness remained high, at 1.37 ± 0.06, 1.36 ± 0.07 and 1.41 ± 0.03 for AN1, AN4 and AN6 anaerobic lineages in the anaerobic environment. Within the first 2,000 generations, the mean fitness of all three anaerobic lineages in the aerobic environment was similar for all populations (**Figure 3.7 b**).

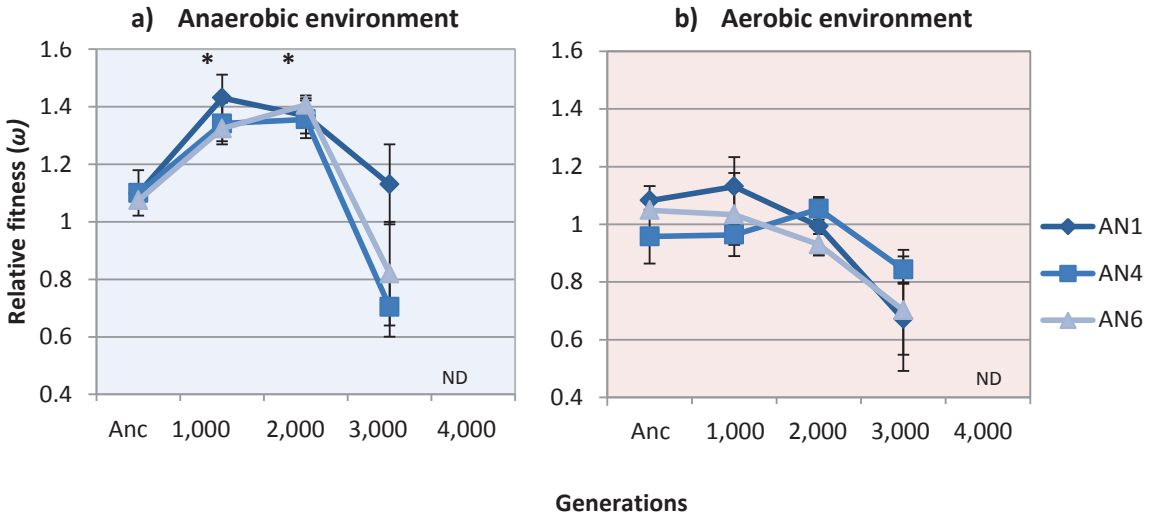

Figure 3.7: Relative fitness trajectories of anaerobic lineages over 4,000 generations. Displayed are the mean relative fitness values of three individual populations (AN1, AN4 and AN6) relative to the ancestor in a) the anaerobic (blue background) and b) aerobic (red background) environments. Mean values were obtained from averages of five biological replicates with error bars representing the standard error of the mean. ND represents not determined, see text for details.

Puentez Téllez *et al.* 2013 studied the adaptation of *E. coli* MC1000 to oxygen rich and oxygen limited conditions over 1,000 generations and provides a good point of comparison for the present study. The results reported in this chapter regarding fitness changes during the evolution of *E. coli* REL4536 to narrow aerobic and anaerobic niches are in line with those of the Puentez Téllez *et al.* 2013 study (101). Although media and strains used in both studies differ, values of fitness are broadly similar.

The highest fitness increase in my data and that of Puentes-Téllez *et al.* (2013) was found among the oxygen limited populations (87). These authors reported an increase in mean fitness within 1,000 generations of 1.35 ± 0.03 for oxygen limited lineages, and in my study, anaerobic lineages increased in fitness by 1.36 ± 0.03 in the anaerobic environment. Fitness increases among lineages evolving in aerobic conditions were also similar to the findings of my study. Puentes-Téllez reported a fitness increase of 1.16 ± 0.04 in aerobically evolving lineages, while a more modest increase of 1.11 ± 0.06 for aerobically evolved lineages was reported in this thesis, which was not statistically significant. This difference is most likely due to the pre-adapted state of the ancestor used in this thesis. No information is given as to the history of the MC1000 strain used in the Puentes-Téllez study, but it is unlikely the MC1000 ancestral strain would have pre-adapted to

oxygen rich LB media to the extent that the ancestral REL4536 strain had pre-adapted to aerobic DM25 (as described in **Section 3.3.3.2**). Interestingly, when comparing the fitness response of specialists in the alternative environment more discrepancies arise from the Puentes-Tellez study. In this study, lineages evolved in an oxygen rich environment were found to increase in fitness in the oxygen limited environment by 1.07 ± 0.01, while the same is true for lineages subjected to the oxygen limited treatment in the aerobic environment 1.10 ± 0.01. A plausible explanation for these increases in fitness in the alternative environments could be that they represent fitness increases in response to adaptation to the media rather than the treatment.

3.3.3.3. Adaptation to a broad niche

3.3.3.3.1. Evolution of the fluctuating lineages in aerobic and anaerobic environments

As discussed in **Section 3.3.3.2**, adaptation to a narrow niche or constant environment tends to promote the evolution of specialists. However, adaptation to a broad niche, or non-constant environment, tends to promote the evolution of generalists. The ability to thrive in both an aerobic and anaerobic environment is an important consideration for a facultative anaerobic organism, such as *E. coli*. To understand the fitness response of *E. coli* to growth alternating between aerobic and anaerobic conditions, the fitness response of fluctuating lineages was tested in both aerobic and anaerobic environments (**Figure 3.8**). On average there was a gradual increase in fitness among the fluctuating lineages in the aerobic environment (**Figure 3.8 a**). However, increases were only statistically significant at the later time-points, at 3,000 and 4,000 generations increasing to 1.08 ± 0.03 and 1.10 ± 0.03, respectively (two sample t-test, $P < 0.05$ in both cases). Under the anaerobic condition (**Figure 3.8 b**), a much higher rate of fitness increase was seen. Unlike the fitness response in the aerobic environment, these increases were significant at all time-points (two sample t-test, $P < 0.05$), with an average increase in fitness to 1.31 ± 0.08 by 4,000 generations (two sample t-test, $P < 0.05$).

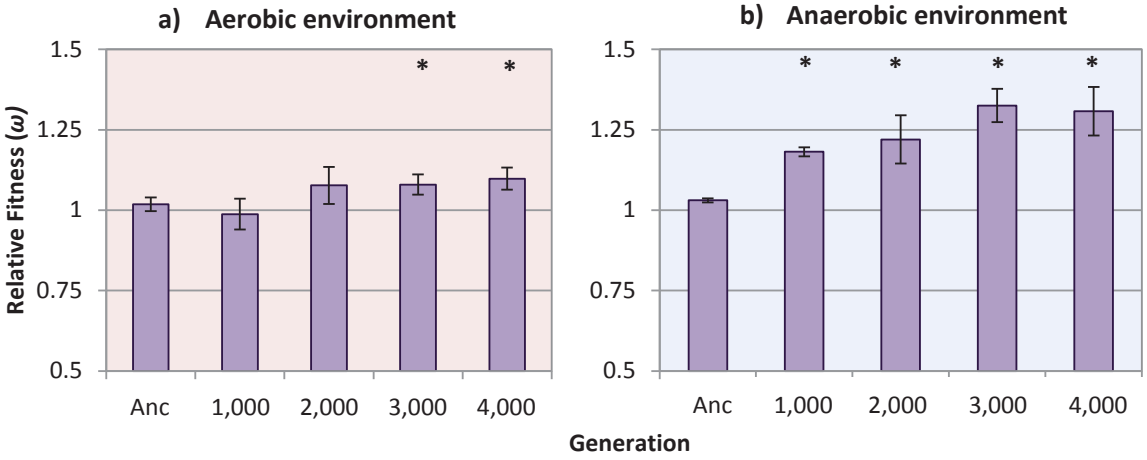

Figure 3.8: Average fitness of fluctuating lineages over 4,000 generations. Displayed are the mean relative fitness values of three fluctuating lineages FL2, FL3 and FL 7 relative to the ancestor as found in a) the aerobic environment (red background) b) the anaerobic environment (blue background), with error bars representing the standard error of the mean.

This general trend of a greater rate of fitness increase in the anaerobic environment as compared to the aerobic environment could be a result of two factors. Firstly, the pre-adapted state of the ancestral REL4536 to aerobic DM25 prior to this experiment has been shown to reduce the rate at which aerobic lineages increase in fitness in the aerobic environment (**Section 3.3.3.2.1.1**). This is also likely to have reduced the rate at which the fluctuating lineages adapted to the aerobic environment. Secondly, the rate of spontaneous mutation has been shown to be higher in the anaerobic environment as described in **Section 3.3.3.2.2** (32), potentially mediating a more rapid rate of adaptation of fluctuating lineages to the anaerobic environment, and may explain the different rates of fitness increase of fluctuating lineages in aerobic and anaerobic environments. Unlike populations that were exposed to constant aerobic conditions (**Figure 3.4**) or constant anaerobic conditions (**Figure 3.6),** there was no cost of fitness during adaptation to aerobic or anaerobic conditions among the fluctuating lineages. This is typical of adaptation to a non-constant environment and evidence of generalist evolution (88, 97, 209, 210).

3.3.3.3.1.1. Comparing of fitness responses of lineages adapted to narrow or broad niches

As outlined in **Chapter 1 Sections 1.3.2**, evolutionary theory predicts that adaptation to a constant environment is likely to result in the evolution of specialists, while exposure to a non-constant environment promotes the evolution of generalists. Specialists typically have the highest mean

fitness relative to their ancestor in the environment of selection, but perform poorly in the alternative environment (97, 209, 210). Generalists are expected to have reduced fitness, as compared to specialists, but perform well in environments to which they have been exposed (92). I wanted to compare the fitness response of lineages that had evolved through different regimes, i.e. those that had adapted to a narrow versus broad niche. The 2,000 generation time point was selected as direct comparisons could be made between the aerobic, anaerobic and fluctuating lineages. At 2,000 generations, the fitness of populations evolved in narrow aerobic and anaerobic niches was compared to the fitness of populations that were evolved to a broad niche (fluctuating lineages) in both the aerobic and anaerobic environments (**Figure 3.9**).

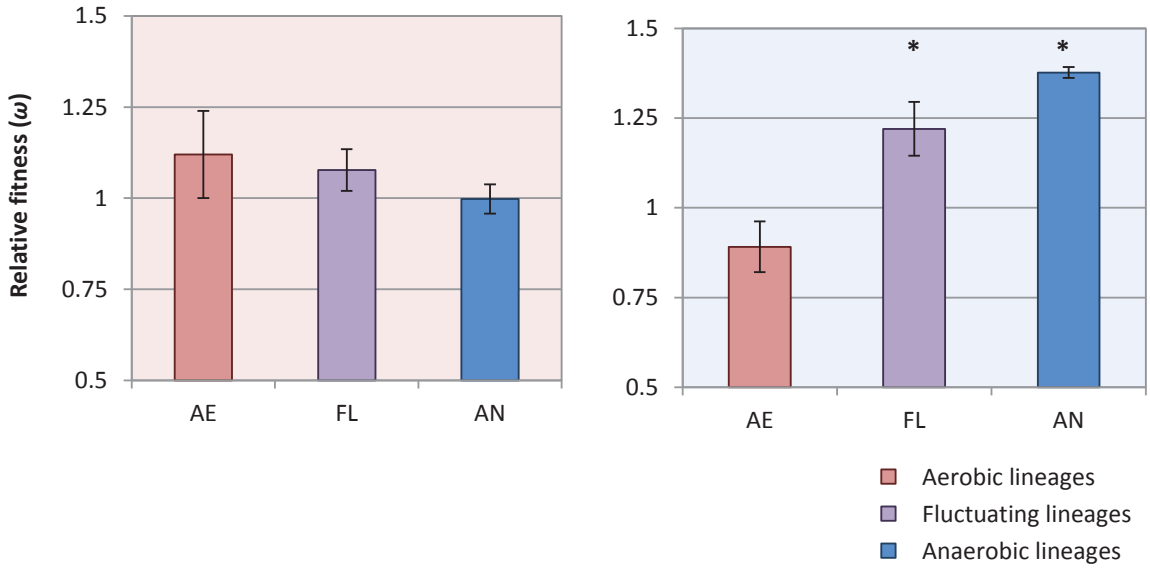

Figure 3.9: Generalist adaptation of fluctuating lineages as compared to aerobic and anaerobic lineages at 2,000 generations. Displayed are the mean relative fitness values of three aerobic, fluctuating and anaerobic populations after 2,000 generations of evolution in a) the aerobic (red background) and b) the anaerobic environment (blue background). Error bars represent standard error of the means.

Fitness values of fluctuating lineages were lower than those of lineages that had evolved in a constant aerobic or anaerobic environment. However, unlike the specialists, generalists displayed increased fitness in both environments. In the aerobic environment (**Figure 3.9 a**) lineages that had adapted to the aerobic conditions reported the highest mean fitness response of 1.12 ± 0.11, despite a lot of variability (as described in **Section 3.3.3.2.1.1**). Lineages that had the lowest mean fitness in the aerobic environment were the anaerobically evolved populations, with a fitness value of 1.00 ± 0.04 (two sample t-test, $P < 0.05$). Fluctuators displayed an intermediate fitness

response of 1.08 ± 0.05 (P = 0.06) in the aerobic environment. While the general trend is observed, these observations are not statistically significant. This is likely due to three main reasons i) the large variability in the fitness response in the aerobic environment (**Section 3.3.3.2.1.1**), ii) the small increases in fitness due to the pre-adapted state of REL4536 to aerobic DM25 (**Section 3.3.3.2.1.1**) and iii) the lack of fitness cost of the anaerobic specialists in the aerobic environment (**Section 3.3.3.2.1**). Conversely, this was not the case in the anaerobic environment.

In the anaerobic environment (**Figure 3.9 b**), these increases in fitness were statistically significant, presumably because the anaerobic environment was a novel environment for REL4536, and thus fitness responses were more pronounced. In the anaerobic environment, the lineages that evolved only in the anaerobic environment showed the highest fitness increase of 1.38 ± 0.02 (P < 0.05). When lineages adapting to the aerobic environment were competed in the anaerobic environment, a significant decline in fitness was observed, with a mean fitness of 0.89 ± 0.07 (two sample t-test, P < 0.05). However, fluctuating lineages that were evolved in both aerobic and anaerobic environments displayed increase in mean fitness. Specifically, the fluctuating lineages had a mean relative fitness increase of 1.22 ± 0.07 (two sample *t*-test, P < 0.05) after 2,000 generations. Based on the fitness data outlined above, lineages that evolved in the fluctuating treatment represent generalist lineages that performed well in both treatments to which they were exposed.

Generalist lineages that were exposure to a broad niche, displayed increases in fitness in both aerobic and anaerobic environments, which conforms to evolutionary theory. Puentes-Téllez *et al.* (2013) also developed generalists in the adaptation of *E. coli* to oxygen rich and oxygen limited conditions. Puentes-Tellez also found a greater response of generalist lineages in the oxygen limited environment as compared to that of the oxygen rich environment. In the oxygen limited environment, fluctuating evolved lineages from MC1000 achieved a mean fitness increase to 1.12 ± 0.02, as compared to 1.08 ± 0.05 in this thesis. In the anaerobic environment fluctuating MC1000, the mean fitness increased to 1.27 ± 0.02, as compared to 1.22 ± 0.07 in the present study. In both conditions fitness increases reported by Puentes-Tellez were larger than those reported in this study. This is likely due to the differences in media or pre-adapted state of REL4536 in this thesis (**Section 3.3.3.2**). The fitness response of *E. coli* to the anaerobic environment is also supported by Puentes-Téllez *et al.* 2013 (101).

3.4. Summary

The aim of this chapter was to address three main objectives. The first objective was to determine the growth dynamics of *E. coli* REL4536 in batch cultures in both the aerobic and anaerobic environments. The growth characteristics of REL4536 were determined in 1 mL total culture volumes in 24-well tissue plates. In the aerobic environment REL4536 grew to similar cell yields and timing of stationary phase to that reported by Lenski *et al.* (1991) (6). Growth dynamics of REL4536 anaerobically were found to be much slower, when compared to aerobic growth. However, anaerobically, REL4536 was still capable of attaining similar cell yields after 24 hours of growth. This is likely due to the presence of citrate in the media and the innate ability of *E. coli* to use this carbon source in the anaerobic condition due to the presence of co-substrates.

The second objective of this chapter was to establish long-term adaptive lineages in aerobic, anaerobic and fluctuating treatments. As such, on 11 February 2011, adaptive lineages were set up as aerobic, anaerobic and fluctuating lineages and continued for 4,000 generations. The long-term adaptive experiment was completed in September 2012. The final objective for this chapter was to assess the fitness of the lineages as they had evolved within their respective treatments. However, before competitive fitness assays could be performed, an appropriate neutrally marked strain was required. Initially, it was expected that the commonly employed Ara+ gene marker could be used. However, repeated experiments indicated this marker was not appropriate under anaerobic conditions. Many alternative strategies were trialled and tested but were not successful. Finally, competition assays using replicate plating based techniques with a rifampicin resistant marked ancestral clone against evolved populations was found to reliably determine fitness under both aerobic and anaerobic conditions. This meant that assessment of the fitness of the lineages in this study could be achieved.

The fitness of three randomly selected populations from each of the three treatments was assessed using competitive fitness assays to determine how the populations adapted over time. In general, aerobic lineages showed no significant increase in fitness under the aerobic environment. This was expected and likely due to the pre-adapted state of REL4536 to aerobic growth in DM25. There was no significant decrease in the fitness of aerobic lineages under anaerobic conditions. This was found to be a result of varied responses of different lineages in both the aerobic and anaerobic environments. This is likely to represent different mutational pathways taken by the different lineages. Competitions with the anaerobic lineages at the 3,000 generation time-point all

display dramatic decreases in fitness, for reasons explained in **Chapter 5** of this thesis. However, within 2,000 generations common trends were observed. Fitness of anaerobic lineages in the anaerobic environment displayed highly significant increases in fitness by 1,000 generations and remained high at 2,000 generations. In the aerobic environment, all three anaerobic lineages displayed no decrease in fitness, which is characteristic of independent adaptation. Generalist adaptation to aerobic and anaerobic environments was examined for the fluctuating lineages. In the aerobic environment there was a very gradual increase in fitness among the fluctuating lineages, which only became significant in the 3,000 and 4,000 generation. In anaerobic conditions there was a much more dramatic increase in fitness. The differences in rate of adaptation of generalists to the different environments were likely attributed to the pre-adapted state of REL4536. In general, lineages adapted to a broad niche width displayed less dramatic fitness responses to lineages adapted to a narrow niche as evolutionary theory would predict.

In this chapter, long-term lineages were established and evidence for multiple pathways to adaptation for aerobic and anaerobic conditions were identified. These lineages have been studied in further detail at the genome level to identify the mutations that resulted in the fitness changes discussed in this chapter, the results of which will be discussed in **Chapter 4**.

Chapter Four : Genotypic analysis of lineages adapting to aerobic and anaerobic environments

4.1. Introduction

In the previous chapter, I investigated the effect that long-term exposure of *E. coli* REL4536 to aerobic and anaerobic conditions had on the competitive fitness of adapting lineages. The aim of this chapter was to uncover the genetic basis underlying the evolution of *E. coli* REL4536 in conditions that were aerobic, anaerobic and a treatment that fluctuated between the two conditions. LTEE combined with whole genome re-sequencing is a powerful tool which allows the identification of the full suite of mutations that have occurred during the course of evolution. From the genome sequences of evolved lineages, putative mutations underlying adaptation to different environments can be uncovered.

Mutations that arise during evolution under aerobic, anaerobic and fluctuating conditions are expected to include both adaptive mutations that are beneficial to the cell, and therefore, have been selected for; as well as neutral mutations that do not affect fitness, but have spread by random drift. Additionally, neutral and deleterious mutations that have arisen in lineages that contain adaptive mutations may also be present, by hitchhiking with adaptive mutations if the overall effect on fitness is positive. For the purpose of this thesis, the extent of parallel evolution of traits among lineages was used as evidence to suggest that mutations were acted on by natural selection, and therefore adaptive. Parallelism is the observation of traits mutated independently across replicate lineages (105). At the genetic level, this is reported as genes involved in similar functional pathways acquiring mutations which lead to the same or similar phenotypic trait. In the present study, such occurrences would indicate traits that are likely to be beneficial and selected for during adaptation (104, 105, 111, 211).

Adaptation is primarily caused by natural selection. Adaptation to a given environment may happen through a variety of mechanisms due to the random processes that underlie adaptation. The variety of mechanisms by which populations become more fit within their environments are typically referred to as adaptive pathways. Different pathways are mediated by different mutation

events that are key in driving a population towards a peak in the adaptive landscape. By combining genomic data from the present chapter with fitness data from **Chapter 3**, mutations and adaptive pathways will be identified that are responsible for movement of lineages towards the peaks within the adaptive landscapes in each treatment. To this end, randomly selected clones from the 21 adaptive lineages that had evolved in their respective treatments (aerobic, anaerobic and a treatment that fluctuated between these environments) for 4,000 generations (**Chapter 3**), were sequenced at the whole-genome level to uncover the genetic changes that had occurred during the course of this study. Mutations that had arisen in each evolved lineage were identified at 2,000 and 4,000 generations.

4.2. Objectives

The objectives of this chapter are:

1. To identify the mutations that had arisen during the evolution of *E. coli* REL4536 during long-term growth in aerobic, anaerobic and the fluctuating treatments.
2. To determine which mutations are likely to be beneficial and contribute to adaptation to each treatment.
3. Determine the diversity of traits and pathways that are undertaken for adaptation to each treatment.

4.3. Results and discussion

4.3.1. Observation of polymorphism among evolved lineages

To identify the genetic changes that arose during the evolution of lineages that had been subjected to long-term growth under strict aerobic, strict anaerobic and a treatment that fluctuated between both conditions (**Chapter 3**), 21 individual colonies, from each of the 21 populations, were randomly selected and cultivated. Genomic DNA was extracted from all 21 cultures for genome sequencing (**Section 2.2.12**). At 2,000 generations, it was found that the maximum cell densities of clones (from lineages AN-2K-4, AN-2K-6 and AN-2K-7) were considerably lower as compared to the clones from the other anaerobic lineages. Specifically, stationary phase 10 mL LB broth cultures obtained after 24 hours of incubation for AN-2K-4, AN-2K-6 and AN-2K-7 yielded OD_{600} of 0.107, 0.054 and 0.022 respectively as compared to an OD_{600} of ~0.500 for the remaining four anaerobic lineages. Culture volumes were therefore increased to 200 mL in order to obtain the yields of DNA required for whole genome sequencing for these three clones. The limited growth of the clones from the AN-2K-4, AN-2K-6 and AN-2K-7 lineages was

hypothesised to be a result of specialisation to anaerobic DM25 media, rather than anaerobic LB media used for genomic DNA extractions. Further analysis showed that when grown aerobically on LB agar plates, these cultures gave rise to colonies that were morphologically distinct to those from the ancestral population, and are referred to as small colony morphotypes (SCMs). At 4,000 generations, all seven anaerobic lineages were comprised solely of SCM. Investigations into the evolution and biology of the SCM variants are presented in **Chapter 5**.

4.3.2. Modification of the ancestral genome sequence

To identify mutations in the genomes of the evolved strains, evolved genomes were compared to the genome of the ancestor, REL4536. However, the REL4536 genome is not publicly available. Thus, the mutational differences identified by Barrick *et al.* 2009 between REL606 and REL4536 (52) were incorporated into the publically available *E. coli* REL606 genome sequence (212) to manually create a REL4536 reference genome sequence. Furthermore, the ancestral REL4536 culture (**Section 2.2.5.1**) used to generate the lineages in this thesis was re-sequenced. This ancestral REL4536 genome was used to determine the accuracy of the manually created genome sequence described above by mapping the genome re-sequenced ancestral REL4536 genome to the manually created sequence. Four discrepancies were reported between the re-sequenced REL4536 strain, and the manually created REL4536 genome sequence (**Table 4.1**).

Table 4.1: Mutations detected from the genome re-sequencing of REL4536 as compared to the manually created REL4536 genome sequence.

Gene/region affected	Position (bps)[a]	Mutation[b]
insA-9/*gatA*	626,523	2 bps insertion
ECB_01992	651,204	20 bps ×2 duplication
pykF	998,182	IS150 insertion
xasA	188,987	IS150 deletion

[a] Position on the manually created REL4536 GenBank reference file

[b] Mutation detected in REL4536 sequenced culture, as compared to the REL4536 reference file.

To verify these four differences, the REL4536 genome sequence data was *de novo* assembled (S. Shewaramani, unpublished data). The results of this assembly indicated that the 2 bps insertion in the intergenic space between the *insA-9* and *gatA* genes (*insA-9/gatA*) and the 20 bps duplication in *ECB_01992* gene were not present in the ancestral genome, and were likely artefacts of the breseq program. Furthermore, the *de novo* assembly indicated that the orientation of the IS150 element in the *pykF gene* was incorrect, as compared with the manually created REL4536 genome

sequence. This orientation change was confirmed by PCR (S. Shewaramani, unpublished data) and updated according to the genome sequence. Finally, the reported IS150 deletion in *xasA* is due to the absence of an IS150 element in the gene in the ancestral REL4536 genome used in this thesis. The reported IS150 deletion in *xasA* was further confirmed by PCR. The IS150 insertion event was identified by Barrick *et al.* (2009) in only reported in the 10,000 descendent clone of REL606 and was not reported again in further descendent lineages (52). Thus, it is feasible that the ancestral genome used in this thesis differed from that published by Barrick *et al.* (2009) at this *xasA* locus. This difference was accounted for on the manually created reference sequence. Re-sequencing and *de novo* assembly were important to ensure the correct genome sequence of the ancestral REL4536 was used for all downstream analysis in this experiment.

To further confirm that the re-sequenced REL4536 genome identified had the same 28 mutations as reported by Barrick *et al.* (2009), the genome sequence reads of REL4536 were mapped to the complete REL606 genome sequence of Joeng *et al.* 2009 (212). Results from the assembly and mutation detection confirmed 27 mutations. However, a large inversion of 1,493,854 bps between *citC* and *gatZ* was expected, but this was not detected by breseq. This inversion is mediated by recombination between two IS1 elements (768 bps in length). The sequencing strategy employed for the ancestral REL4536 genome, the 2,000 evolved clones and the 4,000 aerobic and fluctuator clones utilised a 500 bps paired end insert library. This approach was unable to detect large repeat-sequence mediated rearrangements. A higher incidence of IS movement has been reported in anaerobically grown *E. coli* cultures (32) and, thus, was expected in anaerobically evolving lineages. Taking the finding that 500 bps insert is not ideal for the identification of large scale genomic rearrangements into consideration, the anaerobic 4,000 lineages were sequenced using an alternative approach, utilising 2 kb inserts. This sequencing strategy was expected to capture any movement of all *E. coli* REL4536 IS elements (with the largest approximately 1,443 bps in length) to be detected and assembled with confidence. Confirmation of this sequencing strategy to identify IS element mediated mutations has subsequently been confirmed by PCR (S. Shewaramani, unpublished data), and confirmed breseq's ability to accurately identify IS-mediated mutations when the appropriate library is used.

4.3.3. Mutation analysis

4.3.3.1. Classification of mutations

Genomes were re-sequenced, as outlined in **Section 2.2.14**. Sequence data statistics are shown in **Appendix Table 7.2**. Mutations identified in all 42 clones from the evolved aerobic, anaerobic and fluctuator lineages at 2,000 and 4,000 generations were classified based on four types, with further sub-classifications as listed:

i) SNPs; mutations were further classified into transitions or transversions.

ii) Insertions; including duplications, where regions of the genome were copied and inserted in tandem or where sequences were inserted.

iii) Deletions; deletions were classified as small if less than a gene was deleted, and large if one or more genes were deleted.

iv) Mobile elements; mobile element mediated mutations which were further classified into insertions or deletions of IS elements.

Mutations identified in evolved genomes are displayed in **Appendix Table 7.3** for aerobic clones, **Appendix Table 7.4** for anaerobic clones and **Appendix Table 7.5** for fluctuating clones. In total, there were 588 mutations reported among the 42 evolved genomes. A general summary of the mutation types and sub-classifications identified is displayed in **Table 4.2**. This table lists the count data of the four different mutation types, each with two sub-classifications, for all 42 evolved clones. The total number of mutations in each class is given. Note that the independence of the anaerobic lineages could not be assured at 4,000 generations (**Section 3.3.2**), thus a direct comparison of the frequency of mutations occurring in each treatment in **Table 4.2** is not valid.

Table 4.2: Mutation types and classes as reported within all 42 evolved clones when compared to the ancestral *E. coli* REL4536 strain.

Condition	SNP		Insertions		Deletions		IS elements		Total
	Trans-ition	Trans-version	Sequence insertion	Duplication	Small	Large	Insertion	Deletion	
Aerobic	18	14	3	6	4	12	47	3	**107**
Anaerobic	12	17	24	41	41	12	103	4	**227**
Fluctuating	16	12	1	18	18	26	166	10	**254**
Total	**46**	**43**	**32**	**21**	**63**	**50**	**316**	**17**	**588**

4.3.3.2. Evidence of cross contamination

An unexpected finding from analysis of the mutation dataset was the high degree of identical mutations that were detected between the different lineages within a treatment, and even

between treatments. This issue was first highlighted due to the striking similarity of mutations found in genomes from the 4,000 generation anaerobic clones. In general, such high genetic similarity may be due to selection for adaptive traits that have arisen, however cross contamination of lineages is also a likely possibility.

In light of this finding, sequences from all clones were re-examined for the presence of identical mutations. While the extent of identical mutation abundance between lineages was not as pronounced as for the AN-4K clones, identical mutations were observed in the aerobic and fluctuating treatments. The number of mutations and mutational events are reported (**Table 4.3**). Specifically, a mutation indicates the occurrence of a genomic change compared to the ancestral genome. Mutations have been classified as unique mutation events and identical mutation events. Unique mutation events are mutations that were observed only once per treatment. In contrast, identical mutation events refer to instances when identical mutations, two or more, were observed among genomes of the same treatment. Mutations were further classified as common if identical mutations were found between treatments. A summary of the number of mutation events observed is indicated in **Table 4.3**.

Table 4.3: Number of shared mutations

Treatment	Total mutations	Total mutation events	Unique mutation events	Identical mutation events	Common events	Reference
Aerobic	107	67	49	18	13	**Table 7.6**
Anaerobic	227	72	31	41	9	**Table 7.7**
Fluctuating	254	139	104	35	16	**Table 7.8**

From **Table 4.3**, the proportion of identical mutation events between treatments varies, where for the aerobic and fluctuating clones, approximately a quarter of the mutation events observed were comprised of identical mutations, whereas for the anaerobic clones, the incidence of identical mutation events was over half, suggesting that the degree of genetic similarity among the anaerobic lineages was much greater than for the aerobic and fluctuating lineages (**Appendix Table 7.9**).

Overall, two main lines of evidence led me to believe that there was cross contamination within the different treatments during this experiment.

i) High similarity among lineages of the same generation, but not within lineages between generations

For independently evolved lineages, it is expected that clones from the same lineage will be more similar to each other than clones from different lineages. Furthermore, descendent clones are likely to share identical mutations with clones from preceding generations as they have been selected for and inherited, or are linked to such adaptive mutations. Such identical mutations have been termed "online mutations" by Barrick *et al.* (52). In the present study there was a lack of genetic similarity between the 2,000 and 4,000 generation clones within the same lineage. Moreover, evolved clones from different lineages at the same generation were more similar to each other, than to those from earlier generations. Of the 21 lineages in this study, 13 online mutations were observed, and these were not distributed evenly across the three treatments. Unexpectedly, no online mutations were observed among the aerobic lineages (**Appendix Table 7.12**). In the anaerobic environment, 11 online mutations were reported, with six found in the AN7 lineage, and the remaining five distributed across multiple lineages. Within the fluctuating treatment, only two online mutations were seen. While there are several remote possibilities that may explain this pattern of data for independently evolved lineages, the most likely explanation is that cross contamination between lineages has taken place. Furthermore, it would appear that within the anaerobic treatment, the AN7 lineage displays the largest degree of heredity, and this is further explored in **Chapter 5**.

ii) Presence of identical synonymous mutations among lineages

Synonymous mutations are not expected to have an effect on phenotype, and therefore, impact on selection. Thus the chance that identical synonymous mutations had independenly arisen across different lineages is highly unlikely. In the 4,000 generation aerobic genome data, it was noted that 13 cases of synonymous mutations were observed (**Appendix Figure 7.10**). A T → C mutation was found to occur in five aerobic lineages at 1,905,307 bps in AE-4K-1, AE-4K-2, AE-4K-3, AE-4K-4 and AE-4K-7 clones. Cross contamination (**Section 2.2.5.4**) was reported in the blank wells of the anaerobic lineages, but not in the aerobic or fluctuating lineages. The availability of genome sequence data further revealed that cross contamination also likely took place in the aerobic and fluctuating lineages, though was undetected during routine contamination checks. This most likely took place during sub-culturing. This calls into question the use of 24-well tissue culture plates as appropriate culturing vessels for the maintenance of independent cultures,

especially for long term studies. While it was likely that a proportion of identical mutations observed in this genomic dataset had arisen due to cross contamination, adaptive traits and pathways may still be identified from the data, and are explored in **Section 4.3.3.3**.

4.3.3.3. Critical analysis of mutations

Genome sequencing data have revealed that evolution of the lineages within the aerobic, anaerobic and fluctuating treatments may not have been independent (**Section 4.3.3.2**). Nonetheless, adaptive mutations are still likely to have arisen and been selected for within this experiment. This unexpected result has introduced a considerable challenge to the accurate identification of adaptive mutations and necessitated the development of a strategy to determine if a mutation identified was likely to be adaptive and acted on by natural selection. This process is illustrated as a flowchart (**Figure 4.1)**. A detailed explanation of the considerations for assessing if a mutation was adaptive is described below. This is insaine. This is insane

Q1) Considering the biological context of the mutation, would it likely confer a fitness advantage in the treatment of selection?

The effect that each mutation was likely to have on gene function, and hence, biological function, was considered. If the mutation was presumed to be beneficial to growth in the environment in which it arose, it is more likely to be adaptive. For example, mutations in *adhE* occurred only in lineages adapting to the anaerobic environment, and these are presumed to be beneficial via the diversion of anaerobic fermentation pathways to the production of acetate, a growth substrate for *E. coli* (discussed further in **Section 4.3.4.1.1.4**). If a mutation was likely to confer a benefit, then it is further considered under Q3. However, for the majority of mutations, fitness advantages arising from them were unclear, and these are further considered in Q2.

Q2) Do identical mutations occur within other lineages of the same treatment?

The presence of identical mutations within lineages of a treatment may indicate cross contamination of genotypes between lineages. However, specific gain of function mutations that have arisen independently between lineages, and are selected for, will also result in the same outcome. It is generally not possible to distinguish between these possibilities, however further considerations (e.g. from Q3) will help to determine if the mutation confers an adaptive advantage. Unique mutations within the dataset (**Appendix Tables 7.6, 7.7** and **7.8**) were frequently encountered. Without functional characterisation, it could not be determined if such

mutations were adaptive,and therefore these were generally not considered further as adaptive mutation candidates within this study. However, it is acknowledged that such unique mutations may be adaptive, particularly if multiple independent adaptive pathways within that treatment exist (see Q3).

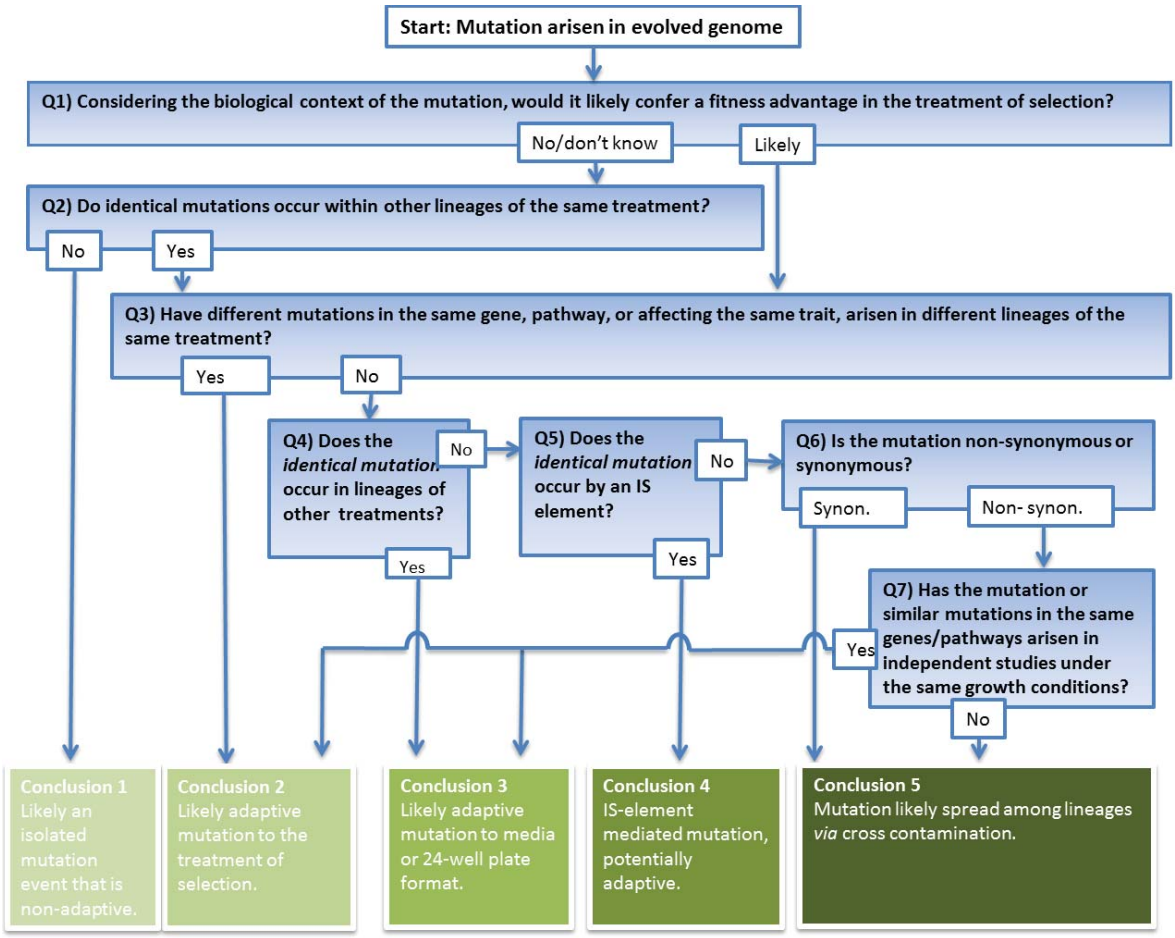

Figure 4.1: Flow chart for identification of likely adaptive mutations. This flow chart consists of seven questions (blue boxes) and five conclusions (from light green to dark green) to determine if each mutation was likely to be adaptive to the treatment conditions. Refer to text for details.

Q3) Have different mutations in the same gene, pathway, or affecting the same trait, arisen in different lineages of the same treatment?

Mutation events that have arisen in parallel between independently evolved lineages provide strong evidence that natural selection has acted upon such mutations. However, since the independence of lineages in this study cannot be relied upon, independent mutation events that

affect the same traits need to be considered. If such instances occur, mutations are likely to be adaptive within the treatment of selection (Conclusion 2). However, if the mutation occurred only once (Q2), and no related mutations are observed (Q3), this is considered to be a likely isolated mutation that is non-adaptive (Conclusion 1). In contrast, if identical mutations are present (Q2), yet no related mutations are evident (Q3), the possibility that the mutation is adaptive and is of a highly specific nature (for example, confers a gain of function) and has arisen independently, still cannot be ruled out. Thus, further considerations must be made (Q4).

Q4) Does the identical mutation occur in lineages of other treatments?

It is deemed highly unlikely that contamination could have occurred across the three different treatments, as many precautions were taken during the sub-culturing procedures. Thus, identical mutations in common between different treatments are not likely to have arisen by cross contamination, but rather independently, and selected for within the affected treatments. Thus, identical mutations that had occurred between treatments were grouped into Conclusion 3 - likely adaptive mutation to media or 24-well plate format. However, if identical mutations arose that were specific to one treatment, further questions regarding the classification of the mutation event were considered (Q5).

Q5) Does the identical mutation involve an IS element?

The non-random insertion of IS elements within the genome is well documented in the literature (39, 213-216). Many instances of identical IS insertions were reported within treatments, between treatments and also across independent studies, particularly Barrick *et al.* 2008 and S. Shewaramani, unpublished data. These examples act as additional support for the independent nature of identical IS element mediated mutations (214, 217-219). Mobile elements are known to integrate into hotspots (214), thus the possibility that identical mutations, that are mobile element-mediated, have arisen independently is far greater than for other classes of mutation. If mutations were not mediated by mobile elements, Q6 was asked. However, identical mutations mediated by IS elements potentially had arisen independently and were adaptive (Conclusion 4), however the possibility that they had spread via contamination still could not be ruled out.

Q6) Is the mutation non-synonymous or synonymous?

As mentioned in **Section 4.3.3.1**, instances of identical SNP mutations that were synonymous are highly unlikely to have arisen independently. Despite recent evidence that synonymous mutations may represent an underappreciated set of mutations important for the adaptive process (220), cases of identical synonymous mutations across lineages were considered to have spread due to cross contamination in this thesis. Identical occurrences of synonymous mutations were thus classified into Conclusion 5: Mutation likely spread among lineages via cross contamination. Examples of identical synonymous mutations, such as those in *ybiI* in aerobic lineages as mentioned earlier (**Section 4.3.3.2**). If the mutation was non-synonymous, this led to Q7.

Q7) Has the mutation or different mutations in the same genes, pathway, affecting the same trait, occurred in independent studies under similar growth conditions?

The final question considered data from other independent LTEE studies, similar in design and set up as outlined in this thesis, to see if similar mutations had arisen in parallel, and were likely to be adaptive. The three studies considered were:

i) Puentes Télles *et al.* (101): As discussed in **Section 1.4.2.1.6**.

ii) S. Shewaramani, (unpublished data): Two experimental evolution experiments conducted in our lab with identical operating procedures and technical equipment. One was an independent adaptive selection experiment that was conducted in our laboratory using an *sbcC* knockout of *E. coli* REL4536. The second varied slightly, in that *E. coli* REL4536 lineages were passaged through population bottlenecks via serial streaking on DM agar plates. Both were conducted under aerobic and anaerobic environments.

iii) Barrick *et al.* (52): This is an adaptive selection experiment, which contains an account of the mutations arising in the ancestral strain of REL4536, REL606, until the 40,000 generation time-point in DM25 in aerobic conditions.

If also found in independent LTEE studies, mutations were grouped as either likely to be adaptive to the treatment of selection (Conclusion 2), or the media in which they were grown, or to growth within the 24-well plate format (Conclusion 3). If mutations were not reported in other selection experiments, they were grouped into Conclusion 5.

In total, 588 mutations were identified from the 42 genomes in the present study. Due to the unforeseen cross contamination that occurred, careful examination and classification of each of

these 588 mutations was required. Thus, each mutation was classified into one of five conclusions described. This approach allowed for the identification of mutations that had likely arisen as an adaptive response to the environments tested.

4.3.3.4. Increased occurrance of IS elements within evolved clones

Due to the extent of cross contamination in this experiment, the ability to compare the prevalence of different mutation types between treatments is limited. However, in general, mutation data from evolved clones does support reports of the increased occurrance of IS element activity within the anaerobic environment, as seen by Sakai *et al.* (32). Furthermore, increased activity of IS element movement found within the fluctuating treatment.

IS element activity is a key feature of the *E. coli* B genome (52, 53, 221). In the ancestral *E. coli* REL4536 genome, there are nine different insertion elements, each with their own size, transposase activity and frequency within the genome (as indicated in **Table 4.4**), as well as unique modes of transposition (39, 48, 53, 212)*.*

Table 4.4: The IS elements in the *E. coli* REL4536 genome. A list of each insertion element size and frequency within the REL4536 genome is indicated.

IS element	Size	Number of copies in REL4536
IS1	768	26
IS2	1,331	1
IS3	1,258	5
IS4	1,426	1
IS30	1,258	1
IS150	1,443	7
IS186	1,343	5
IS600	1,264	2
IS911	1,250	1

From the genomic data of evolved clones, the occurrance of IS proliferation was not equal among the nine different IS elements, with transposition activity reported in only IS1, IS150 and IS186 among evolved clones in the three treatments at 2,000 generations (**Figure 4.5**). The vast majority of IS mediated activity was attributed to the insertion of IS150 in evolved genomes, particularly within the fluctuating treatment.

Table 4.5: Average number of IS element insertion of IS1, IS150 and IS186 mutations per clone in all treatments at 2,000 generations.

IS element	Aerobic clones	Anaerobic clones	Fluctuator clones
IS1	0.28	1.00	0.42
IS150	0.28	7.00	7.85
IS186	0.00	0.57	0.14

In general, there was a high number of IS150 insertions reported in the dataset of the present study. Within individual treatments, there was a 27-fold and 24-fold higher number of IS150 insertions in fluctuating and anaerobic clones than in the aerobic clones. Traditionally, IS elements have been viewed as genetic parasites relocating and proliferating within host genomes (222). Widespread dispersal of IS elements throughout the genome is predicted to largely result in deleterious mutations, that would be selected against according to natural selection. However, more recently, growing evidence cites IS element transposition can also have profound effects on the adaptation of organisms to their environments (48, 51, 111, 223, 224), depending on the nature of the resulting mutation. The greater number of IS150-associated mutations observed may be the result of greater selection for these mutations in the novel (anaerobic) or non-constant (fluctuating) treatments, and thus facilitating adaptation to these environments (46). IS element mediated mutation, due to their transient nature may offer some form of genome plasticity in adaptation to non-constant environments (48). The role that these and other mutation types play during the evolution of experimental lineages is explored in the remainder of this chapter.

4.3.4. Identifying evolutionary pathways within the adaptive landscapes

Adaptation of evolving lineages took place over 4,000 generations by natural selection acting on the adaptive mutations that had arisen in each population. The adaptive landscape is a representation of how fitness changes as a function of genotype or phenotype in a given environment (81). In this thesis, three different selective regimes were considered, corresponding to the three treatments under investigation. Evolutionary pathways in each environment were inferred by whole genome sequencing of individual clones, randomly isolated from each population to identify mutations that had arisen during adaptation to the different treatments (225). In this thesis, the identification of adaptive pathways with whole genome data was done by grouping mutated genes that have similar functional activities, or operate within a shared

functional network. The accumulation of a range of different adaptive mutations allows for alternative evolutionary pathways within each landscape (226). Furthermore, different routes along these pathways were taken, as identified by different genes mutated within the same functional network. In this way an insight into the evolutionary pathways that are present in the adaptive landscape of *E. coli* adapting to each environment was obtained. A summary of the distribution of the putative adaptive mutations (**Appendix Table 7.13**) in the different treatments of selection is shown in **Figure 4.2**.

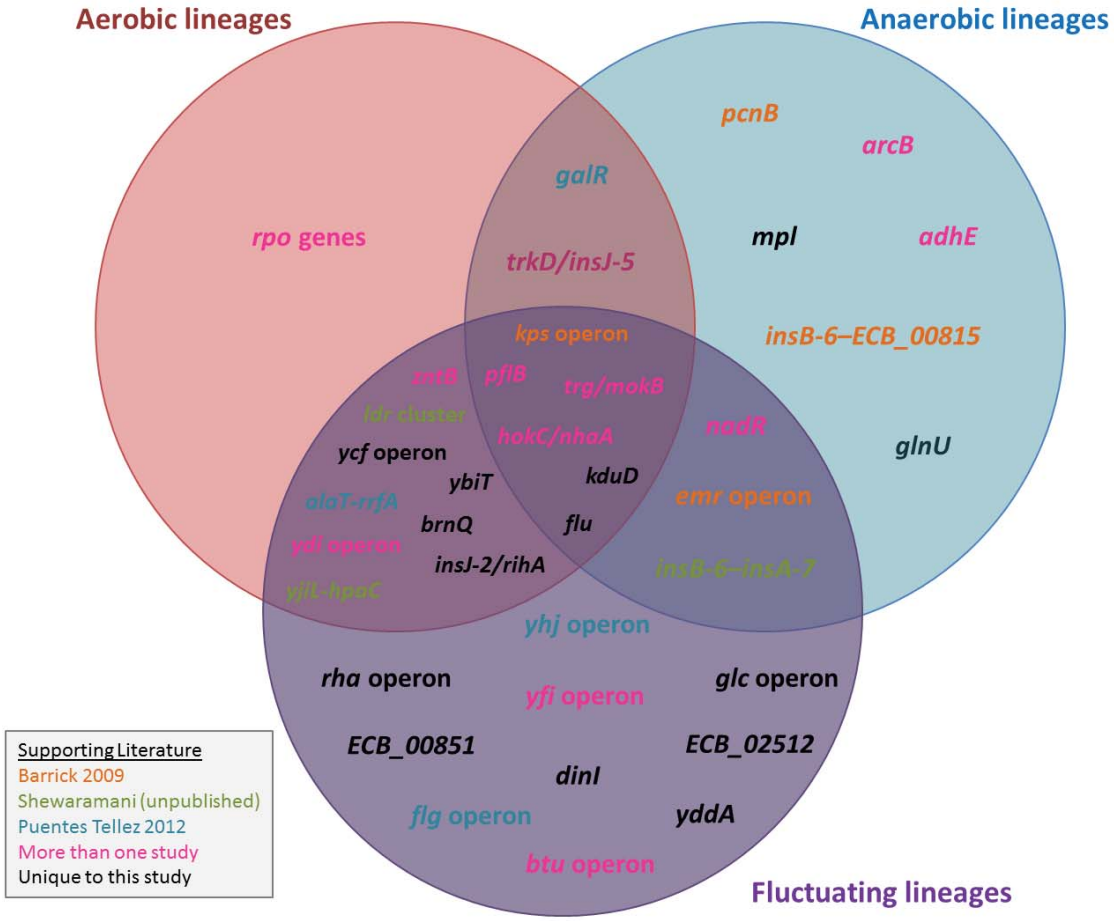

Figure 4.2: Venn diagram of genes and operons with putative adaptive mutations that arose during evolution under the aerobic, anaerobic and fluctuating treatments. The Venn diagram consists of mutations that are likely to be adaptive mutations, based on examination as described in **Appendix Table 7.13.** Each mutation is classified based on the treatment or treatments under which it had evolved. The mutated loci are also colour coded based on their occurrence in other studies listed in the figure legend.

Genes that were independently mutated multiple times, transcribed in the same operon or work in similar functional networks were deemed to have occurred at high frequency by selection acting on these mutations in response to their effects in the various treatments. These mutations are listed in **Appendix Table 7.13** and also displayed in **Figure 4.2** where they are grouped according to the treatments under which they were reported. The numbers of putative independently adaptive mutated genes that were unique to each treatment varied considerably; one in the aerobic environment, five in the anaerobic environment and ten in the fluctuating treatments. Many mutated loci were found to be shared between two or more treatments, i.e. nine were shared between only aerobic and fluctuating lineages; three were shared between only anaerobic and fluctuating lineages; and two were shared between only aerobic and anaerobic lineages, while six were shared between all three treatments.

Of note is the high incidence of adaptive mutations shared between fluctuating lineages that were in common with adaptive mutations found in the aerobic and anaerobic lineages. This is consistent with the expectation of generalists, which have adapted broadly to both anaerobic and aerobic environments, and have accumulated more mutations that are beneficial in each of these environments. Adaptive mutations present in all three treatments are hypothesised to confer adaptations to conditions in common between all three treatments, such as the media type, and the 24-well plate culture vessel. Common functions or characteristics were identified among the mutated loci in each treatment, that suggested eight likely modes by which adaptation may have taken place through the 4,000 generations of evolution reported in this thesis. The hypothesised effects and evolutionary significance of these mutations are discussed below.

4.3.4.1. Mechanisms of adaptation to different treatments

4.3.4.1.1. Modification of anaerobic fermentation pathways

Under aerobic conditions glucose is catabolised and leads to the production of CO_2 and H_2O during the generation of ATP. Under anaerobic conditions, glucose is fermented, producing a mixture of five fermentation end products: formate, acetate, succinate, ethanol and lactate (227) (**Figure 4.3**). Genes involved in the various fermentation pathways of *E. coli* were found to be mutated several times in the anaerobic environment. In both anaerobic and fluctuating lineages, four genes in particular were repeatedly and independently mutated, and were predicted to enable the evolved lineages to increase the rate of energy production during anaerobic growth.

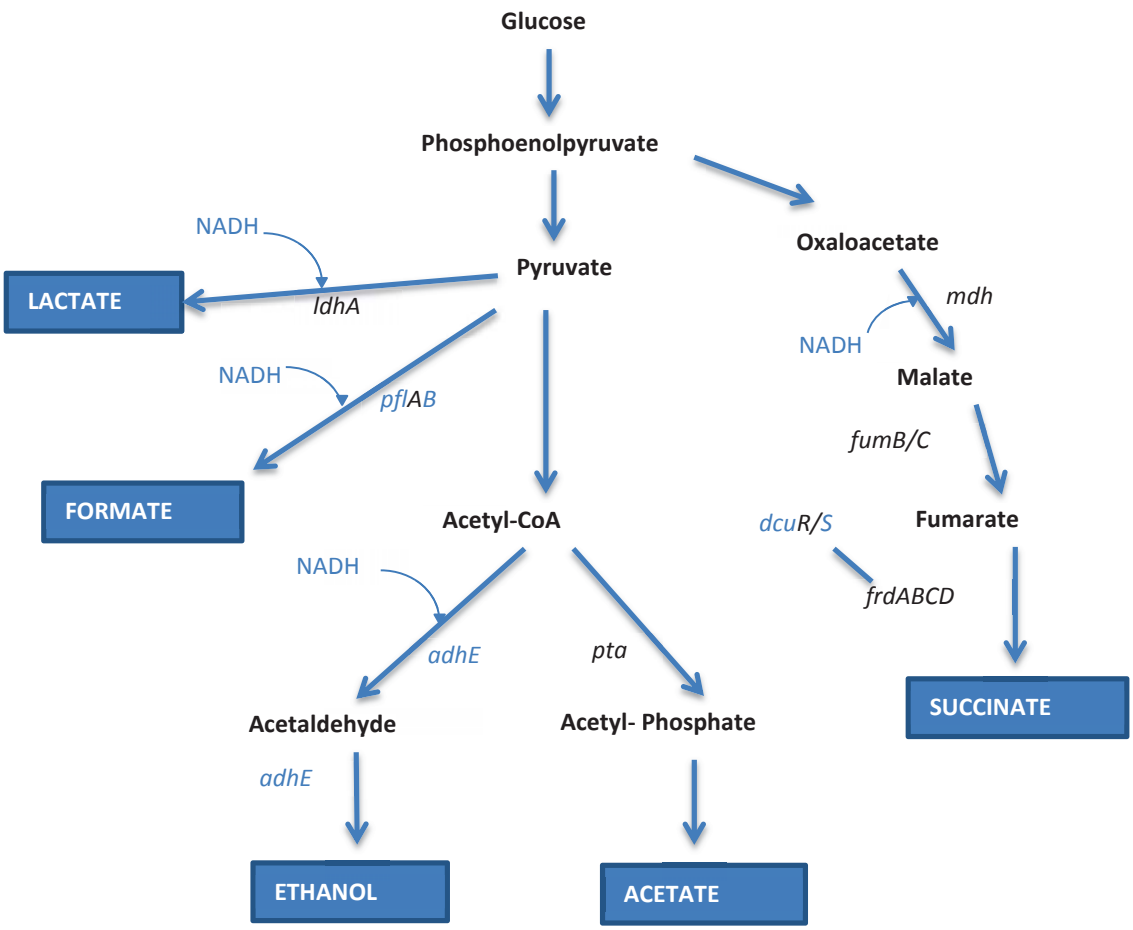

Figure 4.3: Diagram of the anaerobic fermentation pathways in *E. coli.* Three of the eight genes involved in the anaerobic fermentation network were mutated in the present study (*adhE, pflB* and *dcuR/S*) and are coloured blue. Figure adapted from Jiang *et al.* 2001 (228).

More specifically, the mutations and genes affected are listed in **Table 4.6** in order of those mutations which have the highest support based on parallelism.

Table 4.6: Genes mutated in anaerobic fermentation pathways.

Mutated anaerobic fermentation pathways					
Gene [a]	**Function** [b]	**Treatment**[c]	**Mutation type** [d]	**Position** [e]	**Lineages affected** [f]
nadR	Upregulates fermentation network via NAD	FL and AN	IS150 insertion	4581545	AN-2K-1, 3, 4, 5 and 6
			IS150 insertion	4581546	AN-4K-1, 2, 4 and 5
			IS150 insertion	4581547	FL-4K-1, 2, 3, 4, 5, 6 and 7
			IS150 insertion	4581549	AN-4K-3
pflB	Formate production	FL and AN	IS150 deletion	1764886	AN-2K-2,3 and FL-

					2K-1,3,7
			IS150 deletion	1764888	FL-2K-2,4,5 and 6 and FL-4K-1,2,3,4,5,6 and 7
dcuS	Induces succinate production	FL and AN	Deletion	4295377	AN-2K-1,2,3,4,5 and 6 and FL-4K-2,3,4,5,6 and 7
adhE	Disrupts ethanol production pathway	AN	SNP	1438030	AN-2K-1, 3, 4 and 5
			SNP	1439673	AN-2K-2, 7 and AN-4K-1, 2, 3, 4, 5, 6 and 7

[a] The gene mutated.

[b] The function of the wild type gene.

[c] The treatment under which the mutations were reported.

[d] The mutation types affecting the genes.

[e] The position the mutation occurred in the evolved strain – coordinates given are the position in the ancestral reference location in all cases (bps).

[f] The specific clones that possess the given mutation.

4.3.4.1.1.1. Mutations in nadR

There were four separate IS150 insertions at different positions within *nadR* in this present study (**Table 4.6**). NadR negatively regulates the transcription of NAD biosynthetic genes, particularly *nadA* and *nadB,* whose products catalyse the initial steps of the NAD biosynthetic pathway (229). Furthermore, NadR controls the transport of exogenous nicotinamide mononucleotide (NMN), the immediate NAD precursor, across the cytoplasmic membrane, and can catalyse NAD from nicotinamide adenosine mononucleotide and ATP (230) .

NadR is a bifunctional protein that regulates the levels of nicotinamide adenosine dinucleotide, (NAD^+ and NADH) (230). NAD and NADH are essential co-enzymens, required for the activity of over 300 proteins in the cell (231, 232). Co-enzymes play a central role in metabolism due to their ability to function as electron carriers, acting as electron acceptors in their oxidized state (NAD^+) and electron donors in their reduced state (NADH) in many biochemical redox reactions (232). During anaerobic growth, NADH is used to reduce metabolic intermediates produced during fermentation, generating NAD+ in the process.

Mutations in the *nadR* gene have been reported at a high frequency in LTEE with *E. coli* B strains under aerobic conditions (211). In the present study, however, mutations in the *nadR* gene only occurred in either the anaerobic or fluctuating lineages, suggesting there is strong selection for

nadR disruption in the anaerobic environment. Each mutation within the *nadR* gene was localized to the N-terminus, which is responsible for the DNA binding transcriptional activator activity. These mutations are likely to disrupt the negative repressor function of the NadR regulator. Therefore, disruption of *nadR* results in constitutive expression of *nadA* and *nadB* and is likely to increase the intracellular concentration of NAD (105). Fermentation pathways are highly responsive to the intracellular NADH/NAD^+ ratio (233, 234). This is likely to be an effective and global means for the cell to increase the activity of fermentation under anaerobic growth conditions (229).

4.3.4.1.1.2. Mutations in pflB

The use of REL4536 as the ancestral strain for this study allowed us insights into the recent (10,000 generations) evolutionary history of REL4536 that was the focus of Barrick *et al.* (52). Two loci involved in fermentation pathways, *pflB,* and the *dcuR/yjdI* locus, are of particular interest. Barrick,*et al.* (2009) reported an IS150 insertion in the *pflB* gene in REL4536 as compared to the genome of REL606. The *pflB* gene encodes pyruvate formate lyase, which is a central enzyme in the anaerobic fermentation network. PflB converts pyruvate to formate, thus inactivation of *pflB* is likely to be beneficial or neutral during long-term aerobic growth, as was the case in the LTEE study (52). However, a functional *pflB* gene is highly beneficial for growth in an anaerobic environment. In the anaerobic and fluctuating lineages only, the IS150 insertion in *pflB* was deleted, leaving a functional *pflB* gene. This mutation was also found among the anaerobically evolved *E. coli* lineages of S. Shewaramani (unpublished data). Interestingly, there was a higher instance of *pflB* reactivation in fluctuating lineages which may be the result of higher IS activity in the non-constant environment (**Table 4.6**).

4.3.4.1.1.3. Re-activation of dcuS

Breseq indicated a 5 bps deletion (GCGGC) in the *dcuR/yjdI* locus. Specifically, this occurred at 755 bps upstream of *dcuR* and 1,055 bps upstream of *yjdI* in six out of seven anaerobic lineages at 2,000, and six out of seven fluctuating lineages at 4,000 (**Table 4.6**). This region contained a *dcuS* pseudogene, and the 5 bps deletion resulted in a -2 frame-shift which restored the full length *dcuS* gene. The *dcuS* gene originally appeared to have become disrupted by a 5 bps slippage event at position 4,295,376 bps, resulting in the insertion of GCGGC downstream of a 15 base long GC tandem repeat string (GGGGCGCGGCGCGGC).

DcuS and DcuR function as a two component regulatory system. DcuS is the membrane-bound histidine kinase component and is responsible for sensing the concentration of external four-carbon dicarboxylates, such as malate and fumarate. Once active it phosphorylates the cytoplasmic response regulator DcuR (235). The DcuR protein contains a helix-turn-helix DNA binding domain repressor which binds to target promoters of genes including the fumarate reductase genes *frdABCD* (236). Activation of these genes is likely to reactivate the fermentation pathway that converts phosphoenolpyruvate to succinate (**Figure 4.3**), thereby reactivating one of the two fermentation pathways that had accumulated deleterious mutations during the 10,000 generations of evolution previous to this study.

4.3.4.1.1.4. Mutations in adhE

Two independent mutations occurred within the *adhE* gene in 13 clones sequenced (**Table 4.6**) and these mutations were found only in the anaerobic lineages (**Appendix Table 7.13**). An A → G transition at position 1,438,030 bps resulted in a cysteine residue being substituted for a tyrosine in codon 75 of the protein. This mutation was found in four separate clones at the 2,000 generation (AN-2K-1, 3, 4 and 5, see **Table 4.6**). Cysteine is a much smaller amino acid than the aromatic tyrosine. This substitution occurs within the N-terminal aldehyde domain of the protein and is likely to confer a conformational change, most likely resulting in a non-functional or disrupted protein. The second of the SNP mutations occurred within the C-terminal alcohol dehydrogenase domain at position 623, with a G → A transition point mutation. The G → A transition in *adhE* was reported among nine clones in the anaerobic environment (**Table 4.6**). It represents one of the two online mutations as it was reported at 2,000 and again at 4,000 generations in AN2 and AN7 (AN-2K-2, 7 and AN-4K-1,2,3,4,5,6,7 **Appendix Table 7.7**). This mutation results in the replacement of a threonine residue for an alanine residue. While similar in size, threonine has an attached alcohol in the R-group of the amino acid which is likely to lead to disruptions to the alcohol dehydrogenase domain. Furthermore, mutations in *adhE* were observed independently in an additional LTEE study of an *E. coli sbcC* knockout strain under anaerobic conditions (S. Shewaramani, unpublished data) and also by Puentes Tellez *et al.* 2013.

The *adhE* gene in *E. coli* encodes an 891 amino acid alcohol dehydrogenase protein, and functions in the fermentation pathway leading to the production of ethanol when glucose is metabolised under anaerobic conditions (see **Figure 4.3**). This protein works in a two-step process by reducing acetyl-CoA to acetaldehyde, and further reduces this to form the fermentation end-product

ethanol (233). AdhE is composed of three main functional domains, an N-terminal aldehyde dehydrogenase domain, a central aldehyde dehydrogenase domain and a C-terminal alcohol dehydrogenase domain.

There are two likely advantages to mutations in *adhE* under anaerobic conditions, both of which involve the diversion of the fermentation pathways of *E. coli* to obtain a more desirable end product. In the biotechnology sector, molecular genetic manipulation of genes involved in fermentation pathways is commonplace. Gene knock outs of *adhE* lead to desirable end-products by blocking the ethanol production pathway, resulting in the elevated production of acetate typically (213). Loss of function mutations in the *adhE* gene are likely to have had a similar effect in this study, diverting the fermentation pathways in *E. coli* to an end product that may continue or support growth, such as succinate or acetate (45, 237). *E. coli* B has been shown to reach high levels of biomass when grown on acetate but also to produce high volumes of acetate when grown on glucose, as compared to K-12 derived *E. coli* strains (238). Furthermore, deleterious mutations in the *adhE* gene would presumably reduce the concentration of extracellular ethanol, which in high quantities is toxic to the cell. High performance liquid chromatography (HPLC) analysis of filter sterilised anaerobic media after 24 hours of growth would address this hypothesis.

4.3.4.1.2. Modification of the toxin-antitoxin systems

The second broad mechanism in which multiple independent mutations arose was in toxin-antitoxin producing genes commonly used in post-segregational killing (239). Many bacteria contain extrachromosomal DNA, in the form of one or more plasmids, capable of autonomous replication. Plasmids may contain groups of genes which ensure their persistence in the cell during host replication, known as plasmid addiction systems. Generally, these systems comprise of a toxin producing module and a counteracting antitoxin module, collectively known as toxin-antitoxin (TA) sytems. If the plasmid is not passed on to a daughter cell, unstable antitoxin is degraded more rapidly than the toxin, resulting in death of the daughter cell. The systems have been classified into five different types, depending on their mode of activity (240). Some forms of plasmid addiction modules can be maintained on the bacterial chromosome, acquired through transfer from plasmid to chromosome (241, 242). Due to their apparent lack of benefit to the host, they have been termed selfish elements. In the present study, multiple mutations were found in genes

encoding two types of TA systems in REL4535, the *hok*/*sok* and *ldr* systems, it is noted that these genes are chromosomally located, and REL4536 does not possess a plasmid. The genomic distribution of genes encoding *hok/sok* and *ldr* systems in REL4536 is displayed in **Figure 4.4**, and the mutations observed in this study are presented in **Table 4.7**.

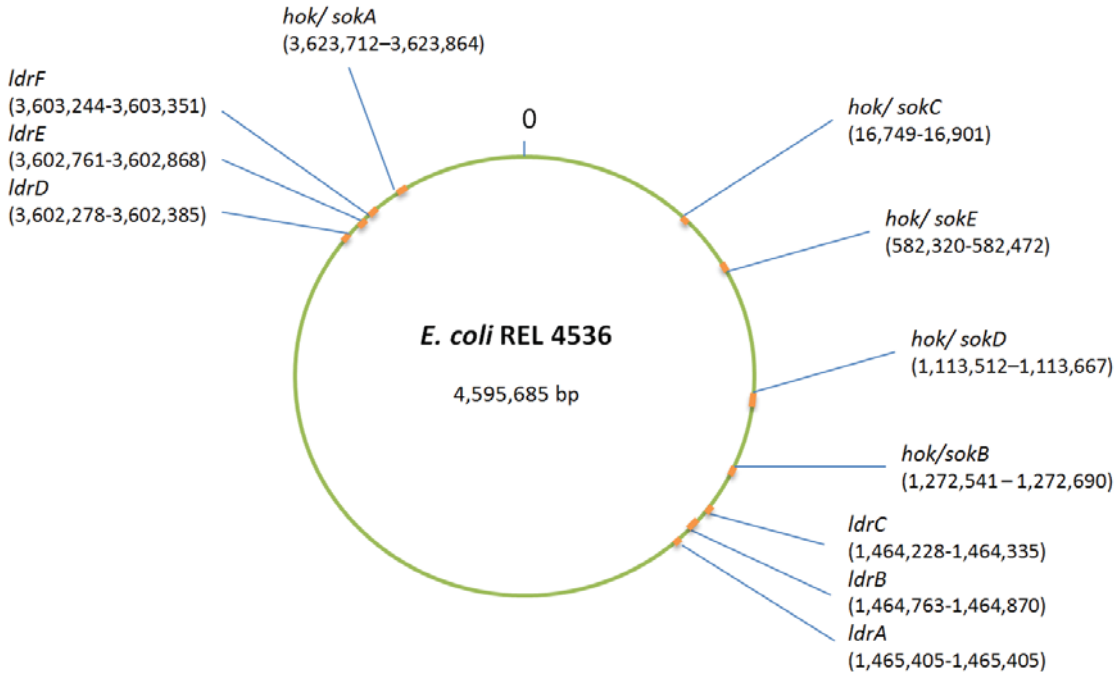

Figure 4.4: Location of *hok/sok* and *ldr* toxin-antitoxin system genes within the *E. coli* REL4536 genome.

Table 4.7: Mutations in TA systems within evolved lineages.

Toxin-antitoxin systems					
TA system [a]	**Locus** [b]	**Treatment** [c]	**Mutation events**		
			Mutation Type [d]	**Position** [e]	**Lineages affected** [f]
hok/sok	*hokC/nhaA*	All	IS150 insertion	16972	AE-4K-1 and 6, AN-2K-1 and 4, FL-2K-4 and FL-4K-2
			IS150 insertion	16989	FL-2K-7 and FL-4K-3
			IS150 insertion	16992	FL-4K-6
	insA-7/hokE	FL	IS150 insertion	582237	FL-2K-7
	ECB_01533/ hokD	FL	Deletion	1113343	FL-4K-5
			IS150 insertion	1113403	FL-4K-2, 3, 4, 6 and 7
	trg/mokB	AE and FL	Deletion	1272262	AE-4K-6
			IS150 insertion	1272399	AE-4K-1

			IS150 insertion	1272400	FL-4K-2, 3, 5 and 6
			Deletion	1272401	FL-4K-2, 3, 5 and 6
			IS150 insertion	1272468	AE-4K-2, AE-4K-5, AN-2K-1 and 4, AN-4K-2, 3, 4, 5 & 6 and FL-2K-2, 6 & 7. FL-4K-1, 4 & 7
ldr	*chaA/ldrC*	AE and FL	IS150 insertion	1464061	FL-4K-2, 3, 4, 5 and 6, and FL-2K-7
			IS150 insertion	1464143	AE-4K-6
	[ldrC] – ldrB		2 gene deletion	1464062	FL-4K-2
	[ldrC]		Deletion	1464062	FL-4K-5
	ldrC/ldrB		IS150 insertion	1464595	FL-4K-1
			IS150 insertion	1464672	FL-4K-6
			IS150 insertion	1464678	FL-2K-2
			IS150 insertion	1464679	AE-2K-1
	ldrB/ldrA		IS150 insertion	1465130	AE-4K-3
	ldrF		IS150 insertion	3603153	FL-4K-2
	ldrE		IS150 insertion	3602958	FL-4K-3, 5 and 6

[a] Type of the TA system mutated.

[b] Region affected by mutations.

[c] The treatment(s) under which each mutation was reported.

[d] The type of mutation that occurred within each gene.

[e] The position the mutation occurred in the evolved strain – coordinates given are the position in the ancestral reference location in all cases (bps).

[f] The specific clones that possess the given mutation.

4.3.4.1.2.1. Mutations in the hokC/nhaA *locus*

The *hokC/nhaA* intergenic region was mutated by the insertion of IS150 elements in three different locations (**Figure 4.4** and **Table 4.7**). The *nhaA* gene encodes a sodium ion/proton antiporter (243), while *hokC* is responsible for TA production, a common mechanism for a plasmid stability system (244, 245). The *hokC* gene and its homologs encode plasmid stability systems (244, 245). As the IS150 insertions occurred upstream of *hokC* these are likely to affect *hokC* expression. There is an IS186 element present in this region downstream of *hokC*. This may disrupt the function of the *hokC* gene as it interferes with the fbi (fold-back inhibitory) element (246). However, it is likely not to be as effective at disrupting the *hokC* gene as disruptions to the upstream *ucb* (upstream complementary box) region, and may explain why four IS element mutations were reported in this region (see **Figure 4.5**). Interestingly, the insertion at 16,972 bps which occurred in the aerobic, anaerobic and fluctuating lineages, was also reported in the

anaerobically grown *E. coli* by Shewaramani *et al.*, and also by Barrick *et al.* 2009, lending further support to a role in fitness enhancement at 2,000 and 4,000 generations. As the *hokC/nhaA* mutations occurred in all treatments it is likely that these mutations are adaptive to the general conditions in common between the treatments, such as the media, rather than specifically the treatments to which they were exposed.

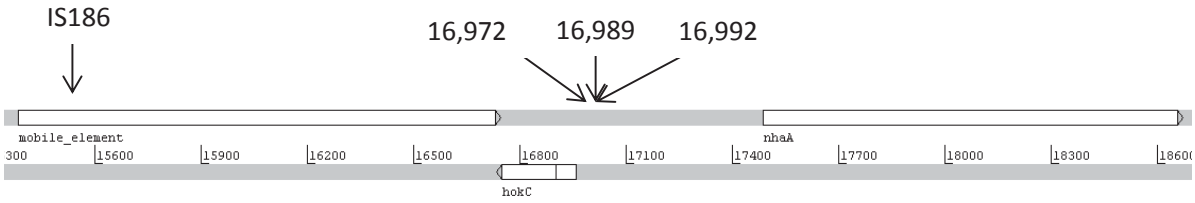

Figure 4.5: Mutation events located between *hokC* and *nhaA* genes in evolved lineages. The location of the three IS150 insertion sites are indicated, as is the pre-existing IS186 insertion element downstream of the *hokC* gene in the *E. coli* REL4536 genome as viewed in Artemis. Black arrows represent the locations of mutations arising in aerobic, anaerobic and fluctuating lineages as listed in **Table 4.7.**

4.3.4.1.2.2. Mutations in the trg/mokB *locus*

Mutations that arose in *hok/sok* TA systems were not limited to *hok*C. Between the *trg* and *mokB* genes, five mutation events were found among lineages in all three treatments, including two deletions (of 67 and 205 bases) and three separate IS150 insertions (13, 82 and 83 bases upstream of *mokB*) (**Figure 4.6**). There are many *cis* and *trans* signature sequences located upstream of *mokB* (246), suggesting that these mutations affect *mokB* expression. The IS150 insertion at 83 bps upstream of *mokB* disrupts the translational activator (*tac*) element, thereby affecting protein expression. The two deletions of 67 or 205 bps are just 15 bps upstream of the transcriptional start site and interupt the upstream complementary box (*ucb*), likely terminating the transcription of the *mokB* gene. Unlike the *hokC/nhaA* locus*,* there are no IS elements present in the *trg/mokB* locus, thus IS element insertion in this area is likely to affect *mokB* expression, leading to enhanced fitness (as described in **Section 4.3.4.1.2.4**).

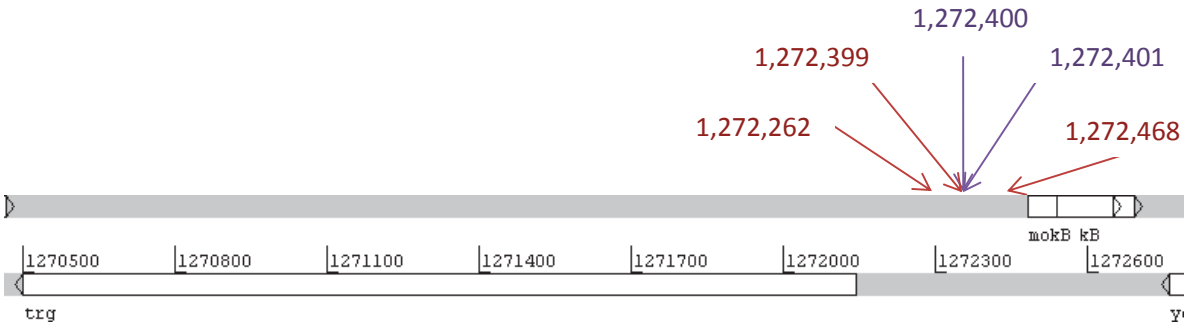

Figure 4.6: Mutation events located between the *trg* and *mokB* genes in evolved lineages. The location of the five mutation events in the *E. coli* REL4536 genome as viewed in Artemis. Red or purple arrows and font represent mutations in aerobic or fluctuating lineages, respectively, as listed in **Table 4.7.**

4.3.4.1.2.3. Mutations in the ECB_01533/hokD *locus*

The *ECB_01533/hokD* intergenic region was mutated by two events, a deletion of 60 bases and an IS150 insertion as seen in **Figure 4.7** and **Table 4.7**. The 1,103,326 to 1,133,222 bps within which the *hokD* lies, is the area in the *E. coli* genome that contains a cryptic prophage known as Qin. In addition to the events displayed in **Figure 4.7** and **Table 4.8**, a large, 25-gene deletion ranging from *ECB_01536* to *insE-3* took place in AE-2K-6 lineage. The *hokD* gene was among the genes excised from the evolved genome via this deletion event.

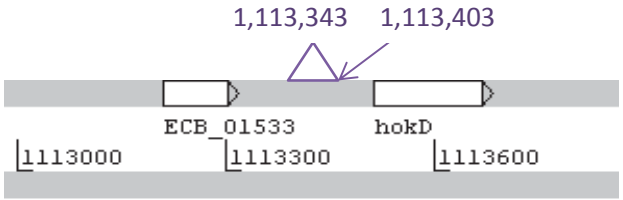

Figure 4.7: Mutation events located between the *ECB_01533* and *hokD* genes in evolved lineages. The location of the mutation events in the *E. coli* REL4536 genome as viewed. The purple arrow, triangle and font represent a mutation in the fluctuating lineages as listed in **Table 4.7.**

4.3.4.1.2.4. Mutation in the insA-7/hokE *locus*

The *insA-7/hokE* intergenic region was mutated by the insertion of a single IS150 element at 582,237 bps in FL-2K-7 (**Figure 4.8** and **Table 4.7**). Although not immediately close in proximity to the *hokE* gene, this mutation lies within the upstream regulatory sequence of the *hok/sok* gene family which can extend ~150 bps upstream of the gene (246). At this location, the single IS element insertion in FL-2K-7 is likely to disrupt the translational activator element (*tae* motif) of the *hokE* gene.

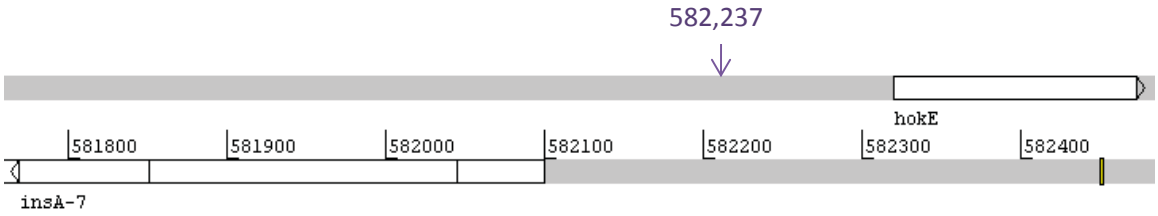

Figure 4.8: Mutation event located between the *insA-7* and *hokE* genes in evolved lineages. The location of the mutation event in the *E. coli* REL4536 genome as viewed in Artemis. The purple arrow and font represent a mutation in the fluctuating lineages as listed in **Table 4.7.**

4.3.4.1.2.5. Mutations in the ldr *locus*

There were 11 distinct mutation events reported in the long direct repeat (*ldr*) gene cluster regions (**Figure 4.9** and listed in **Table 4.7**). The majority of instances appeared to occur in the fluctuating lineages, with eight of the 11 reported events occurring in the fluctuating lineages at both 2,000 and 4,000 generations. Nine mutations were IS150 element insertions and two were deletions of IS150 elements of different sizes, occurring in FL-4K-2 and FL-4K-5. The deletion event in FL-4K-2 was the larger of the two, resulting in a deletion of *ldrC* and *ldrB* genes. The deletion in FL-4K-2 resulted in the deletion of *ldrC* and *ldrB*, whereas the deletion in FL-4K-5 resulted in a partial deletion of the *ldrC* gene.

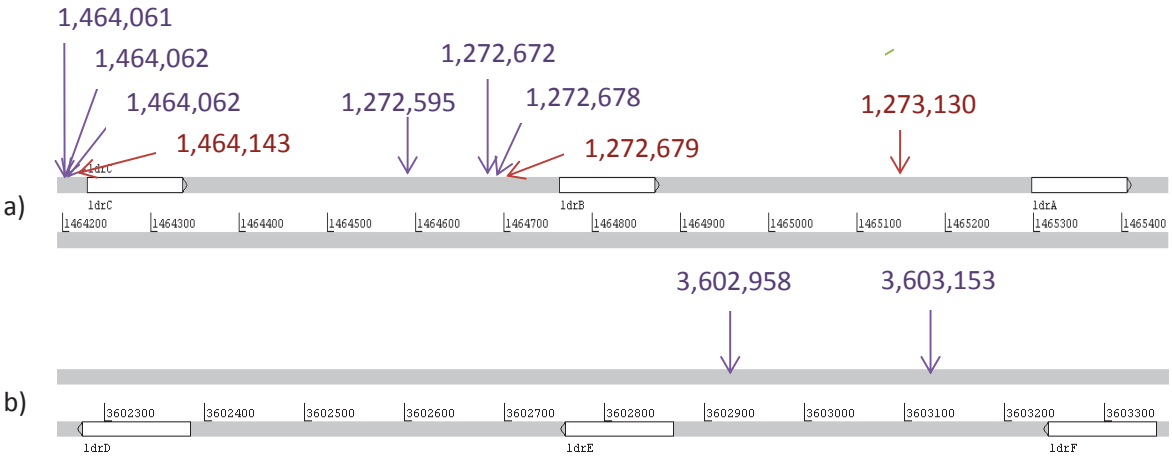

Figure 4.9: Mutation events located near the *ldr* gene clusters in *E. coli* REL4536. Figure of a) the *ldrABC* and b) the *ldrDEF* genes in the *E. coli* REL4536 genome as viewed in Artemis. Red or purple arrows and font represent mutations reported in the aerobic or fluctuating lineages, respectively, as listed in **Table 4.7.**

The *ldr* gene clusters are present in two parts, each containing three genes *ldrABC* and *ldrDEF* (247) (see **Figure 4.4**). Each gene is 108 bps in length, encoding a 35 amino acid protein, and genes are clustered together in repeats (247). The genes of the *ldr* clusters are symmetrically distributed on opposing ends of the *E. coli* genome (**Figure 4.7**), with *ldrABC* located between 1,464,228 and 1,465,405 bps and *ldrDEF* between 3,602,278 and 3,603,351 bps. *ldrABC* has 65% sequence similarity to *ldrDEF*. The clusters encode a TA system in *E. coli* and is the second example of a toxin-antitoxin phenotype being modified in this study.

4.3.4.1.2.6. Adaptation through mutation of toxin and antitoxin systems

Mutations in four of the five genes involved in the *hok/sok* TA system, and five of the six genes involved in the *ldr* system, were detected independently at a high rate in this LTEE study. While these two systems are not related at a sequence level, they represent two of the five known toxin-antitoxin systems associated with post segregation killing of host cells as a result of plasmid loss (240, 242, 248).

Mutations in chromosomally located TA genes have been reported in other LTEE studies (224, 244, 249). Possible reasons for TA gene inactivation are discussed. Firstly, post segregational killing is no longer the role of these TA systems, and such chromosomally located genes may be considered selfish DNA (248). However, a growing body of research suggests chromosomal TA systems may have other functions such as cell cycle arrest and programmed cell death, triggered under certain stressful conditions such as oxygen limitation (239, 248). The high prevalence of loss of function mutations in TA genes reported in this study may promote adaption by protecting against programmed cell death.. This may have been more prevalent in the fluctuating lineages, where a high instance of TA gene mutations were found. Here, cells would have had frequent exposure to anaerobic conditions . Alternatively, . loss of function mutations may eliminate any potentially lethal activity due to expression of the genes, or reduce the metabolic burden on the cell of expressing the TA system of the genes..

4.3.4.1.3. Deletions of cryptic prophages

Of the large deletions observed in this study, 20 out of the 50 events arose in genomic regions containing cryptic or remnants of prophages. There are nine regions of historic prophage integration in the REL4536 genome (**Figure 4.10**). They range in size from 5 to 37 kb and together, account for 4% of the genome. Prophages are bacterial viruses that have integrated into the bacterial chromosome. These can cause profound metabolic changes in a cell, including lysis, cell

disintegration and can ultimately kill off the population (250). Despite this, prophages contain many genes that may allow bacteria to cope with adverse environments (251). Once integrated into the host genome, an equilibrium between cost and benefit to both phage and host must be achieved (252, 253). Four of the largest deletions identified in the evolved genomes in the present study were all found to occur within regions of homology to the bacterial prophages 186, Qin, CP4-44 and to the highly diverged cryptic prophage P22 (**Table 4.8**).

Some of *E. coli* B's characteristic traits include the loss of flagella and associated mobility through the deletion of a 41kb region containing a cluster of flagella related genes present in most other *E. coli* strains (53). Deletions from 0.2 to 41 kb have been reported due to homologous recombination between adjacent IS elements (212).

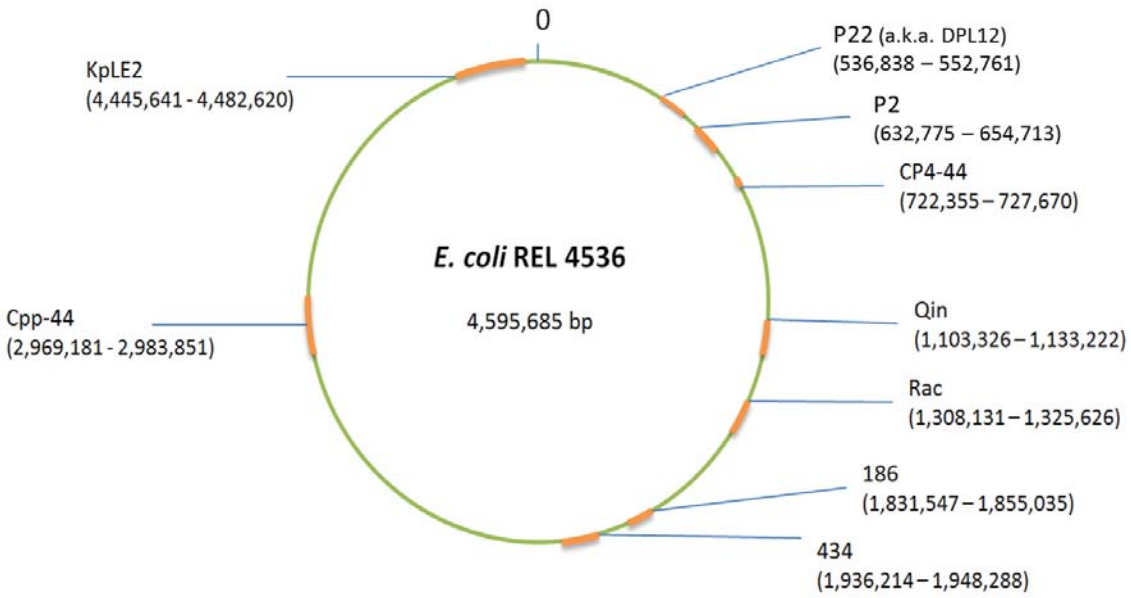

Figure 4.10: Locations of nine cryptic prophages within the *E. coli* REL4536 genome.

Table 4.8: List of prophage excisions reported in this study.

Prophage excisions					
Prophage [a]	**Locus** [b]	**Treatment** [c]	**Mutation events**		
			Number of genes affected [d]	**Gene range** [e]	**Lineages affected** [f]
P22 a.k.a. DLP12	536,838 – 552,761	AN and FL	29	*[insB-6] - ybdK*	AN-2K-7 and AN-4K-1, 2, 3, 4, 6 and 7
			7	*[insB-6] – [ompY]*	FL-2K-7

			33	*[insB-6] – [insA-7]*	AN-4K-5
186	1831547 - 1855035	FL	38	*ECB_00851 – ECB_00814*	FL-4K-2, 3, 5, 6
			37	*ECB_00851 – ECB_00815*	FL-4K-7
Qin	1,103,326 – 1,133,222	AE and FL	36	*[tqsA] – ECB_01533*	FL-4K-4
			5	*[ydfX] – ECB_01533*	FL-2K-2
			25	*[ECB_01536] – insE-3*	AE-2K-6
			1 (60 bps)	*ECB_01533-hokD*	FL-4K-5
P2	632,775 – 654,713	FL	27	*[ECB_02013]-ogrK*	FL-4K-4

[a] Prophage responsible for adjacent partial or complete deletion.

[b] The location of the resident prophage in the reference REL4536 genome sequence (bps).

[c] The treatment(s) under which each mutation was reported.

[d] Number of genes deleted from as a result of the prophage deletion event.

[e] The first and the last gene affected by the deletion event

[f] The specific clones that possess the given mutation, noting that independence of the anaerobic lineages could not be guaranteed after 2,000 generations.

4.3.4.1.3.1. Deletion of cryptic P22 prophage

Reports in the literature suggest that the 21 kb region from *argU* (536,838 bps) corresponds to a cryptic prophage similar to phage P22 (254-256), thus the deletions observed in the present study appear to result in the partial deletion of the prophage. Interestingly, Barrick *et al.* (2009) reported a large deletion that arose by 10,000 in the LTEE (the REL4536 strain), where a deletion event at position 547,700 bps excised 8,224 bps from the original REL606 genome. This deletion was just downstream of *insB-6,* suggesting high activity of this transposable element.

In the present study, several deletion events from *insB-6,* the transposase gene of an IS1 element, at position 547,451 bps were observed. The smallest of these deletions (4,874 bps containing seven genes) occurred in one fluctuating lineage, FL-2K-7. This deletion event excised *insB-6*, four genes of unknown function likely to be P22-phage specific *ECB_00514, 00515, 00516* and *00517, appY* and a partial deletion of *ompT.* The *ompT* gene is an outer membrane protein hypothesised to confer protection against the antimicrobial peptide protamine (257) and has been implicated in a range of processes, including virulence in urinary tract diseases, resistance to certain anti-microbial peptides and DNA excision repair (258). The second to last gene in the seven gene

deletion was *appY*. AppY is a member of the AraC/XylS family of positive transcriptional regulators (259). Induction of the *appY* gene is mediated through the global regulator ArcA. Under anaerobic conditions ArcA represses the transcription of genes involved in aerobic respiration and promotes the transcription of *appY*. AppY then activates known anaerobic energy metabolism genes such as pyruvate-formate lyase and hydrogenase 1 (254).

In the anaerobically evolved lineages, the P22 deletion events were much larger and encompass the same seven genes deleted in FL-2K-7. The largest spanned 33,293 bps, including 33 genes from *insB-6* to *insA-7*, while the other spanned 32,570 bps, including 30 genes from *insB-6* to *ybdK* (**Table 4.9**). Due to the presence of *insB-6* at the deletion junction in all cases, it is likely these deletion events were mediated through recombination of the IS1 mobile element to the downstream genes.

Within the 29 and 33 gene deletion regions is the *cus* locus, which encodes a copper ion efflux system (260). The locus is comprised of a two component sensory encoding operon, *cusSR,* which activates the expression of the *cusCBAF* operon in response to elevated concentrations of copper or silver. Copper is both an essential nutrient and a toxic substance to *E. coli,* and as such, *E. coli* has evolved many ways in which to monitor the intracellular concentration of the metal. Reports suggest that copper sensitivity is higher in anaerobically growing cultures (261). The lack of added copper in DM25 media may render the *cus* locus unnecessary under the anaerobic conditions of the current study, and its deletion may have been selected for.

Apart from *appY,* many of the downstream genes in the 29 and 33 gene deletions do not have an obvious function in the anaerobic environment. Following *appY* is *ompT*. Downstream of *ompT* is *envY* which encodes the EnvY DNA-binding transcriptional regulator. EnvY functions in cellular envelope assembly during stationary phase and at low temperatures (262). Nine genes were of unknown function, however appear to be related to phage. Two of these (*ECB_00529* and *ECB_00530*) were homologous to phage type IV secretion systems, which may mediate bacterial pathogenesis. Sequence similarity between the *attB* genes and tRNA genes have been shown to promote phage integration at these points (263-265). Other phage like proteins in the area include *nfrA*, important for bacteriophage N4 adsorption (266).

Given the importance of the *appY* gene in the anaerobic environment, it is curious why the P22 deletion events occurred so frequently in the anaerobic and fluctuating lineages and its deletion is

investigated further in **Chapter 5**. However, over-expression of *appY* has led to an elevated spontaneous mutation rate (267). Deletion of the *appY* gene may be selected for as a means of reducing the risk of evolving a mutator phenotype.

4.3.4.1.3.2. Deletion of cryptic 186 prophage

The two largest mutation events identified in this study occurred in the fluctuating lineages. The 37 gene deletion was from *ECB_00851* and *ECB_00814,* while the 38 gene deletion was from *ECB_00851* to *ECB_00815.* These mutations resulted in the deletion of prophage 186, which is exclusive to the *E. coli* B genome compared to other *E. coli* strains (53). The 186-like region contains only genes of unknown function, which are likely to be phage related. This coliphage has been shown to have negative effects on the growth of the host when introduced into *E. coli* K-12 strains (268). In the present study, it is likely that excision of this large region of DNA was beneficial to the host, though the deletion was only reported in the fluctuating lineages. It is likely that continuous variation to the environment mediated the excision of the 186 coliphage related genes in the host (251).

4.3.4.1.3.3. Deletion of cryptic Qin prophage

The Qin prophage region between the *ynfP* and *ydfJ* genes spans 26,896 bps of the *E. coli* REL4536 genome. There are 43 genes, of which 21 are of unknown function or display high homology with phage genes. In this study, three large deletion events ranging in size from five to 36 genes (see **Table 4.8**) occurred in the Qin region. These events all involved the deletion of *ECB_01536*, *ECB_01535, ECB_01534* and *ECB_01533*. In addition, there was a small deletion of 60 bps, within the region, specifically between *insA-7* and *ECB_01533* as reported in **Section 4.3.4.1.2.3**). That multiple independent deletions of portions of the Qin-like prophage were found among aerobic and fluctuating lineages suggests that deletion of this region is likely to be adaptive under these conditions.

4.3.4.1.3.4. Deletion of cryptic P2 prophage

The final example of a prophage excision mutation is the 27 gene deletion which occurred in the primary P2 attachment site (**Table 4.8**) in FL-4K-4. The 22 kb region between 632,775 bps and 654,713 bps shares very high sequence similarity to the 33 kb P2 prophage. Integration of this coliphage was likely via recombination between the *attB* site in *E. coli* and the P2 phage attachment gene *attP* (269). The prophage is thought to have become defective due to the loss of three parts of the P2 genome previous to this study and that of Lenski and Barrick (53). The P2

prophage was excised completely in one lineage (FL-4K-4) in the fluctuating condition (**Appendix Chapter 8.4**). The P2 prophage-like region is unlikely to contain any essential genes. Of the 27 genes, all are likely phage related genes such as *ogrK*, which codes for bacteriophage P2 late transcription (270). The excision of this region may be mechanism used by the host to streamline the genome of non-essential genes (271).

4.3.4.1.3.5. Adaptation through prophage excisions

Phage related sequences in bacterial genomes can account for a significant proportion of sequence diversity between isolates and strains (272). Resident phages have been shown to be important for processes such as growth, protection against quinolone and β-lactam antibiotics, increased resistance to osmotic, acidic and oxidative stress (251, 273, 274). Beneficial genes on the host chromosome of prophage origin are often maintained due to the fitness increases they impart in some environments. However, many resident prophages do not contain beneficial genes to the host but rather comprise a collection of phage related genes. The presence of the prophage can also impair the fitness of the host in a novel or non-constant environment, for example when in an environment where the metabolic cost of prophage mediated motility or maintenance outweighs the benefit.

Prophage deletions have recently been identified as responsible for a high percentage of large scale deletions in LTEE with the *E. coli* B strains by Raeside *et al.* (149). Here, 34% of the large deletions were attributed to phage excision from the evolved lineage genomes. Similar to the findings of this thesis, Raeside *et al.* (149) report a high rate of Qin, P2 and P22 prophage excisions, accounting for 21, 32 and 35% of all prophage excision events. Also in agreement were the lack of excisions of three prophages in particular; Rac, Cpp-44 and KpLE2. Investigation of genes associated with these prophages did not uncover any genes or operons with apparent essential function that may explain their lack of excision. It is feasible that large deletion of non-essential DNA is likely to be selected for in evolving populations. In this way a compact and streamlined bacterial chromosome will be the outcome (253). A particularly high prevalence of prophage excision events found within the non-constant fluctuating environment supports this view. A potential reason for this may be that frequent changes in cell physiology as the environment is changed are likely to impose a considerably greater metabolic burden on the cell, and any kind of streamlining will help make the cell more efficient and thus be selected for.

4.3.4.1.4. Inactivation of virulence determining genes

This classification refers to a potential adaptive pathway that was identified based on numerous mutations reported in the *agn43* gene and also genes within the *kps* operons. In total, these two locations reported 19 different mutation events outlined in **Table 4.8**.

Table 4.9: Inactivation of virulence genes.

Inactivation of virulence genes					
Gene [a]	**Function** [b]	**Treatment**[c]	**Mutation events**		
			Mutation type [d]	**Position** [e]	**Lineages affected** [f]
flu a.k.a. *agn43*	Outer membrane auto-transporter	AN	Duplication	2972604	AN-4K-4
		AN	SNP	2972858	AN-2K-7 and AN-4K-1,2,3,4,5,6 and 7
		AE	IS1 insertion	2972936	AE-2K-5 and 7
		AN	SNP	2973574	AN-2K-7 and AN-4K-1,2,3,4,5,6 and 7
		AN	Deletion	2973807	AN-4K-4
		AN	SNP	2974237	AN-4K-1, 2, 4, 5, 6 and 7
		FL	IS150 insertion	2974776	FL-4K-7
		AE	IS150 insertion	2974778	AE-4K-6 and 7
		AE	IS150 insertion	2975127	AE-4K-1
kpsT	Inner membrane polysaccharide transport	AE	Deletion	2999898	AE-2K-6
		AE	SNP	3000095	AE-2K-3
		AE	SNP	3000161	AE-2K-1
		AE	SNP	3000346	AE-2K-5
kpsM	Inner membrane polysaccharide transport	AE	IS3 insertion	3000514	AE-4K-6
		AE and FL	IS3 insertion	3000519	AE-4K-5 and FL-2K-3
		AE	IS3 insertion	3000527	AE-4K-1,2,3,4,5 and 7
kpsE	Polymer translocation	AE	SNP	2987334	AE-2K-2
kpsD	Polymer translocation	AE	SNP	2988653	AE-2K-3
kpsS	Post translational modification	AN	IS1 insertion	2992382	AN-2K-1,3,4,5 and 6

[a] Gene mutated.

[b] The function of the wild type gene.

[c] The treatment(s) under which each mutation was reported.

[d] The type of mutation that occurred within each gene.

[e] The position the mutation occurred in the evolved strain – coordinates given are the position in the

ancestral reference location in all cases (bps).
[f] The specific clones that possess the given mutation, noting that independence of the anaerobic lineages could not be guaranteed after 2,000 generations .

4.3.4.1.4.1. Mutations in agn43

Antigen 43 (Ag43), encoded by *agn43* (previously known as *flu*) is an abundant 948 amino acid outer membrane auto-transporter in *E. coli.* Auto-transporters are typically virulence factors associated with infection or virulence in pathogenic bacteria. Consistent with other outer membrane auto-transporters, the protein is composed of an N-terminal signal or α domain processing site, and a C-terminal β barrel domain translocation unit that, together, form the transporter (275). Ag43 is a multifunctional protein and promotes bacterial binding to some human cells as well as promoting biofilm formation on various substances (276). The *agn43* gene was mutated among lineages in all three treatments via nine independent mutation events.

Mutations in *agn43* occurred most frequently in the anaerobic lineages, accounting for five of the nine events. Four of these mutations were localised to the α domain of the protein. The four mutations were i) a duplication of nine bases at 218 amino acids into the protein, ii) a T → G SNP at amino acid 158, substituting an arginine for methionine, iii) a G → T SNP at amino acid 397, substituting a tryptophan for a glycine, iv) a six base deletion at 2,973,807 bps removing an aspartic acid and isoleucine from within the α domain. The remaining mutation that arose in the anaerobic environment but not in the α domain was reported in the β translocation unit and is a G → A SNP, resulting in a threonine substitution for a glutamine at position 618. In the aerobic environment there were two IS insertions in the β translocation unit (IS150 at 2,974,778 bps and another IS150 insertion at 2,975,127 bps), while the IS1 insertion occurred in the α domain at 2,972,936. Finally, in the fluctuating environment, one IS150 insertion occurred within the β translocation unit at 2,974,776 bps.

The α passenger domain of Ag43 consists of many amino acid motifs but, interestingly, mutagenesis and subsequent functional analysis of the mutant proteins has shown that none of these individual motifs are important for the proteins' function (275). Similarly, mutations in the β translocation unit do not appear to impact significantly on the known functions of the Ag43 protein. In the aerobic and fluctuating environments, the four mutation events all consisted of IS insertions, and thus are likely to disrupt protein function. However, mutation events in the anaerobic environment resulted only in changes to single amino acids, or an in-frame duplication.

This suggests that minor changes in protein folding may have resulted in minor changes in the function of the autotransporter, which may have been beneficial in the anaerobic environment.

4.3.4.1.4.2. Mutations in the kps *cluster*

The *kps* cluster is organised into three regions in REL4536, located between 2,986,242 and 3,001,283 bps. The cluster is involved in the assembly of capsular subunits that is comprised of K antigen polysaccharides, which are frequently implicated in pathogenesis (277, 278), particularly in the uropathogenic (UPEC) *E. coli* strains. The *kps* cluster is divided into three regions; region one and three contain *kpsFEDUCS* and *kpsMT* respectively and are separated by region two containing region *kfiABCD* (279) (**Figure 4.11**). Region one and three are highly conserved and involved in the assembly and transport of the capsular material. In contrast, region two is highly variable and specific to each serotype. The *kfi* genes are directly responsible for the biosynthesis of the capsular material.

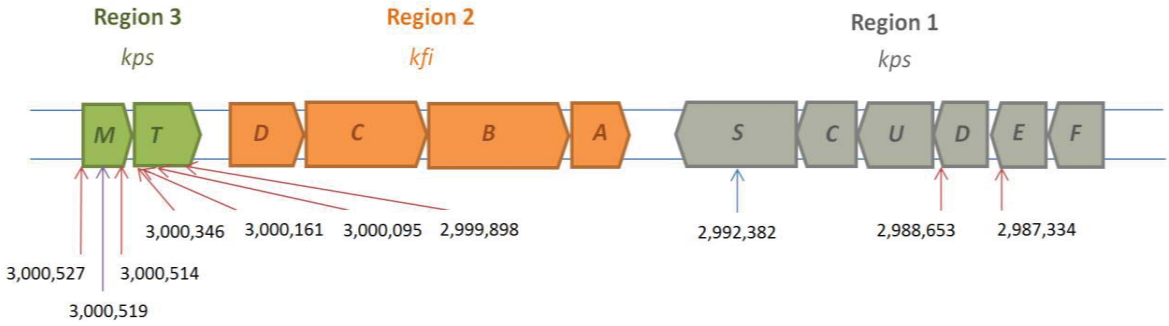

Figure 4.11: The *kps/kfi* operons. The three regions of the *kps/kfi* gene cluster are indicated in green, orange and grey (Region 3, Region 2 and Region 1 respectively), with the orientation of each gene depicted by large filled arrows. Ten mutation events identified in the evolved lineages are indicated. The position of each mutation is shown, and the treatments under which they originated by arrow colours (red for aerobic lineages, light blue for anaerobic and purple for one fluctuator mutation). Figure adapted from Silver *et al.* (280).

Mutations were found in the *kpsE, D, S, M* and *T* genes (as shown in detail in **Figure 4.11**). Both KpsE and D are involved in translocation of the K antigen polymer across the periplasm and onto the cell surface (281, 282). KpsS is involved in post-translational modification of the K antigens (279, 283) and *kpsS* was the only gene mutated in anaerobically evolved lineages. Both *kpsM* and *kpsT* were mutated in aerobic lineages, with the former also being mutated in the fluctuating

condition. Both are directly involved in the transport of the K antigen polysaccharide (284). KpsM is the trans membrane domain (285), and KpsT provides the ATP hydrolysis necessary for the ATP-binding cassette system to work (286).

This *kps* cluster was subject to the highest number of independent mutation events reported in this study. An IS150 insertion has been reported in the Ara-1 population of *E. coli* by Barrick *et al.* (52) in the *kpsD* gene. However, we report that mutations in this gene cluster are not limited to the aerobic environment, but also may have an adaptive role in the anaerobic and fluctuating treatments. The majority of mutation events (eight out of 10 events) were reported in the aerobic environment while *kpsS* and *kpsM* were mutated in the anaerobic and fluctuating lineages (**Table 4.9**).

4.3.4.1.4.3. Adaptation through loss of function of virulence gene

Many roles have been attributed to Ag43, from biofilm formation, aggregation, adhesion and virulence in UPEC strains. Virulence traits are likely to be beneficial to pathogenic strains of *E. coli* in the wild. There is evidence to show that the K capsules may confer resistance to some phages ,such as bacteriophage T7 (277), which presumably would be heavily selected for in the wild. However, in the controlled conditions of the laboratory-based LTEE study, it is hypothesised that these genes are a costly metabolic burden and thus accumulate either alteration or loss of function mutations during evolution in DM25 minimal media under different treatments. In the present study, these virulence associated genes may accumulate mutations as a result.

4.3.4.1.5. Mutations of the *brnQ* gene

BrnQ is an amino acid Na^+ symporter specific for the uptake of branched chain amino acids such as L-leucine, L-isoleucine and L-valine, collectively known as the LIV family (287, 288). Eight independent mutations arose in *BrnQ* during the course of this study, six of which occurred in the fluctuating lineages with the remaining two arising in aerobic lineages (**Table 4.10**).

Figure 4.10: Modification of the *brnQ* gene.

BrnQ					
Gene [a]	**Function [b]**	**Treatment [c]**	**Mutation events**		
			Mutation type [d]	**Position [e]**	**Lineages affected [f]**
brnQ	Branched chain amino acid transporter	FL	Deletion	380365	FL-4K-1
		FL	Insertion	388020	FL-2K-3
		AE	IS150 insertion	388234	AE-4K-1, 2, 3, 4 and 6
		AE and FL	IS150 insertion	388275	AE-4K-6 and FL-2K-6, FL-2K-4 and FL-4K-1
		FL	IS150 insertion	388527	FL-4K-2,3,4,5 and 6
		FL	IS150 insertion	388530	FL-4K-7
		FL	IS150 insertion	388543	FL-2K-2
		AE	SNP	389105	AE-4K-5

[a] Gene mutated

[b] The function of the wild type gene

[c] The treatment(s) under which each mutation was reported

[d] The type of mutation that occurred within each gene.

[e] The position (bps) where the mutation event occurred as reported on the ancestral genome sequence.

[f] The specific clones that possess the given mutation.

Of the eight mutation events, five were mediated by IS150 insertions at various locations. An ATCA insertion at 388,020 bps resulted in a frame-shift leading to a premature stop codon. In FL-2K-6, *brnQ* was partially deleted by a large deletion of six genes starting at the *araJ* gene. The final mutation was a SNP in aerobic lineage AE-4K-5 which resulted in a nonsense mutation. These eight mutations are all likely loss of function mutations due to the disruptive activity of IS element insertions.

4.3.4.1.6. Mutations of the *cycA* gene

A single mutation in the *cycA* gene is reported in each environment. An IS150 insertion occurred at 4,381,583 bps within all seven of the 4,000 aerobic lineages, 13 anaerobic lineages and all 14 fluctuating lineages. The *cycA* gene encodes a transporter involved in the uptake of glycine, D-serine and D-alanine and thus cells with a disrupted *cycA* show decreased transport of these amino acids. Deletion mutants are reported to have a decreased rate of glycine uptake. The CycA protein is a H^+ symporter and a member of the amino acid/ polyamine/ organocation (APC) superfamily. Mutated CycA confers resistance to D-cycloserine inhibition (289). Why this would be selected for, or even be advantageous in each of the three treatments in an antibiotic-free and

amino acid-free media, is unknown. The *cycA* and *brnQ* mutations are most likely a result of adaptation to the growth format, rather than adaptation to the treatments under study.

4.3.4.1.7. Mutations of *rpo* genes

The *rpo* genes were the only functionally similar genes to have been mutated independently multiple times specific to the aerobic condition. Members of the *rpo* gene family are involved in the synthesis of each of the five core subunits of the RNA polymerase holoenzyme. Crucial components of the holoenzyme are the sigma factors, which enable specific binding of the RNA polymerase core to gene promoters. Sigma factors used to initiate transcription vary depending on the environment, allowing flexibility in changing growth conditions.

In this study, two independent SNPs mutations were reported in *rpoD.* The *rpoD* gene encodes sigma factor 70, which enables targeted specificity of the RNA polymerase core to promoters of genes transcribed during exponential growth phase (290). One of the two *rpoD* mutations was a T → G transversion in AE-4K-5 at position 3,119,112 bps, which resulted in an aspartic acid substitution of tyrosine at the 143rd amino acid of the peptide chain. The second *rpoD* SNP mutation was a C → A transversion occurring in AE-2K-6 at 3,119,520 bps, which resulted in a serine replacement of an arginine at position 279 of the polypeptide. Both of these substitutions occurred within the RNA polymerase sigma factor domain region two, which is involved in the binding of the sigma factor to the core RNA polymerase. This region has been referred to as the non-essential domain as it may be mutated without apparent loss of function (291).

The other *rpo* gene to be mutated was *rpoS.* The mutation, a C → A transversion at 2,734,340 bps within the AE-4K-4 genome, resulted in the substitution of a cysteine in place of an arginine at the 229th amino acid position. This places the mutation in the sigma factor domain region three, which is between the regions that recognise the -35 and -10 regions of transcription promoters (292). RpoS encodes the alternative sigma factor σ^S, known as the master regulator for the general stress response. However, mutations in *rpoS* are common among both natural and laboratory populations, particularly those involving nutrient limiting conditions (293-296).

4.3.4.1.7.1. Adaptation through GASP mutations

Mutations have been reported in the *rpo* genes in many adaptive evolution studies, particularly in *rpoB* (52), *rpoC* (107, 297, 298) and *rpoS* (155). The *rpo* genes encode key components of the *E. coli* regulatory networks, and their mutation can lead to changes in global transcription control.

Mutants of RpoC have been shown to increase the growth rate of K-12 strains growing in glycerol minimal media (299).

Mutations in the *rpoS* and *rpoD* genes may result in the growth advantage in stationary phase (GASP) response (120, 122, 300). GASP mutations confer a scavenging phenotype on cells which allows for nutrient recycling during periods of prolonged stationary phase (122). Four different GASP mutations have been reported, each differing markedly at the molecular level. The first identified was among the *rpo* gene family, specifically the *rpoS* gene. Mutations are generally not loss of function, but described as attenuated phenotypes, with decreased *rpoS* activity (125). As binding of the various sigma factors to the RNA polymerase holoenzyme is on a competitive basis, variants of alternative sigma factors are thought to be a means of the cell shifting the balance in favour of the most ideal sigma factor (120). Mutations in the *rpoS* gene have been shown to confer to the cell the ability to catabolise amino acids, nucleic acids, fatty acids and peptides which may be present in the media in later stationary phase due to the accumulation of dead and lysed cells (300).

The occurrence of a GASP mutation specific to the aerobic environment is not unexpected. Aerobic lineages spend on average 16 hours in stationary phrase as compared to lineages grown in anaerobic conditions, which spend a maximum of four hours in stationary phase and thus, are unlikely to develop GASP mutations. Furthermore, RpoS mutated populations had a competitive advantage over wild type populations under glucose limited conditions in other studies (293, 301).

4.3.4.1.8. Mutations of the *pcnB* gene

Another set of mutations that affected polymerase enzymes but arose solely in the anaerobic environment, were found in *pcnB*. The *pcnB* gene encodes poly(A) polymerase 1, which is responsible for polyadenylation of 3' ends of mRNA. In prokaryotes, polyadenylation of mRNA leads to decreased mRNA stability (302). Two mutations were found in *pcnB*, the first of which occurred at position 161,094 bps, a deletion of two bases (AT) leading to a frame-shift mutation that resulted in a premature stop in transcription of the gene, disrupting its function. This mutation was reported in eight anaerobic lineages (AN-2K-7 and AN-4K-1, 2, 3, 4, 5, 6 and 7). The second mutation occurred in AN-2K-3, a G → A transition mutation resulting in arginine substitution by cysteine at position 59 of PcnB. This mutation is presumed to be detrimental to PcnB function, as it occurs in the conserved head domain. The incorporation of cysteine is likely to

be associated with changes in protein structure, as cysteine residues commonly form covalently bonded disulphide bridges that affect protein folding and stability.

Polyadenylation of the mRNA has been suggested to act as an extended binding site for multi-protein RNA degradation complexes (303, 304). A disrupted *pcnB* gene leads to a decrease in mRNA decay (302). With the slower growth rate of REL4536 in the anaerobic environment, a reduced rate of RNA turnover may be advantageous and would likely be acted on and selected for in this environment.

4.3.5. Exploring the adaptive landscapes

From Section 4.3.4, there were eight putative adaptive pathways identified across all treatments during the 4,000 generations of this thesis (identified in **Section 4.3.4**). Each of the putative adaptive pathways are expected to have accompanying changes in fitness, with different routes along the same pathways also likely to have characteristic mean fitness responses. In **Chapter 3**, relative fitness of the evolved lineages was assessed at the population level. Thus, the putative evolutionary pathways were examined in light of the fitness responses that the adaptive mutations may have caused. Analysis was focused on 2,000 generation data as both fitness and genomic data were available at this time, and cross-contamination between lineages (as described in **Section 4.3.3.2**) was presumed to be very limited at the 2,000 generation stage.

4.3.5.1 Evolutionary pathways undertaken by aerobic lineages

Lineages that had adapted to the aerobic environment were observed to have varied fitness responses (**Figure 3.5**) by 2,000 generations. These ranged from 1.31 ± 0.09, 1.15 ± 0.12 and 0.91 ± 0.05 in AE-2K-2, AE-2K-3 and AE-2K-7 respectively. During the adaptation of lineages to the aerobic landscape, different mutations among the *kps* gene cluster (such as *kpsE, D* and *T*), as well as the *agn43* gene, resulted in alternative evolutionary pathways with different mean fitnesses in the aerobic environment.

Among the three lineages tested for competitive fitness, the only lineage to have significantly increased in fitness within the 2,000 generations was AE-2K-2, with a relative mean fitness of 1.31 (two sample t-test, $P < 0.05$). In total, AE-2K-2 had acquired four SNP mutations (**Appendix Chapter 7.6**). Based on the extent of parallelism among the *kps* operon in the present study, it is reasonable to attribute the increase in fitness to the non-synonymous A → G transversion in *kpsE,* rather than the other three SNP mutations. As described in **Section 4.3.4.1.4.2,** this gene is

responsible for K antigen transport to the cell surface and may be the most effective of the virulence gene mutations which confers the greatest fitness response. However, it seems unlikely that this mutation alone would be responsible for the 31% increase in relative mean fitness (77, 209, 305). Thus, it is likely that there may have been some positive epistatic relationship with the *kpsE* mutation and the other three mutations in the AE-2K-2 lineage (**Appendix Chapter 7.6**) (136).

While AE-2K-3 did increase in fitness, this increase was not significant (1.15 ± 0.12, two sample t-test, P = 0.14) (**Table 4.11**). However, this lineage also reported mutations in the *kps* gene cluster (non-synonymous SNPs in *kpsD* and *kpsT* (**Appendix Chapter 7.3**) were identified in the AE-2K-3 sequenced clone).

Genomic analysis of AE-2K-7 revealed five mutations (**Appendix Table 7.3**). A decline in fitness was observed in AE-2K-7 (0.91 ± 0.07) (**Table 4.11**), however this was not significant (two sample t-test, P = 0.14). Two of the five mutations were associated with mutated virulence genes, a G → C transversion in *kpsT* and an IS1 insertion in *ag43*. The *kpsT* mutation suggests another route in the modification of virulence genes was taken. However, with the lack of increased fitness in this lineage, it is likely this particular pathway is of neutral effect to the cell, or alternatively some form of strong negative epistatic interaction occurs with the *agn43* mutation.

Table 4.11: Putative adaptive traits occurring in lineages at 2,000 generation during this LTEE.

Environment of selection [a]	Lineage tested [b]	Putative adaptive trait [c]	Genes affected [d]	Aerobic fitness *(ω)* [e]	Anaerobic fitness *(ω)* [f]
Aerobic	AE-2K-2	Inactivation of virulence genes	*kpsE*	1.31 *	0.83 *
	AE-2K-3	Inactivation of virulence genes	*kpsD & T*	1.14	1.02
	AE-2K-7	Inactivation of virulence genes	*kpsT, agn43*	0.91	0.79 *
Anaerobic	AN-2K-1	Modification of fermentation pathways	*adhE, dcuR/S, nadR*	0.99	1.37 *
		Inactivation of virulence genes	*kpsS,*		
		TA system inactivation	*trg/mokB*		
	AN-2K-4	Modification of fermentation pathways	*adhE, dcuR/S*	1.06	1.36 *
		Inactivation of virulence genes	*kpsS*		
		TA system inactivation	*hocC/nhaA, trg/mokB*		
	AN-2K-6	Modification of fermentation pathways	*dcuR/S, nadR*	0.93 *	1.41 *
		Inactivation of virulence genes	*kpsS*		
Fluctuating	FL-2K-2	BrnQ	*brnQ*	1.08	1.37 *
		Prophage excision	*Qin prophage*		

			excision		
		TA system inactivation	*ldrC/ldrB, ECB_01533/hokD, trg/mokB*		
		Modification of fermentation pathways	*pflB*		
	FL-2K-3	BrnQ	*brnQ*	1.19 *	1.23 *
		Inactivation of virulence genes	*kpsM*		
		Modification of fermentation pathways	*pflB*		
	FL-2K-7	Modification of fermentation pathways	*pflB*	0.96	1.07
		BrnQ	*brnQ*		
		Prophage excision	*P22 prophage excision*		
		TA system	*hokC/nhaA, trg/mokB chaA/ldrC, ECB_01533/hokD, insA-7/hokE*		

[a] Growth condition under which the lineage was subjected in LTEE.

[b] Lineage with relative fitness and genomic data.

[c] Hypothesised functional network modified by various means during adaptation to the respective environment.

[d] Genes mutated in the functional network in each lineage during adaptation.

[e] Mean relative fitness of lineages determined at the population level in the aerobic environment.

[f] Mean relative fitness of lineage determined at the population level in the anaerobic environment.

* Statistical significance as determined by two sample t-tests with $P < 0.05$.

Barrick *et al.* 2009 observed 15 new mutations in the Ara-1 population between the 10,000 to 15,000 generations (52). One such mutation occurred in a gene identified as a putative adaptive trait in this thesis. Specifically, inactivation of the *kpsD* virulence gene by an IS150 insertion was an online mutation that originated between 10,000 and 15,000 generations of adaptation to aerobic DM25 in the Barrick study. However, between these time-points, mean relative fitness of 1.40 at 10,000 declined to 1.38 at 15,000, suggesting only very low increases in fitness as a result of this mutation or that epistasis may have played a role, similar to the findings of this thesis (**Section 3.3.3.2.1**). Other putative adaptive mutations that occurred in parallel between this thesis and Barrick's study beyond 15,000 generations of evolved lineages included mutations in the *rpo* genes.

4.3.5.2 Evolutionary pathways undertaken by anaerobic lineages

The fitness responses of the anaerobically evolved lineages were much greater than those of the aerobically evolved lineages (**Figure 3.6**). As REL4536 had previously adapted to the aerobic environment, and the anaerobic environment was a novel environment for the organism, rapid adaptation was expected in this environment. Variability in fitness responses among anaerobic lineages was far less extensive than that for aerobic lineages adapting to the aerobic environment (**Figure 3.5**). Among the three lineages tested, AN-2K-1, AN-2K-4 and AN-2K-6 all exhibited significant increases in mean relative fitness when compared to the ancestor at the 2,000 time-point (1.37 ± 0.06, 1.36 ± 0.07 and 1.41 ± 0.03 respectively with $P < 0.01$ in all cases) (**Figure 3.7** and **Table 4.12**).

A higher number of mutations was identified in the anaerobic lineages than the aerobic lineages within the first 2,000 generations. On average, 14 mutations were identified in each of the anaerobic lineages, compared to four mutations within the aerobic lineages. Despite the larger number of mutations, the evolutionary pathways taken by the anaerobic lineages during adaptation were broadly similar. AN-2K-6 had the highest relative fitness increase, and acquired mutations that modified the fermentation pathways (with mutations that disrupted *nadR* and restored *dcuS*) and modified virulence genes (an IS insertion in the *kpsS* gene) (**Table 4.12**). Modifications of virulence genes among the anaerobic lineages represent an adaptive pathway also reported in aerobic lineages. Despite similar pathways found among lineages in the different environments, different routes were likely taken in each treatment. For example, in the aerobic lineages, *kpsE, D* or *T* were mutated, while in the anaerobic lineages, only mutations in *kpsS* were found (**Appendix Chapter 7.9**). Both AN-2K-4 and AN-2K-6 possessed the two putative adaptive pathways reported in AN-2K-6 while also having the TA system inactivation pathway. Both lineages had mutations in the *hokC/nhaA* locus, where AN-2K-4 had an additional mutation in the *trg/mokB* locus (**Appendix Chapter 8.12**). Despite this, AN-2K-1 and AN-2K-4 both displayed slightly lower fitness increases than AN-2K-6 with relative mean fitness of 1.37 and 1.36 respectively. This may be evidence of negative epistasis (136) among mutations in the TA inactivation and/or the fermentation pathway modification systems.

4.3.5.3 Evolutionary pathways undertaken by fluctuating lineages

Evolutionary pathways undertaken by lineages adapting to the fluctuating treatment were more similar to those of the anaerobic environment than the aerobic environment. Of the three fluctuating lineages tested for competitive fitness in both environments (FL-2K-2, FL-2K-3 and FL-

2K-7), FL-2K-2 and FL-2K-3 displayed significant increases in fitness in both environments (**Table 4.11**). In both cases fitness increases were greater in the anaerobic environment (1.37 ± 0.12 and 1.23 ± 0.02 in anaerobic conditions versus 1.08 ± 0.10 and 1.19 ± 0.10 in aerobic conditions for FL-2K-2 and FL-2K-3 respectively). In addition, the average number of mutations in each lineage adapting to the fluctuating environment was 13, similar to that found for the anaerobically evolved lineages (14 mutations per lineage on average), and much higher than the aerobic environment (four mutations per lineage on average).

Of the three lineages tested, FL-2K-2 and FL-2K-3 displayed significant increases in fitness in the anaerobic environment and both lineages possessed mutations in five of the putative evolutionary pathways identified in **Section 4.3.4.1**. Two of these pathways were identified as predominantly occurring in the fluctuating lineages; BrnQ and prophage excision. All three of the fluctuating lineages examined displayed independent putative loss of function *brnQ* mutations (**Appendix Table 7.5**). As discussed in **Section 4.3.2.1.4.1**, BrnQ is a branched chain amino acid transporter responsible for the transport of leucine, valine and isoleucine and how this is likely to be beneficial in a non-constant environment is currently unclear. The other adaptive trait predominantly identified in fluctuating lineages was that of prophage excision, which among the lineages tested for competitive fitness occurred in FL-2K-2 and FL-2K-7 (**Tables 4.11**). As with the BrnQ mutations, the mutations specific to each lineage and therefore routes of adaptation were not similar between the two lineages. In FL-2K-2, five genes of the cryptic Qin prophage were deleted from the genome (**Appendix Chapter 7.5**). In FL-2K-7, a partial deletion of a P22-like cryptic prophage took place at 546,986 to 551,143 bps (**Appendix Chapter 7.5**).

Of the five different adaptive pathways taken by the fluctuating lineages, a common pathway taken with the aerobically evolved lineages was that of the virulence gene mutation pathway. Among the fluctuating lineages this was found in FL-2K-3 (**Appendix Chapter 7.5**). While the pathways were similar, the genes mutated within each pathway were not. Adaptation to aerobic conditions resulted in mutations in the *kpsE, D* and *T* genes. In the anaerobic environment *kpsS* was primarily mutated, while in the fluctuating treatment, *kpsM* mutations were reported (**Appendix Chapter 7.5**). Thus, different routes of the virulence gene mutation pathway were taken in the three different treatments.

In comparison with the anaerobic environment, different routes along the same pathway were taken. In each of the three fluctuating lineages, modification of the fermentation network

occurred. However, the means by which the network was modified differed in fluctuating lineages, via an IS150 deletion within the *pflB* gene (**Appendix Chapter 7.5**). TA system inactivation was seen in all three fluctuating lineages. FL-2K-2 reported a mutation in the *ldrC/ldrB* and *ECB_01533/hokD* loci, FL-2K-3 reported a mutation between *ECB_01533/hokD*, while FL-2K-7 reported five separate TA inactivations including mutations in *hocC/nhaA, trg/mokB, ECB_01533/hokD, insA-7/hokE* and *chaA/ ldrC* (**Appendix Chapter 7.5**).

The relative mean fitness of the fluctuating lineages increased to 1.22 ± 0.07 on average in the anaerobic condition at 2,000 generations. These increases did not come at a cost to fitness in the aerobic environment, with an average aerobic fitness increase of 1.07 ± 0.05 at 2,000 generations. Results from mutation analysis indicated that the fluctuating lineages undertook similar evolutionary pathways during adaptation as anaerobic lineages during specialisation to the anaerobic environment. This was feasible as there was little cost in fitness in the aerobic environment during anaerobic specialisation (**Figure 3.8**). Pathways taken by anaerobic specialists during adaptation to the anaerobic environment generally did not come at a cost to fitness in the aerobic environment (**Figure 3.6**). Generalist evolution may have also occurred due to additional pathways such as BrnQ mutations and prophage excisions, which were identified as having a higher occurrence in the fluctuating lineages. Most likely a combination of these two factors is responsible but further investigation would shed light on this issue. Also of note in the fluctuating lineages was a high incidence of mutations in genes of unknown function. This suggests that at least some of these genes of unknown function may have a role in regulating cell responses to non-constant environments.

4.4. Summary

The genetic changes underlying the changes in fitness of *E. coli* populations that evolved in aerobic, anaerobic and fluctuating growth treatments were investigated via whole genome re-sequencing. Sequence data indicated that cross contamination between replicate lineages had occurred amongst the anaerobically evolving populations, likely via a low level of cross contamination of lineages, which complicated the interpretation of the data for identifying independently arising mutations between lineages. However, a set of caveats were developed to distinguish adaptive mutations that were likely to have been selected for during evolution in each treatment.

Mutation analysis also indicated that IS element movement may have contributed considerably to adaptation. Movement of IS elements can affect both genome structure and gene expression. IS elements can eliminate the function of many genes via inserting into regulatory elements or within genes or alternatively, activate neighbouring genes by acting as jumping promoters due to the presence of outward acting transcriptional promoters on their sequence. Gene inactivation through IS element insertion is likely to be more effective than SNP mutations in disrupting gene function, and effects may be transient, offering some genome plasticity during adaptation. This is likely to be advantageous to lineages that are exposed to a non-constant environment (48).

The relative rates of IS activity among the 12 LTEE lineages established by Richard Lenski is known to vary (221), however, the lineage that REL4536 was derived from did not have a high rate of IS movement. Within the aerobic lineages, the average copy number (3) of IS elements remained constant throughout the 4,000 generations. However, in the anaerobic and fluctuating lineages, the average copy number of IS150 in particular increased by 7 and 12 copies per lineage. Thus, it is likely that the observed incidence of IS150 mutations in the anaerobic and fluctuating environments maybe the result of adaptation to these new environments, and selection for mutations that promote this. Possibly exposure to the anaerobic environment, where energy generation is more limited as compared to growth in an aerobic environment, may result in stresses that activate IS150 element activity. .

The evolutionary pathways that lineages undertook during adaptation to the different treatments were investigated. Evidence for adaptive mutations was primarily reliant on the occurrence of parallel mutations arising in different treatments and in independent studies, as well as independent mutations arising in the same genes/metabolic pathways within lineages of the same treatments. For the aerobically evolved lineages, there was limited increase in fitness due to the pre-adapted state of REL4536 ancestral cells in this environment. However, mutations responsible for this adaptation were through inactivation of virulence genes. Furthermore, by 4,000 generations, mutations in the *rpo* genes seemed likely to be adaptive through a GASP effect.

The anaerobically evolved lineages exhibited the greatest increases in fitness of the three treatments. There were three distinct functional groups of mutations through which adaptation to the anaerobic environment occurred: i) modification of fermentation pathways, ii) TA system inactivation and iii) inactivation of virulence genes. Anaerobic specialisations arose by taking different routes along similar pathways. While the pathway itself was similar to that of the aerobic

lineages adapting to the aerobic environment, the specific genes mutated in the pathway differed. Thus, similar pathways with mutations in different genes resulted in adaptation to the two environments.

Two pathways occurred at a higher frequency within the fluctuating lineages. The first pathway is that of prophage excision, which potentially streamlines the bacterial chromosome of accessory genes which could be selected for in a non-constant environment. The second adaptive pathway reported at high frequency in the fluctuating lineages was the mutated BrnQ pathway. The evolutionary significance of mutations in the *brnQ* gene is unclear at present. In addition, fluctuating lineages adapted to non-constant aerobic and anaerobic conditions via different routes, along shared pathways with both aerobic and anaerobically evolved lineages, including inactivation of virulence genes, modification of fermentation pathways, and TA system inactivation.

To confirm the fitness enhancements from the putative adaptive mutations identified in this chapter, further experimental verification is desirable. Allelic replacement studies, in which a hypothesised adaptive mutation is recreated in the ancestral strain and subsequently competed against the ancestor, would allow for direct measurements of the fitness enhancement due to the introduced mutation. Where multiple putatively adaptive mutations are apparent in evolved strains, mutant strains recreated with combinations of multiple adaptive mutations would identify the nature of the epistatic interactions between them (61, 136). However, overall we have identified eight putative adaptive pathways taken by *E. coli* REL4536 in the aerobic, anaerobic and fluctuating adaptive landscapes. The underlying adaptive mutations responsible were uncovered and their resulting fitness responses were inferred.

Chapter Five : Investigating the origin and maintenance of diversity in the anaerobic lineages

5.1. Introduction

Experimental evolutionary lineages were established in aerobic, anaerobic and fluctuating treatments with the expectation that the relative fitness of each evolving lineage would increase due to adaptation to each treatment. From Chapter 3, the fitness responses of the lineages within each treatment varied. Aerobic lineages displayed marginal increases in fitness in the environment of selection, while fluctuating lineages experienced fitness increases in both aerobic and anaerobic conditions by 4,000 generations. Anaerobic lineages however, initially displayed an increase in fitness within 2,000 generations, but fitness (relative to the ancestral type) declined dramatically thereafter. The fixation of genotypes with significantly reduced growth rates in the anaerobically evolved lineages was not anticipated.

In addition to the decline in fitness of the strictly anaerobically grown lineages, phenotypically distinct colony polymorphisms arose within the first 1,000 generations of evolution. Diversity in colony morphology was specific to the anaerobic lineages in this LTEE. Polymorphic colonies were distinguished based on colony size when the evolved populations were plated on LB agar plates and grown aerobically, where typical colony morphotypes (TCM) which were similar in size to the ancestor, and small colony morphotypes (SCM), were observed.

As reported in **Chapter 3**, all anaerobic populations were comprised of only SCM during the later stages of the experiment, with each displaying very low population densities (~10^4 CFU/mL at 3,000 and ~10^3 CFU/mL at 4,000 populations). Throughout the 4,000 generations, populations which contained SCM were associated with low population densities. However, populations containing SCM were not consistently associated with low fitness. By 2,000 generations, four of the seven populations were polymorphic (AN1, AN2, AN4 and AN6), displaying both TCM and SCM colonies, with one population (AN7) comprised of only SCM. Of the seven anaerobic populations, competition assays were performed on AN-2K-1, AN-2K-4 and AN-2K-6 with all populations

displaying significant increases in fitness at the 2,000 generation time-point. Despite this, a number of the single clones isolated from these populations for whole-genome sequencing were difficult to grow (**Chapter 4**).

The aim of this chapter is to understand the molecular basis of the fitness decline that occurred in the anaerobic environment. Furthermore, I aimed to determine if there was a correlation between the declining fitness and origin of polymorphic populations observed in the anaerobically evolved lineages.

5.2. Objectives

The objectives of this chapter are:

1. To gain insight into the genetic basis of the origin and maintenance of diversity within anaerobically evolving populations.
2. To better understand how fitness declined in the anaerobically evolved populations.

5.3. Results and discussion

5.3.1. SCM within anaerobic lineages

In this LTEE study, morphological diversification among evolving populations was only found in the anaerobic environment, and not in the aerobic or fluctuating treatments. SCM were first observed in the AN2 lineage at approximately generation 375 (**Figure 5.1**). To confirm the SCM were not external contaminants three checks were performed; i) cultivation on TAra agar plates to confirm that cells were unable to utilise arabinose, consistent with the ancestor, REL4536, ii) T5 and T6 phage phenotypes were examined (**Section 2.2.5.4**) to confirm the characteristic *E. coli* B strain phenotype and iii) the 16S rRNA gene was sequenced to ensure colonies were not external contaminants. In all cases tests proved negative for contamination. Over the 4,000 generations of this LTEE, the occurrence of SCM was observed in other anaerobic lineages, and by 4,000 generations, all anaerobic populations were SCM. The extent of SCM abundance through the various anaerobic populations is listed in **Table 5.1**.

The appearance of divergent/polymorphic colonies within an initially homogenous culture is not unique to this study (45, 306, 307). The appearance of SCM in this LTEE is not unique to LTEEs nor to other non-long-term based experiments. Other small colony sub-populations have previously

been described as small colony variants (SCVs), first reported in *Salmonella typhi* in 1910, and subsequently identified in *E. coli* in 1946 (308). They are described as a slow-growing sub-population of bacteria, visualised as pin-point colonies after 24 to 72 hours of growth on agar plates. Since their first description they have been identified in a wide range of bacterial genera and species. Biochemically, they are distinct from their wild type counterparts by a number of properties other than growth rate. One factor thought to explain the reduced rate of growth is that defective pathways involved in electron transport among SCV, such that metabolic processes like oxidative phosphorylation are impaired. Furthermore, SCV are auxotrophs, dependant on compounds including CO_2 and thiamine in order to synthesise menadione (309, 310). Clinically, they have been recovered from specimens such as abscesses, blood, bones, and joints (311-313). SCVs from *S. aureus* and *P. aeruginosa* are frequently identified in the airways of sufferers of cystic fibrosis (314, 315). Typically, the SCV sub-population is more resistant to antibiotics than the parent population from which they arose (316).

In addition, polymorphic colonies were also reported in one of the 12 populations established by Prof Lenski in 1988. The Ara-2 population had developed two distinct clones designated LG (large) and SL (small) by 6,500 generations. Based on similarities between the strains, media, overall experimental design and description of the LG and SL morphotypes SCM in this LTEE are analogous to those reported in the LTEEs of Lenski and colleagues. Each clone was morphologically and genetically distinct and the two morphotypes have remained in a balanced co-existing state for more than 30,000 generations (71). Plucain *et al.* (108) found that at least three mutations that displayed epistatic interactions were responsible for the evolution of this balanced co-existence. In other studies, single point mutations have been identified as responsible for maintaining balanced polymorphic populations. These include diversity seen among *E. coli* K-12 in LTEE performed in chemostats (45, 317).

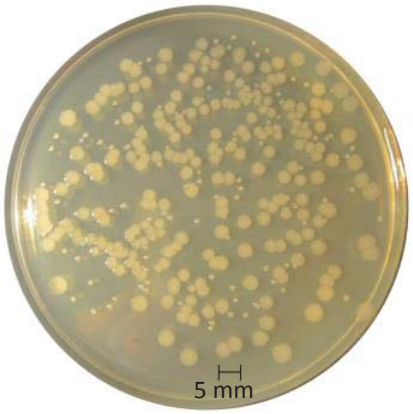

Figure 5.1: Agar plate containing typical and TCM and SCM morphotypes. The picture indicates an LB agar plate onto which 100 µL of AN-2K-2, diluted to 10^{-4}, was spread. Plates were incubated aerobically at 37^{o}C for 30 hours to readily distinguish between the different morphotypes. Scale bar indicates 5 mm.

Table 5.1: Population composition of anaerobic lineages throughout 4,000 generations.

Lineage	1,000 generations	2,000 generations	3,000 generations	4,000 generations
AN1	Mix	Mix	Mix	SCM
AN2	Mix	Mix	SCM	SCM
AN3	TCM	TCM	Mix	SCM
AN4	TCM	Mix	SCM	SCM
AN5	TCM	TCM	WT	SCM
AN6	TCM	Mix	SCM	SCM
AN7	SCM	SCM	SCM	SCM

5.3.2. Genetic basis for SCM in the anaerobic environment

To determine the genetic basis of the SCM in this study, whole genome sequence data obtained at 2,000 generations were analysed. At the 2,000 generation time-point, three of the seven clones sent for sequencing were phenotypically confirmed to be SCM, AN-2K-4, AN-2K-6 and AN-2K-7, with AN-2K-1, AN-2K-2, AN-2K-3 and AN-2K-5 displaying TCM phenotype (**Table 5.1**). Mutations that had arisen in each clone were listed and grouped according to the colony phenotype of the clone, either TCM or SCM colonies. A further sub-classification of the SCM group was made based on the proportion of SCM to TCM colonies within the populations from which they were isolated. As such, clones isolated from populations that were of mixed morphology (AN-2K-4 and AN-2K-6) were grouped together (SCM Type 1), while AN-2K-7 was grouped separately as it was isolated from a population that was comprised fully of SCM (SCM Type 2). The three groups (SCM Type 1, SCM Type 2, and TCM) were compared in an analysis to find common mutations between them (**Figure 5.2**). In total, 50 mutations from the seven anaerobic 2,000 generation clones were included. The lists of each mutation for each group were compared using Venny (318) and results are displayed in **Figure 5.2**.

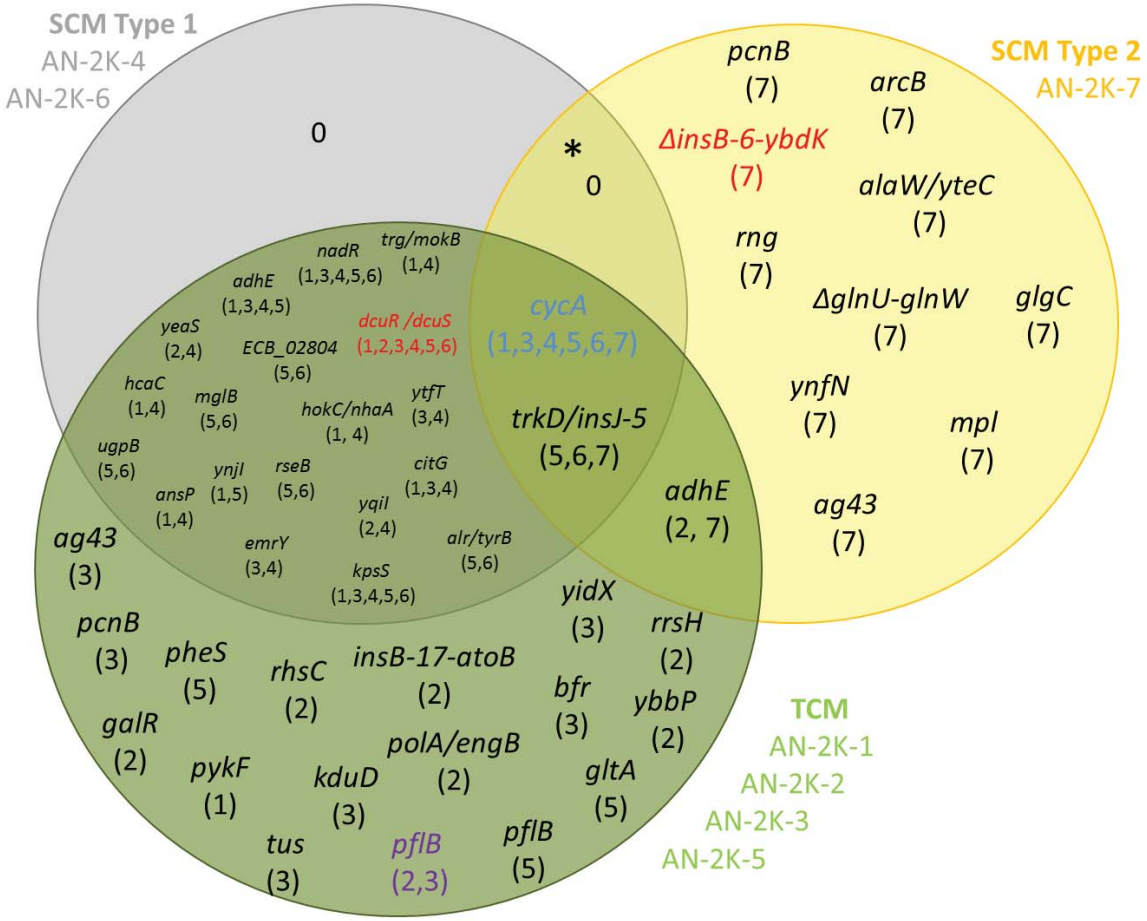

Figure 5.2: Venn diagram of collective mutations in each of the three colony morphotype groups (SCM Type 1, SCM Type 2, and TCM) from 2,000 generation anaerobically evolved clones. Mutations and loci from clones that were SCM Type 1 and 2, and TCM are displayed as grey, yellow and green sets, respectively. The lineage number(s) in which each mutation was found is indicated in parentheses under each mutation. Mutations that were also found in a similar study by S. Shewaramani, unpubliushed data are indicated in red (for mutations from SCM clones), purple (from TCM clones) and blue (from both SCM and TCM clones). Figure adapted from Venny output (318).

In the closely related LTEE study by Shewaramani *et al.* (S. Shewaramani, unpublished data) (**Section 4.3.3.3**), SCM also arose under anaerobic conditions within 1,000 generations of sub-culturing in anaerobic DM25. This alternative LTEE study was similar in design, a batch serial sub-culture experiment and also performed under aerobic and anaerobic conditions. Identical operating protocols, procedures and equipment such as 24-well plates for culture growth and aerobic and anaerobically prepared DM25 media were used. However, the ancestral strain for the LTEE study was *E. coli* REL4536 with a disruption in the *sbcC* gene, a gene responsible for DNA

repair. Whole genome sequencing from the *sbcC* LTEE experiment uncovered similar mutations to those identified in this thesis across multiple lineages. Mutations that were identified in parallel among the two studies are indicated in **Figure 5.2.**

The lack of mutations in the yellow-grey set (marked with "*" in **Figure 5.2**), representing exclusively common mutations between SCM Type 1 and Type 2, indicated that no single mutation was correlated with clones from both SCM Types, and not in clones with TCM clones. Therefore, it was likely that more than one pathway was responsible for the SCM phenotype in the anaerobic condition in this LTEE. Thus, rather than looking for mutations in common between SCM Type 1 and SCM Type 2, mutations were investigated that were exclusive to each group, as these mutations were most likely to be responsible for different pathways for the generation of the SCM phenotype.

For SCM Type 1, mutations were examined that were exclusive to the grey set (mutations that occurred exclusively to the AN-2K-4 and AN-2K-6 clones) (**Figure 5.2**). Due to the lack of mutations in this set, mutations were considered from sets that included mutations also reported in TCM clones, denoted with the light and dark green sets (**Figure 5.2**) and are discussed in **Section 5.3.2.1**. Additionally, mutations were examined that were exclusive to the yellow set for SCM Type 2 (**Figure 5.2**). There were 10 mutations that occurred exclusively in the yellow set, and these are discussed further in **Section 5.3.2.2**.

5.3.2.1. Colony polymorphism due to multiple mutations

Due to the lack of mutations that occurred exclusively to AN-2K-4 and AN-2K-6, the genetic cause of the SCM Type 1 phenotype remains ambiguous. Both SCM Type 1 clones shared mutations in common in five loci (*dcuR/dcuS*, *cycA*, *nadR* and *kpsS*), but these mutations were also found in TCM clone AN-2K-5. It is likely that particular combinations of mutations are responsible for the SCM Type 1 phenotype, but the combination is unclear from this dataset.

Similar studies in the literature, which attempt to infer the mutations responsible for a phenotype, have reported the important role that epistasis between multiple pleiotropic mutations may have on the development of the observed phenotype (58, 108, 132, 319). Plucain *et al.* (108) investigated the genetic basis of two populations that had diverged into LG and SL lineages from Prof. Lenski's Ara-2 population. Mutations in three regulatory genes (*spoT*, *arcA* and *gntR*) were hypothesised to be responsible for the mechanism whereby the population polymorphism

evolved. Individual mutations were introduced separately, and in combination, to the ancestor and the phenotypes confirmed based on whether the two morphs were maintained via frequency-dependent fitness effects, allowing the morphs to stably co-exist. Introducing the mutations individually displayed very different results on the fitness of the ancestor; the *spoT* mutation increased fitness, as did the *arcA* mutation, while the *gntR* mutation was deleterious. However, in combination, the recreated strain was reported to stably co-exist with LG lineages. Plucain *et al.* (108) also found that other lineages had the same mutations in all three genes but did not display the SL phenotype. Thus, the authors hypothesised that the establishment of the SL lineages was a multistep process with each step dependent on its ecological and genetic context.

E. coli B populations repeatedly evolved into metabolically distinct phenotypes known as slow switchers (SS) and fast switchers (FS) when grown in media containing both glucose and acetate as a carbon source (132). Herron *et al.* (9) focused on uncovering the genetic mechanism responsible for this diversity in metabolic switching and is particularly relevant due to the hypothesised high concentration of acetate in the anaerobic media in this thesis (**Chapter 4.3.4.1**.). Mutations in three genes were identified as responsible for the generation of diversity, namely *spoT*, *rbs* deletion and *nadR*. Among *spoT*, *rbs* and *nadR*, mutations in both *spoT* and *rbs* were present in REL4536, relative to REL606, and thus present in the evolved strains in this LTEE. An A → T transversion mutation occurred within *spoT* at position 3,762,741 bps in the REL606 genome (position 3,735,547 bps in REL4536). The *rbsDACBKR* operon was deleted from the REL606 genome in a deletion event from *Δkup-yieO* containing 6,934 bp (52). This also took place within the first 2,000 generations of evolution of REL606 (111) previous to the origin of REL4536. The mutation in *spoT* has been shown to be highly beneficial (104). This gene encodes a protein that controls the levels of ppGpp, an important effector molecule, and when mutated, is known to reduce lag phase during growth with glucose as the carbon source. Three different *nadR* mutations were reported in 11 of the 14 anaerobic lineages in this thesis, described in **Section 4.3.4.1.1.1**. While none of these three genes are directly involved in acetate metabolism, these mutations were hypothesised to allow for some genome flexibility which allowed for diversification. Both *spoT* and *nadR* are global regulators which are known to have highly pleiotropic effects once mutated (232) and may have had a role in the generation of the SCM phenotype in this thesis also.

5.3.2.2. *Colony polymorphism due to* ΔinsB-6-ybdK

Biodiversity arising from clonal populations may be mediated by as little as a single mutation. For example, the generation of polymorphic colonies due to a single mutation within a subset of the population was found in *E. coli* K-12 evolving in chemostat cultures (45). A single T → A point mutation in the promoter region of the *acs* gene allowed for acetate scavenging and the maintenance of diversity. From **Figure 5.2** there were 10 candidate mutations to choose from that were responsible for the SCM Type 2 phenotype (yellow set in **Figure 5.2**).

The SCM Type 2 group was comprised solely of mutations from the AN-2K-7 clone. Among the mutations found in the SCM Type 2 clone, 10 mutations were specific to the clone, and three mutations were common to clones with TCM or SCM Type 1 phenotypes (**Figure 5.2**). Of the mutations not occurring in other groups, *ΔinsB-6-ybdK* was deemed the most likely to be responsible for the generation of the SCM Type 2. Three lines of evidence support this view. i) This mutation occurred exclusively in SCM clones, and was not observed in any TCM clones. The deletion event in this area was reported to have occurred as a deletion of a cryptic prophage (**Chapter 4.3.4.1.3.1**). ii) The same mutation was also identified in parallel among two clones that displayed the SCM morphotype in anaerobic lineages of the *sbcC*-disrupted LTEE study (S. Shewaramani, unpublished data). iii) This same mutation event was reported by Le Gac *et al.* (320) in the SL lineages of Prof. Lenski's Ara-2 population when investigating the co-existence of the LG and SL cell types. mRNA transcription profiling found that the SL genotypes reported a dramatic decline in transcription rates in the region between 547,701 and 588,493 bps when comparing gene transcription data to LG genotypes. Genomic analysis confirmed that the reduced transcription rate between 547,701 and 588,493 bps was due to a large deletion event, encompassing 30 genes between *insB-6* and *ybdK* (Δ*insB-6-ybdK*), the same mutation event as indicated in this thesis. Thus, the Δ*insB-6-ybdK* is a strong candidate for being responsible for the population divergence that led to the SCM Type 2 sub-population.

Further identification of the genetic causes underlying the different SCM types is required. This could be achieved by screening the LTEE populations to identify when SCM first appeared, and from this time point, whole genome sequencing of both the SCM and TCM colonies from one population could reveal the mutations specific to the SCM colonies relative to the TCM. Increased acetate production during anaerobic growth is hypothesised to be at least partially responsible for anaerobic population diversification into the TCM and SCM morphotypes. A serial sub-culture

experiment with anaerobic DM25 with both glucose and acetate as carbon sources may uncover if an isogenic ancestral population would diversify into TCM and SCM morphotypes again in the anaerobic environment.

5.3.3. Evolutionary dynamics of polymorphic populations within the anaerobic environment

5.3.3.1. Existence of a stable equilibrium

To investigate if a stable equilibrium could be maintained between the TCM and SCM type clones within polymorphic populations, reciprocal invasion experiments were performed (**Chapter 2.2.9**) on AN-2K-4 and AN-2K-7 populations. These populations were selected as representatives of the two types of SCM identified in **Section 5.3.2** SCM Type 1 (AN-2K-4) and SCM Type 2 (AN-2K-7). As there were no TCM colonies in the AN-2K-7 population, the ancestral REL4536 culture was used as a representative of the TCM sub-population, and its interaction with the SCM Type 2 colony was observed. Results from reciprocal invasion experiments are displayed in **Figure 5.3**.

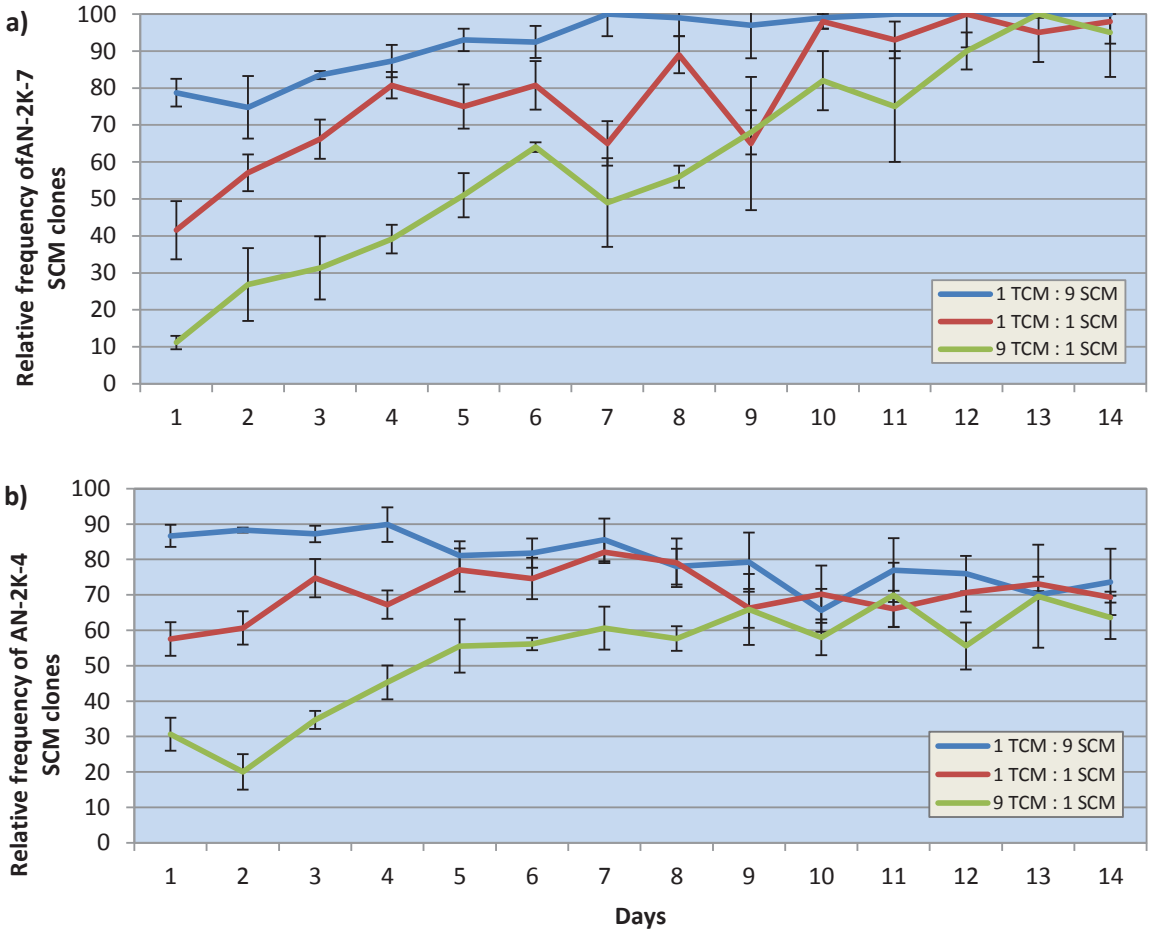

Figure 5.3: Investigation of co-existence between SCM and TCM. a) TCM and SCM from AN-2K-4 (SCM Type 1) and b) TCM ancestral REL4536 and SCM from AN-2K-7 (SCM Type 2). Displayed in both graphs are the relative frequencies of SCM morphotypes within the populations, initiated at three different starting ratios: 1 TCM: 9 SCM, 1 TCM: 1 TCM and 9 WT: 1 SCM. Serial transfers into fresh anaerobic DM25 medium took place every 24 hours.

From results in **Figure 5.3 a** and **b**, differences between the frequencies of the two different SCM converged to different means. In **Figure 5.3 a** both SCM and TCM morphotypes from AN-2K-4 could invade each other when each was initially rare. The frequency of the SCM declined slightly when it was initially common, and increased when initially rare such that, within 11 days, all populations with different starting ratios, converged to a stable equilibrium of 71.01 ± 3.19 ($F_{2,9}$ = 4.24, $P > 0.05$, single factor ANOVA), based on nine populations at day 11, which remained equal up to day 14. This is in contrast to results from the AN-2K-7 populations (**Figure 5.3 b**)**.** Instead of converging to an equilibrium, all three SCM and TCM populations, despite their different starting ratios, converged to 100% SCM frequency. The rate of fixation differed for populations with differing initial ratios of SCM to TCM cells, and was inversely proportional to the amount of SCM initially present in the populations; the lower the starting concentration of SCM, the longer it took to reach fixation. By 13 days the relative frequency of SCM morphotypes for all populations, regardless of initial starting ratios, had converged to fixation ($P > 0.05$, two tailed t-test) at day 13.

The capacity of each type to increase in frequency when rare indicates the operation of negative frequency dependent selection. Such interactions can maintain diversity. The existence of such interactions indicates that the two genotypes must occupy different niches. It is possible that such opportunity is provided by resource partitioning. It is likely that negative frequency dependent selection maintained a stable polymorphism between SCM and TCM within the AN-2K-4 anaerobic population, likely via the production of acetate by the TCM sub-population. However this was not the case with the AN-2K-7 population. In contrast, when TCM ancestral cells were mixed at different ratios with AN-2K-7, SCM and TCM failed to reach a stable equilibrium and instead out-competed the TCM cells in all cases within 13 days. Thus, in contrast to the populations containing SCM Type 1, it is unlikely that negative frequency dependent selection was at play in the AN-2K-7 population, but rather natural selection lead to the fixation of the AN-2K-7. It was not unexpected for AN-2K-7 SCM to fail to reach a stable equilibrium as AN-2K-7 was the first population to reach 100% SCM among the seven anaerobic populations. Further analysis was undertaken to

investigate if cross feeding or toxin production could explain the difference between the two different SCM types.

5.3.3.2. Evidence of cross-feeding

A typical scenario leading to the generation of polymorphic populations occurs when more than one metabolisable energy source is available to a starting population. Polymorphic populations can stably co-exist through resource partitioning and are maintained by negative frequency dependent selection. To investigate if the SCM and TCM morphotypes were sustained or inhibited by metabolites produced by either morphotype, a cross-feeding experiment was performed, where each sub-population was grown in media derived from its co-evolved counterpart (**Section 2.2.10.1**). As before, there were no TCM colony clones in the AN-2K-7 population, thus the ancestor, REL4536, was used. The three populations in which SCM clones were sequenced (AN-2K-4, AN-2K-6 and AN-2K-7) were selected for investigation. Cell densities in the differently treated media were compared and results displayed in **Figure 5.4.**

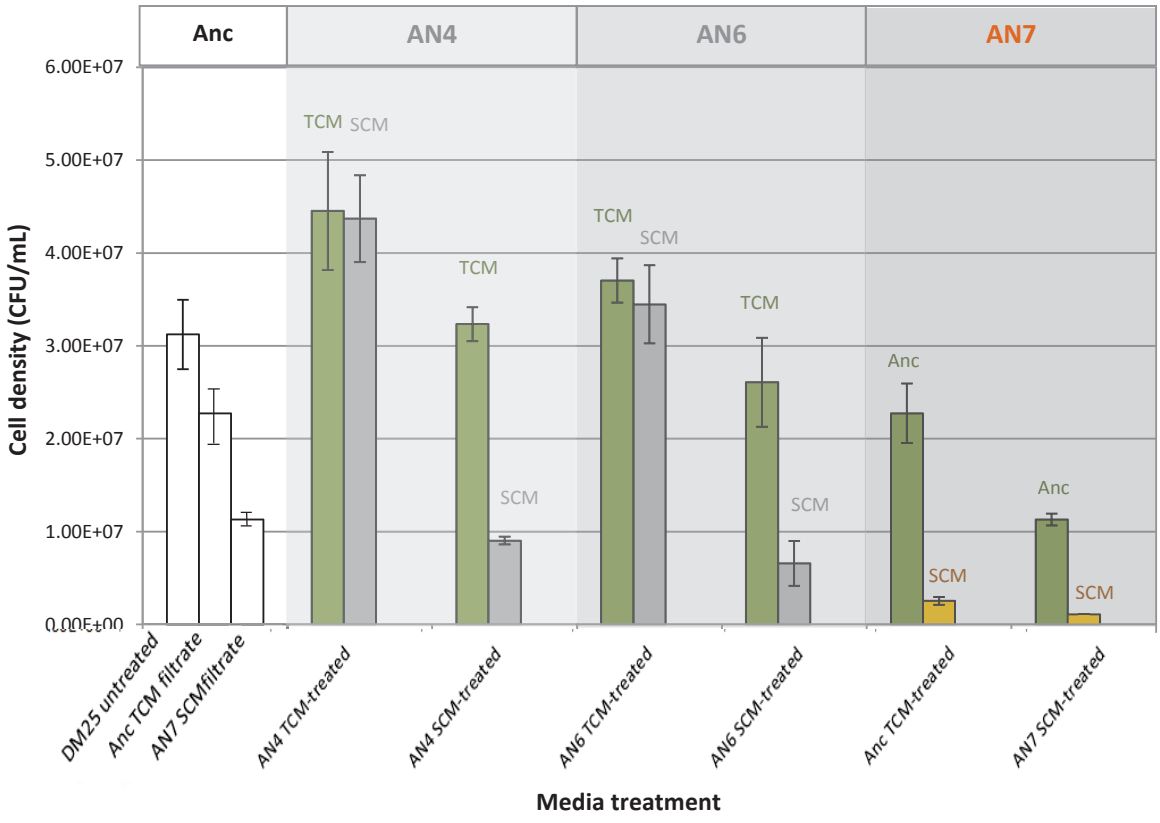

Figure 5.4: Cross-feeding between TCM- and SCM-treated cultures for the three populations in which SCM clones were isolated. Cultures that were inoculated with ancestral, TCM or SCM cells are displayed as white,

dark green and grey bars respectively. All clones were obtained from 2,000 generation populations. See text for details.

In TCM-treated DM25, TCM morphotypes from both AN-2K-4 and AN-2K-6 populations achieved higher cell densities than those in untreated media (Figure 5.4). Furthermore, SCM morphotypes from AN-2K-4 and AN-2K-6 populations reached similar cell densities to the TCM morphotypes when grown in TCM-treated DM25. In SCM-treated DM25, TCM cells from both AN-2K-4 and AN-2K-6 populations reached similar cell densities to those obtained for TCM-treated DM25. However, cell densities of SCM morphotypes isolated from AN-2K-4 and AN-2K-6 populations decreased significantly ($P < 0.05$ in both cases, two sample t-test) as compared to AN-2K-4 and AN-2K-6 SCM cells grown in TCM-treated DM25.

The trends observed with TCM and SCM morphotypes from AN-2K-4 and AN-2K-6 populations were not found in AN-2K-7 SCMs. Firstly, AN-2K-7 SCM cultures inoculated in REL4536-treated DM25 displayed a much lower cell density of 3.19×10^6 CFU/mL compared to those reached by SCM from both AN-2K-4 and AN-2K-6 populations. Furthermore, REL4536 cell density decreased when grown in either TCM-treated or SCM-treated DM25 ($P < 0.05$, two sample t-test in both cases) (Figure 5.4).

Two main conclusions can be made from this experiment. Firstly, TCM-treated DM25 promoted the growth of AN-2K-4 and AN-2K-6 SCMs to increased cell densities from 2.27×10^7 CFU/mL in untreated DM25, to 4.37×10^7 CFU/mL and 3.45×10^7 CFU/mL in TCM-treated DM25 respectively. Contrastingly, REL4536-treated DM25 did not promote the growth of AN-2K-7 SCM. It is likely that a component or components of the TCM-treated media supports the growth of the SCM morphotypes from the same population, which supports the hypothesis that these populations stably co-exist via a cross-feeding dynamic between the TCM and SCM morphotypes.

Where populations with colony polymorphisms have been characterised in the literature, the presence of acetate is frequently cited as a key metabolite that drives the divergence of populations with differing colony sizes and metabolic activities (45, 58, 307, 319). In addition, organisms may modify their environment (131) such that ecological opportunities arise which can lead to population diversification (73) and co-existing sub-populations (321). In some cases, metabolites accumulate in the media as a by-product of glucose catabolism and can be utilised by the sub-population as a sole carbon source. It is likely that during the 24 hours of growth, acetate is secreted into the media due to fermentation by *E. coli*. By 2,000 generations, a form of niche

construction may have evolved as all seven anaerobic lineages had evolved one of two SNP mutations in the *adhE* gene, likely to divert fermentation pathways towards production of acetate (131, 322) (**Chapter 3.3.1.1**). In addition, from **Chapter 3**, acetate is the end product of citrate metabolism in the anaerobic environment (237). Kinnersley *et al.* (58) identified many different mutations in glucose and acetate metabolism allowing for multiple different mechanisms by which cross-feeding through metabolic switching scenarios may arise. However, acetate metabolism in the anaerobic environment does not happen at a high rate and, as a result, is likely to be excreted from the cell under anaerobic conditions (323) and thus, further experiments are required to address if acetate cross-feeding could explain the morphological divergence observed within populations consisting of TCMs and Type 1 SCMs. This could be addressed by high performance liquid chromatography experiments monitoring for changes in acetate concentration in populations which reported morphological divergence.

The second conclusion is that cross-feeding did not appear to explain the relationship between REL4536 and AN-2K-7 SCM. Unlike in the case of both SCM Type 1 colonies, the cell density achieved by AN-2K-7 SCM did not increase, but rather decreased significantly from 1.13×10^7 CFU/mL in untreated DM25 to 2.53×10^6 CFU/mL in Anc-treated DM25, while TCM morphotypes were not observed in the AN7 population within 1,000 generations. This observation, combined with data from **Figure 5.4,** led me to believe that there may have been a different dynamic among the morphotypes in AN7 than in the other anaerobic lineages. However, it is recognised that the cross-feeding experiment for SCM Type 2 was not conducted with a co-existing AN7 TCM morphotype. However, due to the apparent complete replacement of TCM morphotypes by SCM Type 2 at some time prior to 1,000 generations, it is reasonable to assume that TCM morphotypes were not capable of long-term co-existence with SCM Type 2. Further characterisation of the fitness of the AN7 SCM Type 2 clone was undertaken.

5.3.3.3. Relative fitness of AN7

Fitness data for both AN4 and AN6 at the population level were obtained and discussed in **Section 3.3.3.2.2.1**. From **Section 5.3.3.1**, there was positive cross-feeding in AN4 and AN6, which is likely to have contributed towards the increase in fitness in these lineages at the 2,000 generation time-point. To investigate the fitness trajectory of AN7, competition assays were performed (**Section 2.2.8**) on clones randomly isolated from the AN7 lineage at 1,000 to 4,000 generations. Results from the competition assays are indicated in **Figure 5.5**.

The mean fitness of AN7 declined from 1,000 to 3,000 generations to ~0.8, with declines significant at 1,000 and 2,000 generations ($P < 0.05$ in both cases). Relative mean fitness of AN7 at 4,000 generations could not be determined ("ND") due to low cell densities of AN7 at 4,000 generations, similar to fitness assays reported in **Section 3.3.3.2.2**. Results from **Figure 5.5** indicate that the AN7 clone had decreased in fitness relative to the ancestor within the first 1,000 generations of this LTEE.

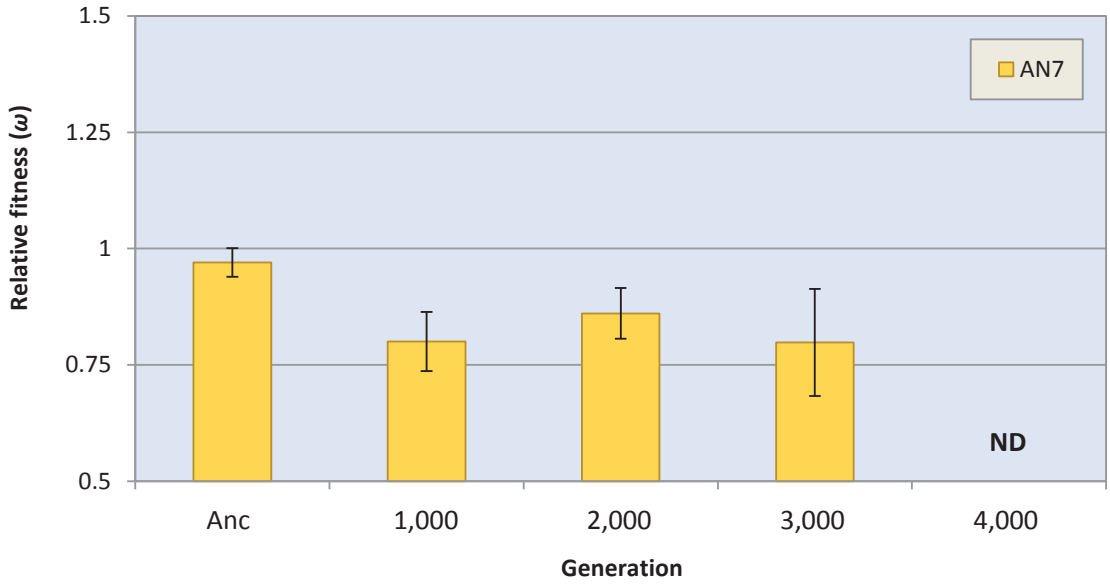

Figure 5.5: Relative fitness of AN7 clones over 4,000 generations. Displayed are the mean relative fitness values of AN7 clones as compared to the ancestor. Mean values were obtained from four biological replicates, with error bars representing the standard error of the mean. "ND" represents not determined and applies to fitness values for the AN-4K-7 clone.

An overview of the dominant morphotypes, cell densities, relative fitness of clones isolated from populations, and mutations identified within the AN7 populations over 4,000 generations is shown in **Figure 5.6**. i) The timeline in which the relative frequencies of the initially abundant TCM morphotypes were replaced by the SCM morphotypes is indicated (**Figure 5.5**). This replacement occurred during the first 1,000 generations. ii) Average cell densities (**Chapter 3.3.2**) declined dramatically throughout the 4,000 generations, with final densities of 1.5×10^3 CFU/mL. iii) Average relative fitness declined to 0.79 by 3,000 generations, and was not able to be determined at 4,000 generations since the cell density of the population at that point was too low to perform the competition experiment. iv) Within the first 1,000 generations, 11 mutations were detected,

however in the following 3,000 generations, the rate at which mutations had accumulated had decreased, with only nine found in this period.

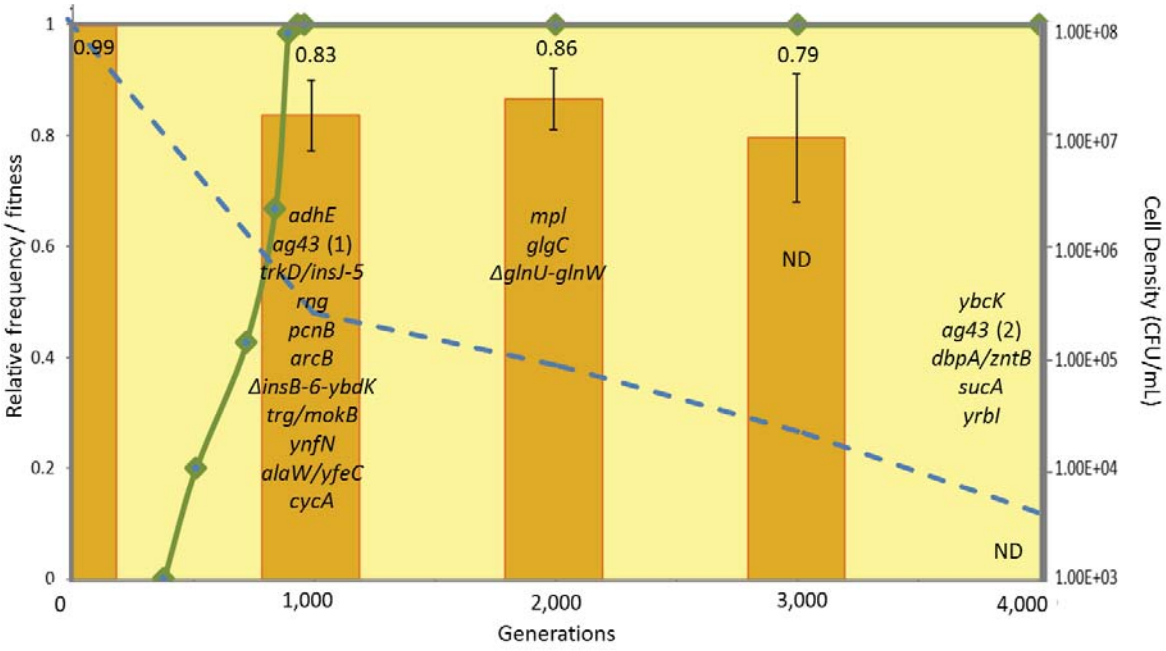

Figure 5.6: Representation of population morphotype frequency, relative fitness, mutations and average cell densities in anaerobically evolving lineages over 4,000 generations. i) The relative frequency of the SCM in the AN7 populations is depicted by the green line (axis on left). ii) Average relative fitness values of clones isolated at 1,000, 2,000 and 3,000 generations are indicated above bars. iii) Mutations that were newly detected at 1,000, 2,000 and 4,000 generations are listed on each bar. Note, a 3,000 generation AN7 clone was not sequenced. Furthermore, the two *ag43* mutations represent different mutations that arose at 1,000 and 4,000 generations, represented in parentheses. iv) Average cell densities of AN7 populations throughout the 4,000 generations are indicated with dashed line (axis on right). ND; "not determined" for newly arising mutations in AN-3K-7 clone and relative fitness of AN-4K-7 clone. See text for details.

What is striking is that the AN7 genotype, which has less fitness (and a slower growth rate) than the ancestor, has effectively been able to outcompete the ancestor and reach fixation within the population. Moreover, due to the cross-contamination that occurred between the anaerobic lineages within this LTEE, the AN7 genotype has outcompeted the various genotypes that were present in the other six lineages, which emphasises the effectiveness with which this genotype has been able to outcompete alternative evolved genotypes. Based purely upon relative growth rate, this outcome is unexpected. However, a likely possibility that will allow such a slower growing genotype to reach fixation is if it is able to inhibit or kill its competitors. Indeed, such phenomena

have been demonstrated in the literature, for example (324, 325). In the present study, an inhibition/killer phenotype also appears to most likely explain the data, given that the 2,000 generation clone (AN-2K-7) possessed a mutation in *glgC*, which is associated with a stationary phase contact dependent inhibition (SCDI) (325). Thus, how the *glgC* mutation has enabled the AN7 genotype, despite its relatively low fitness, to reach fixation is discussed further in the next section. Furthermore, it is considered that the killing phenotype associated with *glgC* mutations may be independent of the decreased competitive fitness of AN7, however, candidate deleterious mutations may have arisen to fixation through hitchhiking with the other mutations.

5.3.4. Fixation of a *glgC* mutation in anaerobic lineages

There are many cases in the literature of deleterious mutations reaching fixation, ranging from theoretical studies with computer simulations and digital organisms (65, 326, 327) to empirical studies with *E. coli* and yeast (84, 206, 328, 329). Hitchhiking occurs in genomes where a beneficial (driver) mutation and deleterious (passenger) mutation occur in the genome within the same cell (65, 326). Deleterious mutations are continuously generated, and provide a multitude of potential routes to adaptive peaks (326).

5.3.4.1. Mutation in glgC

A mutation in *glgC* was found in the sequenced clones of all 4,000 generation anaerobic lineages as well as in AN7 at 2,000 generations. Specifically, the mutation was a single base deletion of a C nucleotide at position 3,473,571 bps, resulting in a frame shift within the *glgC* gene. The *glgBXCAP* operon is responsible for the production of glycogen in *E. coli* REL4536. Glycogen is a branched polymer of glucose, and the presence of glycogen storage in *E. coli* is thought to increase the viability of the bacteria under adverse conditions, or in specific ecological niches (330). The *glgC* gene encodes the first enzyme in the glycogen biosynthesis pathway (**Figure 5.7**), ADP-glucose pyrophosphorylase. This enzyme converts glucose-1-phosphate into ADP glucose, and is also the rate limiting step in the biosynthesis of glycogen.

Disruption of *glgC* is of particular interest as mutations have been identified in the gene which result in the SCDI phenotype (325), where strains carrying these mutations appeared to inhibit the growth of the bacteria from which they were derived. The SCDI phenotype has been observed to decrease the cell density of the ancestral strain (with wild type *glgC*) from ~10^9 CFU/mL to 10^4 CFU/mL in 15 hours in mixed cultures, coinciding with the beginning of stationary phase. It is presumed that the *glgC* mutation that arose in the present LTEE study imparts a similar phenotype

and therefore, has enabled *glgC* mutant strains to inhibit non *glgC* mutants, and drive the fixation of the *glgC* mutation.

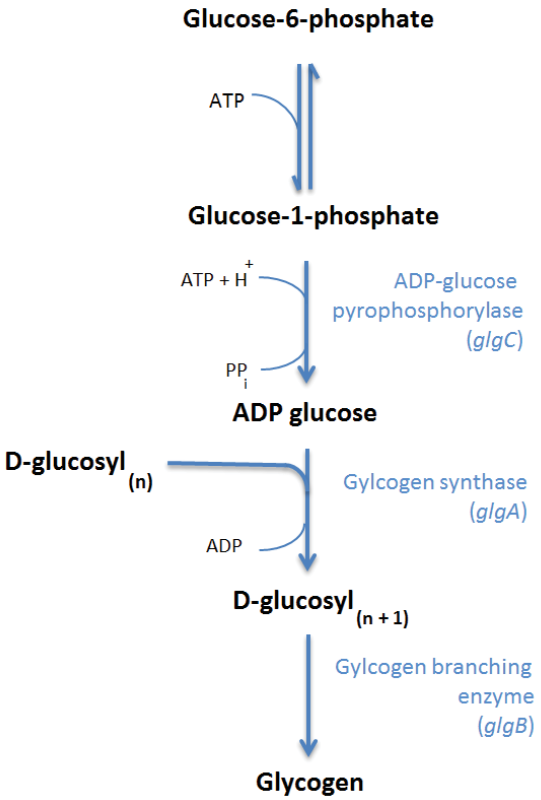

Figure 5.7: Glycogen synthesis pathway. The intermediates of the glycogen synthesis pathway are indicated. Three enzymes involved in the pathways are indicated, as are the genes from which they are encoded, in parentheses. Figure adapted from Keseler *et al.* (160).

5.3.4.1.1. Characterisation of the *glgC* mutation

The mutation conferring the SCDI phenotype was reported by the LTEE study of Lemmonier *et al.* (2008) which evolved *E. coli* K-12 AB1157 in rich LB media. The *glgC* mutations arose in eight replicate evolving populations within 412 generations (325). The authors indicate that the effect of SCDI varied widely between experiments, with declines in cell density ranging from 200 to 5,000 fold when ancestral and *glgC* mutant cells were grown together, with inhibition occurring once cells entered the stationary phase of growth. However, this extent of inhibition was not found in this thesis, as competition assays (**Section 5.3.3.3**) and reciprocal invasion experiments (**Section 5.3.3.1**) in which ancestral cells and evolved cells containing the *glgC* mutation were performed without such dramatic declines in cell density. There are two differences between the phenotype

reported in Lemonnier *et al.* (325) and that hypothesised to be take place in this thesis, which may explain the discrepancy in the extent of inhibition observed.

5.3.4.1.2. Implications of genetic background on *glgC* mediated inhibition

The first difference between this LTEE and the Lemonnier study is the nature of the genetic background in which the mutations arose. This applies on two levels. This first point is that Lemonnier *et al.* (325) found mutations in *glgC* in a different strain, *E. coli* K-12 AB1157, and thus under a different ancestral background. Furthermore, the ancestor in the Lemonnier study was a mutator, which had a deletion (*ΔmutS*) conferring an increased mutation rate of the strain. This is in contrast to the ancestral strain *E. coli* REL4536 used in this thesis. However, K-12 SCDI cells and *E. coli* B cells were grown together in a mixed 10 mL culture, which confirmed that *E. coli* B cells were also susceptible to SCDI mediated inhibition.

The second important consideration is the nature of mutations found in the *glgC* gene. There were eight mutations which mediated the SCDI phenotype from Lemonnier *et al.* (325), with a non-synonymous SNP mutation in the *glgC* gene responsible in all cases (**Figure 5.8**). In this thesis, a single base deletion at position 3,473,571 bps, leading to a frame shift mutation from the 13th codon in the *glgC* gene, is likely to encode a missense non-functional protein.

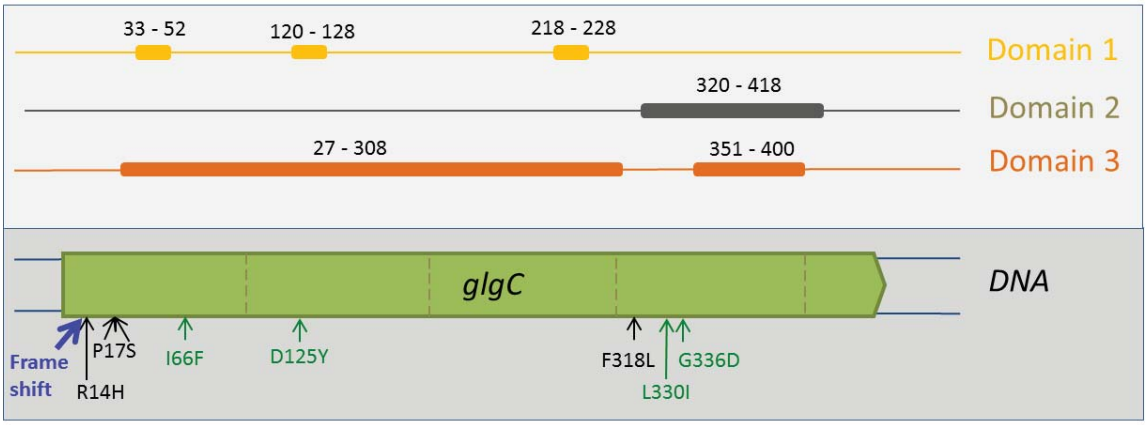

Figure 5.8: The mutations in the *glgC* gene as reported in this thesis and by Lemonnier *et al.* (325). Indicated is the gene and the three domain types contained within the ADP-glucose pyrophosphorylase protein. Domain 1 shows the three instances of ADP-glucose pyrophosphorylase conserved sites, domain 2 represents a trimeric LpxA-like superfamily domain and domain 3 indicates the two instances of nucleotidyl transferase domains. The location of the eight SNP non-synonymous point mutations from (325) are indicated. Note that the mutation that occurred at codon 17 was reported twice among replicate lineages.

The four mutations that are in green font are those that are located in domain coding regions of the gene, with those occurring outside conserved domains indicated in black. Also indicated by the boldface blue arrow is the frame-shift location reported in this thesis within the 13th codon of the gene. Figure adapted from Hunter *et al* 2011. (331).

In the Lemonnier *et al.* study show that each of the eight mutations increased the rate of glycogen production and were thus hypothesised to be involved in the mode of action for the SCDI phenotype. Increased glycogen production was confirmed by phenotypic tests via iodine vapour assays. A *glgC* deletion mutant did not lead to an increase in the production of glycogen. The frame-shift mutation reported in this thesis is likely to encode a missense protein and thus not likely to increase in glycogen activity. Furthermore, it is highly likely that four of the eight mutations reported by Lemonnier *et al.* would cause gene disruption, leading to a non-functioning GlgC protein as four of the eight mutations are located in at least one of three conserved domains, essential for the protein's proper functioning (**Figure 5.8**).

However, the exact mechanism of SCDI remains unknown. Despite phenotypic links to glycogen production, addition of glycogen to the growing cultures by Lemonierr *et al.* tested negative to bactericidal activity in solution and thus, to date, there is no clear mechanism linking the over production of glycogen to the bactericidal activity of SCDI. Furthermore, recent work (332) has suggested the mutation identified in the Lemonierr study may lead to epistatic interference with the regulation of chromosomal toxin-antitoxin systems (239).

5.3.4.1.3. Implications of growth conditions on *glgC* mediated inhibition

The second difference between this LTEE study and that of Lemonnier *et al.* 2008 is that the evolving strains were not under similar growth conditions. This is an important consideration for the contact dependent aspect of the SCDI phenotype. Different media were used in each LTEE study, with cells grown in LB media in Lemonnier *et al.* 2008, and in anaerobically prepared DM25 minimal media in this thesis. The different nutritional composition of media allowed for different growth rates and maximum cell densities between the two studies, with both differences likely to have implications on the extent of inhibition produced from mutated *glgC* cells.

Considering differences in growth rates, rich aerobic LB and nutrient poor anaerobic DM25 allow for different growth rates of inoculated cultures. The SCDI phenotype was observed as a sharp decrease in cell density within 15 hours, coinciding with the beginning of stationary phase of cultures in LB. Anaerobically grown cultures in DM25 do not reach stationary phase until ~20 hrs

due to extended lag and log phases (**Chapter 3.3.1**). The longer time taken for cells to reach stationary phase in anaerobic DM25 is likely to delay the time at which the SCDI phenotype is observed. With regard to different cell densities achieved, the SCDI phenomenon is reported to work maximally when cell densities are high (~10^9 CFU/mL), to allow for maximum contact between cells, which in turn mediates the inhibiting phenotype. However, for *E. coli* in anaerobic glucose-limited DM25, maximum cell densities are much lower, rarely higher than ~10^7 CFU/mL. Nevertheless, at this level of cell density, inhibition was demonstrated by Lemonnier *et al.* (325), albeit to a lesser extent.

5.3.4.1.3.1. Potential non-contact inhibition of REL4536 by AN7

Lemmonier *et al.* 2008 showed that cell-cell contact was critical to the SCDI mode of action. However, results of the cross-feeding experiment (**Section 5.3.3.2)** show that inhibition of REL4536 was obtained when grown in the culture filtrate of AN-2K-7. It is possible that cell fragments or contents in the stationary phase culture are present in the filtrate and allow a contact-dependent inhibition effect. However, I cannot rule out the possibility that a non-contact inhibition mechanism was also at play, for example the production of a toxic compound. However, analysis of genomic data did not indicate any mutations which were likely to result in toxin production. Alternatively, the nature of the *glgC* mutation in this study, as compared to in Lemmonier *et al.* 2008, may have resulted in modification of the SCDI phenotype. Further characterisation of the *glgC* mutation, and link to an SCDI-type phenotype in the anaerobic environment, is required. In addition, analysis of culture filtrates from AN7 would help identify if this lineage is capable of bacteriocin activity.

5.3.5. Loss of fitness within AN7

A significant decline in relative mean fitness ($P < 0.05$, two sample t-test) was reported in AN7 clones (**Figure 5.5**). An unexpected finding was observed with clone AN-1K-7, which is presumed to be a rare genotype within the 1,000 generation population, as it exhibited reduced competitive fitness relative to the ancestor (**Figure 5.5**); however, it did not possess the *glgC* mutation that would allow it to outcompete alternative genotypes. This less fit genotype was predicted to have been lost from the AN7 lineage over time. However, a derivative of it likely acquired the mutation in *glgC* (as seen in AN-2K-7), which was then able to inhibit alternative genotypes and reach

fixation. With regard to AN-1K-7, mutations that were responsible for its reduced relative fitness may be identified from genomic data (**Table 5.2)**.

Table 5.2: List of mutations in AN-1K-7

Mutation type	Mutation classification	Number of mutations	Position [a]	Locus/gene affected	Function
SNP	Transition	1	1,439,673	*adhE*	Encodes alcohol dehydrogenase
	Transversion	1	2,972,858	*ag43*	Involved in bacterial virulence
Insertion	Single base	2	3,866,357	*trkD/insJ-7*	-
			3,298,183	*rng*	Encodes ribonuclease G involved in RNA turnover
Deletion	Small	2	161,094	*pcnB*	Encodes poly(A) polymerase I
			3,260,806	*arcB*	Encodes aerobic respiration control sensor protein ArcB
	Large	1	546,986	*ΔinsB-6-ybdK*	Cryptic prophage P2- like region (30 gene deletion)
IS element	Insertion	4	1,272,468	*trg/mokB*	-
			1,123,058	*ynfN*	Hypothetical protein
			2,424,083	*alaW/yfeC*	-
			4,381,583	*cycA*	D-alanine, D-serine, glycine permease

[a] Position of mutation in the AN-1K-7 genome (bps)

In total, 11 mutations were identified in AN-1K-7 relative to REL4536. Among the 11 mutations, three were likely to be adaptive (as identified in **Chapter 4**): i) the transition event in *adhE,* likely resulting in modification of the fermentation network, ii) the transversion event in *ag43* and iii) the IS element insertion in *trg/mokB*, both predicted to modify virulence genes. However, two mutations, affecting *appY* and *arcB*, may explain the negative fitness of the AN-1K-7 clone in the anaerobic conditions and are discussed below.

5.3.5.1. Loss of appY

The first candidate mutation that was likely to negatively impact the fitness of the organism under the anaerobic environment was the large deletion *ΔinsB-6-ybdK*. This is the same mutation event described in **Section 5.3.2.2**, in which there was evidence implicating the mutation as potentially responsible for the origin of one type of SCM. As described in **Section 4.3.4.1.3**, this region is close to cryptic prophage P22 and is likely to be the cell's means of excising these phage-like genes. This event is likely to have occurred via recombination between *insB-6* and a homologous region downstream of *ybdK* and upstream of another IS element, *insA-7*. The deletion event spanned 32,573 bps and 30 genes. All 30 genes deleted from the genome of AN7 are listed in **Table 5.3** below.

Figure 5.3: Table of genes deleted in the *ΔinsB-6-ybdK* deletion event in AN-1K-7

No.	Gene name	Description
1	*[insB-6]*	After frame-shift event, transcribed with InsA and mediates IS1 transposition
2	*insA-6*	Transcriptional repressor of IS1
3	*ECB_00514*	Unknown
4	*ECB_00515*	Unknown
5	*ECB_00516*	Unknown
6	*ECB_00517*	Unknown
7	*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis
8	*ompT*	Outer membrane protease, with specificity for paired basic residues
9	*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode cellular envelope proteins at low temperatures and during stationary phase
10	*ybcH*	Unknown
11	*nfrA*	Important in N4 absorption
12	*ECB_00524*	Unknown
13	*yhhI-2*	H repeat-associated protein, RhsE-linked, function unknown
14	*ECB_00526*	Unknown
15	*ECB_00527*	Unknown
16	*ECB_00528*	Unknown
17	*ECB_00529*	Unknown
18	*ECB_00530*	Unknown
19	*cusS*	Detoxification of copper and silver ions
20	*cusR*	Detoxification of copper and silver ions
21	*cusC*	Detoxification of copper and silver ions
22	*cusF*	Detoxification of copper and silver ions
23	*cusB*	Detoxification of copper and silver ions
24	*cusA*	Phenylalanine transporter
25	*pheP*	Low abundance mechanosensitive channel of miniconductance
26	*ybdG*	Unknown
27	*nfnB*	*nfsB*-encoded nitroreductase is the minor oxygen-insensitive nitroreductase. NfsB reduces a broad range of nitroaromatic compounds, including the antibiotics nitrofurazone and nitrofurantoin a.k.a. *nfsB*
28	*ybdF*	Unknown
29	*ybdJ*	Unknown
30	*ybdK*	Unknown

Large deletions are likely to have wide ranging pleiotropic effects within the environment in which they arise. There are two genes involved in IS element movement (*insB-6* and *insA-6*). Two genes are known to be associated with phage integration, *nfrA* and *yhhI-2*. Fifteen of the genes are of unknown function with 10 likely to be associated with cryptic prophage P22 activity. Five belong to the *cus* operon involved in the detoxification of copper and silver ions. Of the deleted genes (**Table 5.2)**, the loss of *appY* was the most likely candidate for causing deleterious effects on growth in the anaerobic environment.

Under anaerobic conditions, the global transcriptional controller ArcBA promotes the expression of the *appY* gene. AppY is a member of the ArcA/XylS superfamily which consists exclusively of positive transcriptional regulators (259). More specifically, AppY positively regulates genes involved in energy metabolism during anaerobic growth. The increased rate of synthesis of AppY leads to the increase in the expression of approximately 30 proteins belonging to the fermentative anaerobic stimulon (333) and is transcribed as an immediate response to anaerobic growth. The *appY* gene has an unusually high AT nucleotide content, and located within the P22 prophage region (**Chapter 4.3.4.1.3),** most likely has been introduced into the *E. coli* genome by horizontal gene transfer from an outside source (334). Brøndsted and Atlung (254) have suggested AppY as the third global transcriptional regulator involved in anaerobic gene regulation (in addition to ArcBA and Fnr – see **Chapter 1.1.3**).

Two major operons affected by AppY activation are the *hyaABCDEF* and *cbd-appA* operons. Under anaerobic conditions, AppY alleviates the constitutive repression mediated by *lscR* on the *hya* operon (335), which allows for the transcription of hydrogenase 1 isoenzyme. Hydrogenase 1 reduces fumarate generated from fermentation, and is conserved among many facultative anaerobic organisms (336). Hydrogenase 1 can also catalyse the uptake and oxidation of H_2 to produce protons and electrons (337). The second major pathway directly regulated by AppY during anaerobic growth involves the *cbd*-encoded cytochrome oxidases. Generally, cytochrome oxidases are involved in the aerobic respiration chain. However, the *cbd* promoter for the acid phosphatase cytochrome, which is induced by anaerobic growth, encodes a fermentation specific acid phosphatase that is induced on entry into stationary phase, and during phosphate starvation under anaerobic conditions (254).

Mutations in global regulators such as AppY are known to have large effects on the cell due to their central roles in extensive functional networks (58, 60). Loss of the chromosomal region surrounding *appY* has been reported among lineages with co-existing sub-populations (320). However this is the first time this has occurred in the anaerobic environment, in which AppY is one of the major DNA-binding transcriptional activators for energy generation under anaerobic conditions.

However, this same mutation was found to have occurred in parallel in two independent lineages in the *sbcC* adaptive lineages of S. Shewaramani, (unpublished data) and was not associated with a decline in fitness. Potentially, differences in the ancestral background with the disrupted *sbcC* may

have played a role in generating the mutation. In addition, there was strong evidence that this mutation could have been responsible for the origin of the SCM colonies and thus may have had strong pleiotropic effects, due to the number of genes affected by the deletion event. From this it is likely that this was not the only mutation that may have impaired fitness under the anaerobic condition.

5.3.5.2. Partial deletion in arcB

The second mutation likely to be deleterious in the anaerobic environment was the small deletion of 42 bps in *arcB* at 3,260,806 bps. The ArcB protein functions as part of the two component ArcBA anoxic redox control system (**Chapter 1.1.3**). More specifically, the ArcB protein acts as the membrane bound sensory component for anoxic redox control. Once ArcB is activated under anaerobic conditions, a phosphorylation transfer system results in the activation of the ArcA protein. The cytosolic ArcA is the transcription factor component of the system. Once ArcA is activated, via phosphorylation from ArcB, it negatively regulates the transcription of genes involved in respiration (338) and positively promotes the transcription of genes involved in anaerobic energy metabolism, including *appY* (14, 339).

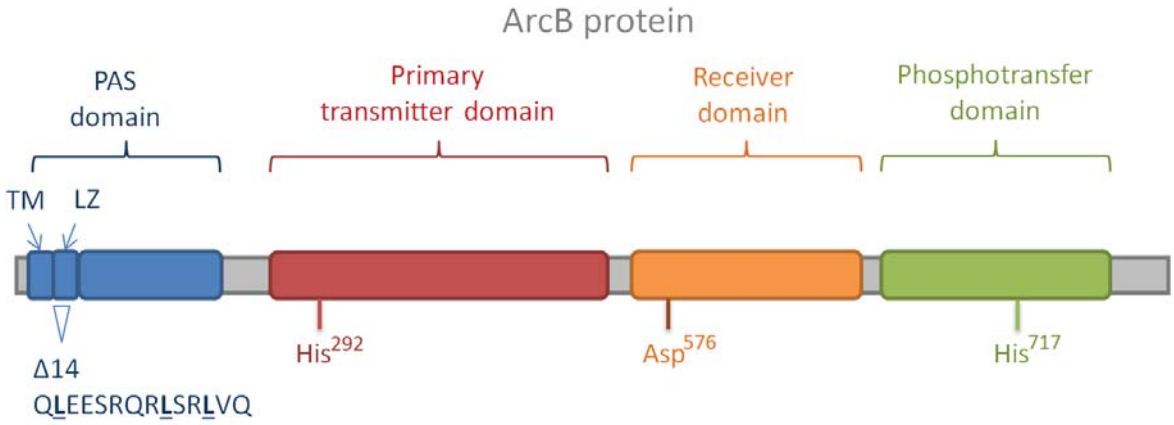

Figure 5.9: The domain structure of ArcB. The four domains of the ArcB protein are shown. Also indicated are conserved amino acid residues in three of the four domains and the region of the deletion reported in this thesis. TM, transmembrane module; LZ, leucine zipper. Figure modified from Bidart *et al.* 2012 (340).

Common to most sensor molecules of two component pathways, ArcB has four highly conserved domains. First, a hydrophobic N-terminal PAS domain, named after three proteins contained within the domain (periodic circadian proteins, Ah receptor nuclear translocator protein and single-minded protein) involved in many signaling proteins (341). Contained within the PAS domain spanning 268 amino acids is a 73 amino acid long transmembrane module and a 104

amino acid leucine zipper module from amino acid 74 to 177. Second, the primary transmitter domain contains a highly conserved histidine residue (His^{292}) crucial for phosphorelay. Third is the receiver domain with an Asp^{576}, which relays the phosphorylation signal from His^{292} the next domain. Finally, located in the C-terminus of the protein is the phosphotransfer domain, which is a signal receiver domain containing a conserved His^{717} which transfers the phosphorylation system to the ArcA transcription factor protein (340).

The deletion of 42 bps at 3,260,806 bps lies within the region that codes for the leucine zipper of the PAS domain. The deletion event encompassing 14 amino acids (QLEESRQRLSRLVQ), including three leucine amino acids presumably important in the zipper module, is likely to result in a loss of function mutation in the sensory capacity of the ArcBA global anaerobic transcriptional control network. Further support for this gene being responsible for the decline in fitness is that it was online, found in all subsequent re-sequencing time-points which continued to decline in fitness. In addition, this mutation was not found in the adaptive lineages of S. Shewaramani (unpublished data) which did not decrease in fitness.

5.3.5.2.1. Effect of deleterious mutations in AN-1K-7

Both the *appY* loss and deletion in *arcB* are hypothesised to be the deleterious mutations that occurred in the AN-1K-7 clone. The central role played by both genes in energy production in the anaerobic environment makes this a likely possibility. However, due to a lack of functional characterisation, we cannot rule out the possibility that the loss of *appY* or partial deletion of *arcB* mutations could be of neutral effect or even weakly beneficial to the cell in the anaerobic environment. If either mutation was found to be neutral or weakly beneficial, the decline in fitness could still be explained by a strong negative epistatic interaction between the two mutations. This phenomenon, where mutations are individually beneficial but collectively can have deleterious results, has been reported in the literature (61, 108, 136). As both the *ΔinsB-6-ybdK* deletion and *arcB* deletion occurred within 1,000 generations it is difficult to determine if these mutations are deleterious, neutral or weakly beneficial. To confirm these assumptions, characterisation of the *arcB* and *ΔinsB-6-ybdK* mutations, by recreating these mutations in the ancestral genetic background, is required. Subsequent competition and reciprocal invasion experiments could confirm if one or both of these mutations were responsible for the decline in fitness.

5.3.5.3. A model for the evolution of the AN7 lineage

Given the data, a model for how the AN7 lineage evolved is proposed (**Figure 5.10**). It is thought that the IS1-mediated deletion, *ΔinsB-6-ybdK*, which resulted in deletion of *appY*, resulted in individuals that possess the SCM phenotype, and may have contributed to the lower fitness relative to the ancestor (**Section 5.3.3.3**). A mutation in *arcB* may also have contributed to the lower fitness observed (**Section 5.3.5.3**). It is expected that such genotypes of reduced fitness would be lost from a population, through being outcompeted by ancestral types. However, if the mutation in *glgC* arose, it would enable the deleterious mutation(s) to hitchhike, while inhibiting and outcompeting alternative genotypes.

The mutation reported in *glgC* was first observed in the 2,000 generation clone (AN-2K-7), which is later than expected with regard to TCM clone inhibition, given SCM fixation was observed prior to 1,000 generations (see **Figure 5.5**). A likely explanation for this is that the *glgC* mutation was present within the population earlier than 2,000 generations, and possibly as early as around 400 generations, when the relative abundance of SCM began to rise sharply (**Figure 5.5**). I hypothesise that the reason the *glgC* mutation was not observed in AN-1K-7 is that the 1,000 generation population, despite being fully SCM, was genetically heterogeneous with regard to the *glgC* mutation, and the clone that was randomly selected for genome sequencing was off the line of descent to AN-2K-7 (**Figure 5.10**). To test whether the *glgC* mutation was indeed present at 1,000 generations (or earlier), PCR detection or genome sequencing should be performed.

Further support for the potent "killing/inhibition" ability of AN7 after 2,000 generations was provided with the likely cross contamination between wells in the 24-well plates in the anaerobic condition. Initially, anaerobically evolved populations were expected to increase in fitness, however this was not found. Sometime after 2,000 generations but before 3,000 generations, relative mean fitness of the anaerobic populations began to decrease (as seen in **Figure 3.3** from **Chapter 3.3.2**). At this point, cell densities had reduced from ~10^7 to ~10^5 CFU/mL. By 4,000 generations, population densities were so low that competition assays could not be conducted in practice. Also at this point, all of the anaerobically evolved populations had become comprised solely of SCM, with no TCM cells present in the populations.

The most likely explanation was that cross-contamination between populations had taken place, consistent with the contamination that was detected after 2,000 generations (**Section 3.3.2**). Whole genome sequencing of the 4,000 generation anaerobic lineages revealed that all 4,000

generation genome sequences were almost identical to that of AN-2K-7 (**Appendix Table 7.11**). Cross contamination of a lineage by a genotype with reduced fitness linked with a mutation conferring an inhibition mechanism would likely explain the declines in fitness in the anaerobic lineages after 2,000 generations, such that eventually all lineages were supplanted by AN7.

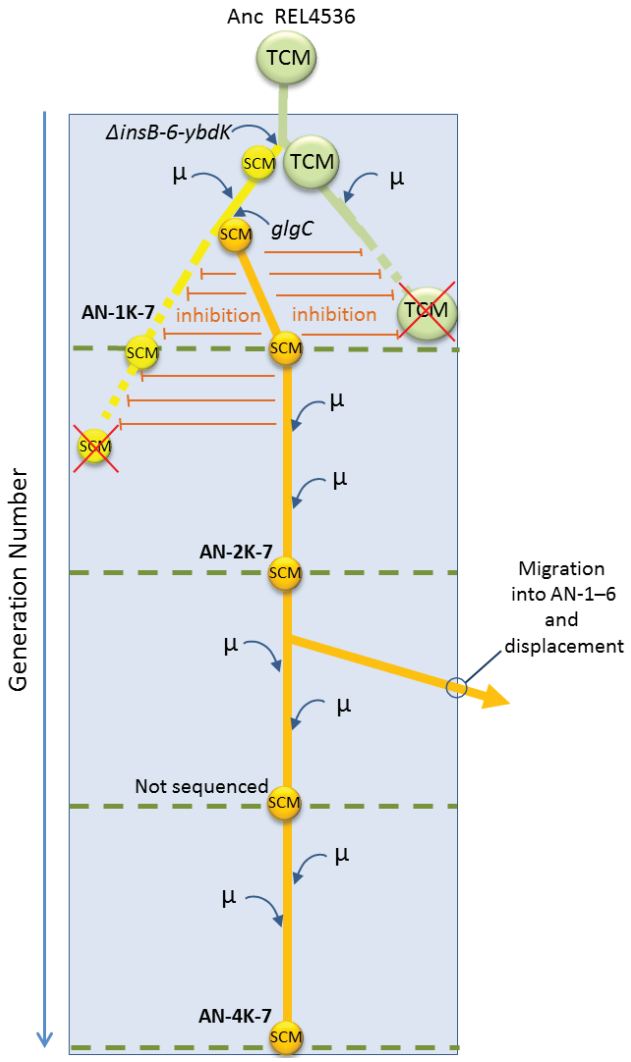

Figure 5.9: Proposed evolution of the AN7 lineage. The SCM morphotype is predicted to be due to deletion of the *insB6-ybdK* region prior to 400 generations. Mutation in *glgC* is predicted to take place within an SCM clone around 400 generations, and descendents of this clone have inhibited both TCM cells, and SCM cells with the wild type *glgC* allele (decline of the TCM and non-mutant *glgC* SCM sub-populations are denoted by dashed lines). By 1,000 generations, the TCM cells have been out-competed. However, the clone randomly selected for genome sequencing (AN-1K-7) is SCM, but contains the wild type *glgC* allele. At 2,000 generations, AN-2K-7, which contained the mutant *glgC*, was sequenced. After 2,000 generations, migration

of the AN7 population into other anaerobic lineages (AN1-6) eventually results in their displacement by AN7. The generic accumulation of mutations within the divergent sub-populations is denoted by the "μ" symbols.

5.4. Summary

The main goal of this chapter was to understand the dynamics that had occurred in the anaerobic lineages during the 4,000 generation of this LTEE. Three observations in particular were explored as outlined in **Section 5.2**.

The first observation was the origin of different sized colonies. Small colonies were first reported among the anaerobic lineages after ca. 400 generations. The origin and maintenance of SCM colonies was specific to populations grown in the anaerobic environment in this thesis. It is hypothesised that the polymorphism may have resulted in a form of cross-feeding. Through whole genome analysis, I attempted to identify the genetic basis underlying this phenotype. Two different mechanisms leading to the divergence of the two morphotypes were considered: one through a combination of undefined mutations (SCM Type 1) and the other through a 30 gene deletion event (SCM Type 2). These two different types of SCM proved to differ phenotypically, SCM Type 1 were able to co-exist stably with TCM cells in their respective populations of origin via negative frequency dependent selection. In contrast, SCM Type 2 cells out-competed ancestral TCM cells within 12 days of co-culturing.

The second observation was the decrease in relative mean fitness of clones within the AN7 lineage concurrent with the fixation of the genotype within the population. It is counter intuitive that a genotype of reduced fitness could out-compete established populations. However, this is speculated to have arisen by the lineage acquiring deleterious mutations, involving the loss of *appY*, a gene known to be critical for energy production under anaerobiosis, and *arcB*, which encodes the membrane component of the global anaerobic control two component pathways ArcBA. Soon thereafter, it is hypothesised that the AN7 strain had developed a SCDI-like generating mutation in the *glgC* gene, and was able to inhibit other lineages that do not have the *glgC* mutation. Further examination of this hypothesis could be conducted by recreating the same *glgC* mutation in the ancestral strain, with subsequent comparative growth assays with the ancestor and *glgC* mutant, both separately and in combination to observe if the killing phenotype is observed. The cross-contamination of the anaerobic lineages, evidenced by the similarity of the genomic data in the 4,000 generation genomes, provides additional evidence of the effectiveness

of the AN7 genotype in outcompeting alternative genotypes that had evolved in the anaerobic environment.

This chapter has provided further insights into the evolutionary dynamics that were observed during the 4,000 generations of this LTEE study in the anaerobic environment.

Chapter Six : Final discussion

6.1 Further discussion and conclusions

This thesis set out to understand the genetics and dynamics of bacterial adaptation to an anaerobic environment. While environmental conditions are known to play a large role in adaptation (48, 73, 95, 97, 126, 127, 317), key traits and their underlying mutations that are selected for during growth in anaerobic environments have been little studied.. Thus, the aim of this thesis was to understand and compare evolutionary change in facultative anaerobes adapting to aerobic and anaerobic environments. To do this, experimental evolution of the facultative anaerobic bacterium, *E. coli* REL4536 (6) was performed to observe adaptation to strict aerobic and strict anaerobic environments, as well exposure to both environments in a treatment that regularly fluctuated between the two. Competitive fitness assays revealed changes in fitness over time of the evolving lineages, while whole genome sequencing of individual clones from the lineages allowed for the identification of genome-wide genetic changes that had occurred within the genome during experimental evolution. Genes acted on by natural selection in their respective environments were identified by the occurrence of parallel mutations, and thus, were deemed important for adaptation within the aerobic and anaerobic environments. As outlined in **Section 1.7**, four research questions were specifically addressed in this thesis, and a summary of the findings in this thesis is given below.

- What are the dynamics of adaptation under aerobic and anaerobic conditions?

The dynamics of adaptation under aerobic and anaerobic conditions were investigated by measuring the competitive fitness of lineages evolved in aerobic and anaerobic environments, relative to the ancestral strain (**Chapter 3**). For lineages that had evolved in the aerobic environment, no increases in fitness were observed on average over 4,000 generations of evolution. This was likely a consequence of the ancestral strain having already been pre-adapted state for 10,000 generations to the aerobic growth conditions that were similar to those used in this study (52). However, in the anaerobic environment, an environment that the ancestor had not

been exposed to for 10,000 generations, increases in fitness were observed within 1,000 generations. The 25% increase in average fitness in the anaerobic environment within 2,000 generations, is likely a result of adaptation to a novel environment, and constant exposure to this environment led to such dramatic increases. Such observations have been reported by other studies that expose experimental strains to novel environments in a long-term basis, e.g. experimental lineages of *Pseudomonas fluorescens* evolved in the novel sugar xylose for 500 generation reported increases in fitness by 30% (81). The fitness of lineages that were regularly fluctuated between aerobic and anaerobic growth varied greatly between the aerobic and anaerobic environments. During competition in the aerobic environment, as before, no difference in fitness was observed within 4,000 generations. In the anaerobic environment however, significant increases in fitness were observed from as early as 1,000 generations, again likely due to adaptation to the novel environment. Comparing the fitness of fluctuating lineages to both aerobic and anaerobic lineages in both environments, by 2,000 generations fitness increased in both environments but were of intermediate levels as compared to lineages exposed to the constant environments, indicating the evolution of generalists under the fluctuating treatment, Aerobic and anaerobic generalists, of intermediate fitness to lineages exposed to the constant aerobic and oxygen-limited environments were also reported by Puentes-Tellez (101).

- What genes and mutations are more important for adaptation to aerobic, anaerobic and a temporally heterogeneous aerobic to anaerobic treatment?

To identify the genes that were important for adaptation to the aerobic and anaerobic environments, a combination of fitness data, whole genome sequence data and inferences based on parallelism was used (**Chapter 4**). Despite an overall lack of fitness increases in the aerobic environment, analysis of genome sequencing data indicated one likely evolutionary pathway specific lineages evolving in the aerobic landscape. Mutations in the *rpo* genes, likely conferred a GASP-like phenotype in aerobically evolving lineages. This is thought to confer a selective advantage to the aerobic lineages which during aerobic growth, spent over 15 of the 24 hours in stationary phase in the serial subculture regime employed in this study. In the anaerobic environment, increases in fitness were found within the first 1,000 generations of evolution and mutations specific to the anaerobic environment involved those in genes for fermentation pathways. These mutations were likely to have increased the concentration of coenzyme

molecules, reactivated the functions of pseudo-genes which were directly involved in fermentation, or the diversion of fermentation to the production of alternative end products, as outlined in **Chapter 4**. Additionally, many genes were mutated in both the aerobic and anaerobic lineages and were mostly associated TA systems, or with virulence activity. They are presumed to have accounted for the majority of adaptation in the aerobically evolving lineages and thus, likely to account for minor fitness increases.

From the analysis in **Chapter 4**, some inferences on the underlying topology of the aerobic and anaerobic adaptive landscapes can be made. The lack of specialisation in the aerobic lineages suggests that the ancestral lineage was already on, or close to, a fitness peak. An adaptive peak that may specific to the aerobic landscape in this study may represent increased fitness due to GASP mutations. For the anaerobic landscape, enhanced energy production through mutations in genes involved in anaerobic fermentation were adaptive pathways commonly undertaken that were specific to lineages exposed to the anaerobic environment. Between the aerobic and anaerobic landscapes, many similar peaks were identified. These represent mutations in genes like TA systems and virulence genes. They are presumed to inactivate both TA systems and virulence genes, and may represent adaptations to features in common between the aerobic and anaerobic treatments, such as the media used and the growth format in this experiment. Although many peaks were in common, they likely differed in height, as fitness increases were minimal among aerobic lineages as compared to anaerobic lineages. These shared peaks may also represent adaptations to the shared experimental conditions between the two environments (media and growth format), which may not directly reflect adaptation to the presence or absence of oxygen in each environment.

- How does niche width affect adaptive pathways?

Adaptive pathways taken by fluctuating lineages, that had a wider niche width than either the aerobic or anaerobic only grown lineages, were similar to a combination of those undertaken by both the aerobic and anaerobic lineages. These included modification of fermentation, inactivation of virulence genes and TA genes. This may account for the ability of the generalists to grow well within both environments. Additionally, specific to the fluctuating lineages, there was a high rate of prophage excisions and mutations in *brnQ*, which may represent a unique pathway taken by generalists on both the aerobic and anaerobic landscapes. Overall, mutations likely to be

advantageous in the anaerobic environment, such as modification of the fermentation genes, appeared to have little or no antagonistic pleiotropy effect in the aerobic environment, as seen by the increase in fitness of fluctuating lineages in both environments (63, 88, 101).

- Do multiple pathways exist for bacterial adaptation to aerobic and anaerobic conditions, and if so how do these pathways differ?

Multiple pathways for aerobic and anaerobic adaptation were identified in this study (as discussed previously, and in **Chapter 4**). Moreover, evidence for multiple pathways in anaerobic adaptation were described in **Chapter 5** that resulted in distinct colony morphological differences when cells were grown aerobically on solid media. These included a balanced polymorphism that between co-existing TCM and SCM cells, and an SCM that outcompeted the ancestral strain. Genomic investigation identified that the two types of SCM differed genetically, having a combination of mutations and a 30 gene deletion event for SCM1 and SCM2 respectively. The balanced co-existence between SCM1 and the co-evolved TCM sub-population, was potentially maintained through a cross-feeding dynamic, and is thought to have led to the strong population-level fitness enhancement at 1,000 and 2,000 generations (**Chapter 3**). In contrast, SCM2 colonies did not stably co-exist with the TCM sub-population, and resulted in the loss of TCM in this population within 1,000 generations. A putative hypotheses accounting for the origin of the colony morphology biodiversity relate to the presence of acetate in the media, and subsequent generation of acetate cross-feeding among TCM and SCM of the same population. Acetate is a product of glucose fermentation under anaerobic metabolism, and is presumed to be in high concentrations after 24 hours in media containing fermenting bacteria. Further genome analysis revealed evidence of mutations that were likely to have negatively impacted fitness of the SCM2 population, a 42 bp deletion in *arcB* and a deletion of 30 genes from *insB-6-ybdK*. However, a subsequent mutation in *glgC*, which has previously been shown to inhibit the wild-type form (325), may have allowed this genotype with deleterious mutations to reach fixation within the population.

6.2 Future perspectives

The general approaches and insights obtained from this thesis could support further investigations in three main areas; elucidating evolutionary pathways for adaptation within and to anaerobic environments by other medically and industrially relevant model organisms; further

characterisation of the adaptive pathways found in this thesis and further characterisation of the biodiversity that arose in the anaerobic environment in this study.

This thesis used a facultative anaerobe as a model to understand adaptation to the anaerobic environment. Many medically and industrially relevant microbes are facultative anaerobes, thus insights obtained may allow for predictions of adaptation to the anaerobic environment of other facultative anaerobes. One such medically important organism is the opportunistic human pathogen *Pseudomonas aeruginosa*, which infects the lungs of patients with cystic fibrosis, and quickly become infectious during adaptation to the anaerobic environment (63, 155, 342, 343). A better understanding it's adaptation to anaerobiosis may allow the identification or development of drugs that are effective during this critical stage of infection. With regard to organisms frequently used in the biotechnology sector, a greater understanding of the adaptation of the facultative anaerobe, *Saccharomyces cerevisiae* – a unicellular eukaryote - to the anaerobic environment, could be of interest. The improvement of test strains for greater fitness, or enhanced production of compounds of interest (344), in the anaerobic environment may be beneficial for various downstream applications. Similar strategies could be applied to other yeast strains which produce bioethanol through anaerobic fermentation of glycerol in the biofuel industry (345-347).

Many different adaptive pathways were found in *E. coli* during adaption to aerobic and anaerobic environments in this study. These ranged from GASP, to fermentation pathways, to inactivation of TA and virulence genes. To verify if the underlying mutations led to increases in fitness, and the extent to which fitness was increased as a result, re-construction of defined mutants via allelic replacements are desirable. In addition, experiments that identify the molecular mechanisms underpinning these increases in fitness should also be considered. In general, GASP phenotypes are mediated by the cells acquired ability to catabolise amino acids and nucleic acids released by dead or lysed cells. To confirm GASP-based mutations occurred in aerobic lineages, microbial phenotypic assays (e.g. Biolog) (348) with numerous amino acids and nucleic acids as informative substrates could be a measure of the GASP phenotype. Furthermore, such assays could also address if loss of function mutations in certain amino acid transport genes, such as *brnQ* allowed for increased fitness in many fluctuating lineages. As *brnQ* is responsible for the uptake of the LIV family amino acids (L-leucine, L-isoleucine and L-valine), evolved cells with the *brnQ* mutations may display reduced uptake of these particular amino acids when compared to the ancestral

strain. Experimental verification of mutations which are predicted lead modified activities of fermentation pathways could be performed by detecting differences in the type and production rate of fermentation end-products using high performance liquid chromatography or gas chromatography of spent culture media of the ancestral and evolved strains. For the investigation of the role of prophage activity or presence in adaptation to the anaerobic environment, removals of each prophage, individually, and in combination, could identify if their deletion resulted in increased fitness among lineages. This could also identify which prophage led to the highest increase in fitness when lost from the genome. Furthermore, it was hypothesised that frequent mutations in TA associated genes may have led to increased fitness in lineages by eliminating unnecessary functions in the cell. This could be tested by qPCR of lineages that acquired TA mutations to confirm if a decrease in transcription of TA genes resulted from loss of function mutations that occurred during the course of this experiment. This strategy may also be used for the investigation of the inactivation of virulence genes.

Within the anaerobic environment, the origin and maintenance of biodiversity of SCM and TCM in this thesis remains unclear and further experiments are required to understand this phenomenon. Experiments would ideally include more in-depth sequencing with a combination of whole genome sequencing of clones, as well as whole population sequencing of the co-evolved anaerobic populations. The isolation of clones is required to confirm the SCM phenotype and sequencing of the clonal genome would identify the mutation(s) responsible. In addition, subsequent sequencing of the whole population at frequent intervals, before and after the origin of the variants, could potentially determine when the mutation responsible first arose in the population, as well as the frequency changes of the mutation within the population through time. Furthermore, to understand the component of the media which was responsible for the biodiversity, additional LTEE with modified anaerobic media, particularly acetate may allow for insights. The establishment of lineages in anaerobic media with different concentrations of acetate, and subsequent monitoring of the extent of SCM generation in replicate populations, may identify if this compound was responsible for morphological diversification in the anaerobic environment.

6.3 Summary

The application of WGS to long-term *E. coli* experimental lineages evolved under strict aerobic, strict anaerobic environments, and a treatment that fluctuated between the two conditions provided insights into the molecular basis of the evolutionary process in the anaerobic

environment in a level of detail not previously realised. Results from this thesis offer insight into the pathways that are associated with adaptation to the anaerobic environment. The large role played by IS element movement within genomes during adaptation to the anaerobic environment supports the view that mobile genetic elements, including prophage excision from genomes, may be important for adaptation to new environments.

Understanding pathways that facultative anaerobic organisms utilise while adapting to anaerobic environments may have several practical implications. For example, the molecular mechanisms underpinning the adaptation of microbes to industrial fermentative processes may facilitate productivity enhancement or the generation of new characteristics for industrial application. Furthermore, bacterial pathogens that become infectious under anaerobic conditions are able to diversify and adapt to broaden their host range. Knowledge of the pathways involved in pathogenicity and divergence may enable new targeted treatments to be developed.

In conclusion, this study attempted to understand adaptation to the anaerobic environment. The evolutionary dynamics in the anaerobic environment were more complex than in the aerobic environment. Findings included the generation of biodiversity likely through a form of niche construction, maintenance of diversity through negative frequency dependent selection, and declines in fitness due to fixation of maladaptive traits and the evolution of a putative killing mechanism, all within 4,000 generations. These findings highlight the importance of investigating evolution in a wide range of environments, a key factor in the overall goal of understanding evolution in the wild.

Chapter Seven : Appendix

Table 7.1: List of primers used in this study

Name	5' → 3' sequence	Application
fD1	aga gtt tga tcc tgg ctc ag	16S rRNA gene amplification
rD1	aag gag gtg atc cag cc	16S rRNA gene amplification
adhE m	cca cag gtt ttt tca tct tta tag gcg ttg tag ata c	Mutation screening
adhE c	cca cag gtt ttt tca tct tta tag gcg ttg tag ata t	Mutation screening
adhE p	ggt atc gtc gaa gat aaa gtg atc aaa aac cac ttt g	Mutation screening
ansP m	gcg ttt cgc ggc gtg tta	Mutation screening
ansP c	gcg ttt cgc ggc gtg ttg	Mutation screening
ansP p	cgc cag gat gaa taa aca ttg ttc atg gc	Mutation screening
fadD m	gat cgc cca tac ggg tca gc	Mutation screening
fadD c	gat cgc cca tac ggg tca ga	Mutation screening
fadD p	atc agc tta acg ata gcg gcg c	Mutation screening
kpsD m	ttc gcc aca aag cgg ctg aa	Mutation screening
kpsD c	ttc gcc aca aag cgg ctg at	Mutation screening
kpsD p	gcc tca ctt tct gcc cag tca tc	Mutation screening
kpsT 100 m	tga aga att tgc cga gct cgg taa gtg	Mutation screening
kpsT 100 c	tga aga att tgc cga gct cgg taa gta	Mutation screening
kpsT 100 p	ctt aaa tgc cat gct taa acc aaa gcc taa ac	Mutation screening
kpsT 140 m	gat cgc gtt tag gct ttg gtt taa gca c	Mutation screening
kpsT 140 c	gat cgc gtt tag gct ttg gtt taa gca t	Mutation screening
kps 140 p	ctt tat gcc ttt ctt taa aca att gag cgc	Mutation screening
kpsT 180 m	caa cgt tgc taa gaa tga ttg gtg gta ttg ag	Mutation screening
kpsT 180 c	caa cgt tgc taa gaa tga ttg gtg gta ttg ac	Mutation screening
kps 180 p	gtt ctt ctt gct tcg cgt ata acc gc	Mutation screening
kpsE m	cag cca gca aca gtt tca ggg g	Mutation screening
kpsE c	cag cca gca aca gtt tca ggg t	Mutation screening
kpsE p	gag ttc aac acc gag ctg tac aaa ctg	Mutation screening
rpoD 160 m	aga cgt tct tgc gta cga acg ct	Mutation screening
rpoD 160 c	aga cgt tct tgc gta cga acg cg	Mutation screening
rpoD 160 p	tca aag cga aag gtc gca gtc acg	Mutation screening
rpoD 120 m	caa cac gat cgt act gtt cca gca gat c	Mutation screening
rpoD 120 c	caa cac gat cgt act gtt cca gca gat a	Mutation screening
rpoD 120 p	gac ccg cga agg cga aat tg	Mutation screening

Table 7.2: Raw genome sequence data:

Lineage	Generation	Library construction type	Insert size (bp)	Mapped reads	Read length (bp)	Fold coverage
Anc	0	90 bp paired end	500	11,177,776	90	219
Rif	0	90 bp mate pair	2,000	8,544,607	90	167
AE-1	2K	90 bp paired end	500	11,788,889	90	231
AE-2	2K	90 bp paired end	500	12,211,111	90	239
AE-3	2K	90 bp paired end	500	11,555,555	90	226
AE-4	2K	90 bp paired end	500	11,988,888	90	235
AE-5	2K	90 bp paired end	500	11,877,778	90	233
AE-6	2K	90 bp paired end	500	11,777,778	90	231
AE-7	2K	90 bp paired end	500	11,433,334	90	224
AN-1	2K	90 bp paired end	500	11,622,223	90	228
AN-2	2K	90 bp mate pair	2,000	8,815,835	90	173
AN-3	2K	90 bp paired end	500	11,288,890	90	221
AN-4	2K	90 bp paired end	500	11,188,889	90	219
AN-5	2K	90 bp paired end	500	11,688,889	90	229
AN-6	2K	90 bp paired end	500	11,655,555	90	228
AN-7	2K	90 bp paired end	500	11,711,112	90	229
FL-1	2K	90 bp paired end	500	11,477,778	90	225
FL-2	2K	90 bp paired end	500	11,422,224	90	224
FL-3	2K	90 bp paired end	500	11,344,444	90	222
FL-4	2K	90 bp paired end	500	11,499,999	90	225
FL-5	2K	90 bp paired end	500	11,444,436	90	224
FL-6	2K	90 bp paired end	500	11,599,998	90	227
FL-7	2K	90 bp paired end	500	11,244,445	90	220
AE-1	4K	90 bp paired end	500	9,396,633	90	184
AE-2	4K	90 bp paired end	500	11,146,350	90	218
AE-3	4K	90 bp paired end	500	10,154,982	90	199
AE-4	4K	90 bp paired end	500	9,629,376	90	189
AE-5	4K	90 bp paired end	500	8,073,132	90	158
AE-6	4K	90 bp paired end	500	10,105,743	90	198
AE-7	4K	90 bp paired end	500	11,916,726	90	233
AN-1	4K	90 bp mate pair	2,000	10,210,487	90	200
AN-2	4K	90 bp mate pair	2,000	9,285,394	90	182
AN-3	4K	90 bp mate pair	2,000	9,275,265	90	182
AN-4	4K	90 bp mate pair	2,000	10,659,088	90	209
AN-5	4K	90 bp mate pair	2,000	8,387,144	90	164
AN-6	4K	90 bp mate pair	2,000	8,655,605	90	170
AN-7	4K	90 bp mate pair	2,000	9,857,860	90	193
FL-1	4K	90 bp paired end	500	12,889,134	90	252
FL-2	4K	90 bp paired end	500	11,889,134	90	233
FL-3	4K	90 bp paired end	500	11,979,834	90	235
FL-4	4K	90 bp paired end	500	11,236,998	90	220
FL-5	4K	90 bp paired end	500	9,717,158	90	190
FL-6	4K	90 bp paired end	500	9,678,437	90	190
FL-7	4K	90 bp paired end	500	12,010,478	90	235

Figure 7.3: List of all mutation in aerobically evolved genomes

Lineage[a]	Type[b]	Class[c]	Reference Position 1[d]	Change[e]	Amino Acid[f]	Reference Gene 1[g]	Description Gene 1[h]	Reference Gene 2[i]	Description Gene 2[j]	Comment[k]
AE-2K-1	SNP	Transversion	700,702	G-->T	Gly (48) --> Val	*gnd*	6-phosphogluconate dehydrogenase	-	-	-
AE-2K-1	Mobile element	Insertion	1,464,679	-	-	*IdrC* (+344)	Long Direct Repeats, toxin-antitoxin system	*IdrB* (-84)	Long direct repeats, toxin-antitoxin system	IS150 mediated, intergenic
AE-2K-1	SNP	Transversion	1,943,095	A-->T	Ile (355) --> Thr	*ECB_00733*	Hypothetical protein	-	-	-
AE-2K-1	SNP	Transition	3,000,161	T-->C	Tyr (111) --> Cys	*kpsT*	Polysialic acid transport ATP binding protein KpsT	-	-	-
AE-2K-1	Deletion	Small	3,866,358	-G	-	*trkD* (+6)	Potassium transport protein Kup	*insJ-5* (-51)	IS150 putative transposase	Intergenic
AE-2K-1	Deletion	Small	3,524,253 - 3,525,072	-	-	*rhsB*	rhsB element core protein RshB	-	-	-
AE-4K-1	Mobile element	Insertion	16,972	-	-	*hokC* (-71)	Small toxic membrane polypeptide	*nhaA* (-514)	pH dependent sodium or proton antiporter	IS150 mediated, intergenic
AE-4K-1	Mobile element	Insertion	388,234	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
AE-4K-1	Mobile element	Insertion	550,063	-	-	*ECB_00516*	Hypothetical protein	-	-	IS3 mediated
AE-4K-1	Mobile element	Insertion	963,716	-	-	*ydiU*	Hypothetical protein	-	-	IS150 mediated
AE-4K-1	Mobile element	Insertion	1,272,399	-	-	*trg* (-257)	Methyl accepting chemotaxis protein III, ribose and galactose sensor	*mokB* (-82)	Regulatory peptide	IS150 mediated Intergenic

							receptor			
AE-4K-1	Deletion	Small	1,389,511	-	-	*rnb* (+137)	Exoribonuclease II	*yciR* (-98)	RNase II stability modulator	Intergenic
AE-4K-1	Mobile element	Deletion	1,764,886	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
AE-4K-1	SNP	Transition	1,905,307	T --> C	Tyr (74) --> Tyr	*ybiI*	Hypothetical protein	-	-	-
AE-4K-1	Insertion	Duplication	2,844,904	×2	-	*galR*	DNA binding transcriptional regulator	-	-	-
AE-4K-1	Mobile element	Insertion	2,975,127	-	-	*flu*	Antigen 43 (Ag43) phase−variable biofilm formation autotransporter	-	-	IS150 mediated
AE-4K-1	Mobile element	Insertion	3,000,527	-	-	*kpsM*	Polysialic acid transport protein KpsM	-	-	IS3 mediated
AE-4K-1	Mobile element	Insertion	4,015,454	-	-	*yihS*	Putative glucosamine isomerase	-	-	IS150 mediated
AE-4K-1	Mobile element	Insertion	4,076,128	-	-	*metL*	Bifunctional aspartate kinase II and homoserine dyhydrogenase	-	-	IS150 mediated
AE-4K-1	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated
AE-2K-2	SNP	Transition	109,384	A-->G	Lys (8) --> Lys	*lpxC*	N-acetylglucosamine deacetylase	-	-	-
AE-2K-2	SNP	Transition	2,052,242	G-->A	Gly (308) --> Ser	*nagA*	N-acetylglucosamine -6- phosphate deacetylase	-	-	-
AE-2K-2	SNP	Transversion	2,987,334	A-->C	Thr (365) --> Pro	*kpsE*	Capsule polysaccharide export protein KpsE	-	-	-

AE-2K-2	SNP	Transition	4,217,932	G-->A	Ala (249) --> Thr	*uvrA*	UvrABC DNA lesion repair protein	-	-	-
AE-4K-2	Mobile element	Insertion	388,234	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
AE-4K-2	Mobile element	Insertion	550,063	-	-	*ECB_00516*	Hypothetical protein	-	-	IS3 mediated
AE-4K-2	Mobile element	Insertion	963,716	-	-	*ydiU*	Hypothetical protein	-	-	IS150 mediated
AE-4K-2	Mobile element	Insertion	1,272,468	-	-	*trg* (-326)	Methyl accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB* (-13)	Regulatory peptide	IS150 mediated, intergenic
AE-4K-2	SNP	Transition	1,905,307	T --> C	Tyr (74) --> Tyr	*ybiI*	Hypothetical protein	-	-	-
AE-4K-2	Insertion	Duplication	2,844,904	×2	-	*galR*	DNA binding transcriptional regulator	-	-	-
AE-4K-2	Mobile element	Insertion	3,000,527	-	-	*kpsM*	Polysialic acid transport protein KpsM	-	-	IS3 mediated
AE-4K-2	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated
AE-2K-3	SNP	Transversion	2,988,653	A-->T	Ile (418) --> Phe	*kpsD*	Polysialic acid transporter	-	-	-
AE-2K-3	SNP	Transition	3,000,095	T-->C	Met (133) --> Thr	*kpsT*	Polysialic acid transport ATP binding protein KpsT	-	-	-
AE-2K-3	Deletion	Small	3,915,605	-	-	*hemX*	Putative uroporphyrinogen III C-methyltransferase	-	-	-

AE-2K-3	SNP	Transversion	4,107,509	A-->C	Lys (2) --> Thr	*yijD*	Hypothetical protein	-	-	-
AE-4K-3	Mobile element	Deletion	261,860	-	-	*[frsA]*	Fermentation respiratory switch protein	*crl*	RNA polymerase holoenzyme assembly factor	IS1 mediated, intergenic
AE-4K-3	Mobile element	Insertion	388,234	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
AE-4K-3	Mobile element	Insertion	963,716	-	-	*ydiU*	Hypothetical protein	-	-	IS150 mediated
AE-4K-3	Mobile element	Insertion	1,465,130	-	-	*IdrB (+260)*	Long Direct Repeats, toxin-antitoxin system	*IdrA* (-167)	Long direct repeats, toxin-antitoxin system	IS150 mediated, intergenic
AE-4K-3	SNP	Transition	1,905,307	T --> C	Tyr (74) --> Tyr	*ybiI*	Hypothetical protein	-	-	-
AE-4K-3	Deletion	Small	2,071,483	-	-	*insJ-2* (-51)	IS150 protein	*rihA* (-486)	Ribonucleoside hydrolase 1	IS mediated, intergenic
AE-4K-3	Insertion	Duplication	2,844,904	×2	-	*galR*	DNA binding transcriptional regulator	-	-	-
AE-4K-3	Mobile element	Insertion	3,000,527	-	-	*kpsM*	Polysialic acid transport protein KpsM	-	-	IS3 mediated
AE-4K-3	SNP	Transversion	3,399,123	C --> A	Asn (195) --> Lys	*cysG*	Siroheme synthase	-	-	-
AE-4K-3	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated
AE-2K-4	SNP	Transition	323,923	T-->C	Val (20) --> Ala	*prpD*	2- methylcitrate dehydratase	-	-	-
AE-2K-4	SNP	Transversion	2,988,653	A-->T	Ile (418) --> Phe	*kpsD*	Polysialic acid transporter	-	-	-

AE-2K-4	SNP	Transition	3,000,095	A-->G	Met (133) --> Thr	*kpsT*	Polysialic acid transport ATP binding protein KpsT	-	-	-
AE-4K-4	Mobile element	Deletion	388,234	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
AE-4K-4	Mobile element	Insertion	550,063	-	-	*ECB_00516*	Hypothetical protein	-	-	IS3 mediated
AE-4K-4	SNP	Transition	1,905,307	T --> C	Tyr (74) --> Tyr	*ybiI*	Hypothetical protein	-	-	-
AE-4K-4	SNP	Transversion	2,734,340	C --> A	Arg (299) --> Cys	*rpoS*	RNA pol sigma factor RpoS	-	-	-
AE-4K-4	Insertion	Duplication	2,844,904	×2	-	*galR*	DNA binding transcriptional regulator	-	-	-
AE-4K-4	Mobile element	Insertion	3,000,527	-	-	*kpsM*	Polysialic acid transport protein KpsM	-	-	IS3 mediated
AE-4K-4	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated
AE-2K-5	SNP	Transversion	863,948	T -->G	Ile (148) --> Met	*fadD*	Long chain fatty acid CoA ligase	-	-	-
AE-2K-5	Deletion	Small	892,201	-	-	*yeaA*	Methionine sulfoxide reductase B	-	-	-
AE-2K-5	Mobile element	Insertion	2,972,936	-	-	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	IS1 mediated
AE-2K-5	SNP	Transversion	3,000,346	G-->C	Asp (133) --> Thr	*kpsT*	Polysialic acid transport ATP binding protein KpsT	-	-	-
AE-2K-5	Insertion	Single base	3,866,357	+G	-	*trkD* (+5)	Potassium transport	*insJ-5* (-	IS150 putative	Intergenic

							protein Kup	52)	transposase	
AE-4K-5	SNP	Transition	389,105	G --> A	Trp (364) --> Stop	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
AE-4K-5	Insertion	Base	818,462	+ACAG	-	*zwf*	Glucose-6-phosphate 1-dehydrogenase	*edd*	Phosphogluconate dehydrogenase	Intergenic
AE-4K-5	SNP	Transition	1,046,106	C --> T	Ala (278) --> Ala	*rnfD*	Electron transport complex protein RnfD	-	-	-
AE-4K-5	Mobile element	Insertion	1,272,468	-	-	*trg* (-326)	Methyl accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB* (-13)	Regulatory peptide	IS150 mediated, intergenic
AE-4K-5	Mobile element	Insertion	1,551,946	-	-	*ycfQ* (+19)	Putative DNA-binding transcriptional regulator	*ycfJ* (+42)	Hypothetical protein	IS150 mediated, intergenic
AE-4K-5	Deletion	Small	1,716,607	-	-	*ycbS*	Putative outer membrane usher protein	-	-	-
AE-4K-5	SNP	Transition	2,119,116	-	Lys (266) --> Lys	*metG*	Mthionyl –tRNA synthetase	-	-	-
AE-4K-5	Mobile element	Insertion	2,138,008	-	-	*yehU*	Putative sensory kinase in two-component system with YehT	-	-	IS150 mediated
AE-4K-5	Mobile element	Insertion	3,000,519	-	-	*kpsM*	Polysialic acid transport protein KpsM	-	-	IS3 mediated
AE-4K-5	SNP	Transversion	3,119,112	T --> G	Tyr (143) --> Asp	*rpoD*	RNA polymerase sigma factor D	-	-	-

AE-4K-5	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated
AE-2K-6	SNP	Transition	92,892	T --> C	Leu (152) --> Pro	*yabB*	Cell division protein MraZ	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*[ECB_01536]*	25 genes involved in 8 unknown	*insE-3*	-	IS3 mediated
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*[ECB_01536]*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01535*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01534*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01533*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*hokD*	Host killing	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*rem*	Removal of superoxide radicals	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01530*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01528*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01527*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*insF-4*	InsF component of InsFE transposase of IS3	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*insE-4*	InsE component of InsFE transposase of IS3	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01523*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01522*	Unknown	-	-	-

AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*essQ*	Phage lambda S lysis holin protein homolog, Qin prophage	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ydfR*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ydfQ-2*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ydfP*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*cspI*	Cold shock protein wioth complex regulation - affect memebrane fluidity, Quin prophage	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ECB_01516*	Unknown	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ynfN*	Cold shock-induced protein, function unknown, Qin prophage	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*gnsB*	GnsA and GnsB affect unsaturated fatty acid abundance and membrane fluidity	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*ybcW*	Function unknown, DLP12 prophage	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*insF-3*	InsF component of InsFE transposase of IS3	-	-	-
AE-2K-6	Deletion	Large	1,111,336	25 genes	-	*insE-3*	InsE component of InsFE transposase of IS3	-	-	-
AE-2K-6	Deletion	Small	2,999,898	-	-	*kpsT*	Polysialic acid transport ATP	-	-	-

							binding protein KpsT			
AE-2K-6	SNP	Transversion	3,119,520	C --> A	Arg (279) --> Ser	*rpoD*	RNA polymerase sigma factor RpoD	-	-	-
AE-4K-6	Mobile element	Insertion	16,972	-	-	*hokC* (-71)	Small toxic membrane polypeptide	*nhaA* (-514)	pH dependent sodium or proton antiporter	IS150 mediated, intergenic
AE-4K-6	Mobile element	Insertion	360,203	-	-	*yaiT*	Hypothetical protein	-		IS150 mediated
AE-4K-6	Mobile element	Insertion	388,275	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
AE-4K-6	Mobile element	Insertion	785,037	-	-	*cheB*	Chemotaxis-specific methylesterase	-	-	IS150 mediated
AE-4K-6	Deletion	Small	1,272,262	-	-	*trg*	Methyl accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB*	Regulatory peptide	Deletion mediated by 2 IS150 insertions at 1272262 and 1272468, resulting in 67 bp deletion, intergenic
AE-4K-6	Insertion	Base	1,328,521	+ A	-	*zntB*	Trans-membrane zinc transporter	-	-	-
AE-4K-6	Mobile element	Insertion	1,464,143	-	-	*chaA* (+320)	Calcium/sodium proton antiporter	*IdrC* (-83)	Long direct repeats, toxin-antitoxin system	IS150 mediated, intergenic
AE-4K-6	Mobile element	Insertion	1,675,380	-	-	*yccC*	Cryptic autophosphorylating protein tyrosine kinase Etk	-	-	IS150 mediated

AE-4K-6	Mobile element	Insertion	2,127,047	-	-	*yehI*	Hypothetical	-	-	IS150 mediated
AE-4K-6	Mobile element	Insertion	2,378,921	-	-	*fadL*	Long−chain fatty acid outer membrane transporter	-	-	IS150 mediated
AE-4K-6	Deletion	-	2,748,700	-	-	*insJ-3 (-160)*	IS150 hypothetical protein	*cysH (-231)*	Phosphoadenosine phosphosulfate reductase	Intergenic
AE-4K-6	Mobile element	Insertion	2,850,135	-	-	*kduD*	2-deoxy-D-gluconate 3-dehydrogenase	-	-	IS150 mediated
AE-4K-6	Mobile element	Insertion	2,974,778	-	-	*flu*	Antigen 43 (Ag43) phase−variable biofilm formation autotransporter	-	-	IS150 mediated
AE-4K-6	Mobile element	Insertion	3,000,514	-	-	*flu*	Polysialic acid transport protein KpsM	-	-	IS3 mediated
AE-4K-6	Mobile element	Insertion	3,009,254	-	-	*gspE*	Type II secretion protein GspE	-	-	IS150 mediated
AE-4K-6	Deletion	-	3,164,090	-	-	*yhaO (-288)*	Putative transporter	*tdcG (+51)*	L-serine dehydratase 3	Intergenic
AE-4K-6	Mobile element	Insertion	3,445,026	-	-	*yhgA*	Hypothetical protein	-	-	IS150 mediated
AE-4K-6	Deletion	-	3,866,358	-G	-	*trkD (+6)*	Potassium transport protein Kup	*insJ-5 (-51)*	IS150 putative transposase	Intergenic
AE-4K-6	Mobile element	Insertion	4,015,454	-	-	*yihS*	Putative glucosamine isomerase	-	-	IS150 mediated
AE-4K-6	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated
AE-4K-6	Deletion	IS150	4,532,961	16 genes		*yjiI a.k.a. kptA*	tRNA phosphotransferase activity	*[mdoB] a.k.a. opgB*	Transfers phosphoglycerol to membrane derived	IS150 mediated, intergenic

									oligosaccharides	
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*yjiY*	Induced in stationary phase when cells are growing on amino acids or peptides	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaC*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaB*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaA*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaX*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaI*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hok*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaF*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hok*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaE*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaG*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*hpaR*	Bacterial pathogenicity	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*tsr*	Chemotactic response to changes in pH	-	-	-
AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*yjiZ*	Proton-driven L-galactonate uptake transporter	-	-	-

AE-4K-6	Deletion	Large	4,532,961	16 genes	-	*yjjN*	L-galactonate oxidoreductase that is required for growth on L-galactonate as the sole carbon source under high-throughput growth conditions with limited aeration	-	-	-
AE-4K-6	Mobile element	Insertion	4,581,547	-	-	*[mdoB]*	phosphoglycerol transferases	-	-	-
AE-2K-7	Mobile element	Insertion	2,972,936	-	-	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	IS1 mediated
AE-2K-7	SNP	Transition	3,000,346	G --> C	Asp (133) --> Thr	*kpsT*	Polysialic acid transport ATP binding protein KpsT	-	-	-
AE-2K-7	Insertion	-	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup	*insJ-5 (-52)*	IS150 putative transposase	Intergenic
AE-4K-7	Mobile element	Insertion	388,234	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 medaited
AE-4K-7	Mobile element	Insertion	550,063	-	-	*ECB_00516*	Hypothetical protein	-	-	IS3 mediated
AE-4K-7	SNP	Transition	1,573,697	A --> G	Arg (47) --> Arg	*yceF*	Maf-like protein	-	-	-
AE-4K-7	SNP	Transition	1,905,307	T --> C	Tyr (74) --> Tyr	*ybiI*	Hypothetical protein	-	-	-
AE-4K-7	SNP	Transversion	2,155,627	T --> A	Thr (138) --> Thr	*yohK*	Hypothetical protein	-	-	-
AE-4K-7	Insertion	Duplication	2,844,904	×2	-	*galR*	DNA binding	-	-	-

							transcriptional regulator			
AE-4K-7	Mobile element	Insertion	2,974,778	-	-	*flu*	Antigen 43 (Ag43) phase−variable biofilm formation autotransporter	-	-	IS150 mediated
AE-4K-7	Mobile element	Insertion	3,000,527	-	-	*kpsM*	Polysialic acid transport protein	-	-	IS3 mediated
AE-4K-7	SNP	Transversion	3,778,620	A --> T	Lys (429)--> Gln	*uhpT*	Sugar phosphate antiporter	-	-	-
AE-4K-7	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes	-	*ECB_01523*	Hypothetical protein	*ybcW*	Hypothetical protein	No evidence for IS mediation
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*ECB_01523*	Hypothetical protein	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*ECB_01522*	Hypothetical protein	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*essQ*	Phage lambda S lysis holin protein homolog, Qin prophage	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*ydfR*	Unknown	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*ydfQ-2*	Unknown	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*ydfP*	Unknown	-	-	-

AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*cspI*	Cold shock protein with complex regulation - affect membrane fluidity, Qin prophage	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*ECB_01516*	Unknown	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*ynfN*	Cold shock-induced protein, function unknown, Qin prophage	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*gnsB*	GnsA and GnsB affect unsaturated fatty acid abundance and membrane fluidity	-	-	-
AE-4K-7	Deletion	-	1,117,805 - 1,117,806	11 genes		*ybcW*	Function unknown, DLP12 prophage	-	-	-
AE-4K-7	Deletion	-	3,981,502 - 3,983,238	2 genes	-	*alaT*	Alanine tRNAs	-	-	-
AE-4K-7	Deletion	-	3,981,502 - 3,983,238	2 genes		*alaT*	One of five alanine tRNAs	-	-	-
AE-4K-7	Deletion	-	3,981,502 - 3,983,238	2 genes		*[rrlA]*	One of seven rRNAs	-	-	-

[a] The lineage in which the mutation arose.

[b] The type of mutation that had occurred.

[c] Further sub-classification of the mutation type that occurred. Note for large deletions, the first and last gene affected is listed as the first event, followed by each in between listed underneath.

[d] The position on the reference genome at which the mutation occurred. In some cases, particularly with large deletions, the exact location of the beginning of the mutation could not be specified from breseq output and instead the likely range is given.

[e] The change that occurred at the DNA level as a result of the mutation. In the case of large deletions, the number of genes deleted is indicated.
[f] The amino acid change that had occurred at the protein level as a result of the mutation. Also indicated was the position along the protein in which the change occurred.
[g] The gene mutated. In the cases of genes partially affected, the gene name was enclosed with parenthesis []. Intergenic mutations are identified by the closest gene to the mutation with the number of bps from the beginning of the gene (+ for downstream and – for upstream).
[h] A brief functional description of the protein encoded by Reference Gene 1.
[i] Intergenic mutations are identified by the closest gene to the mutation with the number of bps from the beginning of the gene
[j] A brief functional description of the protein encoded by Reference Gene 2.
[k] Additional information if required.

Figure 7.4: List of all mutation in anaerobically evolved genomes

Lineage[a]	Type[b]	Class[c]	Reference Position 1[d]	Change[e]	Amino Acid[f]	Reference Gene 1[g]	Description Gene 1[h]	Reference Gene 2[i]	Description Gene 2[j]	Comment[k]
AN-2K-1	Mobile element	Insertion	16,972	-	-	*hokC (-71)*	Small toxic membrane polypeptide	*nhaA (-514)*	pH dependent sodium or proton antiporter	IS150 mediated, intergenic
AN-2K-1	Mobile element	Insertion	621,252	-	-	*citG*	Triphosphoriboseyl-dephospho-CoA synthase	-	-	IS150 mediated
AN-2K-1	Mobile element	Insertion	910,345	-	-	*ynjI*	Hypothetical protein	-	-	IS150 mediated
AN-2K-1	SNP	Transition	1,238,659	C --> T	Gln (11) --> Stop	*ansP*	L-asparagine transporter	-	-	-
AN-2K-1	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, intergenic, online
AN-2K-1	SNP	Transition	1,439,030	A --> G	Tyr (75) --> Cys	*adhE*	Alcohol dehydrogenase	-	-	-
AN-2K-1	Mobile element	Insertion	2,564,470	-	-	*hcaC*	3-phenylpropionate dioxygenase ferredoxin subunit	-	-	IS150 mediated
AN-2K-1	Mobile element	Insertion	2,992,382	-	-	*kpsS*	KpsS protein capsular biosynthesis	-	-	IS1 mediated
AN-2K-1	Deletion	Small	4,295,377	-	-	*dcuR (-755)*	DNA binding transcriptional activator DcuR	*yjdI (-1,055)*	Hypothetical protein	-
AN-2K-1	Mobile element	Insertion	4,381,58	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated,

			3							online
AN-2K-1	Mobile element	Insertion	4,581,545	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
AN-2K-1	Deletion	Large	998,264 -999,628	4 genes	-	*[pykF]*	Pyruvate kinase I	*[ydhV]*	Oxidoreductase system	IS150 mediated insertion in pykF and ydhV
AN-2K-1	Deletion	Large	998,264 -999,628	4 genes	-	*[pykF]*	Pyruvate kinase I	-	-	-
AN-2K-1	Deletion	Large	998,264 -999,628	4 genes	-	*ydhZ*	Unknown	-	-	-
AN-2K-1	Deletion	Large	998,264 -999,628	4 genes	-	*ydhY*	Unknown	-	-	-
AN-2K-1	Deletion	Large	998,264 -999,628	4 genes	-	*[ydhV]*	Unknown	-	-	-
AN-4K-1	Deletion	Small	161,094	-	-	*pcnB*	Poly(A) polymerase I	-	-	-
AN-4K-1	Deletion	Small	838,221	-	-	*proQ*	Putative solute DNA competence effector	-	-	-
AN-4K-1	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, intergenic, online
AN-4K-1	Insertion	Duplication	1,328,413	×2	-	*dbpA*	ATP-dependent RNA helicase DbpA	*zntB*	Zinc transporter	Intergenic
AN-4K-1	SNP	Transition	1,439,673	G --> A	Ala (623) --> Thr	*adhE*	Alcohol dehydrogenase	-	-	-
AN-4K-1	Mobile element	Insertion	1,984,267	-	-	*cydA (-440)*	Cytochrome D terminal oxidase subunit I	*ybgG (+407)*	Alpha mannosidase	IS1 mediated, intergenic

AN-4K-1	Deletion	Large	2,057,189	2 genes	-	*[glnU]*	Glutamine tNRA	*[glnW]*	Glutamine tRNA	Intergenic
AN-4K-1	Deletion	Large	2,057,189	2 genes		*[glnU]*	Four glutamine tRNAs.	-	-	-
AN-4K-1	Deletion	Large	2,057,189	2 genes		*[glnW]*	One of four glutmamine tRNAs	-	-	-
AN-4K-1	Mobile element	Insertion	2,424,083	-	-	*alaW (-154)*	tRNA-Ala	*yfeC (-82)*	Putative DNA binding transcription factor	IS150 mediated, intergenic
AN-4K-1	SNP	Transversion	2,972,858	T --> G	Met (158) --> Arg	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-1	SNP	Transversion	2,974,237	G --> A	Glu (618) --> Thr	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-1	Deletion	Small	3,260,806	-	-	*arcB*	Aerobic respiration control sensor protein ArcB	-	-	-
AN-4K-1	Insertion	Base	3,298,183	+C	-	*rng*	Ribonuclease G	-	-	-
AN-4K-1	Deletion	Small	3,473,571	-C	-	*glgC*	Glucose-1-phosphate adenylyltransferase	-	-	-
AN-4K-1	Mobile element	Insertion	3,522,963	-	-	*rhsB*	rhsB element core protein RhsB	-	-	IS150 mediated
AN-4K-1	Insertion	Base	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup/	*insJ-5 (-52)*	IS150 putative transposase	Intergenic
AN-4K-1	Mobile	Insertion		-	-	*cycA*	D-alanine/D-	-	-	IS150

	element		4,381,583				serine/glycine permease			mediated, online
AN-4K-1	Mobile element	Insertion	4,581,546	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes	-	*[insB-6]*	-	*ybdK*	-	IS1 mediated
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*[insB-6]*	After frameshift event,transcribed with InsA mediates IS1 transposition	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*insA-6*	Transcriptional repressor of IS1	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00514*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00515*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00516*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00517*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ompT*	Outer membrane protease, with specificity for paired basic residues	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode	-	-	-

							cellular envelope proteins at low temperatures and during stationary phase			
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ybcH*	Hypothetical prtein	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*nfrA*	Important in N4 absorption	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00524*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*yhhI-2*	H repeat-associated protein, RhsE-linked, function unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00526*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00527*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00528*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00529*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ECB_00530*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*cusS*	Detoxification of copper and silver ions	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*cusR*	Detoxification of copper and silver ions	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*cusC*	Detoxification of copper and silver ions	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*cusB*	Detoxification of copper and silver ions	-	-	-

AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*cusA*	Detoxification of copper and silver ions	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ylcC*	Phenylalanine transporter	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*pheP*	Low abundance mechanosensitive channel of miniconductance	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ybdG*	Nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*nfnB*	aka nfsB The nfsB-encoded nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12. NfsB reduces a broad range of nitroaromatic compounds [Zenno96], including the antibiotics nitrofurazone and nitrofurantoin	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ybdF*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ybdJ*	Unknown	-	-	-
AN-4K-1	Deletion	Large	546,975 - 547,703	30 genes		*ybdK*	Pyruvate:flavodoxin oxidoreductase and/or pyruvate	-	-	-

							synthase			
AN-2K-2	Mobile element	Insertion	227,515	-	-	*rrsH*	16S ribosomal RNA	-	-	IS150 mediated
AN-2K-2	Mobile element	Insertion	494,056	-	-	*ybbP*	Putative inner membrane protein	-	-	IS150 mediated
AN-2K-2	SNP	Transition	1,439,673	G --> A	Ala (623) --> Thr	*adhE*	Alcohol dehydrogenase	-	-	Online
AN-2K-2	Mobile element	Deletion	1,764,886	-	-	*pflB*	Pyruvate formate lyase 1	-	-	IS150 mediated
AN-2K-2	Mobile element	Insertion	2,024,212	-	-	*rhsC*	rhsC element core protein RshC	-	-	IS150 mediated
AN-2K-2	SNP	Transversion	2,844,146	T --> A	Ala (35) --> Ala	*galR*	DNA binding transcriptional regulator GalR	-	-	-
AN-2K-2	Mobile element	Insertion	872,829	-	-	*yeaS*	Leucine export protein LeuE	-	-	IS150 mediated
AN-2K-2	Mobile element	Insertion	3,994,146	-	-	*polA (+223)*	DNA polymerase I	*engB (+156)*	Ribosome biogenesis GTP-binding protein YsxC	IS150 mediated
AN-2K-2	Mobile element	Insertion	3,096,895	-	-	*yqiI*	Hypothetical protein	-	-	IS1 mediated
AN-2K-2	Deletion	Small	4,295,377	-	-	*dcuR (-755)*	DNA binding transcriptional activator DcuR	*yjdI (-1,055)*	Hypothetical protein	-
AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*[insB-17]*	-	*[atoB]*	-	IS1 mediated

AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*[insB-17]*	After frameshift event,transcribed with InsA mediates IS1 transposition	-	-	-
AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*insA-17*	Transcriptional repressor of IS1	-	-	-
AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*rcsC*	Controlling the temporal expression of genes related to biofilm formation, a trc factor that responds to a Changing env	-	-	-
AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*atoS*	Short chain (C4 to C6) catabolism	-	-	-
AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*atoC*	Short chain (C4 to C6) catabolism	-	-	-
AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*atoD*	Short chain (C4 to C6) catabolism	-	-	-
AN-2K-2	Deletion	Large	2,235,631 - 2,236,33	9 genes	-	*atoA*	Short chain (C4 to C6) catabolism	-	-	-

			2							
AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*atoE*	Short chain (C4 to C6) catabolism	-	-	-
AN-2K-2	Deletion	Large	2,235,631 - 2,236,332	9 genes	-	*[atoB]*	Responds to changes in env conditions	-	-	-
AN-4K-2	Insertion	Multi	26,953	+ CAT	-	*ileS*	Isoleucyl-tRNA synthetase	-	-	-
AN-4K-2	Deletion	Small	161,094	-	-	*pcnB*	Poly(A) polymerase I	-	-	-
AN-4K-2	Mobile element	Insertion	546,599	-	-	*ybcQ (+369)*	Putative anti-termination protein	*insB-6 (+347)*	IS1 protein InsB	Intergenic
AN-4K-2	Deletion	Small	838,221			*proQ*	Putative solute DNA competence effector	-	-	-
AN-4K-2	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Ribose and galactose sensor receptor and regulatory peptide	IS150 mediated, intergenic
AN-4K-2	Insertion	Duplication	1,328,413	×2		*dbpA*	ATP-dependent RNA helicase DbpA	*zntB*	Zinc transporter	Intergenic
AN-4K-2	SNP	Transition	1,439,673	G --> A	Ala (623) --> Thr	*adhE*	Alcohol dehydrogenase	-	-	Online
AN-4K-2	Insertion	Duplication	1,995,122	×2		*sucA*	2-oxoglutarate dehydrogenase E1 component	-	-	-
AN-4K-2	Deletion	Small	2,057,181	-	-	*[glnU]*	Glutamine tNRA	*[glnW]*	Glutamine tRNA	Intergenic

AN-4K-2	SNP	Transversion	2,972,858	T --> G	Met (158) --> Arg	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-2	SNP	Transversion	2,974,237	G --> A	Glu (618) --> Thr	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-2	Deletion	Small	3,260,806	-	-	*arcB*	Aerobic respiration control sensor protein ArcB	-	-	-
AN-4K-2	Insertion	Small	3,298,183	+C	-	*rng*	Ribonuclease G	-	-	-
AN-4K-2	Mobile element	Insertion	3,250,708	-	-	*yrbI*	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	-	-	IS3 mediated
AN-4K-2	Deletion	Small	3,473,571	-C	-	*glgC*	Glucose-1-phosphate adenylyltransferase	-	-	-
AN-4K-2	Mobile element	Insertion	3,669,960	-	-	*rhsA*	rhsA element core protein RhsA	-	-	IS150 mediated
AN-4K-2	Insertion	Base	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup/	*insJ-5 (-52)*	IS150 putative transposase	Intergenic
AN-4K-2	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine/D-serine/glycine permease	-	-	IS150 mediated
AN-4K-2	Mobile element	Insertion	4,581,546	-	-	*nadR*	nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes	-	*insB-6*	-	*ybdK*	-	IS1 mediated
AN-4K-2	Deletion	Large	546,982	30 genes		*[insB-6]*	After frameshift			

			- 547,703				event,transcribed with InsA mediates IS1 transposition			
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*insA-6*	Transcriptional repressor of IS1	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00514*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00515*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00516*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00517*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ompT*	Outer membrane protease, with specificity for paired basic residues	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode cellular envelope proteins at low temperatures and during stationary phase	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ybcH*	Hypothetical prtein	-	-	-
AN-4K-2	Deletion	Large	546,982	30 genes		*nfrA*	Important in N4	-	-	-

			- 547,703				absorption			
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00524*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*yhhI-2*	H repeat-associated protein, RhsE-linked, function unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00526*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00527*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00528*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00529*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ECB_00530*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*cusS*	Detoxification of copper and silver ions	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*cusR*	Detoxification of copper and silver ions	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*cusC*	Detoxification of copper and silver ions	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*cusB*	Detoxification of copper and silver ions	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*cusA*	Detoxification of copper and silver ions	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ylcC*	Phenylalanine transporter	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*pheP*	Low abundance mechanosensitive	-	-	-

							channel of miniconductance			
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ybdG*	Nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*nfnB*	aka nfsB The nfsB-encoded nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12. NfsB reduces a broad range of nitroaromatic compounds [Zenno96], including the antibiotics nitrofurazone and nitrofurantoin	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ybdF*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ybdJ*	Unknown	-	-	-
AN-4K-2	Deletion	Large	546,982 - 547,703	30 genes		*ybdK*	Pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase	-	-	-
AN-2K-3	SNP	Transition	161,770	G --> A	Arg (59) --> Cys	*pcnB*	Poly(A) polymerase I	-	-	-
AN-2K-3	Mobile element	Insertion	621,252	-	-	*citG*	Triphosphoribosyl-dephospho-CoA synthase	-	-	IS150 mediated

AN-2K-3	Mobile element	Insertion	1,071,398	-	-	*tus*	DNA replication terminus site-binding protein	-	-	IS150 mediated
AN-2K-3	SNP	Transition	1,438,030	A --> G	Tyr (75) --> Cys	*adhE*	Alcohol dehydrogenase	-	-	-
AN-2K-3	Mobile element	Deletion	1,764,886	-	-	*pflB*	Pyruvate formate lyase 1	-	-	IS150 mediated
AN-2K-3	Mobile element	Insertion	2,387,885	-	-	*emrY*	Putative multidrug efflux system	-	-	IS3 mediated
AN-2K-3	Mobile element	Insertion	2,564,470	-	-	*hcaC*	3-phenylpropionate dioxygenase ferredoxin subunit	-	-	IS150 mediated
AN-2K-3	Mobile element	Insertion	2,850,178	-	-	*kduD*	2-deoxy-D-gluconate 3-dehydrogenase	-	-	S150 mediated
AN-2K-3	SNP	Transversion	2,973,574	G --> T	Gly (397) --> Trp	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-2K-3	Mobile element	Insertion	2,992,382	-	-	*kpsS*	KpsS protein capsular biosynthesis	-	-	IS1 mediated
AN-2K-3	Mobile element	Insertion	3,367,380	-	-	*bfr*	Bacterioferritin	-	-	IS3 mediated
AN-2K-3	Mobile element	Insertion	3,808,730	-	-	*yidX*	Hypothetical protein	-	-	S150 mediated
AN-2K-3	Deletion	Small	4,295,377	-	-	*dcuR (-755)*	DNA binding transcriptional activator DcuR	*yjdI (-1,055)*	Hypothetical protein	Intergenic
AN-2K-3	Mobile element	Insertion	4,381,58	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated,

			3							online
AN-2K-3	Mobile element	Insertion	4,401,824	-	-	*ytfT*	Putative sugar transporter subunit membrane component of ABC superfamily	-	-	IS150 mediated
AN-2K-3	Mobile element	Insertion	4,581,545	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
AN-4K-3	Deletion	Small	161,094	-	-	*pcnB*	Poly(A) polymerase I	-	-	-
AN-4K-3	Insertion	Duplication	462,234	×2	-	*ybaM*	Hypothetical protein	-	-	-
AN-4K-3	Mobile element	Insertion	1,123,058	-	-	*ynfN*	Hypothetical protein	-	-	IS3 mediated
AN-4K-3	Mobile element	Deletion	1,188,988	-	-	*[ydcA]*	Hypothetical protein	*cybB*	Electron transferring component of cytochrome b	IS1 mediated, intergenic
AN-4K-3	Mobile element	Insertion	1,272,468	-		*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, intergenic
AN-4K-3	SNP	Transition	1,439,673	G --> A	Ala (623) --> Thr	*adhE*	Alcohol dehydrogenase	-	-	-
AN-4K-3	Mobile element	Insertion	1,984,267	-	-	*cydA (-448)*	Cytochrome D terminal oxidase subunit I	*ybgG (+407)*	Alpha mannosidase	IS1 mediated, intergenic
AN-4K-3	Deletion	Small	2,057,176	-	-	*[glnU]*	Glutamine tNRA	*[glnW]*	Glutamine tRNA	Intergenic
AN-4K-3	SNP	Transversion	2,648,318	G --> A	Ala (33) --> Asp	*yfjF*	Hypothetical protein	-	-	-

AN-4K-3	SNP	Transversion	2,972,858	T --> G	Met (158) --> Arg	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-3	Deletion	Small	2,973,807	-	-	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-3	Deletion	Small	3,260,770	-	-	*arcB*	Aerobic respiration control sensor protein ArcB	-	-	-
AN-4K-3	Deletion	Small	3524265 - 3525071	-	-	*rhsB*	RhsB element core protein RshB	-	-	-
AN-4K-3	Insertion	Small	3,298,183	+C	-	*rng*	Ribonuclease G	-	-	-
AN-4K-3	Deletion	Small	3,473,571	-C	-	*glgC*	Glucose-1-phosphate adenylyltransferase	-	-	-
AN-4K-3	Insertion	Small	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup/	*insJ-5 (-52)*	IS150 putative transposase	Intergenic
AN-4K-3	SNP	Transition	3,869,168	-	-	*yieP*	Putative transcriptional regulator	*rrsC*	16S ribosomal RNA	Intergenic
AN-4K-3	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine/D-serine/glycine permease	-	-	IS150 mediated, online
AN-4K-3	Mobile element	Insertion	4,581,549	-	-	*nadR*	nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*[insB-6]*	After frameshift event,transcribed with InsA mediates IS1 transposition	-	-	-

AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*insA-6*	Transcriptional repressor of IS1	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00514*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00515*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00516*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00517*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ompT*	Outer membrane protease, with specificity for paired basic residues	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode cellular envelope proteins at low temperatures and during stationary phase	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ybcH*	Hypothetical prtein	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*nfrA*	Important in N4 absorption	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00524*	Unknown	-	-	-

AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*yhhI-2*	H repeat-associated protein, RhsE-linked, function unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00526*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00527*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00528*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00529*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ECB_00530*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*cusS*	Detoxification of copper and silver ions	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*cusR*	Detoxification of copper and silver ions	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*cusC*	Detoxification of copper and silver ions	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*cusB*	Detoxification of copper and silver ions	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*cusA*	Detoxification of copper and silver ions	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ylcC*	Phenylalanine transporter	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*pheP*	Low abundance mechanosensitive channel of miniconductance	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ybdG*	Nitroreductase is the minor oxygen-	-	-	-

							insensitive nitroreductase present in E. coli K-12			
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*nfnB*	aka nfsB The nfsB-encoded nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12. NfsB reduces a broad range of nitroaromatic compounds [Zenno96], including the antibiotics nitrofurazone and nitrofurantoin	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ybdF*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ybdJ*	Unknown	-	-	-
AN-4K-3	Deletion	Large	546,986 - 547,702	30 genes		*ybdK*	Pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase	-	-	-
AN-2K-4	Mobile element	Insertion	16,972	-	-	*hokC (-71)*	Small toxic membrane polypeptide	*nhaA (-514)*	pH dependent sodium or proton antiporter	IS150 mediated, intergenic
AN-2K-4	Mobile element	Insertion	621,252	-	-	*citG*	Triphosphoriboseyl-dephospho-CoA synthase	-	-	IS150 mediated
AN-2K-4	Mobile element	Insertion	872,829	-	-	*yeaS*	Leucine export protein LeuE	-	-	IS150 mediated
AN-2K-4	Mobile element	Insertion	910,345	-	-	*ynjI*	Hypothetical protein	-	-	IS150 mediated

AN-2K-4	SNP	Transition	1,238,659	C --> T	Gln (11) --> Stop	*ansP*	L-asparagine transporter	-	-	-
AN-2K-4	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, intergenic, online
AN-2K-4	SNP	Transition	1,439,030	A --> G	Tyr (75) --> Cys	*adhE*	Alcohol dehydrogenase	-	-	-
AN-2K-4	Mobile element	Insertion	2,387,885	-	-	*emrY*	Putative multidrug efflux system	-	-	IS3 mediated
AN-2K-4	Mobile element	Insertion	2,564,470	-	-	*hcaC*	3-phenylpropionate dioxygenase ferredoxin subunit	-	-	IS150 mediated
AN-2K-4	Mobile element	Insertion	2,992,382	-	-	*kpsS*	KpsS protein capsular biosynthesis	-	-	IS1 mediated
AN-2K-4	Mobile element	Insertion	3,096,895	-	-	*yqiI*	Hypothetical protein	-	-	IS1 mediated
AN-2K-4	Deletion	Small	4,295,377	-	-	*dcuR (-755)*	DNA binding transcriptional activator DcuR	*yjdI (-1,055)*	Hypothetical protein	interngenic
AN-2K-4	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated , online
AN-2K-4	Mobile element	Insertion	4,401,824	-	-	*ytfT*	Putative sugar transporter subunit membrane component of ABC superfamily	-	-	IS150 mediated
AN-2K-4	Mobile element	Insertion	4,581,54	-	-	*nadR*	Nicotinamide-nucleot	-	-	IS150 mediated

			5				ide adenylyltransferase			
AN-4K-4	Deletion	Small	161,094	-	-	*pcnB*	Poly(A) polymerase I	-	-	-
AN-4K-4	Deletion	Small	838,221	-	-	*proQ*	Putative solute DNA competence effector	-	-	-
AN-4K-4	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, intergenic online
AN-4K-4	Insertion	Duplication	1,328,413	×2		*dbpA*	ATP dependent RNA helicase DbpA	*zntB*	zinc transporter	Intergenic
AN-4K-4	SNP	Transition	1,439,673	G --> A	Ala (623) --> Thr	*adhE*	Alcohol dehydrogenase	-	-	-
AN-4K-4	Mobile element	Insertion	1,984,267	-	-	*cydA (-440)*	Cytochrome D terminal oxidase subunit I	*ybgG (+407)*	Alpha mannosidase	IS1 mediated, intergenic
AN-4K-4	Insertion	Duplication	1,995,122	×2		*sucA*	2-oxoglutarate dehydrogenase E1 component	-	-	-
AN-4K-4	Mobile element	Insertion	2,424,083	-	-	*alaW (-154)*	tRNA-Alanine	*yteC (-82)*	Putative DNA-binding transcriptional regulator	IS150 mediated, intergenic
AN-4K-4	Insertion	Duplication	2,972,604	×2	-	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-4	SNP	Transversion	2,972,858	T --> G	Met (158) --> Arg	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-4	SNP	Transversi		G --> A	Glu	*flu*	Antigen 43 (Ag43)	-	-	-

		on	2,974,237		(618) --> Thr		phase-variable biofilm formation autotransporter			
AN-4K-4	Mobile element	Insertion	3,250,708			*yrbI*	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	-	-	IS3 mediated
AN-4K-4	Deletion	Small	3,260,806	-	-	*arcB*	Aerobic respiration control sensor protein ArcB	-	-	-
AN-4K-4	Insertion	Small	3,298,183	+C	-	*rng*	Ribonuclease G	-	-	-
AN-4K-4	Deletion	Small	3,473,571	-C	-	*glgC*	Glucose-1-phosphate adenylyltransferase	-	-	-
AN-4K-4	Mobile element	Insertion	3,669,960	-	-	*rhsA*	rhsA element core protein RhsA	-	-	IS150 mediated
AN-4K-4	Insertion	Small	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup	*insJ-5 (-52)*	IS150 putative transposase	Intergenic
AN-4K-4	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine/D-serine/glycine permease	-	-	IS150 mediated, online
AN-4K-4	Insertion	Multi	4,406,085	+CCGTGGCAG	-	*mpl*	Murein recycling and cell wall remodelling	-	-	-
AN-4K-4	Mobile element	Insertion	4,581,546	-	-	*nadR*	nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*insB-6*	-	*ybdK*	-	IS1 mediated
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*[insB-6]*	After frameshift event,transcribed with InsA mediates	-	-	-

							IS1 transposition			
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*insA-6*	Transcriptional repressor of IS1	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00514*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00515*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00516*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00517*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ompT*	Outer membrane protease, with specificity for paired basic residues	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode cellular envelope proteins at low temperatures and during stationary phase	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ybcH*	Hypothetical prtein	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*nfrA*	Important in N4 absorption	-	-	-

AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00524*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*yhhI-2*	H repeat-associated protein, RhsE-linked, function unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00526*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00527*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00528*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00529*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ECB_00530*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*cusS*	Detoxification of copper and silver ions	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*cusR*	Detoxification of copper and silver ions	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*cusC*	Detoxification of copper and silver ions	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*cusB*	Detoxification of copper and silver ions	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*cusA*	Detoxification of copper and silver ions	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ylcC*	Phenylalanine transporter	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*pheP*	Low abundance mechanosensitive channel of miniconductance	-	-	-

AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ybdG*	Nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*nfnB*	aka nfsB The nfsB-encoded nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12. NfsB reduces a broad range of nitroaromatic compounds [Zenno96], including the antibiotics nitrofurazone and nitrofurantoin	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ybdF*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ybdJ*	Unknown	-	-	-
AN-4K-4	Deletion	Large	546,981 - 547,702	30 genes	-	*ybdK*	Pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase	-	-	-
AN-2K-5	Mobile element	Insertion	910,345	-	-	*ynjI*	Hypothetical protein	-	-	IS150 mediated
AN-2K-5	Mobile element	Insertion	955,693	-	-	*pheS*	Phenylalanyk -tRNA synthesis subunit alpha	-	-	IS150 mediated
AN-2K-5	SNP	Transition	1,438,030	A --> G	Tyr (75) --> Cys	*adhE*	Alcohol dehydrogenase	-	-	-

AN-2K-5	Deletion	Small	1,766,329	-	-	*pflB*	Pyruvate formate lyase I	-	-	-
AN-2K-5	SNP	Transition	2,001,307	G --> A	Val (134) --> Ile	*gltA*	Type II citrate synthase	-	-	-
AN-2K-5	Mobile element	Insertion	2,164,019	-	-	*mglB*	Methyl-galactoside transporter subunit	-	-	IS150 mediated
AN-2K-5	Insertion	Duplication	2,603,060	×2	-	*[rseB]*	Anti-sigma factor, negative regulator of sigma E	*[rseA]*	Anti-sigma factor, negative regulator of sigma E	-
AN-2K-5	Mobile element	Insertion	2,980,273	-	-	*ECB_02804*	Hypothetical protein	-	-	IS150 mediated
AN-2K-5	Mobile element	Insertion	2,992,382	-	-	*kpsS*	KpsS protein capsular biosynthesis	-	-	IS1 mediated
AN-2K-5	Mobile element	Insertion	3,495,511	-	-	*ugpB*	Glycerol-3-phosphate transporter periplasmic binding protein	-	-	IS150 mediated
AN-2K-5	Insertion	Small	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup	*insJ-5 (-52)*	IS150 putative transposase	Intergenic, online
AN-2K-5	Mobile element	Insertion	4,211,817	-	-	*alr (+151)*	Alanine racemase	*tyrB (-102)*	Aromatic amino acid aminotrasferase	IS150 mediated, intergenic
AN-2K-5	Deletion	Small	4,295,377	-	-	*dcuR (-755)*	DNA binding transcriptional activator DcuR	*yjdI (-1,055)*	Hypothetical protein	-
AN-2K-5	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated , online
AN-2K-5	Mobile element	Insertion	4,581,54	-	-	*nadR*	Nicotinamide-nucleotide	-	-	IS150 mediated

			5				adenylyltransferase			
AN-4K-5	Deletion	Small	161,094	-	-	*pcnB*	Poly(A) polymerase I	-	-	-
AN-4K-5	Deletion	Small	838,221	-	-	*proQ*	Putative solute DNA competence effector	-	-	-
AN-4K-5	Mobile element	Insertion	1,123,058			*ynfN*	Hypothetical protein			IS3 mediated
AN-4K-5	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, intergenic
AN-4K-5	Insertion	Duplication	1,328,413	×2	-	*dbpA*	ATP-dependent RNA helicase DbpA	*zntB*	Zinc transporter	Intergenic
AN-4K-5	SNP	Transition	1,439,673	G --> A	Ala (623) --> Thr	*adhE*	Alcohol dehydrogenase	-	-	-
AN-4K-5	Deletion	Small	2,057,185	-	-	*[glnU]*	Glutamine tNRA	*[glnW]*	Glutamine tRNA	Intergenic
AN-4K-5	SNP	Transversion	2,972,858	T --> G	Met (158) --> Arg	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-5	SNP	Transversion	2,974,237	G --> A	Glu (618) --> Thr	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-5	Mobile element	Insertion	3,250,708	-	-	*yrbI*	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	-	-	IS3 mediated
AN-4K-5	Deletion	Small	3,260,806	-	-	*arcB*	Aerobic respiration control sensor protein ArcB	-	-	-

AN-4K-5	Insertion	Small	3,298,183	+C	-	*rng*	Ribonuclease G	-	-	-
AN-4K-5	Deletion	Small	3,473,571	-C	-	*glgC*	Glucose-1-phosphate adenylyltransferase	-	-	-
AN-4K-5	Insertion	Small	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup	*insJ-5 (-52)*	IS150 putative transposase	Intergenic, online
AN-4K-5	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
AN-4K-5	Insertion	Small	4,406,085	+GTGGCAG	-	*mpl*	Murein recucling and cell wall remodelling	-	-	-
AN-4K-5	Mobile element	Insertion	4,581,546	-	-	*nadR*	nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*insB-6*	-	*[insA-17]*	-	IS1 mediated
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*[insB-6]*	After frameshift event,transcribed with InsA mediates IS1 transposition	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*insA-6*	Transcriptional repressor of IS1	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00514*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00515*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 –	33 genes	-	*ECB_00516*	Unknown	-	-	-

			547,702							
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00517*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ompT*	Outer membrane protease, with specificity for paired basic residues	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode cellular envelope proteins at low temperatures and during stationary phase	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ybcH*	Hypothetical prtein	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*nfrA*	Important in N4 absorption	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00524*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*yhhI-2*	H repeat-associated protein, RhsE-linked, function unknown	-	-	-

AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00526*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00527*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00528*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00529*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ECB_00530*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*cusS*	Detoxification of copper and silver ions	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*cusR*	Detoxification of copper and silver ions	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*cusC*	Detoxification of copper and silver ions	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ylcC*	Detoxification of copper and silver ions	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*cusB*	Detoxification of copper and silver ions	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*cusA*	Phenylalanine transporter	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*pheP*	Low abundance mechanosensitive channel of	-	-	-

							miniconductance			
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ybdG*	Nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*nfnB*	aka nfsB The nfsB-encoded nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12. NfsB reduces a broad range of nitroaromatic compounds [Zenno96], including the antibiotics nitrofurazone and nitrofurantoin	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ybdF*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ybdJ*	Unknown	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*ybdK*	Pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase	-	-	-
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*insJ-1*	InsJ component of InsJK transposase of IS150	-	-	-
AN-4K-5	Deletion	Large	546,990 –	33 genes	-	*insB-7*	InsB component of InsAB transposase of	-	-	-

			547,702				IS1			
AN-4K-5	Deletion	Large	546,990 – 547,702	33 genes	-	*[insA-7]*	pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase	-	-	-
AN-2K-6	Mobile element	Insertion	910,345	-	-	*ynjI*	Hypothetical protein	-	-	IS150 mediated
AN-2K-6	Mobile element	Insertion	2,164,019	-	-	*mglB*	Methyl-galactoside transporter subunit	-	-	-
AN-2K-6	Insertion	Duplication	2,603,060	×2	-	*[rseB]*	Anti-sigma factor, negative regulator of sigma E	*[rseA]*	Anti-sigma factor, negative regulator of sigma E	-
AN-2K-6	Mobile element	Insertion	2,980,273	-	-	*ECB_02804*	Hypothetical protein	-	-	IS150 mediated
AN-2K-6	Mobile element	Insertion	2,992,382	-	-	*kpsS*	KpsS protein capsular biosynthesis	-	-	IS1 mediated
AN-2K-6	Mobile element	Insertion	3,495,511	-	-	*ugpB*	Glycerol-3-phosphate transporter periplasmic binding protein	-	-	IS150 mediated
AN-2K-6	Insertion	Single base	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup/	*insJ-5 (-52)*	IS150 putative transposase	Intergenic, online
AN-2K-6	Mobile element	Insertion	4,211,817	-	-	*alr (+151)*	Alanine racemase	*tyrB (-102)*	Aromatic amino acid aminotrasferase	IS150 mediated, intergenic
AN-2K-6	Deletion	Small	4,295,377	-	-	*dcuR (-755)*	DNA binding transcriptional activator DcuR	*yjdI (-1,055)*	Hypothetical protein	-
AN-2K-6	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated , online

AN-2K-6	Mobile element	Insertion	4,581,545	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
AN-4K-6	Deletion	Small	161,094	-	-	*pcnB*	Poly(A) polymerase I	-	-	-
AN-4K-6	Insertion	Small	462,233	+ATCCAGC	-	*ybaM*	Hypothetical protein	-	-	-
AN-4K-6	Deletion	Small	838,221	-	-	*proQ*	Putative solute DNA competence effector	-	-	-
AN-4K-6	Insertion	Duplication	1,137,187	×2	-	*ydfH*	putative DNA-binding transcriptional regulator	*ydfG*	3-hydroxy acid dehydrogenase	Intergenic
AN-4K-6	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, intergenic
AN-4K-6	Insertion	Duplication	1,328,413	×2	-	*dbpA*	ATP-dependent RNA helicase DbpA	*zntB*	Zinc transporter	Intergenic
AN-4K-6	SNP	Transition	1,439,673	G --> A	Ala (623) --> Thr	*adhE*	Alcohol dehydrogenase	-	-	-
AN-4K-6	Mobile element	Insertion	1,984,267	-	-	*cydA (-440)*	Cytochrome D terminal oxidase subunit I	*ybgG (+407)*	Alpha mannosidase	IS1 mediated, intergenic
AN-4K-6	Insertion	Duplication	1,995,122	×2		*sucA*	2-oxoglutarate dehydrogenase E1 component	-	-	-
AN-4K-6	Mobile element	Insertion	2,424,083	-	-	*alaW (-154)*	tRNA-Alanine	*yteC (-82)*	Putative DNA-binding transcriptional regulator	IS150 mediated, intergenic
AN-4K-6	SNP	Transversion	2,972,858	T --> G	Met (158) --> Arg	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-

AN-4K-6	SNP	Transversion	2,974,237	G --> A	Glu (618) --> Thr	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-6	Deletion	Small	3,260,806	-	-	*arcB*	Aerobic respiration control sensor protein ArcB	-	-	-
AN-4K-6	Insertion	Small	3,298,183	+C	-	*rng*	Ribonuclease G	-	-	-
AN-4K-6	Deletion	Small	3,473,571	-C	-	*glgC*	Glucose-1-phosphate adenylyltransferase	-	-	-
AN-4K-6	Insertion	Small	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup	*insJ-5 (-52)*	IS150 putative transposase	Intergenic, online
AN-4K-6	Mobile element	insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
AN-4K-6	Mobile element	Insertion	1,123,058			*ynfN*	Hypothetical protein			IS3 mediated
AN-4K-6	Mobile element	Insertion	3,669,960	-	-	*rhsA*	rhsA element core protein RhsA	-	-	IS150 mediated
AN-4K-6	Insertion	Small	4,406,085	+CCGTGGCAG	-	*mpl*	Murein recucling and cell wall remodelling	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*[insB-6]*	-	*ybdK*	-	IS1 mediated
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*[insB-6]*	After frameshift event,transcribed with InsA mediates IS1 transposition	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*insA-6*	Transcriptional repressor of IS1	-	-	-

AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00514*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00515*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00516*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00517*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ompT*	Outer membrane protease, with specificity for paired basic residues	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode cellular envelope proteins at low temperatures and during stationary phase	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ybcH*	Hypothetical prtein	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*nfrA*	Important in N4 absorption	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00524*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*yhhI-2*	H repeat-associated protein, RhsE-linked,	-	-	-

							function unknown			
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00526*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00527*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00528*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00529*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ECB_00530*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*cusS*	Detoxification of copper and silver ions	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*cusR*	Detoxification of copper and silver ions	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*cusC*	Detoxification of copper and silver ions	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*cusB*	Detoxification of copper and silver ions	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*cusA*	Detoxification of copper and silver ions	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ylcC*	Phenylalanine transporter	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*pheP*	Low abundance mechanosensitive channel of miniconductance	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ybdG*	Nitroreductase is the minor oxygen-insensitive nitroreductase	-	-	-

							present in E. coli K-12			
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*nfnB*	aka nfsB The nfsB-encoded nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12. NfsB reduces a broad range of nitroaromatic compounds [Zenno96], including the antibiotics nitrofurazone and nitrofurantoin	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ybdF*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ybdJ*	Unknown	-	-	-
AN-4K-6	Deletion	Large	546979 - 547703	30 genes	-	*ybdK*	Pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase	-	-	-
AN-4K-6	Mobile element	Insertion	3,250,708	-	-	*yrbI*	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	-	-	IS3 mediated
AN-2K-7	Deletion	Small	161,094	-	-	*pcnB*	Poly(A) polymerase I	-	-	Online
AN-2K-7	Mobile element	Insertion	1,123,058	-	-	*ynfN*	Hypothetical protein	-	-	IS3 mediated, online
AN-2K-7	Insertion	Small	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup	*insJ-5 (-52)*	IS150 putative transposase	Intergenic, online

AN-2K-7	SNP	Transition	1,439,673	G --> A	Ala (623) --> Thr	*adhE*	Alcohol dehydrogenase	-	-	Online
AN-2K-7	Deletion	Large	2,057,167	2 genes	-	*[glnU]*	-	*[glnW]*	-	IS150 mediated
AN-2K-7	Deletion	Large	2,057,167	2 genes		*glnU*	Four glutamine tRNAs	-	-	-
AN-2K-7	Deletion	Large	2,057,167	2 genes		*glnW*	One of four glutmamine tRNAs	-	-	-
AN-2K-7	Mobile element	Insertion	2,424,083	-	-	*alaW (-154)*	tRNA-Ala	*yfeC (-82)*	Putative DNA-binding transcriptional activator	IS150 mediated
AN-2K-7	SNP	Transversion	2,972,858	T --> G	Met (158) --> Arg	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	Online
AN-2K-7	Deletion	Small	3,260,806	-	-	*arcB*	Aerobic respiration control sensor protein ArcB	-	-	Online
AN-2K-7	Insertion	Small	3,298,183	+C	-	*rng*	Ribonuclease G	-	-	Online
AN-2K-7	Deletion	Small	3,473,571	-C	-	*glgC*	Glucose-1-phosphate adenylyltransferase	-	-	Online
AN-2K-7	Insertion	Small	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup	*insJ-5 (-52)-*	IS150	IS150 mediated, intergenic, online
AN-2K-7	Mobile element	Insertion	4,381,58	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated,

			3							online
AN-2K-7	Insertion	Multi base	4,406,085	+GTGGCAG	-	*mpl*	Murein recycling and cell wall remodelling	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*insB-6*	-	*ybdK*	-	IS150 mediated, online
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*[insB-6]*	After frameshift event,transcribed with InsA mediates IS1 transposition	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*insA-6*	Transcriptional repressor of IS1	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00514*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00515*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00516*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00517*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ompT*	Outer membrane protease, with specificity for paired basic residues	-	-	-

AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode cellular envelope proteins at low temperatures and during stationary phase	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ybcH*	Hypothetical prtein	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*nfrA*	Important in N4 absorption	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00524*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*yhhI-2*	H repeat-associated protein, RhsE-linked, function unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00526*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00527*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00528*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ECB_00529*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986	30 genes	-	*ECB_00530*	Unknown	-	-	-

			– 547,702							
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*cusS*	Detoxification of copper and silver ions	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*cusR*	Detoxification of copper and silver ions	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*cusC*	Detoxification of copper and silver ions	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*cusB*	Detoxification of copper and silver ions	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*cusA*	Detoxification of copper and silver ions	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ylcC*	Phenylalanine transporter	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*pheP*	Low abundance mechanosensitive channel of miniconductance	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ybdG*	Nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*nfnB*	aka nfsB The nfsB-encoded nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-	-	-	-

							12. NfsB reduces a broad range of nitroaromatic compounds [Zenno96], including the antibiotics nitrofurazone and nitrofurantoin			
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ybdF*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ybdJ*	Unknown	-	-	-
AN-2K-7	Deletion	Large	546,986 – 547,702	30 genes	-	*ybdK*	Pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase	-	-	-
AN-4K-7	Deletion	Small	161,094	-	-	*pcnB*	Poly(A) polymerase I	-	-	Online
AN-4K-7	SNP	Transversion	542,005	G --> T	Ala (405) --> Ser	*ybcK*	Putative recombinase	-	-	
AN-4K-7	Mobile element	Insertion	1,123,058	-	-	*ynfN*	2-deoxy-D-gluconate 3-dehydrogenase	-	-	IS3 medaited, online
AN-4K-7	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl accepting chemotaxis protein III	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, online
AN-4K-7	Insertion	Duplication	1,328,413	×2	-	*dbpA*	ATP dependent RNA helicase DbpA	*zntB*	zinc transporter	Intergenic
AN-4K-7	SNP	Transition	1,439,673	G --> A	Ala (623) -->	*adhE*	Alcohol dehydrogenase	-	-	Online

					Thr					
AN-4K-7	Insertion	Duplication	1,995,122	×2	-	*sucA*	2-oxoglutarate dehydrogenase E1 component	-	-	-
AN-4K-7	Mobile element	Deletion	2,057,184	-	-	*[glnU]*	Glutamine tNRA	*[glnW]*	Glutamine tRNA	Intergenic
AN-4K-7	SNP	Transversion	2,972,858	T --> G	Met (158) --> Arg	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	Online
AN-4K-7	SNP	Transversion	2,974,237	G --> A	Glu (618) --> Thr	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	-
AN-4K-7	Mobile element	Insertion	3,250,708	-	-	*yrbI*	3-deoxy-D-manno-octulosonate 8-phosphate phosphatase	-	-	IS3 mediated
AN-4K-7	Deletion	Small	3,260,806	-	-	*arcB*	Aerobic respiration control sensor protein ArcB	-	-	Online
AN-4K-7	Insertion	Small	3,298,183	+C	-	*rng*	Ribonuclease G	-	-	Online
AN-4K-7	Deletion	Small	3,473,571	-C	-	*glgC*	Glucose-1-phosphate adenylyltransferase	-	-	Online
AN-4K-7	Insertion	Small	3,866,357	+G	-	*trkD (+5)*	Potassium transport protein Kup	*insJ-5 (-52)*	IS150 putative transposase	Intergenic, online
AN-4K-7	Mobile element	insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
AN-4K-7	Deletion	Large	546,982 -	30 genes	-	*[insB-6]*	-	*ybdK*	-	IS1 mediated, online

			547,703							
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*[insB-6]*	After frameshift event,transcribed with InsA mediates IS1 transposition	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*insA-6*	Transcriptional repressor of IS1	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00514*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00515*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00516*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00517*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ompT*	Outer membrane protease, with specificity for paired basic residues	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*envY*	DNA-binding transcriptional regulator that participates in the control of several genes that encode cellular envelope	-	-	-

							proteins at low temperatures and during stationary phase			
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ybcH*	Hypothetical prtein	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*nfrA*	Important in N4 absorption	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00524*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*yhhI-2*	H repeat-associated protein, RhsE-linked, function unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00526*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00527*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00528*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00529*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ECB_00530*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*cusS*	Detoxification of copper and silver ions	-	-	-
AN-4K-7	Deletion	Large	546,982 to	30 genes	-	*cusR*	Detoxification of copper and silver	-	-	-

			547,703				ions			
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*cusC*	Detoxification of copper and silver ions	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*cusB*	Detoxification of copper and silver ions	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*cusA*	Detoxification of copper and silver ions	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ylcC*	Phenylalanine transporter	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*pheP*	Low abundance mechanosensitive channel of miniconductance	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ybdG*	Nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*nfnB*	aka nfsB The nfsB-encoded nitroreductase is the minor oxygen-insensitive nitroreductase present in E. coli K-12. NfsB reduces a broad range of nitroaromatic compounds [Zenno96], including the antibiotics nitrofurazone and	-	-	-

							nitrofurantoin			
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ybdF*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ybdJ*	Unknown	-	-	-
AN-4K-7	Deletion	Large	546,982 to 547,703	30 genes	-	*ybdK*	Pyruvate:flavodoxin oxidoreductase and/or pyruvate synthase	-	-	-

[a] The lineage in which the mutation arose.

[b] The type of mutation that had occurred.

[c] Further sub-classification of the mutation type that occurred. Note for large deletions, the first and last gene affected is listed as the first event, followed by each in between listed underneath.

[d] The position on the reference genome at which the mutation occurred. In some cases, particularly with large deletions, the exact location of the beginning of the mutation could not be specified from breseq output and instead the likely range is given.

[e] The change that occurred at the DNA level as a result of the mutation. In the case of large deletions, the number of genes deleted is indicated.

[f] The amino acid change that had occurred at the protein level as a result of the mutation. Also indicated was the position along the protein in which the change occurred.

[g] The gene mutated. In the cases of genes partially affected, the gene name was enclosed with parenthesis []. Intergenic mutations are identified by the closest gene to the mutation with the number of bps from the beginning of the gene (+ for downstream and – for upstream).

[h] A brief functional description of the protein encoded by Reference Gene 1.

[i] Intergenic mutations are identified by the closest gene to the mutation with the number of bps from the beginning of the gene

[j] A brief functional description of the protein encoded by Reference Gene 2.

[k] Additional information if required.

Figure 7.5: List of all mutation in fluctuating genomes

Lineage [a]	Type [b]	Class [c]	Reference Position 1 [d]	Change [e]	Amino Acid [f]	Reference Gene 1 [g]	Description Gene 1 [h]	Reference Gene 2 [i]	Description Gene 2 [j]	Comment [k]
FL-2K-1	SNP	Transition	20,661	T --> C	Ile (372) --> Thr	*ECB_00021*	Putative usher protein	-	-	-
FL-2K-1	Mobile element	Insertion	959,668	-	-	*btuC*	Vitamin B12-transporter permease	-	-	IS150 mediated
FL-2K-1	Mobile element	Insertion	1,288,903	-	-	*ydbD*	Hypothetical protein	-	-	IS1 mediated
FL-2K-1	Mobile element	Insertion	1,341,071	-	-	*ynaI*	Inner membrane protein	-	-	IS1 mediated
FL-2K-1	Mobile element	Deletion	1,764,886	-	-	*pflB*	Pyruvate formate lyase 1	-	-	IS150 mediated
FL-2K-1	SNP	Transition	2,229,796	A --> G	-	*mqo (-116)*	Malate:quinone oxidoreductase	*yojI (+102)*	Multidrug transporter membrane component	Intergenic
FL-2K-1	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-2K-1	Deletion	Large	4,533,440	2 genes	-	*[yjiY]*	Putative protein	*[hpaC]*	4-hydroxyphenylacetate 3-monooxygenase reductase component	Deletion is mediated by two IS150 insertions at 4535654 and 4533439 mediated
FL-2K-1	Deletion	- Large	4,533,440	2 genes	-	*[yjiY]*	Putative protein			
FL-2K-1	Deletion	- Large	4,533,440	2 genes	-	*[hpaC]*	4-hydroxyphenylacet			

							ate 3-monooxygenase reductase component			
FL-4K-1	Mobile element	Insertion	388,275	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	360,203	-	-	*yaiT*	Hypothetical protein	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	388,275	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	785,037	-	-	*cheB*	Chemotaxis-specific methylesterase	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	963,731	-	-	*ydiU*	Hypothetical protein	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	974,185	-	-	*ydiQ*	Putative electron transfer flavoprotein YdiQ	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	1,181,682	-	-	*yddA (+15)*	Multidrug ABC transporter membrane ATP binding protein	*yddB (-11)*	Putative porin protein	IS4 mediated, intergenic
FL-4K-1	Mobile element	Insertion	1,272,468	-	-	*trg (-326)*	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB (-13)*	Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-1	Insertion	Base	1,328,521	+A	-	*zntB*	Trans-membrane zinc transporter	-	-	-
FL-4K-1	Mobile element	Insertion	1,464,595	-	-	*IdrC (+260)*	LTR Toxin-antitoxin system	*IdrB (-167)*	LTR Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-1	SNP	Transversion	1,544,946	G --> T	-	*ycfU (-247)*	Outer membrane-specific lipprotein transporter subunit	*ycfT (-15)*	Putative inner membrane protein	Intergenic

							LolC			
FL-4K-1	Mobile element	Insertion	1,635,657	-	-	*ycdO*	Hypothetical protein	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	1,670,113	-	-	*ymcA*	Hypothetical protein	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	1,675,380	-	-	*yccC*	Cryptic autophosphorylating protein tyrosine kinase Etk	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	1,886,677	-	-	*ybiT*	ABC transporter protein	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	2,024,210	-	-	*rhsC*	rhsC element core protein	-	-	IS150 mediated
FL-4K-1	SNP	Transition	2,554,932	G --> A	His (41) --> Tyr	*iscC*	Cysteine desulfurase	-	-	-
FL-4K-1	Deletion	Small	2,748,700	-	-	*insJ-3 (-48)*	IS150 hypothetical protein	*cysH (-201)*	Phosphoadenosine phosphosulfate reductase	Intergenic
FL-4K-1	Mobile element	Insertion	2,872,437	-	-	*yqeB*	Hypothetical protein	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	3,148,353	-	-	*uxaA*	Altronate hydrolase	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	3,156,802	-	-	*yqiF*	Putative quinol oxidase subunit	-	-	IS150 mediated
FL-4K-1	Deletion	Small	3,164,090	-	-	*yhaO (-288)*	Putative transporter	*tdcG (+51)*	L-serine dehydratase 3	Intergenic
FL-4K-1	Mobile element	Insertion	3,205,065	-	-	*yraP (+25)*	Hypothetical protein	*yraQ (+89)*	Putative permease	IS150 mediated,intergenic
FL-4K-1	Mobile element	Insertion	3,385,413	-	-	*prkB*	Putative phosphoribulokinase	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	3,577,565	-	-	*yhjE*	Putative transporter	-	-	IS150 mediated

FL-4K-1	Mobile element	Insertion	3,669,960	-	-	*rhsA*	rhsA element core protein RshA	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	4,015,454	-	-	*yihS*	Putative glucosamine isomerase	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	4,040,290	-	-	*rhaA*	L-rhamnose isomerase	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-4K-1	Mobile element	Insertion	4,552,611	-	-	*mdoB*	Encodes two phosphoglycerol transferase I	-	-	IS150 mediated
FL-4K-1	Mobile element	Insertion	4,581,547	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes	-	*[insA-28]*	-	*[mdoB]*	Genes involved in restriction endonuclease activity, chemotaxis signalling and hydroxyphenylaceti c acid catabolism	IS1 mediated
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes	-	*[insB-28]*	IS1 recombinase	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*mcrC*	Modified cytosine restriction	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*mcrB*	Modified cytosine restriction	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*yjiW*	aka symE SOS response	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hsdS*	Component of restriction modification system	-	-	-
FL-4K-1	Deletion	Large	4516617 -	27		*ECB_00064*	Unknown	-	-	-

			4517320	genes						
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hsdM*	Unknown	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hsdR*	Unknown	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*mrr*	RecB-dependent high pressure induction of the SOS stress response	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*yjiA*	DNA damage response	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*yjiX*	Unknown	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*yjiY*	Unknown	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaC*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaB*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaA*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaX*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaI*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaH*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaF*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaD*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaE*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaG*	Bacterial pathogenicity	-	-	-

FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*hpaR*	Bacterial pathogenicity	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*tsr*	Chemotactic response to changes in pH	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*yjiZ*	Proton-driven L-galactonate uptake transporter	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*yjjN*	L-galactonate oxidoreductase that is required for growth on L-galactonate as the sole carbon source under high-throughput growth conditions with limited aeration	-	-	-
FL-4K-1	Deletion	Large	4516617 - 4517320	27 genes		*[mdoB]*	Phosphoglycerol transferases	-	-	-
FL-2K-2	Mobile element	Insertion	388,543	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
FL-2K-2	Mobile element	Deletion	429,505	-	-	*insL-2*	Transposase of IS186	-	-	IS186 mediated
FL-2K-2	Mobile element	Insertion	462,604	-	-	*priC*	Primosomal replicon protein N for restarting stalled replication forks	-	-	IS150 mediated
FL-2K-2	Mobile element	Insertion	654,734	-	-	*ogrK (+21)*	DNA binding transcriptional regulator	*yegQ (+251)*	Putative peptidase	IS150 mediated, intergenic
FL-2K-2	Deletion	Large	1,110,292	5 genes	-	*[ydfX]*	-	*ECB_01533*	-	IS150 mediated
FL-2K-2	Deletion	Large	1,110,292	5 genes		*[ydfX]*	pseudogene Qin	-	-	-

							prophage			
FL-2K-2	Deletion	Large	1,110,292	5 genes		*ECB_01536*	unknown	-	-	-
FL-2K-2	Deletion	Large	1,110,292	5 genes		*ECB_01535*	unknown	-	-	-
FL-2K-2	Deletion	Large	1,110,292	5 genes		*ECB_01534*	unknown	-	-	-
FL-2K-2	Deletion	Large	1,110,292	5 genes		*ECB_01533*	unknown	-	-	-
FL-2K-2	Mobile element	Insertion	1,272,468	-	-	*trg* (-326)	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB* (-13)	Toxin-antitoxin system	IS150 mediated, intergenic
FL-2K-2	SNP	Transversion	1,290,872	A --> G	Lys (285) -> Glu	*insF-2*	IS3 element protein InsF	-	-	-
FL-2K-2	Mobile element	Insertion	1,464,678	-	-	*ldrC* (-343)	*ldrC* is one of a set of four Long Direct Repeats (LDRs), one of which has been shown to code for the toxin portion of a toxin-antitoxin pair.	*ldrB* (-85)	*ldrB* is one of a set of four Long Direct Repeats (LDRs), one of which has been shown to code for the toxin portion of a toxin-antitoxin pair.	IS150 mediated, intergenic
FL-2K-2	Mobile element	Insertion	1,580,827	-		*flgK*	Flagellar hook associated protein	-	-	IS150 mediated
FL-2K-2	Mobile element	Deletion	1,764,888	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated, online
FL-2K-2	Mobile element	Insertion	3,386,643	-	-	*yhfA* (-114)	Hypothetical protein	*crp* (-186)	cAMP regulatory protein	IS150 mediated, intergenic
FL-2K-2	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-2K-2	Mobile element	Insertion	4,532,958	-	-	*yjiX* (-31)	Hypothetical protein	*yjiY* (+19)	Putative inner membrane protein	IS150 mediated, intergenic

FL-4K-2	Mobile element	Insertion	16,972	-	-	*hokC* (-71)	Small toxic membrane polypeptide	*nhaA* (-514)	pH dependent sodium or proton antiporter	IS150 mediated, intergenic
FL-4K-2	Mobile element	Insertion	388,527	-	-	*brnQ*	Putative branched chain amino acid transporter	-	-	IS150 mediated
FL-4K-2	Mobile element	Insertion	471,785	-	-	*aes*	Acetyl esterase	-	-	IS150 mediated
FL-4K-2	Mobile element	Insertion	649,973	-	-	*ECB_01994* (+102)	Hypothetical protein	*ECB_01993* (+102)	Hypothetical protein	IS150 mediated, intergenic
FL-4K-2	Mobile element	Insertion	743,346	-	-	*yeeJ*	Adhesin	-	-	IS150 mediated
FL-4K-2	Mobile element	Insertion	974,185	-	-	*ydiQ*	Putative electron transfer flavoprotein YdiQ	-	-	IS150 mediated
FL-4K-2	Mobile element	Insertion	1,113,403	-	-	*ECB_01533* (+103)	Hypothetical protein	*hokD* (-107)	Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-2	SNP	Transversion	1,116,591	G --> T	Gln (285) -> Lys	*insF-4*	Putative transposase for IS3	-	-	
FL-4K-2	SNP	Transition	1,206,872	C --> T	-	*sfcA (+61)*	Malate dehydrogenase	*adhP (-73)*	Alcohol dehydrogenase	Intergenic
FL-4K-2	Deletion	Small	1,272,401	-	-	*trg*	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB*	Toxin-antitoxin system	Deletion mediated by 2 IS150 insertions at 1272400 and 1272468, resulting in 67 bp deletion, intergenic
FL-4K-2	Mobile	Insertion	1,272,400	-	-	*trg* (-258)	Methyl-accepting	*mokB* (-83)	Toxin-antitoxin	IS150

	element						chemotaxis protein III, ribose and galactose sensor receptor		system	mediated, intergenic
FL-4K-2	Mobile element	Insertion	1,464,061	-	-	*chaA* (+238)	calcium/sodium:proton antiporter	*ldrC* (-167)	*ldrC* is one of a set of four Long Direct Repeats (LDRs), one of which has been shown to code for the toxin portion of a toxin-antitoxin pair.	IS150 mediated, intergenic
FL-4K-2	Deletion	Large	1,464,062	2 genes	-	*ldrC*	LTR Toxin-antitoxin system	*ldrB*	LTR Toxin-antitoxin system	
FL-4K-2	Deletion	Large	1,464,062	2 genes		*ldrC*	LTR Toxin-antitoxin system	-	-	
FL-4K-2	Deletion	Large	1,464,062	2 genes		*ldrB*	LTR Toxin-antitoxin system	-	-	
FL-4K-2	Mobile element	Insertion	1,579,629	-	-	*flgL*	Flagellar hook-associated protein FlgL	-	-	IS150 mediated
FL-4K-2	Mobile element	Insertion	1,598,688	-	-	*dinI*	DNA damage inducible protein I	-	-	IS150 mediated
FL-4K-2	Mobile element	Deletion	1,764,888	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated, online
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00851*	Hypothetical protein	*ECB_00814*	Hypothetical protein	No IS elements, but posible homologous sequences at *ybjL* and *ybjK* flanking the deletion event

FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00851*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00850*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00849*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00848*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*tfaE*	Tail fiber assembly	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00846*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00845*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00844*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00843*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00842*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00841*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00840*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00839*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00838*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00837*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00836*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00835*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00834*	Unknown	-	-	-

FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00833*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00832*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00831*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00830*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00829*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00828*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00827*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*yegZ*	gpD phage P2-like protein D; C-terminal fragment	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ogrK*	homologous of the phage-encoded Ogr protein that controls the bacteriophage P2 late transcription	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00824*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00823*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00822*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00821*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00820*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00819*	Unknown	-	-	-

FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00818*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00817*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00816*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00815*	Unknown	-	-	-
FL-4K-2	Deletion	Large	1,831,485	38 genes	-	*ECB_00814*	Part of the lambdoid prophage	-	-	-
FL-4K-2	SNP	Transversion	2,364,735	G --> T	Asn (40) --> Lys	*yfcB*	N5-glutamine S-adenosyl-L-methionine-dependent methyltransferase	-	-	-
FL-4K-2	Mobile element	Insertion	2,378,861	-	-	*fadL*	Long chain fatty acid outer membrane transporter	-	-	IS150 mediated
FL-4K-2	SNP	Transition	3,163,949	G --> A	-	*yhaO* (-147)	Putative transporter	*tdcG* (+192)	L-serine dehydratase 3	Intergenic
FL-4K-2	Mobile element	Insertion	3,482,646	-	-	*gntR*	DNA-binding transcriptional repressor	-	-	IS150 mediated
FL-4K-2	Mobile element	Insertion	3,603,153	-	-	*ldrE* (+337)	LTR Toxin-antitoxin system	*ldrF* (-91)	LTR Toxin-antitoxin system	IS150 mediated
FL-4K-2	Mobile element	Insertion	3,625,440	-	-	*insK-4* (+127)	IS150 putative transposase	*glyS* (-152)	Glycyl-tRNA synthetase subunit beta	IS150 mediated
FL-4K-2	Deletion	Small	4,140,414	-	-	*thiC*	Thiamine biosynthesis protein ThiC	-	-	
FL-4K-2	Deletion	Small	4,295,377	-	-	*dcuR* (-755)	DNA binding transcriptional activator DcuR	*yjdI* (-1055)	Hypothetical protein	Intergenic
FL-4K-2	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated

FL-4K-2	Mobile element	Insertion	4,581,547	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
FL-2K-3	Insertion	Bases	388,020	+ATCA	-	*brnQ*	Branched chain amino acid transporter	-	-	-
FL-2K-3	Insertion	Base	422,593	-	-	*cyoA* (-280)	Cytochrome o ubiquinol oxidase subunit II	*ampG* (+182)	Muropeptide transporter	Intergenic
FL-2K-3	Mobile element	Insertion	960,637	-	-	*btuE*	Putative glutathione peroxidase	-	-	IS150 mediated
FL-2K-3	Deletion	Small	1,328,493	-	-	*[zntB]*	Trans-membrane Zinc transporter (see reference)	-	-	-
FL-2K-3	SNP	Transversion	1,329,917	A --> T	Ser (54) --> Ser	*ydaM*	Putative diguanylate cyclase	-	-	-
FL-2K-3	Mobile element	Deletion	1,764,886	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-2K-3	Mobile element	Insertion	3,000,519	-	-	*kpsM*	Polysialic acid transport protein KpsM	-	-	-
FL-2K-3	SNP	Transition	3,829,404	G --> A	Lys (125) --> Lys	*yieG*	Putative inner membrane protein	-	-	-
FL-2K-3	Mobile element	Insertion	4,299,101	-	-	*lysU*	Lysl-tRNA synthetase	-	-	IS150 mediated
FL-2K-3	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-2K-3	Mobile element	Insertion	4,534,750	-	-	*yjiY*	Putative inner membrane protein	-	-	IS150 mediated
FL-2K-3	Deletion	Small	4,240,695 - 4,240,865	-	-	*gltP*	Glutamate and aspartate proton symporter	*yjcO*	Hypothetical protein	Intergenic

FL-4K-3	Mobile element	Insertion	16,989	-	-	*hokC* (-88)	Small toxic membrane polypeptide	*nhaA (*-496)	pH dependent sodium proton antiporter	IS150 mediated, intergenic
FL-4K-3	Mobile element	Insertion	388,527	-	-	*brnQ*	Putative branched chain amino acid transporter	-	-	IS150 mediated
FL-4K-3	Mobile element	Insertion	471,785	-	-	*aes*	Acetyl esterase	-	-	IS150 mediated
FL-4K-3	Mobile element	Insertion	743,346	-	-	*yeeJ*	Adhesin	-	-	IS150 mediated
FL-4K-3	Mobile element	Insertion	974,185	-	-	*ydiQ*	Putative electron transfer flavoprotein YdiQ	-	-	IS150 mediated
FL-4K-3	Mobile element	Insertion	1,113,403	-	-	*ECB_01533* (+103)	Hypothetical protein	*hokD* (-107)	Small toxic polypeptide	IS150 mediated, intergenic
FL-4K-3	Deletion	Small	2,071,479	-	-	*insJ-2* (-47)	IS150 hypothetical protein	*rihA* (-490)	Ribonucleoside hydrolase 1	Intergenic
FL-4K-3	Mobile element	Insertion	1,272,400	-	-	*trg* (-258)	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB* (-83)	Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-3	Deletion	Small	1,272,401	-	-	*trg*	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB*	Toxin-antitoxin system	Deletion mediated by the insertion of IS150 at 1272400 and 1272468 and resulting in 67 bp deletion, intergenic

FL-4K-3	Mobile element	Insertion	1,464,061	-	-	*chaA* (+238)	calcium/sodium:proton antiporter	*ldrC* (-167)	LTR Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-3	Mobile element	Insertion	1,579,629	-	-	*flgL*	Flagellar hook-associated protein FlgL	-	-	IS150 mediated
FL-4K-3	Deletion	Large	1,702,902	3 genes	-	*[pqiB]*	-	*[uup]*	Part of the lambdoid prophage and replication fork progression	IS150 mediated, inserted into upp
FL-4K-3	Deletion	Large	1,702,902	3 genes		*pqiB*	Paraquat-inducible, SoxRS-regulated MCE domain protein; RpoS regulon; predicted N-terminal membrane anchor; C-terminal			
FL-4K-3	Deletion	Large	1,702,902	3 genes		*pqiA*	Paraquat-inducible, SoxRS-regulated inner membrane protein; RpoS regulon; function unknown			
FL-4K-3	Deletion	Large	1,702,902	3 genes		*upp*	memebers of the soxRS regulons and replication fork progression, DNA-binding ATPase involved in replication; cytosolic; ABC-F family protein			
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00851*		*ECB_00814*	-	No IS elements, but posible

										homologous sequences at *ybjL* and *ybjK* flanking the deletion event
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00851*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00850*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00849*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00848*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*tfaE*	Tail fiber assembly	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00846*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00845*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00844*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00843*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00842*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00841*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00840*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00839*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00838*	Unknown	-	-	-

FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00837*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00836*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00835*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00834*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00833*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00832*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00831*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00830*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00829*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00828*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00827*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*yegZ*	gpD phage P2-like protein D; C-terminal fragment	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ogrK*	homologous of the phage-encoded Ogr protein that controls the bacteriophage P2 late transcription	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00824*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00823*	Unknown	-	-	-

FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00822*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00821*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00820*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00819*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00818*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00817*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00816*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00815*	Unknown	-	-	-
FL-4K-3	Deletion	Large	1,831,485	38 genes	-	*ECB_00814*	Part of the lambdoid prophage	-	-	-
FL-4K-3	Deletion	Large	2,071,479	-	-	*insJ-2* (-47)	IS150 hypothetical protein	*rihA* (-490)	Ribonucleoside hydrolase 1	-
FL-4K-3	Mobile element	Insertion	2,073,388	-	-	*hscC*	Hsp70 family chaperone Hsc62, inhibits transcription	-	-	IS150 mediated
FL-4K-3	Mobile element	Insertion	2,615,526	-	-	*yfiQ*	Acyl-CoA synthetase NAD(P)-binding subunit	-	-	IS150 mediated
FL-4K-3	SNP	Transition	3,163,949	G --> A	-	*yhaO* (-147)	Putative transporter	*tdcG* (+192)	L-serine dehydratase 3	Intergenic
FL-4K-3	Mobile element	Insertion	3,602,958	-	-	*ldrE* (-90)	LTR Toxin-antitoxin system	*ldrF* (+286)	LTR Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-3	Deletion	Small	4,295,377	-	-	*dcuR* (-755)	DNA binding transcriptional activator DcuR	*yjdI* (-1055)	Hypothetical protein	Intergenic

FL-4K-3	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-4K-3	Mobile element	Insertion	4,581,547	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
FL-4K-3	Deletion	Small	1,764,944 - 1,766,330	1 gene	-	*pflB*	Pyruvate formate lyase I	-	-	-
FL-4K-3	Deletion	Large	3,981,508 - 3,983,489	3 genes	-	*[alaT]*	-	*rrfA*	tRNA and 5 and 23S rRNA genes	-
FL-4K-3	Deletion	Large	3,981,508 - 3,983,489	3 genes		*[alaT]*	Alanine tRNAs	-	-	-
FL-4K-3	Deletion	Large	3,981,508 - 3,983,489	3 genes		*rrlA*	One of five alanine tRNAs	-	-	-
FL-4K-3	Deletion	Large	3,981,508 - 3,983,489	3 genes		*rrfA*	One of seven rRNAs	-	-	-
FL-4K-3	Mobile element	Insertion	1,766,334	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-2K-4	Mobile element	Insertion	16,972	-	-	*hokC* (-71)	Small toxic membrane polypeptide	*nhaA* (-514)	pH dependent sodium or proton antiporter	IS150 mediated, intergenic
FL-2K-4	Mobile element	Insertion	388,275	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated
FL-2K-4	Deletion	Large	963,078	2 genes	-	*[ydiV]*	Hypothetical proteins	*[ydiU]*	-	IS150 mediated
FL-2K-4	Deletion	Large	963,078	2 genes		*[ydiV]*	Involved in motility and quorom sensing			
FL-2K-4	Deletion	Large	963,078	2 genes		*[ydiU]*	control of motility, regulate expression			

							of flagella and motility in response to nutrient availability			
FL-2K-4	Mobile element	Insertion	1,181,538	-	-	*yddA*	Multidrug ABC transporter membrane ATP binding protein	-	-	IS150 mediated
FL-2K-4	Mobile element	Insertion	1,551,960	-	-	*ycfQ* (+33)	putative DNA-binding transcriptional regulator	*ycfJ* (+29)	Hypothetical protein	IS150 mediated, intergenic
FL-2K-4	Mobile element	Insertion	1,762,790	-	-	*ycaO* (+33)	Hypothetical protein	*focA* (-370)	Formate transporter	Intergenic
FL-2K-4	Mobile element	Deletion	1,764,888	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-2K-4	Mobile element	Insertion	2,138,008	-	-	*yehU*	Putative sensory kinase in two-component system with YehT	-	-	IS150 mediated
FL-2K-4	SNP	Transition	3,041,966	A --> G	Tyr (439) -> His	*pitB*	Phosphate transporter	-	-	-
FL-2K-4	SNP	Transversion	3,153,950	A --> T	-	*yhaO* (-148)	Putative transporter	*tdcG* (+191)	L-serine dehydratase 3	Intergenic
FL-2K-4	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-4K-4	Mobile element	Insertion	388,527	-	-	*brnQ*	Putative branched chain amino acid transporter	-	-	IS150 mediated
FL-4K-4	Deletion	Large	632,692	27 genes	-	*[ECB_02013]*	Hypothetical protein	*ogrK*	phage-encoded Ogr protein that controls the bacteriophage P2 late transcription	No IS elements, but brings ybjL and ybjK

										together
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02013*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02012*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02011*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02010*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02009*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02008*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02007*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02006*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02005*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02004*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02003*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02002*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02001*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_02000*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01999*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01998*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01997*	Unknown	-	-	-

FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01996*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01995*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01994*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01993*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01992*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01991*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01990*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ECB_01989*	Unknown	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*yegZ*	gpD phage P2-like protein D; C-terminal fragment	-	-	-
FL-4K-4	Deletion	Large	632,692	27 genes		*ogrK*	Homologous of the phage-encoded Ogr protein that controls the bacteriophage P2 late transcription			
-4K-4	Mobile element	Insertion	851,286	-	-	*manX (-84)*	PTS system mannose-specific transporter	*yoaE (-371)*	Hypothetical protein	IS1 mediated
FL-4K-4	Mobile element	Insertion	932,998	-	-	*ydiV*	Hypothetical protein	-	-	IS1 mediated
FL-4K-4	Mobile element	Insertion	994,990	-	-	*sufS*	Bifunctional cysteine desulfurase/seleno cysteine lyase	-	-	IS1 mediated
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*[tqsA]*	Controls transport of autoinducer 2	*ECB_01533*	Unknown-	IS150 mediated

FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*[tqsA]*	Controls transport of autoinducer 2	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ydgF*	aka mdtJ spermidine efflux transporter, multidrug resistance	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ydgE*	aka mdtJ spermidine efflux transporter, multidrug resistance	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ydgD*	aka mdtJ spermidine efflux transporter, multidrug resistance	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*asr*	Survival under acidic conditions	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfM*	Member of the major facilitator superfamily (MFS) of transporters	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfL*	Member of the major facilitator superfamily (MFS) of transporters	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*mlc*	Regulates genes involved in the uptake of glucose	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfK*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-

FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfJ*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfI*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfH*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfG*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfF*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfE*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-

FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfD*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfC*	The catalytic subunit of the dimethyl sulfoxide reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*speG*	Sperimidine acetyltransferase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfB*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ynfA*	ynf operon contains genes that are paralogues to those in the dms operon for DMSO reductase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*rspA*	Regulatory in stationary phase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*rspB*	Regulatory in stationary phase	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01546*	Unknown	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ydfE*	Pseudogene Qin prophage	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ydfD*	Pseudogene Qin prophage	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*dicB*	Regulation of cell division	-	-	-

FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01542*	Unknown	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01541*	Unknown	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01540*	Unknown	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01539*	Unknown	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*dicC*	Negatively controls the expression of the gene dicB which encodes the protein involved in cell division inhibition	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ydfX*	Pseudogene Qin prophage	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01536*	Unknown	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01535*	Unknown	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01534*	Unknown	-	-	-
FL-4K-4	Deletion	Large	1,081,310	36 genes	-	*ECB_01533*	Unknown	-	-	-
FL-4K-4	Mobile element	Insertion	1,113,403	-	-	*ECB_01533* (+103)	Hypothetical protein	*hokD* (-107)	Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-4	SNP	Transversion	1,116,591	G --> T	Gln (285) --> Lys	*insF-4*	Putative transposase for IS3	-	-	-
FL-4K-4	Mobile element	Insertion	1,272,468	-	-	*trg* (-326)	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB* (-13)	Regulatory peptide	IS150 mediated, intergenic

FL-4K-4	Mobile element	Deletion	1,292,085	-	-	*[tynA]*	Amine oxidase	-	-	IS3 mediated
FL-4K-4	Mobile element	Insertion	1,464,061	-	-	*chaA* (+238)	calcium/sodium:proton antiporter	*ldrC* (-167)	*ldrC* is one of a set of four Long Direct Repeats (LDRs), one of which has been shown to code for the toxin portion of a toxin-antitoxin pair.	IS150 mediated, intergenic
FL-4K-4	Mobile element	Insertion	1,579,629	-	-	*flgL*	Flagellar hook-associated protein FlgL	-	-	IS150 mediated
FL-4K-4	Mobile element	Insertion	1,766,334	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-4K-4	Mobile element	Insertion	2,343,551	-	-	*argT*	Lysine/serine/ornithine transporter subunit	-	-	IS150 mediated
FL-4K-4	Mobile element	Insertion	2,850,135	-	-	*kduD*	2-deoxy-D-gluconate 3-dehydrogenase	-	-	IS150 mediated
FL-4K-4	Mobile element	Insertion	2,965,217	-	-	*nupG*	Nucleoside transporter	-	-	IS150 mediated
FL-4K-4	Deletion	Small	3,027,366	-	-	*glcD*	Glycolate oxidase subunit GlcD	-	-	-
FL-4K-4	SNP	Transition	3,163,949	G --> A	-	*yhaO* (-147)	Putative transporter	*tdcG* (+192)	L-serine dehydratase 3	Intergenic
FL-4K-4	Deletion	Small	4,295,377	-	-	*dcuR* (-755)	DNA binding transcriptional activator DcuR	*yjdI* (-1055)	Hypothetical protein	Intergenic
FL-4K-4	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-4K-4	Mobile element	Insertion	4,581,547	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated

FL-2K-5	Mobile element	Insertion	959,872	-	-	*btuC*	Vitamin B12-transporter permease	-	-	IS150 mediated
FL-2K-5	SNP	Transversion	1,328,511	C --> A	Ala (57) --> Glu	*trpC*	Indole-3- glycerol phosphate synthase	-	-	-
FL-2K-5	Insertion	Base	1,328,521	+A	-	*zntB*	Trans-membrane zinc transporter (see reference)	-	-	-
FL-2K-5	Mobile element	Deletion	1,764,888	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-2K-5	SNP	Transition	4,343,175	C --> T	Thr (579) --> Ile	*mutL*	DNA mismatch repair	-	-	-
FL-2K-5	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-2K-5	Mobile element	Insertion	4,533,508	-	-	*yjiY*	Putative inner membrane protein	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	113,403	-	-	*ECB_01533* (- 103)	Hypothetical protein	*hokD* (-109)	Toxin-antitoxin system	IS150 mediated
FL-4K-5	Mobile element	Insertion	388,527	-	-	*brnQ*	Putative branched chain amino acid transporter	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	471,785	-	-	*aes*	Acetyl esterase	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	668,255	-	-	*yegM* (-257)	Multidrug efflux system subunit MdtA	*yegL* (-1,285)	Hypothetical protein	IS150 mediated, intergenic
FL-4K-5	Mobile element	Insertion	743,346	-	-	*yeeJ*	Adhesin	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	910,345	-	-	*ynjI*	Hypothetical protein	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	974,185	-	-	*ydiQ*	Putative electron transfer flavoprotein YdiQ	-	-	IS150 mediated

FL-4K-5	Deletion	Small	1,113,343	-	-	*ECB_01533*	Hypothetical protein	*hokD*	Small toxic polypeptide	IS1 insertion mediated, intergenic
FL-4K-5	SNP	Transversion	1,116,591	G --> T	Gln (285) --> Lys	*insF-4*	Putative transposase for IS3	-	-	-
FL-4K-5	Deletion	Small	1,272,401	-	-	*trg*	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB*	Toxin-antitoxin system	Deletion mediated by the insertion of IS150 at 1272400 and 1272468 and resulting in 67 bp deletion, intergenic
FL-4K-5	Mobile element	Insertion	1,272,400	-	-	*trg* (-258)	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB* (-83)	Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-5	Mobile element	Insertion	1,464,061	-	-	*chaA* (+238)	calcium/sodium:proton antiporter	*ldrC* (-167)	LTR Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-5	Deletion	Small	1,464,062	1 gene	-	*[ldrC]*	LTR Toxin-antitoxin system	-	-	IS150 insertion at 1464672 mediates this deletion, intergenic
FL-4K-5	Mobile	Insertion	1,579,629	-	-	*flgL*	Flagellar hook-	-	-	IS150

	element						associated protein FlgL			mediated
FL-4K-5	Deletion	Small	1,766,329	1 gene	-	*[pflB]*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-4K-5	Deletion	Small	1,831,485	38 genes	-	*ECB_00851*	Hypothetical protein	*ECB_00814*	Hypothetical protein	No IS elements, but brings ybjL and ybjK together
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00851*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00850*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00849*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00848*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*tfaE*	Tail fiber assembly	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00846*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00845*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00844*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00843*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00842*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00841*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00840*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00839*	Unknown	-	-	-

FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00838*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00837*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00836*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00835*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00834*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00833*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00832*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00831*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00830*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00829*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00828*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00827*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*yegZ*	gpD phage P2-like protein D; C-terminal fragment	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ogrK*	homologous of the phage-encoded Ogr protein that controls the bacteriophage P2 late transcription	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00824*	Unknown	-	-	-

FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00823*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00822*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00821*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00820*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00819*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00818*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00817*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00816*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00815*	Unknown	-	-	-
FL-4K-5	Deletion	Large	1,831,485	38 genes	-	*ECB_00814*	Part of the lambdoid prophage			
FL-4K-5	Mobile element	Insertion	2,130,703	-	-	*yehM*	Hypothetical protein	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	2,239,354	-	-	*rcsC* (-4)	RcsB two component signal transport	*atoS* (-209)	Sensory histidine kinases	IS150 mediated, intergenic
FL-4K-5	Mobile element	Insertion	2,343,551	-	-	*argT*	Lysine/arginine/ornithine transporter subunit	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	2,378,861	-	-	*fadL*	Long chain fatty acid outer membrane transporter	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	2,400,852	-	-	*yfdX*	Hypothetical protein	-	-	IS1 mediated
FL-4K-5	SNP	Transition	3,163,949	G --> A	-	*yhaO* (-147)	Putative transporter	*tdcG* (+192)	L-serine dehydratase 3	Intergenic

FL-4K-5	Mobile element	Insertion	3,600,372	-	-	*yhjT*	Hypothetical protein	-	-	IS150 mediated
FL-4K-5	Deletion	Large	3,600,373	4 genes	-	*[yhjT]*	-	*ldrE*	-	IS150 mediated
FL-4K-5	Deletion	Large	3,600,373	4 genes		*[yhjT]*	associated with cellulose production	-	-	-
FL-4K-5	Deletion	Large	3,600,373	4 genes		*yhjU*	associated with cellulose production	-	-	-
FL-4K-5	Deletion	Large	3,600,373	4 genes		*ldrD*	LTR Toxin-antitoxin system	-	-	-
FL-4K-5	Deletion	Large	3,600,373	4 genes		*ldrE*	LTR Toxin-antitoxin system	-	-	-
FL-4K-5	Mobile element	Insertion	3,602,958	-	-	*ldrE* (-90)	LTR Toxin-antitoxin system	*ldrF* (+286)	LTR Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-5	Mobile element	Insertion	4,043,794	-	-	*rhaS*	Transcriptional activator RhaS	-	-	IS150 mediated
FL-4K-5	Deletion	Small	4,295,377	-	-	*dcuR* (-755)	DNA binding transcriptional activator DcuR	*yjdI* (-1055)	Hypothetical protein	Intergenic
FL-4K-5	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-4K-5	Mobile element	Insertion	4,554,029	-	-	*mdoB*	Encodes two phosphoglycerol transferase I	-	-	IS150 mediated
FL-4K-5	Mobile element	Insertion	4,581,547	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
FL-2K-6	Mobile element	Insertion	360,203	-	-	*yaiT*	Hypothetical protein	-	-	IS150 mediated
FL-2K-6	Mobile element	Insertion	388,275	-	-	*brnQ*	Branched chain amino acid transporter	-	-	IS150 mediated

FL-2K-6	Mobile element	Insertion	471,785	-	-	*aes*	Acetyl esterase	-	-	IS150 mediated
FL-2K-6	Mobile element	Insertion	785,037	-	-	*cheB*	Chemotaxis-specific methylesterase	-	-	IS150 mediated
FL-2K-6	Mobile element	Insertion	960,637	-	-	*btuE*	Putative glutathione peroxidase	-	-	IS150 mediated
FL-2K-6	Mobile element	Insertion	974,185	-	-	*ydiQ*	Putative electron transfer flavoprotein YdiQ	-	-	IS150 mediated, online
FL-2K-6	Mobile element	Insertion	1,272,468	-	-	*trg* (-326)	Methyl accepting chemotaxis protein III	*mokB* (-13)	Toxin-antitoxin system	IS150 mediated, intergenic
FL-2K-6	Insertion	Base	1,328,521	+A	-	*zntB*	Trans-membrane zinc transporter (see reference)	-	-	-
FL-2K-6	Mobile element	Insertion	1,598,705	-	-	*dinI*	DNA damage-inducible protein	-	-	IS150 mediated
FL-2K-6	Mobile element	Insertion	1,675,380	-	-	*yccC*	Cryptic autophosphorylating protein tyrosine kinase Etk	-	-	IS150 mediated
FL-2K-6	Mobile element	Deletion	1,764,888	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-2K-6	Deletion	Small	3,543,375	-	-	*yhiO* (-160)	Universal stress protein UspB	*uspA* (-231)	Universal stress global regulator	Intergenic
FL-2K-6	Mobile element	Insertion	3,972,154	-	-	*fadA*	3-ketoacyl-CoA thiolase	-	-	IS150 mediated
FL-2K-6	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-2K-6	Deletion	Large	4,532,961	16 genes	-	*yjiY*	-	*[mdoB]*	Toxin-antitoxin system	IS150 mediated
FL-2K-6	Deletion	Large	4,532,961	16 genes		*yjiY*	Induced in stationary phase when cells are			

							growing on amino acids or peptides			
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaC*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaA*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaB*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaX*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaI*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaH*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaF*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaD*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaE*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaG*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*hpaR*	Bacterial pathogenicity	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*tsr*	Methyl-accepting chemotaxis proteins	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*yjiZ*	Proton-driven L-galactonate uptake transporter	-	-	-
FL-2K-6	Deletion	Large	4,532,961	16 genes		*yjjN*	L-galactonate oxidoreductase that is required for growth on L-galactonate as the sole carbon source	-	-	-

							under high-throughput growth conditions with limited aeration			
FL-2K-6	Deletion	Large	4,532,961	16 genes		*[mdoB]*	Encodes two phosphoglycerol transferases	-	-	-
FL-4K-6	Mobile element	insertion	388,527	-	-	*brnQ*	Putative branched chain amino acid transporter	-	-	IS150 mediated
FL-4K-6	Mobile element	Insertion	471,783	-	-	*aes*	Acetyl esterase	-	-	IS150 mediated
FL-4K-6	Mobile element	Insertion	743,346	-	-	*yeeJ*	Adhesin	-	-	IS150 mediated
FL-4K-6	Mobile element	Insertion	974,185	-	-	*ydiQ*	Putative electron transfer flavoprotein YdiQ	-	-	IS150 mediated, online
FL-4K-6	Mobile element	Insertion	1,113,403	-	-	*ECB_01533* (+103)	Hypothetical protein	*hokD* (-107)	Small toxic polypeptide	IS150 mediated, intergenic
FL-4K-6	SNP	Transversion	1,116,591	G --> T	Gln (285) --> Lys	*insF-4*	Putative transposase of IS3	-	-	-
FL-4K-6	Deletion	Small	1,272,401	-	-	*trg*	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB*	Regulatory peptide	Deletion mediated by the insertion of IS150 at 1272400 and 1272468 and resulting in 67 bp deletion, intergenic

FL-4K-6	Mobile element	Insertion	16,992	-	-	*hokC* (-91)	Small toxic membrane polypeptide	*nhaA* (-494)	Toxin-antitoxin system	IS150 mediated , intergenic
FL-4K-6	SNP	Transition	1,325,632	A --> T	Lys (10) --> Lys	*ydaO*	C32 tRNA thiolase	-	-	-
FL-4K-6	Deletion	Small	1,464,062	1 gene	-	*[ldrC]*	LTR Toxin-antitoxin system	-	LTR Toxin-antitoxin system	IS150 insertion at 1464672 mediates this deletion, intergenic
FL-4K-6	Mobile element	Insertion	1,464,672	-	-	*ldrC* (+273)	LTR Toxin-antitoxin system	*ldrB* (-155)	LTR Toxin-antitoxin system	
FL-4K-6	Mobile element	Insertion	1,579,629	-	-	*flgL*	Flagellar hook-associated protein FlgL	-	-	IS150 mediated
FL-4K-6	Deletion	Small	1,766,329	1 gene	-	*[pflB]*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00851*	Hypothetical protein	*ECB_00814*	Hypothetical protein	No IS elements, but brings ybjL and ybjK together
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00851*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00850*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00849*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00848*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*tfaE*	Tail fiber assembly	-	-	-

FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00846*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00845*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00844*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00843*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00842*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00841*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00840*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00839*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00838*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00837*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00836*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00835*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00834*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00833*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00832*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00831*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00830*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00829*	Unknown	-	-	-

FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00828*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00827*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*yegZ*	gpD phage P2-like protein D; C-terminal fragment	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ogrK*	Homologous of the phage-encoded Ogr protein that controls the bacteriophage P2 late transcription	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00824*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00823*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00822*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00821*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00820*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00819*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00818*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00817*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00816*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00815*	Unknown	-	-	-
FL-4K-6	Deletion	Large	1,831,485	38 genes	-	*ECB_00814*	Part of the lambdoid prophage	-	-	-

FL-4K-6	Mobile element	Insertion	2,654,436	-	-	*ECB_02512*	Hypothetical protein	-	-	IS150 mediated
FL-4K-6	Mobile element	Insertion	2,872,437	-	-	*yqeB*	Hypothetical protein	-	-	IS150 mediated
FL-4K-6	SNP	Transition	3,163,949	G --> A	-	*yhaO* (-147)	Putative transporter	*tdcG* (+192)	L-serine dehydratase 3	Intergenic
FL-4K-6	SNP	Transition	3,380,600	G --> A	Ala (240) --> Val	*kefB*	Glutathione-regulated potassium-efflux system dehydratase KefB	-	-	-
FL-4K-6	Mobile element	Insertion	3,602,958	-	-	*ldrE* (-90)	LTR Toxin-antitoxin system	*ldrF* (+286)	LTR Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-6	Mobile element	Insertion	3,669,960	-	-	*rhsA*	rhsA element core protein RshA	-	-	IS150 mediated
FL-4K-6	Mobile element	Insertion	3,896,680	-	-	*rfe*	UDP-GlcNAc:undecaprenylphosphate GlcNAc-1-phosphate transferase	-	-	IS150 mediated
FL-4K-6	Deletion	-	4,295,377	-	-	*dcuR* (-755)	DNA binding transcriptional activator DcuR	*yjdI* (-1055)	Hypothetical protein	Intergenic
FL-4K-6	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-4K-6	Mobile element	Insertion	4,581,547	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated
FL-4K-6	Mobile element	Insertion	1,464,061	-	-	*chaA* (+238)	calcium/sodium:proton antiporter	*ldrC* (-167)	*ldrC* is one of a set of four Long Direct Repeats (LDRs), one of which has been shown to code for the toxin portion of a toxin-antitoxin	IS150 mediated, intergenic

									pair.	
FL-4K-6	Mobile element	Insertion	1,272,400	-	-	*trg* (-258)	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB* (-83)	Regulatory peptide	IS150 mediated, intergenic
FL-2K-7	Mobile element	Insertion	16,989	-	-	*hokC* (-88)	Small toxic membrane polypeptide	*nhaA* (-496)	pH dependent sodium or proton antiporter	IS150 mediated, intergenic
FL-2K-7	Deletion	Large	380,365	6 gene	-	*[araJ]*	Sugar efflux system	*[brnQ]*	Putative branched chain amino acid transporter	-
FL-2K-7	Deletion	Large	380,365	6 gene	-	*[araJ]*	Sugar efflux system	-	-	-
FL-2K-7	Deletion	Large	380,365	6 gene	-	*sbcC*	Required for recombinational repair of dsDNA breaks	-	-	-
FL-2K-7	Deletion	Large	380,365	6 gene	-	*sbcD*	Required for recombinational repair of dsDNA breaks	-	-	-
FL-2K-7	Deletion	Large	380,365	6 gene	-	*phoB*	Positive response regulator for Phosphate regulon, important for pathogenicity, two-component system	-	-	-
FL-2K-7	Deletion	Large	380,365	6 gene	-	*phoR*	Positive response regulator for Phosphate regulon, important for pathogenicity, two-component system	-	-	-
FL-2K-7	Deletion	Large	380,365	6 gene	-	*[brnQ]*	Putative branched chain amino acid	-	-	-

							transporter			
FL-2K-7	Mobile element	Insertion	582,237	-	-	*insA-7* (+193)	IS1 protein	*hokE*	Toxin-antitoxin system	IS150 mediated, intergenic
FL-2K-7	Mobile element	Insertion	963,716	-	-	*ydiU*	Hypothetical protein	-	-	IS150 mediated
FL-2K-7	Mobile element	Insertion	974,185	-	-	*ydiQ*	Putative electron transfer flavoprotein YdiQ	-	-	IS150 mediated, online
FL-2K-7	Mobile element	Insertion	1,272,468	-	-	*trg* (-326)	Methyl accepting chemotaxis protein III	*mokB* (-13)	Regulatory peptide	IS150 mediated, intergenic
FL-2K-7	Mobile element	Insertion	1,464,061	-	-	*chaA* (+238)	calcium/sodium:proton antiporter	*ldrC* (-167)	LTR Toxin-antitoxin system	IS150 mediated, intergenic
FL-2K-7	Mobile element	Deletion	1,764,886	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-2K-7	SNP	Transition	2,069,532	C --> T	Ser (176) --> Phe	*gltK*	Glutamate and aspartate transporter subunit	-	-	-
FL-2K-7	Mobile element	Insertion	2,628,603	-	-	*yfiH*	Hypothetical protein	-	-	IS150 mediated
FL-2K-7	Mobile element	Insertion	2,654,657	-	-	*ECB_02512*	Hypothetical protein	-	-	IS1
FL-2K-7	Mobile element	Insertion	4,239,784	-	-	*gltP*	Glutamate/aspartate proton symporter	-	-	IS150 mediated
FL-2K-7	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-2K-7	Mobile element	Insertion	4,382,959	-	-	*yjiX* (-32)	Hypothetical protein	*yjiY* (+17)	Putative inner membrane protein	IS150 mediated, intergenic
FL-2K-7	Deletion	Large	546,986 – 547,703	7 genes	-	*[insB-6]*	IS150 protein	*[ompT]*	-	IS1 mediated
FL-2K-7	Deletion	Large	546,986 –	7 genes	-	*[insB-6]*	After frameshift	-	-	

			547,703				event,transcribed with InsA mediates IS1 transposition			
FL-2K-7	Deletion	Large	546,986 – 547,703	7 genes	-	*ECB_00514*	Unknown	-	-	-
FL-2K-7	Deletion	Large	546,986 – 547,703	7 genes	-	*ECB_00515*	Unknown	-	-	-
FL-2K-7	Deletion	Large	546,986 – 547,703	7 genes	-	*ECB_00516*	Unknown	-	-	-
FL-2K-7	Deletion	Large	546,986 – 547,703	7 genes	-	*ECB_00517*	Unknown	-	-	-
FL-2K-7	Deletion	Large	546,986 – 547,703	7 genes	-	*appY*	Transcriptional regulator induces the expression of energy metabolism genes under anaerobiosis	-	-	-
FL-2K-7	Deletion	Large	546,986 – 547,703	7 genes	-	*[ompT]*	Outer membrane protease, with specificity for paired basic residues	-	-	-
FL-4K-7	Mobile element	Insertion	388,530	-	-	*brnQ*	Putative branched chain amino acid transporter	-	-	IS150 mediated
FL-4K-7	Mobile element	Insertion	471,783	-	-	*aes*	Acetyl esterase	-	-	IS150 mediated
FL-4K-7	Mobile element	Insertion	743,346	-	-	*yeeJ*	Adhesin	-	-	IS150 mediated
FL-4K-7	Mobile element	Insertion	974,185	-	-	*ydiQ*	Putative electron transfer flavoprotein YdiQ	-	-	IS150 mediated, online
FL-4K-7	Mobile element	Insertion	1,113,403	-	-	*ECB_01533* (+103)	Hypothetical protein	*hokD* (-107)	Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-7	SNP	Transversion	1,116,591	G --> T	Gln	*insF-4*	Putative	-	-	

					(285) -> Lys		transposase of IS3			
FL-4K-7	Mobile element	Insertion	1,272,468	-	-	*trg* (-326)	Methyl-accepting chemotaxis protein III, ribose and galactose sensor receptor	*mokB* (-13)	Toxin-antitoxin system	IS150 mediated, intergenic
FL-4K-7	Mobile element	Insertion	1,579,629	-	-	*flgL*	Flagellar hook-associated protein FlgL	-	-	IS150 mediated
FL-4K-7	Deletion	Small	1,764,888	-	-	*pflB*	Pyruvate formate lyase I	-	-	IS150 mediated
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00851*	Hypothetical protein	*ECB_00814*	Hypothetical protein	No IS elements, but brings ybjL and ybjK together
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00851*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00850*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00849*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00848*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*tfaE*	Tail fiber assembly	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00846*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00845*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00844*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00843*	Unknown	-	-	-

FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00842*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00841*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00840*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00839*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00838*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00837*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00836*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00835*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00834*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00833*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00832*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00831*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00830*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00829*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00828*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00827*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*yegZ*	gpD phage P2-like protein D; C-terminal fragment	-	-	-

FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ogrK*	homologous of the phage-encoded Ogr protein that controls the bacteriophage P2 late transcription	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00824*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00823*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00822*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00821*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00820*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00819*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00818*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00817*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00816*	Unknown	-	-	-
FL-4K-7	Deletion	Large	1,831,485	37 genes	-	*ECB_00815*	Unknown	-	-	-
FL-4K-7	Mobile element	Insertion	2,590,037	-	-	*purl* (-995)	Phosphoribosylformylglycinamidine synthase	*insA-19* (-867)	IS1 protein InsA	IS150 mediated
FL-4K-7	SNP	Transversion	2,678,935	T --> G	Leu (162) --> Arg	*emrR*	Transcriptional repressor MprA (multidrug resistance regulator)	-	-	-
FL-4K-7	Mobile element	Insertion	2,850,178	-	-	*kduD*	2-deoxy-D-gluconate 3-	-		IS150 mediated

							dehydrogenase			
FL-4K-7	Mobile element	Insertion	2,974,776	-	-	*flu*	Antigen 43 (Ag43) phase-variable biofilm formation autotransporter	-	-	IS150 mediated
FL-4K-7	Mobile element	Insertion	3,024,559	-	-	*glcF*	Glycolate oxidase iron-sulfur subunit	-	-	IS150 mediated
FL-4K-7	SNP	Transition	3,163,949	G --> A	-	*yhaO* (-147)	Putative transporter	*tdcG* (+192)	L-serine dehydratase 3	Intergenic
FL-4K-7	Deletion	Small	4,295,377	-	-	*dcuR* (-755)	DNA binding transcriptional activator DcuR	*yjdI* (-1055)	Hypothetical protein	Intergenic
FL-4K-7	Mobile element	Insertion	4,381,583	-	-	*cycA*	D-alanine, D-serine, glycine permease	-	-	IS150 mediated, online
FL-4K-7	Mobile element	Insertion	4,581,547	-	-	*nadR*	Nicotinamide-nucleotide adenylyltransferase	-	-	IS150 mediated

[a] The lineage in which the mutation arose.

[b] The type of mutation that had occurred.

[c] Further sub-classification of the mutation type that occurred. Note for large deletions, the first and last gene affected is listed as the first event, followed by each in between listed underneath.

[d] The position on the reference genome at which the mutation occurred. In some cases, particularly with large deletions, the exact location of the beginning of the mutation could not be specified from breseq output and instead the likely range is given.

[e] The change that occurred at the DNA level as a result of the mutation. In the case of large deletions, the number of genes deleted is indicated.

[f] The amino acid change that had occurred at the protein level as a result of the mutation. Also indicated was the position along the protein in which the change occurred.

[g] The gene mutated. In the cases of genes partially affected, the gene name was enclosed with parenthesis []. Intergenic mutations are identified by the closest gene to the mutation with the number of bps from the beginning of the gene (+ for downstream and – for upstream).

[h] A brief functional description of the protein encoded by Reference Gene 1.

[i] Intergenic mutations are identified by the closest gene to the mutation with the number of bps from the beginning of the gene

[j] A brief functional description of the protein encoded by Reference Gene 2.
[k] Additional information if required

Figure 7.6: List of all mutations arising in the aerobic environment.

Locus [a]	Mutation [b]	Generation [c]	Position [d]	Treatment of occurrence [e]			Conclusion [f]
				AE	AN	FL	
yabB	SNP	2,000	92,892	1			Conclusion 1
lpxC	SNP	2,000	109,384	1			Conclusion 1
prpD	SNP	2,000	323,923	1			Conclusion 1
gnd	SNP	2,000	700,702	1			Conclusion 1
zwf/edd	Insertion	4,000	818,462	1			Conclusion 1
fadD	SNP	2,000	863,948	1			Conclusion 1
rnfD	SNP	4,000	1,046,106	1			Conclusion 1
rnb/yciR	Small deletion	4,000	1,389,511	1			Conclusion 1
yceF	SNP	4,000	1,573,697	1			Conclusion 1
ycbS	Small deletion	4,000	1,716,607	1			Conclusion 1
kps	SNP	2,000	1,943,095	1			Conclusion 1
nagA	SNP	2,000	2,052,242	1			Conclusion 1
metG	SNP	4,000	2,119,116	1			Conclusion 1
yohK	Insertion	4,000	2,155,627	1			Conclusion 1
gspE	IS150 insertion	4,000	3,009,254	1			Conclusion 1
rhsB	Small deletion	2,000	3,524,253	1			Conclusion 1
uhpT	SNP	4,000	3,778,620	1			Conclusion 1
hemX	Deletion	2,000	3,915,605	1			Conclusion 1
ileT/alaT	Small deletion	4,000	3,981,502	1			Conclusion 1
metL	Insertion	4,000	4,076,128	1			Conclusion 1
yijD	SNP	2,000	4,107,509	1			Conclusion 1
uvrA	SNP	2,000	4,217,932	1			Conclusion 1
trg/mokB	SNP	4,000	1,272,262	1			Conclusion 2
trg/mokB	IS150 insertion	4,000	1,272,399	1			Conclusion 2
chaA/ldrC	IS element	4,000	1,464,143	1			Conclusion 2
ldrC/ldrB	IS150 insertion	2,000	1,464,679	1			Conclusion 2
ldrB/ldrA	IS150 insertion	4,000	1,465,130	1			Conclusion 2
yehI	IS150 insertion	4,000	2,127,047	1			Conclusion 2
yehU	IS150 insertion	4,000	2,138,008	1			Conclusion 2
rpoS	SNP	4,000	2,734,340	1			Conclusion 2
flu	IS1 insertion	2,000	2,972,936	2			Conclusion 2
kpsE	SNP	2,000	2,987,334	1			Conclusion 2
kpsT	Deletion	2,000	2,999,898	1			Conclusion 2
kpsT	SNP	2,000	3,000,161	1			Conclusion 2
kpsT	SNP	2,000	3,000,346	2			Conclusion 2
rpoD	SNP	4,000	3,119,112	1			Conclusion 2
rpoD	SNP	2,000	3,119,520	1			Conclusion 2
hokC/nhaA	IS150 insertion	4,000	16,972	2	2	2	Conclusion 3
brnQ	IS150 insertion	4,000	388,275	1		4	Conclusion 3
brnQ	SNP	4,000	389,105	1			Conclusion 3
cheB	IS150 insertion	4,000	785,037	1		2	Conclusion 3
yeaA	Small deletion	2,000	892,201	2			Conclusion 3
ydiU	IS150 insertion	4,000	963,716	3		1	Conclusion 3

trg/mokB	IS150 insertion	4,000	1,272,468	2	9	6	Conclusion 3
zntB	Insertion	4,000	1,328,521	1		3	Conclusion 3
ycfQ/ycfJ	IS150 insertion	4,000	1,551,946	1		2	Conclusion 3
yccC	IS150 insertion	4,000	1,675,380	1		2	Conclusion 3
insJ-2/rihA	Deletion	4,000	2,071,483	1		2	Conclusion 3
fadL	IS150 insertion	4,000	2,378,921	1			Conclusion 3
kduD	IS150 insertion	4,000	2,850,135	1			Conclusion 3
flu	SNP	4,000	2,974,778	2			Conclusion 3
flu	IS150 insertion	4,000	2,975,127	1			Conclusion 3
kpsD	SNP	2,000	2,988,653	2			Conclusion 3
kpsT	SNP	2,000	3,000,095	2			Conclusion 3
flu	IS3 insertion	4,000	3,000,514	1		1	Conclusion 3
yhaO/tdcG	IS150 insertion	4,000	3,164,090	1			Conclusion 3
trkD/insJ-5	Insertion	2,000	3,866,357	4	11		Conclusion 3
trkD/insJ-5	Deletion	2,000	3,866,358	2			Conclusion 3
cycA	IS150 insertion	4,000	4,381,583	7	13	14	Conclusion 3
yjiiX/yjiY	Large deletion	2,000	4,532,961	1			Conclusion 3
yaiT	IS150 insersion	4,000	360,203	1		2	Conclusion 3
brnQ	IS150 insertion	4,000	388,234	5			Conclusion 4
ECB_00516	Deletion	4,000	550,063	4			Conclusion 4
galR	Insertion	4,000	2,844,904	5			Conclusion 4
kpsM	IS150 insertion	4,000	3,000,527	5			Conclusion 4
yihS	IS150 insertion	4,000	4,015,454	2		1	Conclusion 4
ybiI	SNP	4,000	1,905,307	5			Conclusion 5
Total number of mutation in the aerobic environment				**107**			

[a] The gene or position in which the mutation occurred. If a mutation was intergenic, the nearest genes upstream and downstream were indicated with a "/" between them. If the gene was an insertion or a deletion of a range of bases or genes a "-" was indicated between the upstream or downstream of the mutation.

[b] The type of mutation that occurred within that gene or operon

[b] The generation in which the mutation was identified.

[c] The location in which the mutation was identified.

[d] Position in which the mutation occurred. The coordinates are given are of the ancestral file.

[e] The treatment under which the mutation was reported. If the mutation was reported in either anaerobic or fluctuating lineages these events were also reported.

[f] The conclusion category the mutation falls into based on the outcomes of questions outlined in **Figure 4.2.**

Table 7.7: List of all mutations arising in the anaerobic environment.

Locus [a]	Mutation [b]	Generation [c]	Position [d]	Treatment of occurrence [e]			Conclusion [f]
				AE	AN	FL	
ileS	Insertion	4,000	26,953		1		Conclusion 1
rrsH	IS150 insertion	2,000	227,515		1		Conclusion 1
ybaM	Insertion	4,000	462,233		1		Conclusion 1

ybbP	IS150 insertion	2,000	494,056		1		Conclusion 1
ybcK	SNP	4,000	542,005		1		Conclusion 1
ybcQ/insB-6	IS3 insertion	4,000	546,599		1		Conclusion 1
pheS	IS150 insertion	2,000	955,693		1		Conclusion 1
tus	IS150 insertion	2,000	1,071,398		1		Conclusion 1
ansP	SNP	2,000	1,238,659		1		Conclusion 1
gltA	SNP	2,000	2,001,307		1		Conclusion 1
rhsC	IS150 insertion	2,000	2,024,212		1		Conclusion 1
insB-17/[atoB]	Deletion	2,000	2,235,631		1		Conclusion 1
yfjF	SNP	4,000	2,648,318		1		Conclusion 1
galR	SNP	2,000	2,844,146		1		Conclusion 1
flu	Insertion	4,000	2,972,604		1		Conclusion 1
flu	SNP	2,000	2,973,574		1		Conclusion 1
flu	Deletion	4,000	2,973,807		1		Conclusion 1
bfr	IS3 insertion	2,000	3,367,380		1		Conclusion 1
rhsB	IS150 insertion	4,000	3,522,963		1		Conclusion 1
yidX	IS150 insertion	2,000	3,808,730		1		Conclusion 1
yieP	SNP	4,000	3,869,168		1		Conclusion 1
polA/engB	IS150 insertion	2,000	3,994,146		1		Conclusion 1
pcnB	SNP	2,000	161,770		1		Conclusion 2
[insB-6] – [insA-7]	Deletion	4,000	546,990		1		Conclusion 2
adhE	SNP	2,000	1,438,030		4		Conclusion 2
adhE	SNP	2,000 & 4,000	1,439,673		10		Conclusion 2
glnU	Deletion	2,000	2,057,167		1		Conclusion 2
glnU	Deletion	2,000	2,057,176		1		Conclusion 2
glnU	Deletion	2,000	2,057,181		1		Conclusion 2
glnU	Deletion	2,000	2,057,184		1		Conclusion 2
glnU	Deletion	2,000	2,057,185		1		Conclusion 2
glnU	Deletion	2,000	2,057,189		1		Conclusion 2
hokC/nhaA	IS150 insertion	4,000	16,972	2	2	2	Conclusion 3
ynjI	IS150 insertion	2,000	910,345		3	1	Conclusion 3
trg/mokB	IS150 insertion	2,000 & 4,000	1,272,468	2	9	6	Conclusion 3
kduD	IS150 insertion	2,000	2,850,178		1	1	Conclusion 3
flu	SNP	2,000 & 4,000	2,972,858		8		Conclusion 3
flu	SNP	4,000	2,974,237		6		Conclusion 3
trkD/insJ-5	Insertion	2,000 & 4,000	3,866,357	4	11		Conclusion 3
dcuR/yjdI	Deletion	2,000	4,295,377		6	6	Conclusion 3
cycA	IS150 insertion	2,000 & 4,000	4,381,583	7	13	14	Conclusion 3
[insB-6] - ybdK	Deletion	2,000 & 4,000	546,975		7		Conclusion 4
[insB-6] - ybdK	Deletion	2,000 & 4,000	546,975		7		Conclusion 4

citG	IS150 insertion	2,000	621,252		3		Conclusion 4
yeaS	IS150 insertion	2,000	872,829		2		Conclusion 4
ynfN	IS150 insertion	4,000	1,123,058		5		Conclusion 4
pflB	IS150 deletion	4,000	1,764,886		2	3	Conclusion 4
cydA/ybgG	IS1 insertion	4,000	1,984,267		4		Conclusion 4
mglB	IS150 insertion	2,000	2,164,019		2		Conclusion 4
emrY	IS3 insertion	2,000	2,387,885		2		Conclusion 4
alaW/yfeC	IS150 insertion	4,000	2,424,083		4		Conclusion 4
hcaC	IS150 insertion	2,000	2,564,470		2		Conclusion 4
ECB_02804	IS150 insertion	2,000	2,980,273		2		Conclusion 4
kpsS	IS1 insertion	2,000	2,992,382		5		Conclusion 4
yqiI	IS1 insertion	2,000	3,096,895		2		Conclusion 4
yrbI	IS150 insertion	4,000	3,250,708		5		Conclusion 4
ugpB	IS150 insertion	2,000	3,495,511		2		Conclusion 4
rhsA	IS150 insertion	4,000	3,669,960		3	2	Conclusion 4
alr/tyrB	IS150 insertion	2,000	4,211,817		2		Conclusion 4
ytfT	IS150 insertion	2,000	4,401,824		2		Conclusion 4
yjiX	IS150 insertion	2,000	4,581,545		5		Conclusion 4
nadR	IS150 insertion	4,000	4,581,546		4		Conclusion 4
pcnB	Deletion	2,000 & 4,000	161,094		8		Conclusion 5
proQ	Deletion of 4 bases	4,000	838,221		5		Conclusion 5
dbpA	Duplication	4,000	1,328,413		6		Conclusion 5
sucA	Duplication	4,000	1,995,122		4		Conclusion 5
[rseB]/[rseA]	Duplication	2,000	2,603,060		2		Conclusion 5
arcB	Deletion	4,000	3,260,806		7		Conclusion 5
rng	Insertion	2,000 & 4,000	3,298,183		8		Conclusion 5
glgC	Deletion	2,000 & 4,000	3,473,571		8		Conclusion 5
mpl	Insertion (+9)	4,000	4,406,085		2		Conclusion 5
mpl	Insertion (+7)	4,000	4,406,085		2		Conclusion 5
Total number of mutation in the anaerobic environment					227		

[a] The gene or position in which the mutation occurred.

[b] The type of mutation that occurred within that gene or operon

[b] The generation in which the mutation was identified.

[c] The location in which the mutation was identified.

[d] Position in which the mutation occurred. The coordinates are given are of the ancestral file.

[e] The treatment under which the mutation was reported. If the mutation was reported in either aerobic or fluctuating lineages these events were also reported.

[f] The conclusion category the mutation falls into based on the outcomes of questions outlined in **Figure 4.2**.

Table 7.8: List of all mutations arising in the fluctuating environment.

Locus [a]	Mutation [b]	Generation [c]	Position [d]	Treatment of occurrence [e]			Conclusion [f]
				AE	AN	FL	
ECB_00021	SNP	2,000	20,661			1	Conclusion 1
[araJ] – [brnQ]	Large deletion	2,000	380,365			1	Conclusion 1
cyoA/ampG	Insertion	2,000	422,593			1	Conclusion 1
insL-2	Mobile element	2,000	429,505			1	Conclusion 1
priC	IS150 insertion	2,000	462,604			1	Conclusion 1
[insB-6]-ompT	Deletion	2,000	546,986			1	Conclusion 1
insA-7/ hokE	IS150 insertion	2,000	582,237			1	Conclusion 1
[ECB_02013] - ogrK	Deletion	4,000	632,692			1	Conclusion 1
ogrK/yegQ	IS150 insertion	2,000	654,734			1	Conclusion 1
yegM/yegL	IS150 insertion	4,000	668,255			1	Conclusion 1
manX/yaoE	IS1 insertion	4,000	851,286			1	Conclusion 1
ydiV	IS1 insertion	4,000	932,998			1	Conclusion 1
sufS	IS1 insertion	4,000	994,990			1	Conclusion 1
yddA	IS150 insertion	2,000	1,181,538			1	Conclusion 1
yddA/yddB	IS150 insertion	4,000	1,181,682			1	Conclusion 1
sfcA/adhP	SNP	4,000	1,206,872			1	Conclusion 1
ydbD	IS1 insertion	2,000	1,288,903			1	Conclusion 1
insF-2	SNP	2,000	1,290,872			1	Conclusion 1
tynA	IS150 insertion	4,000	1,292,085			1	Conclusion 1
ydaO	SNP	4,000	1,325,632			1	Conclusion 1
ydaM	SNP	2,000	1,329,917			1	Conclusion 1
ynaI	IS1 insertion	2,000	1,341,071			1	Conclusion 1
ycfU/ycfT	IS150 insertion	4,000	1,544,946			1	Conclusion 1
flgK	IS150 insertion	2,000	1,580,827			1	Conclusion 1
ycdO	IS150 insertion	4,000	1,635,657			1	Conclusion 1
ymcA	IS150 insertion	4,000	1,670,113			1	Conclusion 1
[pqiB]	Deletion	4,000	1,702,902			1	Conclusion 1
ycaO/focA	IS150 insertion	2,000	1,762,790			1	Conclusion 1
rhsC	IS150 insertion	4,000	2,024,210			1	Conclusion 1
hscC	IS150 insertion	4,000	2,073,388			1	Conclusion 1
yehM	IS150 insertion	4,000	2,130,703			1	Conclusion 1
yehU	IS150 insertion	2,000	2,138,008			1	Conclusion 1
mqo/yojI	SNP	2,000	2,229,796			1	Conclusion 1
rcsC/atoS	IS150 insertion	4,000	2,239,354			1	Conclusion 1
yfcB	SNP	4,000	2,364,735			1	Conclusion 1
iscC	SNP	4,000	2,554,932			1	Conclusion 1
purI/insA-19	IS150 insertion	4,000	2,590,037			1	Conclusion 1
nupG	IS150 insertion	4,000	2,965,217			1	Conclusion 1
pitB	SNP	2,000	3,041,966			1	Conclusion 1
uxaA	IS150 insertion	4,000	3,148,353			1	Conclusion 1
kefB	SNP	4,000	3,380,600			1	Conclusion 1

prkB	IS150 insertion	4,000	3,385,413			1	Conclusion 1
yhfA/crp	IS150 insertion	2,000	3,386,643			1	Conclusion 1
gntR	IS150 insertion	4,000	3,482,646			1	Conclusion 1
yhiO/uspA	Deletion	2,000	3,543,375			1	Conclusion 1
insK-4/glyS	IS150 insertion	4,000	3,625,440			1	Conclusion 1
yieG	SNP	2,000	3,829,404			1	Conclusion 1
rfe	IS150 insertion	4,000	3,896,680			1	Conclusion 1
fadA	IS150 insertion	2,000	3,972,154			1	Conclusion 1
rhaS	IS150 insertion	4,000	4,043,794			1	Conclusion 1
thiC	Deletion	4,000	4,140,414			1	Conclusion 1
gltP	IS150 insertion	2,000	4,239,784			1	Conclusion 1
lysU	IS150 insertion	2,000	4,299,101			1	Conclusion 1
mutL	SNP	2,000	4,343,175			1	Conclusion 1
yjiX/yjiY	IS150 insertion	2,000	4,382,959			1	Conclusion 1
[yjiY] - [hpaC]	Deletion	2,000	4,533,440			1	Conclusion 1
yjiY	IS150 insertion	2,000	4,534,750			1	Conclusion 1
btuC	Mobile element	2,000	959,668			1	Conclusion 2
btuC	Mobile element	2,000	959,872			1	Conclusion 2
[tqsA] - ECB_01533	Deletion	4,000	1,081,310			1	Conclusion 2
[ydfX] - ECB_01533	Deletion	2,000	1,110,292			1	Conclusion 2
[ldrC] – ldrB	Deletion	4,000	1,464,062			1	Conclusion 2
ldrC/ldrB	Deletion	4,000	1,464,062			1	Conclusion 2
ldrC/ldrB	IS150 insertion	4,000	1,464,595			1	Conclusion 2
ldrC/ldrB	IS150 insertion	2,000	1,464,678			1	Conclusion 2
dinI	IS150 insertion	2,000	1,598,688			1	Conclusion 2
dinI	IS150 insertion	2,000	1,598,705			1	Conclusion 2
pflB	3 base deletion	2,000 & 4,000	1,766,329			1	Conclusion 2
[ECB_00851] – [ECB_00814]	Large deletion	4,000	1,831,485			1	Conclusion 2
ybiT	IS150 insertion	4,000	1,886,677			1	Conclusion 2
yfiQ	IS150 insertion	4,000	2,615,526			1	Conclusion 2
yfiH	IS150 insertion	2,000	2,628,603			1	Conclusion 2
ECB_02512	IS150 insertion	4,000	2,654,436			1	Conclusion 2
ECB_02512	IS1 insertion	2,000	2,654,657			1	Conclusion 2
glcF	IS150 insertion	4,000	3,024,559			1	Conclusion 2
glcD	Small deletion	4,000	3,027,366			1	Conclusion 2
yhaO/tdcG	SNP	4,000	3,163,949			1	Conclusion 2
yhaO/tdcG	Deletion	4,000	3,164,090			1	Conclusion 2
yhjE	IS150 insertion	4,000	3,577,565			1	Conclusion 2
yhjT	IS150 insertion	4,000	3,600,372			1	Conclusion 2
yjiY - [mdo]	Large deletion	2,000 & 4,000	4,532,961			2	Conclusion 2
mdoB	IS150 insertion	4,000	4,552,611			1	Conclusion 2
mdoB	IS150 insertion	4,000	4,554,029			1	Conclusion 2

hokC/nhaA	IS150 insertion	4,000	16,972	2	2	2	Conclusion 3
yaiT	IS150 insertion	2,000 & 4,000	360,203	1		2	Conclusion 3
brnQ	Insertion	2,000	388,020			1	Conclusion 3
brnQ	IS150 insertion	2,000 & 4,000	388,275	1		4	Conclusion 3
brnQ	IS150 insertion	4,000	388,530			1	Conclusion 3
brnQ	IS150 insertion	2,000	388,543			1	Conclusion 3
cheB	IS150 insertion	2,000 & 4,000	785,037	1		2	Conclusion 3
ynjI	IS150 insertion	4,000	910,345		3	1	Conclusion 3
ydiU	IS150 insertion	2,000	963,716	3		1	Conclusion 3
ECB_01533 - hokD	Deletion	4,000	1,113,343			1	Conclusion 3
trg/mokB	IS150 insertion	4,000	1,113,403			5	Conclusion 3
trg/mokB	IS150 insertion	4,000	1,272,400			4	Conclusion 3
trg/mokB	Deletion	4,000	1,272,401			4	Conclusion 3
trg/mokB	IS150 insertion	4,000	1,272,468	2	9	6	Conclusion 3
zntB	Insertion	2,000 & 4000	1,328,521	1		3	Conclusion 3
ldrC/ldrB	IS150 insertion	4,000	1,464,672			1	Conclusion 3
ycfQ/ycfJ	IS150 insertion	2,000	1,551,960	1		1	Conclusion 3
yccC	IS150 insertion	2,000 & 4000	1,675,380	1		2	Conclusion 3
pflB	IS150 deletion	2,000	1,764,886		2	3	Conclusion 3
pflB	5 base deletion	4,000	1,766,329			2	Conclusion 3
pflB	IS150 insertion	4,000	1,766,334			2	Conclusion 3
[ECB_00851] – [ECB_00814]	Deletion	4,000	1,831,485			5	Conclusion 3
insJ-2/rihA	Deletion	4,000	2,071,479	1		2	Conclusion 3
emrR	SNP	4,000	2,678,935			1	Conclusion 3
insJ-3/cysH	Deletion	4,000	2,748,700			1	Conclusion 3
kduD	IS150 insertion	4,000	2,850,135			1	Conclusion 3
kduD	IS150 insertion	4,000	2,850,178		1	1	Conclusion 3
flu	IS150 insertion	4,000	2,974,776			1	Conclusion 3
kpsM	IS150 insertion	2,000	3,000,519	1		1	Conclusion 3
yhaO/tdcG	SNP	4,000	3,153,949			6	Conclusion 3
yqiF	IS150 insertion	4,000	3,156,802			1	Conclusion 3
yhaO/tdcG	SNP	2,000	3,163,950			6	Conclusion 3
yraP/yraQ	IS150 insertion	4,000	3,205,065			1	Conclusion 3
ldrE/ldrF	IS150 insertion	4,000	3,602,958			3	Conclusion 3
ldrE/ldrF	IS150 insertion	4,000	3,603,153			1	Conclusion 3
rhsA	IS150 insertion	4,000	3,669,960		3	2	Conclusion 3
yihS	IS150 insertion	4,000	4,015,454	2		1	Conclusion 3
rhaA	IS150 insertion	4,000	4,040,290			1	Conclusion 3
dcuR/yjdI	Small deletion	2,000	4,295,377		6	6	Conclusion 3
yjiY	IS150 insertion	2,000	4,533,508			1	Conclusion 3
hokC/nhaA	IS150 insertion	4,000	16,989			2	Conclusion 4

brnQ	IS150 insertion	4,000	388,527			5	Conclusion 4
aes	IS150 insertion	4,000	471,783			6	Conclusion 4
yeeJ	IS150 insertion	4,000	743,346			5	Conclusion 4
yeaS	IS150 insertion	2,000	872,829			2	Conclusion 4
btuE	IS150 insertion	2,000	960,637			2	Conclusion 4
ydiQ	IS150 insertion	2,000 & 4,000	974,185			8	Conclusion 4
ECB_01533/ hokD	IS150 insertion	4,000	1,113,403			5	Conclusion 4
chaA/ldrC	IS150 insertion	4,000	1,464,061			6	Conclusion 4
flgL	IS150 insertion	4,000	1,579,629			6	Conclusion 4
pflB	IS150 deletion	2,000 & 4,000	1,764,888			6	Conclusion 4
argT	IS150 insertion	4,000	2,343,551			2	Conclusion 4
fadL	IS150 insertion	4,000	2,378,861	1		2	Conclusion 4
yqeB	IS150 insertion	4,000	2,872,437			2	Conclusion 4
cycA	IS150 insertion	2,000 & 4,000	4,381,583	7	13	14	Conclusion 4
nadR	IS150 insertion	4,000	4,581,547			7	Conclusion 4
Total number of mutation in the fluctuating environment						**254**	

[a] The gene or position in which the mutation occurred.

[b] The type of mutation that occurred within that gene or operon.

[b] The generation in which the mutation was identified.

[c] The location in which the mutation was identified.

[d] Position in which the mutation occurred. The coordinates are given are of the ancestral file.

[e] The treatment under which the mutation was reported. If the mutation was reported in either aerobic or anaerobic lineages these events were also reported.

[f] The conclusion category the mutation falls into based on the outcomes of questions outlined in **Figure 4.2**.

Table 7.9: List of all common mutations arising in the more than one environment.

Locus [a]	Mutation [b]	Generation [c]	Position [d]	Treatment of occurrence [e]		
				AE	AN	FL
hokC/nhaA	IS150 insertion	4,000	16,972	2	2	2
yaiT	IS150 insertion	2,000 & 4,000	360,203	1		2
brnQ	IS150 insertion	2,000 & 4,000	388,275	1		4
cheB	IS150 insertion	2,000 & 4,000	785,037	1		2
ynjI	IS150 insertion	4,000	910,345		3	1
ydiU	IS150 insertion	2,000 & 4,000	963,716	3		1
trg/mokB	IS150 insertion	4,000	1,272,468	2	9	6
zntB	Insertion	2,000 & 4000	1,328,521	1		3
yccC	IS150 insertion	2,000 & 4000	1,675,380	1		2
pflB	IS150 deletion	4,000	1,764,886		2	3
kduD	IS150 insertion	2,000	2,850,178		1	1
kpsM	IS150 insertion	2,000	3,000,519	1		1

rhsA	IS150 insertion	4,000	3,669,960		3	2
trkD/insJ-5	Insertion	2,000 & 4,000	3,866,357	4	11	
yihS	IS150 insertion	4,000	4,015,454	2		1
dcuR/yjdI	Deletion	2,000	4,295,377		6	6
cycA	IS150 insertion	2,000 & 4,000	4,381,583	7	13	14
Common mutations				**26**	**50**	**51**

[a] The gene or position in which the mutation occurred.

[b] The mutation type that occurred.

[c] The generation in which the mutation was identified.

[d] The location in which the mutation was identified. Coordinates are given from the ancestral file.

[e] The treatment under which the mutation was reported. If identical mutations were reported at the same position the number of mutations occurring at the position was indicated.

Table 7.10: Synonymous SNP mutations arising in different treatments and generations.

Gene [a]	Base change [b]	Position [c]	Amino acid [d]	Lineage [e]
ybiI	T → C	1,905,307	Tyr (74) → Tyr	AE-4K-1
lpxC	A → G	109,384	Lys (8) → Lys	AE-2K-2
ybiI	T → C	1,905,307	Tyr (74) → Tyr	AE-4K-2
ybiI	T → C	1,905,307	Tyr (74) → Tyr	AE-4K-3
ybiI	T → C	1,905,307	Tyr (74) → Tyr	AE-4K-4
rnfD	C → T	1,046,106	Ala (278) → Ala	AE-4K-5
metG	G → A	2,119,116	Lys (266) → Lys	AE-4K-5
yceF	A → G	1,573,697	Arg (47) → Arg	AE-4K-7
ybiI	T → C	1,905,307	Tyr (74) → Tyr	AE-4K-7
galR	T → A	2,844,146	Ala (35) → Ala	AN-2K-2
ydaM	A → T	1,329,917	Ser (54) → Ser	FL-2K-3
yieG	G → A	3,829,404	Lys (125) → Lys	FL-2K-3
ydaO	A → T	1,325,632	Lys (10) → Lys	FL-4K-6

[a] The gene in which the synonymous mutation arises.

[b] The base change that occurred as a result of the mutation.

[c] The position at which the mutation occurred (bps).

[d] The amino acid residue and position in polypeptide chain that is substituted.

[e] The lineage in which the synonymous mutation occurred.

Table 7.11: Evidence of identical mutations in the 4,000 generation anaerobic lineages.

Mutation [a]	Position [b]	Lineage [c]						
		AN-4K-1	AN-4K-2	AN-4K-3	AN-4K-4	AN-4K-5	AN-4K-6	AN-4K-7
ileS	26,953							
pcnB	161,094							
ybcK	542,005							
ybcQ/insB-6	546,599							
[insB-6]	546,981							
	546,986							
ybaM	462,234							
proQ	838,221							
ynfN	1,123,058							
ydfH	1,137,187							
[ydcA]	1,188,988							
trg/mokB	1,272,468							
dbpA	1,328,413							
adhE	1,439,673							
cydA/ybgG	1,984,267							
sucA	1,995,122							
[glnU]	2,057,176							
	2,057,181							
	2,057,184							
	2,057,185							
	2,057,189							
alaW/yfeC	2,424,083							
yfjF	2,648,318							
flu	2,972,604							
	2,972,858							
	2,974,237							
yrbI	3,250,708							
arcB	3,260,806							
rhsB	3,524,265							
rng	3,298,183							
glgC	3,473,571							
rhsA	3,669,960							
rhsB	3,522,963							
trkD/insJ-5	3,866,357							
yieP	3,869,168							
cycA	4,381,583							
mpl	4,406,085							
nadR	4,581,546							

[a] The gene or position in which the mutation occurred. If the mutation occurred in between genes, the upstream and downstream genes are given, separated by a "/".
[b] The exact position in which the mutation occurs or begins – in the case of insertions or deletions. All coordinates are given on the reference ancestral strain.
[c] The seven anaerobic lineages at 4,000 generations are indicated. If the mutation from a) and b) was present within the lineage the field was indicated by filled blue colour.

Table 7.12: Online mutations between 2,000 and 4,000 genomic data.

No.[a]	Treatment[b]	Lineages[c]	Gene affected[d]	Mutation Type[e]	Mutation class[f]	Position[g]	Size[h]	Change[i]
1	Anaerobic	AN2 and 7	*adhE*	SNP	Transition	1,439,673	1	G → A
2	Anaerobic	AN7	*arcB*	Deletion	Small	3,260,806	42	-
3	Anaerobic	AN1, AN3, AN4, AN5, AN6, AN7	*cycA*	Mobile element	Insertion	4,381,583	1,472	-
4	Anaerobic	AN7	*flu*	SNP	Transversion	2,972,858	1	T → G
5	Anaerobic	AN7	*glgC*	Deletion	Small	3,473,571	1	-C
6	Anaerobic	AN7	*insB-6*	Deletion	Large	546,986 – 547,702	32,570 - 33,286	30 genes
7	Anaerobic	AN7	*pcnB*	Deletion	Small	161,094	2	-
8	Anaerobic	AN7	*rng*	Insertion	Single base	3,298,183	1	+C
9	Anaerobic	AN1, AN4, AN7	*trg/mokB*	Mobile element	Insertion	1,272,468	1,446	-
10	Anaerobic	AN5, AN6, AN7	*trkD/insJ-5*	Insertion	Small	3,866,357	1	+G
11	Anaerobic	AN7	*ynfN*	Mobile element	Insertion	1,123,058	1,264	-
12	Fluctuator	FL2	*pflB*	Mobile element	Deletion	1,764,888	1446	-
13	Fluctuator	FL1, FL2, FL3, FL4, FL5, FL6, FL7	*cycA*	Mobile element	Insertion	4,381,583	1472	-

[a] Number of online mutation.

[b] Treatment under which the online mutation arose.

[c] Lineages in which online mutations arose.

[d] Gene(s) affected.

[e] The type of mutation.

[f] The classification of mutations within each mutation type

[g] the position the mutation occurred. All coordinates are given on the reference ancestral strain.

[h] the size of the mutation.

[I] The change that occurred in the evolved genome as a result of the mutation.

Table 7.13: List of genes or operons that have acquired multiple mutations among lineages in the three conditions of study.

Gene/ Operon [a]	Mutation [b]	Position [c]	Treatment of occurrence [d]			Support [e]
			AE	AN	FL	
hokC/nhaA	IS150 insertion	16972	2	2	2	Shewaramani, unpublished Barrick 2009
	IS150 insertion	16989			2	
	IS150 insertion	16992			1	
pcnB	Deletion	161094		8		Barrick 2009
	SNP	161770		1		
yaiT	IS150 insertion	360203	1	1	1	
[araJ] – [brnQ]	6 gene deletion	380365			1	
brnQ	Insertion	388020			1	
	IS150 insertion	388234	5			
	IS150 insertion	388275	1		2	
	IS150 insertion	388527			5	
	IS150 insertion	388530			1	
	IS150 insertion	388543			1	
	SNP	389105	1			
ybaM	Insertion	462233		1		Barrick 2009
	Duplication	462234		1		
aes	IS150 insertion	471783			6	Barrick 2009
ECB_00516	IS3 insertion	550063	4			
insA-7/ hokE	IS150 insertion	582237			1	
citG	IS150 insertion	621252		3		Barrick 2009
[ECB_02013] – ogrK	27 gene deletion	632497			1	
[ECB_00851] – [ECB_00814]	38 gene deletion	1831485			5	
[ECB_00851] – [ECB_00815]	37 gene deletion	1831485			1	
yeeJ	IS150 insertion	743346			5	Puentes Tellez 2013 (AE)
cheB	IS150 insertion	785037	1	1	1	Puentes Tellez 2013 (All)
proQ	Deletion of 4 bases	838221		5		
fadD	SNP	863948	1			
yeaS	IS150 insertion	872829		1		
yeaA	Deletion of 1 base	892201	2			
ynjI	IS150 insertion	910345		3	1	Barrick 2009 Shewaramani, unpublished
btuC	IS150 insertion	959668			1	Puentes Tellez 2013 (AE) Shewaramani, unpublished
	IS150 insertion	959872			1	
btuE	IS150 insertion	960637			2	
ydiV	IS1 insertion	932998			1	Puentes Tellez 2013 (All) Shewaramani,
[ydiV] – [ydiU]	2 gene deletion	963078			1	

ydiU	IS150 insertion	963716	3		1	unpublished
ydiU	IS150 insertion	963731			1	
ydiQ	IS150 insertion	974185			8	
[tqsA] – ECB_01533	36 gene deletion	1081310			1	
[ydfX] – ECB_01533	5 gene deletion	1110292			1	
[ECB_01536] – insE-3	25 gene deletion	1111336	1			
ECB_01533/ hokD	Deletion	1113343			1	
	IS150 insertion	1113403			5	
ynfN	IS150 insertion	1123058		5		
yddA	IS150 insertion	1181538			1	
	IS4 insertion	1181682			1	
trg/mokB	Deletion	1272262	1			Shewaramani, unpublished Barrick 2009 Puentes-Tellez 20012
	IS150 insertion	1272399	1			
	IS150 insertion	1272400			4	
	Deletion	1272401			4	
	IS150 insertion	1272468	2	9	6	
dbpA	Duplication	1328413		6		
[zntB]	Deletion	1328493			1	Puentes Tellez 2013 (AN) Shewaramani, unpublished
zntB	Insertion	1328521	1		3	
adhE	SNP	1438030		4		Puentes Tellez 2013 (All) Shewaramani, unpublished
	SNP	1439673		10		
chaA/ldrC	IS150 insertion	1464061			6	Shewaramani, unpublished Barrick 2009
	IS150 insertion	1464143	1			
[ldrC] – ldrB	2 gene deletion	1464062			1	
[ldrC]	Deletion	1464062			1	
ldrC/ldrB	IS150 insertion	1464595			1	
	IS150 insertion	1464672			1	
	IS150 insertion	1464678			1	
	IS150 insertion	1464679	1			
ldrB/ldrA	IS150 insertion	1465130	1			
ycfU	SNP	1544946			1	
ycfQ	IS150 insertion	1551946	1		1	
	IS150 insertion	1551960			1	
flgL	IS150 insertion	1579629			6	Puentes Tellez 2013 (AE)
flgK	IS150 insertion	1580827			1	
dinI	IS150 insertion	1598688			1	
	IS150 insertion	1598705			1	
yccC	IS150 insertion	1675380	1		2	
pflB	IS150 deletion	1764886	1	2	3	Shewaramani, unpublished Barrick 2009
	IS150 deletion	1764888			6	
	3 base deletion	1766329			1	
	5 base deletion	1766329			2	
	IS150 insertion	1766334			2	

[ECB_00851] – [ECB_00814]	38 gene deletion	1831485			5	
[ECB_00851] – [ECB_00815]	37 gene deletion	1831485			1	
ybiT	IS150 insertion	1886677			1	
ybiI	SNP	1905307	5			
cydA/ybgG	IS1 insertion	1984267		4		Puentes Tellez 2013 (AE AN)
sucA	Duplication	1995122		4		
rhsC	IS150 insertion	2024210			1	Puentes Tellez 2013 (AE)
	IS150 insertion	2024212		1		
glnU	2 gene deletion	2057167		1		
	101 base deletion	2057176		1		
	90 base deletion	2057181		1		
	93 base deletion	2057184		1		
	83 base deletion	2057185		1		
	2 gene deletion	2057189		1		
insJ-2/rihA	Deletion	2071478			2	
	IS150 insertion	2071483	1			
yehI	IS150 insertion	2127047	1			Puentes Tellez 2013 (AE)
yehM	IS150 insertion	2130703			1	
yehU	IS150 insertion	2138008	1		1	
mglB	IS150 insertion	2164019		2		Puentes Tellez 2013 (AN)
argT	IS150 insertion	2343551			2	Puentes Tellez 2013 (All)
yfcB	SNP	2364735			1	Puentes Tellez 2013 (AE AN)
fadL	IS150 insertion	2378861			2	
	IS150 insertion	2378921	1			
emrY	IS3 insertion	2387885		2		Barrick 2009
emrR	SNP	2678935			1	
alaW/yfeC	IS150 insertion	2424083		4		Puentes Tellez 2013 (AE)
hcaC	IS150 insertion	2564470		2		Puentes Tellez 2013 (AN) Barrick 2009
[rseB]/[rseA]	Duplication	2603060		2		Puentes Tellez 2013 (AE)
yfiQ	IS150 insertion	2615526			1	Puentes Tellez 2013 (All)
yfiH	IS150 insertion	2628603			1	
ECB_02512	IS150 insertion	2654436			1	
	IS1 insertion	2654657			1	
insJ-3/cysH	Deletion	2748700	1		1	Puentes Tellez 2013 (AN)
galR	SNP	2844146		1		Puentes Tellez 2013 (All)
	Duplication	2844904	5			
kduD	IS150 insertion	2850135	1		1	

	IS150 insertion	2850178		1	1	
yqeB	IS150 insertion	2872437		1	2	Barrick 2009
flu	Duplication	2972604		1		
	SNP	2972858		8		
	IS1 insertion	2972936	2			
	SNP	2973574		1		
	Deletion	2973807		1		
	SNP	2974237		6		
	IS150 insertion	2974776			1	
	IS150 insertion	2974778	2			
	IS150 insertion	2975127	1			
ECB_02804	IS150 insertion	2980273		2		
kpsE	SNP	2987334	1			
kpsD	SNP	2988653	2			
kpsS	IS1 insertion	2992382		5		
kpsT	Deletion	2999898	1			
	SNP	3000095	2			
	SNP	3000161	1			
	SNP	3000346	2			
kpsM	IS3 insertion	3000514	1			
	IS3 insertion	3000519	1		1	
	IS3 insertion	3000527	5			
glcF	IS150 insertion	3024559			1	
glcD	Deletion	3027366			1	
yqiI	IS1 insertion	3096895		2		Barrick 2009
rpoD	SNP	3119112	1			Puentes Tellez 2013 (All) Barrick 2009
	SNP	3119520	1			
rpoS	SNP	2734340	1			
yhaO/tdcG	SNP	3153949			6	Barrick 2009 Shewaramani, unpublished
	SNP	3153950			1	
	SNP	3164090	1		1	
yrbI	IS150 insertion	3250708		5		
arcB	Deletion	3260770		1		Puentes Tellez 2013 (AE FL) Barrick 2009
	Deletion	3260806		7		
rng	Insertion	3298183		8		
glgC	Deletion	3473571		8		
ugpB	IS150 insertion	3495511		2		
yhiO	Deletion	3543375			2	Puentes Tellez 2013 (All)
yhjE	IS150 insertion	3577565			1	Puentes Tellez 2013 (AE)
yhjT	IS150 insertion	3600373			1	
ldrF	IS150 insertion	3603153			1	
ldrE	IS150 insertion	3602958			3	
rhsA	IS150 insertion	3669960		3	2	Puentes Tellez 2013 (AE)
trkD/insJ-5	Insertion	3866357	4	11		Puentes Tellez 2013 (AE FL) Shewaramani, unpublished
	Deletion	3866358	2			
yihS	IS150 insertion	4015454	2		1	Puentes Tellez 2013 (AE)

rhaA	IS150 insertion	4040290			1	
rhaS	IS150 insertion	4043794			1	
alr/tyrB	IS150 insertion	4211817		2		
dcuR/yjdI	Deletion	4295377		6	6	Shewaramani, unpublished Puentes Tellez 2013 (AE)
cycA	IS150 insertion	4381583	7	13	14	Shewaramani, unpublished
mpl	Insertion (+9)	4406085		2		Shewaramani, unpublished
	Insertion (+7)	4406085		2		
yjiX	IS150 insertion	4532958			1	Shewaramani, unpublished
yjiI - yjjN	16 gene deletion	4532961	1			
yjiY – [mdoB]	16 gene deletion	4532961			1	
[insA-28] – mdoB]	27 gene deletion	4516617			1	
[yjiY] – [hpaC]	2 gene deletion	4533440			1	
yjiY	IS150 insertion	4533508			1	
	IS150 insertion	4534750			1	
mdoB	IS150 insertion	4552611			1	
	IS150 insertion	4554029			1	
nadR	IS150 insertion	4581545		5		Puentes Tellez 2013 (All) Barrick 2009
	IS150 insertion	4581546		4		
	IS150 insertion	4581547			7	
	IS150 insertion	4581549		1		
alaT – [rrlA]	2 gene deletion	3981502	1			
[alaT] – rrfA	3 gene deletion	3981508			1	
[insB-6] - ybdK	30 gene deletion	546975		7		Shewaramani, unpublished
[insB-6] – [ompY]	8 genes	546986			1	
[insB-6] – [insA-7]	33 gene deletion	546990		1		

[a] The gene or operon mutated. If a mutation was intergenic, the nearest genes upstream and downstream were indicated with a "/" between them. If the gene was an insertion or a deletion of a range of bases or genes a "-" was indicated between the upstream or downstream of the mutation.

[b] The type of mutation that occurred within that gene or operon

[c] the position at which the mutations occurred.

[d] This is divided into the three treatments under which the mutations were reported (AE – aerobic, AN - anaerobic and FL – fluctuator). The number of times this identical mutation occurred is indicated.

[e] If mutations were reported in three other studies (Shewaramani 2015, (52, 101) this was indicated. Furthermore the treatment under which the mutation was found from Puentes Tellez was also indicated (AE – aerobic, AN - anaerobic, FL – fluctuator, All – all treatments).

Chapter Eight : References

1. Lowe SE, Jain MK, & Zeikus JG (1993) Biology, ecology, and biotechnological applications of anaerobic bacteria adapted to environmental stresses in temperature, pH, salinity, or substrates. *Microbiological reviews* 57(2):451.
2. Kasting JF & Siefert JL (2002) Life and the evolution of Earth's atmosphere. *Science* 296(5570):1066-1068.
3. Berg RD (1996) The indigenous gastrointestinal microflora. *Trends in microbiology* 4(11):430-435.
4. Haagsma J (1991) Pathogenic anaerobic bacteria and the environment. *Rev Sci Tech* 10(3):749-764.
5. Miraglia G (1974) Pathogenic anaerobic bacteria. *Critical Reviews in Microbiology* 3(2):161-181.
6. Lenski RE, Rose MR, Simpson SC, & Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *American Naturalist* 138:1315-1341.
7. Ingledew W & Poole R (1984) The respiratory chains of *Escherichia coli*. *Microbiology and Molecular Biology Reviews* 48(3):222.
8. Pascal M-C, Chippaux M, Abou-Jaoude A, Blaschkowski H, & Knappe J (1981) Mutants of *Escherichia coli* K12 with defects in anaerobic pyruvate metabolism. *Journal of general microbiology* 124(1):35-42.
9. Smith MW & Neidhardt FC (1983) Proteins induced by anaerobiosis in *Escherichia coli*. *Journal of bacteriology* 154(1):336-343.
10. Thomas AD, Doelle H, Westwood A, & Gordon G (1972) Effect of oxygen on several enzymes involved in the aerobic and anaerobic utilization of glucose in *Escherichia coli*. *Journal of bacteriology* 112(3):1099-1105.
11. Haddock BA & Schairer HU (1973) Electron transport chains of *Escherichia coli*. *European Journal of Biochemistry* 35(1):34-45.
12. Hansen R & Henning U (1966) Regulation of pyruvate dehydrogenase activity in *Escherichia coli* K12. *Biochimica et Biophysica Acta (BBA)-Enzymology and Biological Oxidation* 122(2):355-358.
13. Clark DP (1989) The fermentation pathways of *Escherichia coli*. *FEMS Microbiology Letters* 63(3):223-234.
14. Gunsalus RP & Park S-J (1994) Aerobic-anaerobic gene regulation in *Escherichia coli*: control by the ArcAB and Fnr regulons. *Research in Microbiology* 145(5):437-450.
15. Iuchi S & Weiner L (1996) Cellular and molecular physiology of *Escherichia coli* in the adaptation to aerobic environments. *Journal of biochemistry* 120(6):1055-1063.
16. Jovanovic G, Sheng X, Ale A, Feliu E, Harrington HA, Kirk P, Wiuf C, Buck M, & Stumpf MP (2015) Phosphorelay of non-orthodox two component systems functions through a bi-molecular mechanism *in vivo*: the case of ArcB. *Molecular BioSystems* 11(5):1348-1359.
17. Sawers G (1999) The aerobic/anaerobic interface. *Current opinion in microbiology* 2(2):181-187.
18. Henkel SG, Ter Beek A, Steinsiek S, Stagge S, Bettenbrock K, de Mattos MJT, Sauter T, Sawodny O, & Ederer M (2014) Basic regulatory principles of *Escherichia coli*'s electron transport chain for varying oxygen conditions. *PloS one* 9(9):e107640.

19. Denby KJ, Rolfe MD, Crick E, Sanguinetti G, Poole RK, & Green J (2015) Adaptation of anaerobic cultures of *Escherichia coli* K-12 in response to environmental trimethylamine-N-oxide. *Environmental microbiology*.
20. Jiang F, An C, Bao Y, Zhao X, Jernigan RL, Lithio A, Nettleton D, Li L, Wurtele ES, & Nolan LK (2015) ArcA controls metabolism, chemotaxis, and motility contributing to the pathogenicity of vvian pathogenic *Escherichia coli*. *Infection and immunity* 83(9):3545-3554.
21. Unden G, Becker S, Bongaerts J, Holighaus G, Schirawski J, & Six S (1995) O2-sensing and O2-dependent gene regulation in facultatively anaerobic bacteria. *Archives of microbiology* 164(2):81-90.
22. Bekker M, De Mattos M, & Hellingwerf KJ (2006) The role of two-component regulation systems in the physiology of the bacterial cell. *Science progress* 89(3-4):213-242.
23. Ingledew W & Poole R (1984) The respiratory chains of *Escherichia coli*. *Microbiological reviews* 48(3):222.
24. Lazazzera BA, Beinert H, Khoroshilova N, Kennedy MC, & Kiley PJ (1996) DNA binding and dimerization of the Fe− S-containing FNR protein from *Escherichia coli* are regulated by oxygen. *Journal of Biological Chemistry* 271(5):2762-2768.
25. Cabiscol E, Tamarit J, & Ros J (2010) Oxidative stress in bacteria and protein damage by reactive oxygen species. *International Microbiology* 3(1):3-8.
26. Farr SB & Kogoma T (1991) Oxidative stress responses in *Escherichia coli* and *Salmonella typhimurium*. *Microbiological reviews* 55(4):561-585.
27. Imlay JA (2003) Pathways of oxidative damage. *Annual Reviews in Microbiology* 57(1):395-418.
28. Carlioz A & Touati D (1986) Isolation of superoxide dismutase mutants in *Escherichia coli*: is superoxide dismutase necessary for aerobic life? *The EMBO journal* 5(3):623.
29. Benov L & Fridovich I (1996) The rate of adaptive mutagenesis in *Escherichia coli* is enhanced by oxygen (superoxide). *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 357(1):231-236.
30. Maki H (2002) Origins of spontaneous mutations: specificity and directionality of base-substitution, frameshift, and sequence-substitution mutageneses. *Annual Review of Genetics* 36(1):279-303.
31. Bjelland S & Seeberg E (2003) Mutagenicity, toxicity and repair of DNA base damage induced by oxidation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 531(1):37-80.
32. Sakai A, Nakanishi M, Yoshiyama K, & Maki H (2006) Impact of reactive oxygen species on spontaneous mutagenesis in *Escherichia coli*. *Genes to Cells* 11(7):767-778.
33. Lawrence JG & Roth JR (1996) Selfish operons: horizontal transfer may drive the evolution of gene clusters. *Genetics* 143(4):1843-1860.
34. Felsenstein J (1974) The evolutionary advantage of recombination. *Genetics* 78(2):737-756.
35. Dykhuizen DE & Green L (1991) Recombination in *Escherichia coli* and the definition of biological species. *Journal of bacteriology* 173(22):7257-7268.
36. Slatkin M (1985) Gene flow in natural populations. *Annual Review of Ecology and Systematics*:393-430.
37. Friedberg EC, McDaniel LD, & Schultz RA (2004) The role of endogenous and exogenous DNA damage and mutagenesis. *Current opinion in genetics & development* 14(1):5-10.

38. Morley A & Turner D (1999) The contribution of exogenous and endogenous mutagens to in vivo mutations. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 428(1):11-15.
39. Mahillon J & Chandler M (1998) Insertion sequences. *Microbiology and Molecular Biology Reviews* 62(3):725-774.
40. Siguier P, Pérochon J, Lestrade L, Mahillon J, & Chandler M (2006) IS finder: the reference centre for bacterial insertion sequences. *Nucleic acids research* 34(suppl 1):D32-D36.
41. Saedler H, Reif H, Hu S, & Davidson N (1974) IS2, a genetic element for turn-off and turn-on of gene activity in *Escherichia coli*. *Molecular and General Genetics* 132(4):265-289.
42. Ciampi MS, Schmid MB, & Roth JR (1982) Transposon Tn10 provides a promoter for transcription of adjacent sequences. *Proceedings of the National Academy of Sciences* 79(16):5016-5020.
43. Reynolds AE, Felton J, & Wright A (1981) Insertion of DNA activates the cryptic *bgl* operon in *Escherichia coli* K12. *Nature* 293:625-629.
44. Saedler H, Cornelis G, Cullum J, Schumacher B, & Sommer H (1981) IS1 mediated DNA rearrangements. *Cold Spring Harbor symposia on quantitative biology*, (Cold Spring Harbor Laboratory Press), pp 93-98.
45. Treves DS, Manning S, & Adams J (1998) Repeated evolution of an acetate crossfeeding polymorphism in long-term populations of *Escherichia coli*. *Molecular Biology and Evolution* 15(7):789-797.
46. Naas T, Blot M, Fitch W, & Arber W (1994) Insertion sequence-related genetic variation in resting *Escherichia coli* K-12. *Genetics* 136(3):721-730.
47. McGraw JE & Brookfield JF (2006) The interaction between mobile DNAs and their hosts in a fluctuating environment. *Journal of Theoretical Biology* 243(1):13-23.
48. Casacuberta E & González J (2013) The impact of transposable elements in environmental adaptation. *Molecular Ecology* 22(6):1503-1517.
49. Reif H & Saedler H (1975) IS1 is involved in deletion formation in the *gal* region of *Escherichia coli* K12. *Molecular and General Genetics* 137(1):17-28.
50. Louarn J, Bouche J, Legendre F, Louarn J, & Patte J (1985) Characterization and properties of very large inversions of the *Escherichia coli* chromosome along the origin-to-terminus axis. *Molecular and General Genetics* 201(3):467-476.
51. Schneider D, Duperchy E, Coursange E, Lenski RE, & Blot M (2000) Long-term experimental evolution in *Escherichia coli* . IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156(2):477-488.
52. Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, Schneider D, Lenski RE, & Kim JF (2009) Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461(7268):1243-1247.
53. Studier FW, Daegelen P, Lenski RE, Maslov S, & Kim JF (2009) Understanding the differences between genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3) and comparison of the *Escherichia coli* B and K-12 genomes. *Journal of Molecular Biology* 394(4):653-680.
54. Basu AK, Loechler EL, Leadon SA, & Essigmann JM (1989) Genetic effects of thymine glycol: site-specific mutagenesis and molecular modeling studies. *Proceedings of the National Academy of Sciences* 86(20):7677-7681.
55. Orr HA (2005) The genetic theory of adaptation: a brief history. *Nature Reviews Genetics* 6(2):119-127.

56. Sniegowski PD & Gerrish PJ (2010) Beneficial mutations and the dynamics of adaptation in asexual populations. *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1544):1255-1263.
57. Behe MJ (2010) Experimental evolution, loss-of-function mutations, and "the first rule of adaptive evolution". *The Quarterly review of biology* 85(4):419-445.
58. Kinnersley MA, Holben WE, & Rosenzweig F (2009) *E Unibus Plurum*: genomic analysis of an experimentally evolved polymorphism in *Escherichia coli*. *PLoS Genetics* 5(11):e1000713.
59. Carroll SB (2005) Evolution at two levels: on genes and form. *PLoS biology* 3(7):e245.
60. Wray GA (2007) The evolutionary significance of *cis*-regulatory mutations. *Nature Reviews Genetics* 8(3):206-216.
61. Barrick JE & Lenski RE (2013) Genome dynamics during experimental evolution. *Nature Reviews Genetics* 14(12):827-839.
62. Blount ZD, Borland CZ, & Lenski RE (2008) Historical contingency and the evolution of a key innovation in an experimental population of *Escherichia coli*. *Proceedings of the National Academy of Sciences* 105(23):7899-7906.
63. Kassen R (2014) *Experimental evolution and the nature of biodiversity* (Roberts and Company, Greenwood Village, Colorado) p 277.
64. Kimura M (1968) Evolutionary rate at the molecular level. *Nature* 217(5129):624-626.
65. Good BH & Desai MM (2014) Deleterious passengers in adapting populations. *arXiv preprint arXiv:1405.3902*.
66. Schiffels S, Szöllősi GJ, Mustonen V, & Lässig M (2011) Emergent neutrality in adaptive asexual evolution. *Genetics* 189(4):1361-1375.
67. Lewontin RC (1970) The units of selection. *Annual Review of Ecology and Systematics*:1-18.
68. Brandon RN (1978) Adaptation and evolutionary theory. *Studies in History and Philosophy of Science Part A* 9(3):181-206.
69. Rainey PB, Buckling A, Kassen R, & Travisano M (2000) The emergence and maintenance of diversity: insights from experimental bacterial populations. *Trends in ecology & evolution* 15(6):243-247.
70. Buckling A, Kassen R, Bell G, & Rainey PB (2000) Disturbance and diversity in experimental microcosms. *Nature* 408(6815):961-964.
71. Rozen DE & Lenski RE (2000) Long-term experimental evolution in *Escherichia coli*. VIII. Dynamics of a balanced polymorphism. *The American Naturalist* 155(1):24-35.
72. Spiers AJ, Kahn SG, Bohannon J, Travisano M, & Rainey PB (2002) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. I. Genetic and phenotypic bases of wrinkly spreader fitness. *Genetics* 161(1):33-46.
73. Rosenzweig RF, Sharp R, Treves DS, & Adams J (1994) Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* 137(4):903.
74. Spencer CC, Bertrand M, Travisano M, & Doebeli M (2007) Adaptive diversification in genes that regulate resource use in *Escherichia coli*. *PLoS genetics* 3(1):e15.
75. Rainey PB & Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Nature* 394(6688):69-72.
76. Orr HA (2009) Fitness and its role in evolutionary genetics. *Nature Reviews Genetics* 10(8):531-539.
77. Orr HA (2003) The distribution of fitness effects among beneficial mutations. *Genetics* 163(4):1519-1526.
78. Benveniste R & Davies J (1973) Mechanisms of antibiotic resistance in bacteria. *Annual review of biochemistry* 42(1):471-506.

79. Garland T & Rose MR (2009) *Experimental evolution: concepts, methods, and applications of selection experiments* (Univ of California Press).
80. Wright S (1932) *The roles of mutation, inbreeding, crossbreeding and selection in evolution* (na).
81. Melnyk AH & Kassen R (2011) Adaptive landscapes in evolving populations of *Pseudomonas fluorescens*. *Evolution* 65(11):3048-3059.
82. Elena SF & Lenski RE (2003) Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nature Reviews Genetics* 4(6):457-469.
83. Lenski RE & Travisano M (1994) Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *Proceedings of the National Academy of Sciences* 91(15):6808-6814.
84. Barrick JE, Kauth MR, Strelioff CC, & Lenski RE (2010) Escherichia coli rpoB mutants have increased evolvability in proportion to their fitness defects. *Molecular Biology and evolution* 27(6):1338-1347.
85. Peck JR (1994) A ruby in the rubbish: beneficial mutations, deleterious mutations and the evolution of sex. *Genetics* 137(2):597-606.
86. Lenski RE, Ofria C, Pennock RT, & Adami C (2003) The evolutionary origin of complex features. *Nature* 423(6936):139-144.
87. Cooper VS & Lenski RE (2000) The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Earth Planet Science Letters* 12:411-418.
88. Kassen R (2002) The experimental evolution of specialists, generalists, and the maintenance of diversity. *Journal of Evolutionary Biology* 15(2):173-190.
89. Fong DW, Kane TC, & Culver DC (1995) Vestigialization and loss of nonfunctional characters. *Annual Review of Ecology and Systematics*:249-268.
90. Zhang E & Ferenci T (1999) OmpF changes and the complexity of *Escherichia coli* adaptation to prolonged lactose limitation. *FEMS microbiology letters* 176(2):395-401.
91. Lynch M, Conery J, & Burger R (1995) Mutation accumulation and the extinction of small populations. *American Naturalist*:489-518.
92. Kassen R & Bell G (1998) Experimental evolution in *Chlamydomonas*. IV. Selection in environments that vary through time at different scales. *Heredity* 80(6):732-741.
93. Vasi FK & Lenski RE (1999) Ecological strategies and fitness tradeoffs in *Escherichia coli* mutants adapted to prolonged starvation. *Journal of Genetics* 78(1):43-49.
94. Cooper VS (2002) Long-term experimental evolution in *Escherichia coli*. X. Quantifying the fundamental and realized niche. *BMC evolutionary biology* 2(1):12.
95. Reboud X & Bell G (1997) Experimental evolution in *Chlamydomonas*. III. Evolution of specialist and generalist types in environments that vary in space and time. *Heredity* 78(5):507-514.
96. Van Tienderen PH (1997) Generalists, specialists, and the evolution of phenotypic plasticity in sympatric populations of distinct species. *Evolution*:1372-1380.
97. Bennett AF, Lenski RE, & Mittler JE (1992) Evolutionary adaptation to temperature. I. Fitness responses of *Escherichia coli* to changes in its thermal environment. *Evolution*:16-30.
98. Haldane JBS (1927) A mathematical theory of natural and artificial selection, part V: selection and mutation. *Mathematical Proceedings of the Cambridge Philosophical Society*, (Cambridge Univ Press), pp 838-844.
99. Conrad TM, Lewis NE, & Palsson BØ (2011) Microbial laboratory evolution in the era of genome-scale science. *Molecular Systems Biology* 7(1).

100. Lenski RE, Rose MR, Simpson SC, & Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2,000 generations. *American Naturalist*:1315-1341.
101. Puentes-Téllez PE, Hansen MA, Sørensen SJ, & van Elsas JD (2013) Adaptation and heterogeneity of *Escherichia coli* MC1000 growing in complex environments. *Applied and Environmental Microbiology* 79(3):1008-1017.
102. Lenski RE, Mongold JA, Sniegowski PD, Travisano M, Vasi F, Gerrish PJ, & Schmidt TM (1998) Evolution of competitive fitness in experimental populations of *Escherichia coli*: what makes one genotype a better competitor than another? *Antonie van Leeuwenhoek* 73(1):35-47.
103. de Visser JAG & Rozen DE (2006) Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics* 172(4):2093-2100.
104. Cooper TF, Rozen DE, & Lenski RE (2003) Parallel changes in gene expression after 20,000 generations of evolution in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 100(3):1072-1077.
105. Woods R, Schneider D, Winkworth CL, Riley MA, & Lenski RE (2006) Tests of parallel molecular evolution in a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences* 103(24):9107-9112.
106. Applebee MK, Herrgård MJ, & Palsson BØ (2008) Impact of individual mutations on increased fitness in adaptively evolved strains of *Escherichia coli Journal of Bacteriology* 190(14):5087-5094.
107. Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, & Palsson BØ (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):R118.
108. Plucain J, Hindré T, Le Gac M, Tenaillon O, Cruveiller S, Médigue C, Leiby N, Harcombe WR, Marx CJ, & Lenski RE (2014) Epistasis and allele specificity in the emergence of a stable polymorphism in *Escherichia coli*. *Science* 343(6177):1366-1369.
109. Crozat E, Philippe N, Lenski RE, Geiselmann J, & Schneider D (2005) Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169(2):523-532.
110. Stanek MT, Cooper TF, & Lenski RE (2009) Identification and dynamics of a beneficial mutation in a long-term evolution experiment with *Escherichia coli*. *BMC evolutionary biology* 9(1):302.
111. Cooper VS, Schneider D, Blot M, & Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *Journal of Bacteriology* 183(9):2834-2841.
112. Pelosi L, Kühn L, Guetta D, Garin J, Geiselmann J, Lenski RE, & Schneider D (2006) Parallel changes in global protein profiles during long-term experimental evolution in *Escherichia coli*. *Genetics* 173(4):1851-1869.
113. Champoux JJ (2001) DNA topoisomerases: structure, function, and mechanism. *Annual Review of Biochemistry* 70(1):369-413.
114. Gellert M, Mizuuchi K, O'Dea MH, & Nash HA (1976) DNA gyrase: an enzyme that introduces superhelical turns into DNA. *Proceedings of the National Academy of Sciences* 73(11):3872-3876.
115. Wendrich TM, Blaha G, Wilson DN, Marahiel MA, & Nierhaus KH (2002) Dissection of the mechanism for the stringent factor RelA. *Molecular cell* 10(4):779-788.

116. Mongold JA (2000) Cell size, shape, and fitness in evolving populations of bacteria. *Scaling in Biology*:221.
117. Reeves PR, Hobbs M, Valvano MA, Skurnik M, Whitfield C, Coplin D, Kido N, Klena J, Maskell D, & Raetz CR (1996) Bacterial polysaccharide synthesis and gene nomenclature. *Trends in microbiology* 4(12):495-503.
118. Blount ZD, Barrick JE, Davidson CJ, & Lenski RE (2012) Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature* 489(7417):513-518.
119. Leiby N, Harcombe WR, & Marx CJ (2012) Multiple long-term, experimentally-evolved populations of *Escherichia coli* acquire dependence upon citrate as an iron chelator for optimal growth on glucose. *Bio Medical Central Evolutionary Biology* 12(1):151.
120. Zambrano MM & Kolter R (1996) GASPing for life in stationary phase. *Cell* 86(2):181-184.
121. Zinser ER, Schneider D, Blot M, & Kolter R (2003) Bacterial evolution through the selective loss of beneficial genes: trade-offs in expression involving two loci. *Genetics* 164(4):1271-1277.
122. Finkel SE, Zinser ER, & Kolter R (2000) Long-term survival and evolution in the stationary phase. *American Society for Microbiology*:231-238.
123. Zinser ER & Kolter R (2000) Prolonged Stationary-Phase Incubation Selects forlrp Mutations in *Escherichia coli* K-12. *Journal of bacteriology* 182(15):4361-4365.
124. Rozen DE, Philippe N, Arjan de Visser J, Lenski RE, & Schneider D (2009) Death and cannibalism in a seasonal environment facilitate bacterial coexistence. *Ecology Letters* 12(1):34-44.
125. Finkel SE (2006) Long-term survival during stationary phase: evolution and the GASP phenotype. *Nature Reviews Microbiology* 4(2):113-120.
126. Manché K, Notley-McRobb L, & Ferenci T (1999) Mutational adaptation of *Escherichia coli* to glucose limitation involves distinct evolutionary pathways in aerobic and oxygen-limited environments. *Genetics* 153(1):5-12.
127. Rainey PB & Travisano M (1998) Adaptive radiation in a heterogeneous environment. *Trans. R. Soc. B* 280:29-101.
128. Bantinaki E, Kassen R, Knight CG, Robinson Z, Spiers AJ, & Rainey PB (2007) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics* 176(1):441-453.
129. McDonald MJ, Gehrig SM, Meintjes PL, Zhang X-X, & Rainey PB (2009) Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics* 183(3):1041-1053.
130. Ferguson GC, Bertels F, & Rainey PB (2013) Adaptive divergence in experimental populations of *Pseudomonas fluorescens.* V. Insight into the niche specialist fuzzy spreader compels revision of the model *Pseudomonas* radiation. *Genetics* 195(4):1319-1335.
131. Laland KN, Odling-Smee FJ, & Feldman MW (1999) Evolutionary consequences of niche construction and their implications for ecology. *Proceedings of the National Academy of Sciences* 96(18):10242-10247.
132. Herron MD & Doebeli M (2013) Parallel evolutionary dynamics of adaptive diversification in *Escherichia coli*. *PLoS Biology* 11(2):e1001490.
133. Brockhurst MA, Colegrave N, & Rozen DE (2011) Next-generation sequencing as a tool to study microbial evolution. *Molecular Ecology* 20(5):972-980.
134. Bjedov I, Tenaillon O, Gerard B, Souza V, Denamur E, Radman M, Taddei F, & Matic I (2003) Stress-induced mutagenesis in bacteria. *Science* 300(5624):1404-1409.
135. Hegreness M & Kishony R (2007) Analysis of genetic systems using experimental evolution and whole-genome sequencing. *Genome Biol* 8(1):201.

136. Dettman JR, Rodrigue N, Melnyk AH, Wong A, Bailey SF, & Kassen R (2012) Evolutionary insight from whole-genome sequencing of experimentally evolved microbes. *Molecular Ecology* 21(9):2058-2077.
137. Gresham D, Ruderfer DM, Pratt SC, Schacherer J, Dunham MJ, Botstein D, & Kruglyak L (2006) Genome-wide detection of polymorphisms at nucleotide resolution with a single DNA microarray. *Science* 311(5769):1932-1936.
138. Gresham D, Curry B, Ward A, Gordon DB, Brizuela L, Kruglyak L, & Botstein D (2010) Optimized detection of sequence variation in heterozygous genomes using DNA microarrays with isothermal-melting probes. *Proceedings of the National Academy of Sciences* 107(4):1482-1487.
139. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, & Chen Z (2005) Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437(7057):376-380.
140. Bentley DR (2006) Whole-genome re-sequencing. *Current opinion in genetics & development* 16(6):545-552.
141. Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends in Genetics* 24(3):133-141.
142. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, & Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC genomics* 13(1):341.
143. Dicks E, Teague JW, Stephens P, Raine K, Yates A, Mattocks C, Tarpey P, Butler A, Menzies A, & Richardson D (2007) AutoCSA, an algorithm for high throughput DNA sequence variant detection in cancer genomes. *Bioinformatics* 23(13):1689-1691.
144. Deatherage DE & Barrick JE (2014) Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Engineering and Analyzing Multicellular Systems*, (Springer), pp 165-188.
145. Weckx S, Del-Favero J, Rademakers R, Claes L, Cruts M, De Jonghe P, Van Broeckhoven C, & De Rijk P (2005) novoSNP, a novel computational tool for sequence variation discovery. *Genome research* 15(3):436-442.
146. Zhang J, Wheeler DA, Yakub I, Wei S, Sood R, Rowe W, Liu PP, Gibbs RA, & Buetow KH (2005) SNPdetector: a software tool for sensitive and accurate SNP detection. *PLoS computational biology* 1(5):e53.
147. Sarovich DS & Price EP (2014) SPANDx: a genomics pipeline for comparative analysis of large haploid whole genome re-sequencing datasets. *BMC research notes* 7(1):618.
148. Hudson ME (2008) Sequencing breakthroughs for genomic ecology and evolutionary biology. *Molecular ecology resources* 8(1):3-17.
149. Raeside C, Gaffé J, Deatherage DE, Tenaillon O, Briska AM, Ptashkin RN, Cruveiller S, Médigue C, Lenski RE, & Barrick JE (2014) Large chromosomal rearrangements during a long-term evolution experiment with *Escherichia coli*. *mBio* 5(5):e01377-01314.
150. Medvedev P, Stanciu M, & Brudno M (2009) Computational methods for discovering structural variation with next-generation sequencing. *Nature methods* 6:S13-S20.
151. English AC, Richards S, Han Y, Wang M, Vee V, Qu J, Qin X, Muzny DM, Reid JG, & Worley KC (2012) Mind the gap: upgrading genomes with Pacific Biosciences RS long-read sequencing technology. *PloS one* 7(11):e47768.
152. Keightley PD, Trivedi U, Thomson M, Oliver F, Kumar S, & Blaxter M (2009) Analysis of the genome sequences of three *Drosophila melanogaster* spontaneous mutation accumulation lines. *Genome research*:1-38.

153. Denver DR, Dolan PC, Wilhelm LJ, Sung W, Lucas-Lledó JI, Howe DK, Lewis SC, Okamoto K, Thomas WK, & Lynch M (2009) A genome-wide view of Caenorhabditis elegans base-substitution mutation processes. *Proceedings of the National Academy of Sciences* 106(38):16310-16314.
154. Ossowski S, Schneeberger K, Lucas-Lledó JI, Warthmann N, Clark RM, Shaw RG, Weigel D, & Lynch M (2010) The rate and molecular spectrum of spontaneous mutations in Arabidopsis thaliana. *Science* 327(5961):92-94.
155. Riehle MM, Bennett AF, & Long AD (2001) Genetic architecture of thermal adaptation in *Escherichia coli*. *Proceedings of the National Academy of Sciences* 98(2):525-530.
156. Sambrook J, Fritsch EF, & Maniatis T (1989) *Molecular cloning* (Cold spring harbor laboratory press New York).
157. Andrews S (2010) FastQC: A quality control tool for high throughput sequence data. *Reference Source*.
158. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream M-A, & Barrell B (2000) Artemis: sequence visualization and annotation. *Bioinformatics* 16(10):944-945.
159. Rice P, Longden I, & Bleasby A (2000) EMBOSS: the European molecular biology open software suite. *Trends in genetics* 16(6):276-277.
160. Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, Bonavides-Martínez C, Fulcher C, Huerta AM, Kothari A, & Krummenacker M (2013) EcoCyc: fusing model organism databases with systems biology. *Nucleic Acids Research* 41(D1):D605-D612.
161. Huang DW, Sherman BT, & Lempicki RA (2008) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nature protocols* 4(1):44-57.
162. Altschul SF, Gish W, Miller W, Myers EW, & Lipman DJ (1990) Basic local alignment search tool. *Journal of Molecular Biology* 215(3):403-410.
163. Drummond A, Ashton B, Buxton S, Cheung M, Cooper A, Duran C, Field M, Heled J, Kearse M, & Markowitz S (2011) Geneious v5. 4.
164. Cooper VS, Schneider D, Blot M, & Lenski RE (2001) Mechanisms causing rapid and parallel losses of ribose catabolism in evolving populations of *Escherichia coli* B. *Journal of bacteriology* 183(9):2834.
165. Crozat E, Philippe N, Lenski RE, Geiselmann J, & Schneider D (2005) Long-term experimental evolution in *Escherichia coli*. XII. DNA topology as a key target of selection. *Genetics* 169(2):523.
166. Vasi F, Travisano M, & Lenski RE (1994) Long-term experimental evolution in *Escherichia coli*. II. Changes in life-history traits during adaptation to a seasonal environment. *American Naturalist*:432-456.
167. Lenski RE, Rose MR, Simpson SC, & Tadler SC (1991) Long-term experimental evolution in *Escherichia coli*. I. Adaptation and divergence during 2, 000 generations. *American Naturalist* 138(6):1315-1341.
168. Freshney RI (2006) Basic principles of cell culture. *John Wiley & Sons, Inc.*:3-21.
169. Kram KE & Finkel SE (2014) Culture volume and vessel affect long-term survival, mutation frequency, and oxidative stress of *Escherichia coli*. *Applied and Environmental Microbiology* 80(5):1732-1738.
170. Navarro Llorens JM, Tormo A, & Martínez-García E (2010) Stationary phase in Gram-negative bacteria. *FEMS microbiology reviews* 34(4):476-495.
171. Zinser ER & Kolter R (1999) Mutations enhancing amino acid catabolism confer a growth advantage in stationary phase. *Journal of Bacteriology* 181(18):5800-5807.

172. Pin C & Baranyi J (2008) Single-cell and population lag times as a function of cell age. *Applied and Environmental Microbiology* 74(8):2534-2536.
173. Tempest D, Norris JR, & Richmond M (1978) *Dynamics of microbial growth* (John Wiley and Sons).
174. Upadhyay J & Stokes J (1962) Anaerobic growth of psychrophilic bacteria. *Journal of Bacteriology* 83(2):270-275.
175. Bolhuis A, Mathers JE, Thomas JD, Barrett CM, & Robinson C (2001) TatB and TatC form a functional and structural unit of the twin-arginine translocase from *Escherichia coli*. *Journal of Biological Chemistry* 276(23):20213-20219.
176. Yuan Y, Crane DD, & Barry Cr (1996) Stationary phase-associated protein expression in Mycobacterium tuberculosis: function of the mycobacterial alpha-crystallin homolog. *Journal of Bacteriology* 178(15):4484-4492.
177. Gray C, Wimpenny J, & Mossman MR (1966) Regulation of metabolism in facultative bacteria II. Effects of aerobiosis, anaerbiosis and nutrition on the formation of Krebs cycle enzymes in *Escherchia coli*. *Biochimica et Biophysica Acta - General Subjects* 117(1):33-41.
178. Vaughn RH, Osborne JT, Wedding GT, Tabachnick J, Beisel CG, & Braxton T (1950) The utilization of citrate by *Escherichia coli*. *Journal of bacteriology* 60(2):119.
179. Lutgens M & Gottschalk G (1980) Why a co-substrate is required for anaerobic growth of *Escherichia coli* on citrate. *Journal of general Microbiology* 119(1):63-70.
180. Pos KM, Dimroth P, & Bott M (1998) The *Escherichia coli* citrate carrier CitT: a member of a novel eubacterial transporter family related to the 2-oxoglutarate/malate translocator from spinach chloroplasts. *Journal of Bacteriology* 180(16):4160-4165.
181. Thauer RK, Jungermann K, & Decker K (1977) Energy conservation in chemotrophic anaerobic bacteria. *Bacteriological Reviews* 41(1):100.
182. Hall BG (1982) Chromosomal mutation for citrate utilization by *Escherichia coli* K-12. *Journal of Bacteriology* 151(1):269-273.
183. Lara F & Stokes J (1952) Oxidation of citrate by *Escherichia coli Journal of Bacteriology* 63(3):415.
184. Bott M (1997) Anaerobic citrate metabolism and its regulation in enterobacteria. *Archives of Microbiology* 167(2-3):78-88.
185. Yamamoto K, Matsumoto F, Oshima T, Fujita N, Ogasawara N, & Ishihama A (2008) Anaerobic regulation of citrate fermentation by CitAB in *Escherichia coli*. *Bioscience, biotechnology, and biochemistry* 72(11):3011-3014.
186. Lenski RE (1988) Experimental studies of pleiotropy and epistasis in *Escherichia coli*. I. Variation in competitive fitness among mutants resistant to virus T4. *Evolution*:425-432.
187. Schleif R (1969) An L-arabinose binding protein and arabinose permeation in *Escherichia coli*. *Journal of molecular biology* 46(1):185-196.
188. Fuchs BM, Wallner G, Beisker W, Schwippl I, Ludwig W, & Amann R (1998) Flow cytometric analysis of the in situ accessibility of *Escherichia coli* 16S rRNA for fluorescently labeled oligonucleotide probes. *Applied and environmental microbiology* 64(12):4973.
189. Amann R, Fuchs BM, & Behrens S (2001) The identification of microorganisms by fluorescence in situ hybridisation. *Current Opinion in Biotechnology* 12(3):231-236.
190. Bottari B, Ercolini D, Gatti M, & Neviani E (2006) Application of FISH technology for microbiological analysis: current state and prospects. *Applied microbiology and biotechnology* 73(3):485-494.
191. Nebe-von Caron G & Badley R (1995) Viability assessment of bacteria in mixed populations using flow cytometry. *J Microscopy* 179(1):55-66.

192. Hugenholtz P, Tyson GW, & Blackall LL (2002) Design and evaluation of 16S rRNA-targeted oligonucleotide probes for fluorescence in situ hybridization. *Methods in Molecular Biology* 179:29-42.
193. Blank K, Hensel M, & Gerlach RG (2011) Rapid and highly efficient method for scarless mutagenesis within the Salmonella enterica chromosome. *PloS one* 6(1):e15763.
194. Datsenko KA & Wanner BL (2000) One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proceedings of the National Academy of Sciences* 97(12):6640.
195. Shaw W (1975) Chloramphenicol acetyltransferase from chloramphenicol-resistant bacteria. *Methods in enzymology* 43:737-755.
196. Gray GS & Fitch WM (1983) Evolution of antibiotic resistance genes: the DNA sequence of a kanamycin resistance gene from Staphylococcus aureus. *Molecular Biology and evolution* 1(1):57-66.
197. Yoshida H, Bogaki M, Nakamura M, & Nakamura S (1990) Quinolone resistance-determining region in the DNA gyrase *gyrA* gene of *Escherichia coli*. *Antimicrobial Agents and Chemotherapy* 34(6):1271-1272.
198. Jin DJ & Gross CA (1988) Mapping and sequencing of mutations in the *Escherichia coli rpoB* gene that lead to rifampicin resistance. *Journal of molecular biology* 202(1):45-58.
199. Rabussay D & Zillig W (1969) A rifampicin resistent RNA-polymerase from *Escherichia coli* altered in the β-subunit. *FEBS letters* 5(2):104-106.
200. Iwakura Y, Ishihama A, & Yura T (1973) RNA polymerase mutants of *Escherichia coli*. *Molecular and General Genetics* 121(2):181-196.
201. Wehrli W (1983) Rifampin: mechanisms of action and resistance. *Review of Infectious Diseases* 5(Supplement 3):S407-S411.
202. Rodríguez-Verdugo A, Gaut BS, & Tenaillon O (2013) Evolution of *Escherichia coli* rifampicin resistance in an antibiotic-free environment during thermal stress. *BMC evolutionary biology* 13(1):50.
203. Ostrowski EA, Rozen DE, & Lenski RE (2005) Pleiotropic effects of beneficial mutations in *Escherichia coli*. *Evolution* 59(11):2343-2352.
204. Turner PE & Elena SF (2000) Cost of host radiation in an RNA virus. *Genetics* 156(4):1465-1470.
205. Travisano M & Lenski RE (1996) Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. *Genetics* 143(1):15.
206. Paquin CE & Adams J (1983) Relative fitness can decrease in evolving asexual populations of S. cerevisiae.
207. Sniegowski PD, Gerrish PJ, & Lenski RE (1997) Evolution of high mutation rates in experimental populations of *Escherichia coli Nature* 387(6634):703-705.
208. Shaver AC, Dombrowski PG, Sweeney JY, Treis T, Zappala RM, & Sniegowski PD (2002) Fitness evolution and the rise of mutator alleles in experimental *Escherichia coli* populations. *Genetics* 162(2):557-566.
209. Kassen R & Bataillon T (2006) Distribution of fitness effects among beneficial mutations before selection in experimental populations of bacteria. *Nature Genetics* 38(4):484-488.
210. Kassen R (2009) Toward a general theory of adaptive radiation. *Annals of the New York Academy of Sciences* 1168(1):3-22.
211. Ostrowski EA, Woods RJ, & Lenski RE (2008) The genetic basis of parallel and divergent phenotypic responses in evolving populations of *Escherichia coli*. *Proceedings of the Royal Society B: Biological Sciences* 275(1632):277-284.

212. Jeong H, Barbe V, Lee CH, Vallenet D, Yu DS, Choi S-H, Couloux A, Lee S-W, Yoon SH, & Cattolico L (2009) Genome sequences of *Escherichia coli* B strains REL606 and BL21 (DE3). *Journal of molecular biology* 394(4):644-652.
213. Zhu J & Shimizu K (2005) Effect of a single-gene knockout on the metabolic regulation in *Escherichia coli* for D-lactate production under microaerobic condition. *Metabolic engineering* 7(2):104-115.
214. Craigie R (1992) Hotspots and warm spots: integration specificity of retroelements. *Trends in Genetics* 8(6):187-190.
215. Fomukong N, Tang T, Al-Maamary S, Ibrahim W, Ramayah S, Yates M, Zainuddin Z, & Dale J (1994) Insertion sequence typing of *Mycobacterium tuberculosis*: characterization of a widespread subtype with a single copy of IS6110. *Tubercle and Lung Disease* 75(6):435-440.
216. Hermans P, Van Soolingen D, Bik E, De Haas P, Dale J, & Van Embden J (1991) Insertion element IS987 from *Mycobacterium bovis* BCG is located in a hot-spot integration region for insertion elements in *Mycobacterium tuberculosis* complex strains. *Infection and Immunity* 59(8):2695-2705.
217. O'Hare K & Rubin GM (1983) Structures of P transposable elements and their sites of insertion and excision in the *Drosophila melanogaster* genome. *Cell* 34(1):25-35.
218. Halling SM & Kleckner N (1982) A symmetrical six-base-pair target site sequence determines Tn10 insertion specificity. *Cell* 28(1):155-163.
219. Berg DE, Schmandt MA, & Lowe JB (1983) Specificity of transposon Tn5 insertion. *Genetics* 105(4):813-828.
220. Bailey SF, Hinz A, & Kassen R (2014) Adaptive synonymous mutations in an experimentally evolved *Pseudomonas fluorescens* population. *Nat Commun* 5.
221. Papadopoulos D, Schneider D, Meier-Eiss J, Arber W, Lenski RE, & Blot M (1999) Genomic evolution during a 10,000-generation experiment with bacteria. *Proceedings of the National Academy of Sciences* 96(7):3807-3812.
222. Charlesworth B, Sniegowski P, & Stephan W (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371:215-220.
223. Kazazian HH (2004) Mobile elements: drivers of genome evolution. *Science* 303(5664):1626-1632.
224. Schneider D & Lenski RE (2004) Dynamics of insertion sequence elements during experimental evolution of bacteria. *Research in Microbiology* 155(5):319-327.
225. Wichman H, Badgett M, Scott L, Boulianne C, & Bull J (1999) Different trajectories of parallel evolution during viral adaptation. *Science* 285(5426):422-424.
226. Ryall B, Eydallin G, & Ferenci T (2012) Culture history and population heterogeneity as determinants of bacterial adaptation: the adaptomics of a single environmental transition. *Microbiology and Molecular Biology Reviews* 76(3):597-625.
227. Clark D & Cronan J (1980) *Escherichia coli* mutants with altered control of alcohol dehydrogenase and nitrate reductase. *Journal of bacteriology* 141(1):177-183.
228. Jiang GR, Nikolova S, & Clark DP (2001) Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*. *Microbiology* 147(9):2437-2446.
229. San K-Y, Bennett GN, Berríos-Rivera SJ, Vadali RV, Yang Y-T, Horton E, Rudolph FB, Sariyar B, & Blackwood K (2002) Metabolic engineering through cofactor manipulation and its effects on metabolic flux redistribution in *Escherichia coli*. *Metabolic Engineering* 4(2):182-192.

230. Raffaelli N, Lorenzi T, Mariani PL, Emanuelli M, Amici A, Ruggieri S, & Magni G (1999) The *Escherichia coli* NadR regulator is endowed with nicotinamide mononucleotide adenylyltransferase activity. *Journal of Bacteriology* 181(17):5509-5511.
231. Tritz GJ & Chandler JL (1973) Recognition of a gene involved in the regulation of nicotinamide adenine dinucleotide biosynthesis. *Journal of Bacteriology* 114(1):128-136.
232. Gerasimova AV & Gelfand MS (2005) Evolution of the NadR regulon in Enterobacteriaceae. *Journal of Bioinformatics and Computational Biology* 3(04):1007-1019.
233. Leonardo MR, Cunningham P, & Clark D (1993) Anaerobic regulation of the *adhE* gene, encoding the fermentative alcohol dehydrogenase of *Escherichia coli*. *Journal of Bacteriology* 175(3):870-878.
234. Wimpenny JW & Firth A (1972) Levels of nicotinamide adenine dinucleotide and reduced nicotinamide adenine dinucleotide in facultative bacteria and the effect of oxygen. *Journal of Bacteriology* 111(1):24-32.
235. Zientz E, Bongaerts J, & Unden G (1998) Fumarate regulation of gene expression in *Escherichia coli* by the DcuSR (*dcuSR* genes) two-component regulatory system. *Journal of Bacteriology* 180(20):5421-5425.
236. Janausch IG, Garcia-Moreno I, Lehnen D, Zeuner Y, & Unden G (2004) Phosphorylation and DNA binding of the regulator DcuR of the fumarate-responsive two-component system DcuSR of *Escherichia coli*. *Microbiology* 150(4):877-883.
237. Wolfe AJ (2005) The acetate switch. *Microbiology and Molecular Biology Reviews* 69(1):12-50.
238. Luli GW & Strohl WR (1990) Comparison of growth, acetate production, and acetate inhibition of *Escherichia coli* strains in batch and fed-batch fermentations. *Applied and Environmental Microbiology* 56(4):1004-1011.
239. Hayes F (2003) Toxins-antitoxins: plasmid maintenance, programmed cell death, and cell cycle arrest. *Science* 301(5639):1496-1499.
240. Gerdes K & Wagner EGH (2007) RNA antitoxins. *Current Opinion in Microbiology* 10(2):117-124.
241. Magnuson RD (2007) Hypothetical functions of toxin-antitoxin systems. *Journal of bacteriology* 189(17):6089-6092.
242. Van Melderen L (2010) Toxin–antitoxin systems: why so many, what for? *Current Opinion in Microbiology* 13(6):781-785.
243. Padan E & Schuldiner S (1993) Na+/H+ antiporters, molecular devices that couple the Na+ and H+ circulation in cells. *Journal of bioenergetics and biomembranes* 25(6):647-669.
244. Pedersen K & Gerdes K (1999) Multiple hok genes on the chromosome of *Escherichia coli Molecular Microbiology* 32(5):1090-1102.
245. Pedersen K, Christensen SK, & Gerdes K (2002) Rapid induction and reversal of a bacteriostatic condition by controlled expression of toxins and antitoxins. *Molecular Microbiology* 45(2):501-510.
246. Steif A & Meyer IM (2012) The hok mRNA family. *RNA Biology* 9:1399-1404.
247. Kawano M, Oshima T, Kasai H, & Mori H (2002) Molecular characterization of long direct repeat (LDR) sequences expressing a stable mRNA encoding for a 35 amino acid cell killing peptide and a *cis*-encoded small antisense RNA in *Escherichia coli*. *Molecular Microbiology* 45(2):333-349.
248. Mruk I & Kobayashi I (2013) To be or not to be: regulation of restriction–modification systems and other toxin–antitoxin systems. *Nucleic Acids Research*:711.

249. Schneider D, Duperchy E, Coursange E, Lenski RE, & Blot M (2000) Long-term experimental evolution in *Escherichia coli.* IX. Characterization of insertion sequence-mediated mutations and rearrangements. *Genetics* 156(2):477-488.
250. Edlin G, Lin L, & Bitner R (1977) Reproductive fitness of P1, P2, and Mu lysogens of *Escherichia coli*. *Journal of Virology* 21(2):560-564.
251. Wang X, Kim Y, Ma Q, Hong SH, Pokusaeva K, Sturino JM, & Wood TK (2010) Cryptic prophages help bacteria cope with adverse environments. *Nature Communications* 1:147.
252. Chen Y, Golding I, Sawai S, Guo L, & Cox EC (2005) Population fitness and the regulation of *Escherichia coli* genes by bacterial viruses. *PLoS biology* 3(7):e229.
253. Lawrence JG, Hendrix RW, & Casjens S (2001) Where are the pseudogenes in bacterial genomes? *Trends in Microbiology* 9(11):535-540.
254. Brøndsted L & Atlung T (1996) Effect of growth conditions on expression of the acid phosphatase (*cyx-appA*) operon and the *appY* gene, which encodes a transcriptional activator of *Escherichia coli*. *Journal of Bacteriology* 178(6):1556-1564.
255. Nakata N, Tobe T, Fukuda I, Suzuki T, Komatsu K, Yoshikawa M, & Sasakawa C (1993) The absence of a surface protease, OmpT, determines the intercellular spreading ability of Shigella: the relationship between the ompT and kcpA loci. *Molecular microbiology* 9(3):459-468.
256. Lindsey D, Mullin D, & Walker J (1989) Characterization of the cryptic lambdoid prophage DLP12 of *Escherichia coli* and overlap of the DLP12 integrase gene with the tRNA gene *argU*. *Journal of Bacteriology* 171(11):6197-6205.
257. Stumpe S, Schmid R, Stephens DL, Georgiou G, & Bakker EP (1998) Identification of OmpT as the protease that hydrolyzes the antimicrobial peptide protamine before it enters growing cells of *Escherichia coli*. *Journal of Bacteriology* 180(15):4002-4006.
258. Vandeputte‐Rutten L, Kramer RA, Kroon J, Dekker N, Egmond MR, & Gros P (2001) Crystal structure of the outer membrane protease OmpT from *Escherichia coli* suggests a novel catalytic site. *The EMBO journal* 20(18):5033-5039.
259. Gallegos M-T, Michan C, & Ramos JL (1993) The XylS/AraC family of regulators. *Nucleic acids research* 21(4):807-810.
260. Munson GP, Lam DL, Outten FW, & O'Halloran TV (2000) Identification of a copper-responsive two-component system on the chromosome of *Escherichia coli* K-12. *Journal of bacteriology* 182(20):5864-5871.
261. Outten FW, Huffman DL, Hale JA, & O'Halloran TV (2001) The independent *cue* and *cus* systems confer copper tolerance during aerobic and anaerobic growth in *Escherichia coli*. *Journal of Biological Chemistry* 276(33):30670-30677.
262. Lundrigan M & Earhart C (1984) Gene *envY* of *Escherichia col*i K-12 affects thermoregulation of major porin expression. *Journal of Bacteriology* 157(1):262-268.
263. Pierson III L & Kahn M (1987) Integration of satellite bacteriophage P4 in *Escherichia coli*: DNA sequences of the phage and host regions involved in site-specific recombination. *Journal of molecular biology* 196(3):487-496.
264. Reiter W-D, Palm P, & Yeats S (1989) Transfer RNA genes frequently serve as integration sites for prokaryotic genetic elements. *Nucleic acids research* 17(5):1907-1914.
265. Thorbjarnardóttir S, Dingermann T, Rafnar T, Andrésson O, Söll D, & Eggertsson G (1985) Leucine tRNA family of *Escherichia coli*: nucleotide sequence of the *supP* (Am) suppressor gene. *Journal of bacteriology* 161(1):219-222.
266. Kiino DR & Rothman-Denes LB (1989) Genetic analysis of bacteriophage N4 adsorption. *Journal of bacteriology* 171(9):4595-4602.

267. Yang H, Wolff E, Kim M, Diep A, & Miller JH (2004) Identification of mutator genes and mutational pathways in *Escherichia coli* using a multicopy cloning approach. *Molecular Microbiology* 53(1):283-295.
268. Richardson H & Egan JB (1989) DNA replication studies with coliphage 186: II. Depression of host replication by a 186 gene. *Journal of Molecular Biology* 206(1):59-68.
269. Barreiro V & Haggård-Ljungquist E (1992) Attachment sites for bacteriophage P2 on the *Escherichia coli* chromosome: DNA sequences, localization on the physical map, and detection of a P2-like remnant in *Escherichia coli* K-12 derivatives. *Journal of Bacteriology* 174(12):4086-4093.
270. Grambow NJ, Birkeland NK, Anders DL, & Christie GE (1990) Deletion analysis of a bacteriophage P2 late promoter. *Gene* 95(1):9-15.
271. Lee M-C & Marx CJ (2012) Repeated, selection-driven genome reduction of accessory genes in experimental populations. *PLoS Genetics* 8(5):e1002651.
272. Bossi L, Fuentes JA, Mora G, & Figueroa-Bossi N (2003) Prophage contribution to bacterial population dynamics. *Journal of Bacteriology* 185(21):6467-6471.
273. Wang X, Kim Y, & Wood TK (2009) Control and benefits of CP4-57 prophage excision in *Escherichia coli* biofilms. *International Society for Medical Ecology* 3(10):1164-1179.
274. Osman D, Waldron KJ, Denton H, Taylor CM, Grant AJ, Mastroeni P, Robinson NJ, & Cavet JS (2010) Copper homeostasis in *Salmonella* is atypical and copper-CueP is a major periplasmic metal complex. *Journal of Biological Chemistry* 285(33):25259-25268.
275. Van der Woude MW & Henderson IR (2008) Regulation and function of Ag43 (*flu*). *Annual Reviews in Microbiology* 62:153-169.
276. Fexby S, Bjarnsholt T, Jensen PØ, Roos V, Høiby N, Givskov M, & Klemm P (2007) Biological Trojan horse: antigen 43 provides specific bacterial uptake and survival in human neutrophils. *Infection and Immunity* 75(1):30-34.
277. Scholl D, Adhya S, & Merril C (2005) *Escherichia coli* K1's capsule is a barrier to bacteriophage T7. *Applied and Environmental Microbiology* 71(8):4872-4874.
278. Schwan WR, Beck MT, Hultgren SJ, Pinkner J, Woolever NL, & Larson T (2005) Down-regulation of the *kps* region 1 capsular assembly operon following attachment of *Escherichia coli* type 1 fimbriae to D-mannose receptors. *Infection and Immunity* 73(2):1226-1231.
279. Rigg GP & Barrett B (1998) The localization of KpsC, S and T, and KfiA, C and D proteins involved in the biosynthesis of the *Escherichia coli* K5 capsular polysaccharide: evidence for a membrane-bound complex. *Microbiology* 144(10):2905-2914.
280. Silver RP, Prior K, Nsahlai C, & Wright LF (2001) ABC transporters and the export of capsular polysaccharides from Gram-negative bacteria. *Research in Microbiology* 152(3):357-364.
281. Bronner D, Sieberth V, Pazzani C, Smith A, Boulnois G, Roberts I, Jann B, & Jann K (1993) Synthesis of the K5 (group II) capsular polysaccharide in transport-deficient recombinant *Escherichia coli*. *Federation of European Microbiological Societies Microbiology Letters* 113(3):279-284.
282. Wunder DE, Aaronson W, Hayes SF, Bliss JM, & Silver R (1994) Nucleotide sequence and mutational analysis of the gene encoding KpsD, a periplasmic protein involved in transport of polysialic acid in *Escherichia coli* K1. *Journal of Bacteriology* 176(13):4025-4033.
283. Roberts I (1995) Bacterial polysaccharides in sickness and in health. *Microbiology* 141(9):2023-2031.

284. Pavelka M, Hayes SF, & Silver RP (1994) Characterization of KpsT, the ATP-binding component of the ABC-transporter involved with the export of capsular polysialic acid in *Escherichia coli* K1. *Journal of Biological Chemistry* 269(31):20149-20158.
285. Pigeon RP & Silver RP (1994) Topological and mutational analysis of KpsM, the hydrophobic component of the ABC-transporter involved in the export of polysialic acid in *Escherichia coli* K1. *Molecular Microbiology* 14(5):871-881.
286. Smith A, Boulnois G, & Roberts I (1990) Molecular analysis of the *Escherichia coli* K5 *kps* locus: identification and characterization of an inner-membrane capsular polysaccharide transport system. *Molecular Microbiology* 4(11):1863-1869.
287. Saier Jr MH (2000) Families of transmembrane transporters selective for amino acids and their derivatives. *Microbiology* 146(8):1775-1795.
288. Stucky K, Hagting A, Klein JR, Matern H, Henrich B, Konings WN, & Plapp R (1995) Cloning and characterization of *brnQ*, a gene encoding a low-affinity, branched-chain amino acid carrier in *Lactobacillus delbrdückii* subsp. lactic DSM7290. *Molecular and General Genetics MGG* 249(6):682-690.
289. Baisa G, Stabo NJ, & Welch RA (2013) Characterization of *Escherichia coli* D-cycloserine transport and resistant mutants. *Journal of Bacteriology* 195(7):1389-1399.
290. Jishage M, Iwata A, Ueda S, & Ishihama A (1996) Regulation of RNA polymerase sigma subunit synthesis in *Escherichia coli*: intracellular levels of four species of sigma subunit under various growth conditions. *Journal of Bacteriology* 178(18):5447-5451.
291. Bohannon D, Connell N, Keener J, Tormo A, Espinosa-Urgel M, Zambrano MM, & Kolter R (1991) Stationary-phase-inducible "gearbox" promoters: differential effects of *katF* mutations and role of sigma 70. *Journal of Bacteriology* 173(14):4482-4492.
292. Schellhorn HE (2014) Elucidating the function of the RpoS regulon. *Future Microbiology* 9(4):497-507.
293. Notley-McRobb L, King T, & Ferenci T (2002) *rpoS* mutations and loss of general stress resistance in *Escherichia coli* populations as a consequence of conflict between competing stress responses. *Journal of Bacteriology* 184(3):806-811.
294. Wei B, Shin S, LaPorte D, Wolfe AJ, & Romeo T (2000) Global regulatory mutations in *csrA* and *rpoS* cause severe central carbon stress in *Escherichia coli* in the presence of acetate. *Journal of Bacteriology* 182(6):1632-1640.
295. Ivanova A, Renshaw M, Guntaka RV, & Eisenstark A (1992) DNA base sequence variability in *katF* (putative sigma factor) gene of *Escherichia coli*. *Nucleic Acids Research* 20(20):5479.
296. Jishage M & Ishihama A (1997) Variation in RNA polymerase sigma subunit composition within different stocks of *Escherichia coli* W3110. *Journal of Bacteriology* 179(3):959-963.
297. Friedman L, Alder JD, & Silverman JA (2006) Genetic changes that correlate with reduced susceptibility to daptomycin in *Staphylococcus aureus*. *Antimicrobial Agents and Chemotherapy* 50(6):2137-2145.
298. Herring CD, Raghunathan A, Honisch C, Patel T, Applebee MK, Joyce AR, Albert TJ, Blattner FR, Van den Boom D, & Cantor CR (2006) Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nature Genetics* 38(12):1406-1412.
299. Conrad TM, Frazier M, Joyce AR, Cho B-K, Knight EM, Lewis NE, Landick R, & Palsson BØ (2010) RNA polymerase mutants found through adaptive evolution reprogram *Escherichia coli* for optimal growth in minimal media. *Proceedings of the National Academy of Sciences* 107(47):20500-20505.

300. Farrell MJ & Finkel SE (2003) The growth advantage in stationary-phase phenotype conferred by *rpoS* mutations is dependent on the pH and nutrient environment. *Journal of Bacteriology* 185(24):7044-7052.
301. Ferenci T (1999) Regulation by nutrient limitation. *Current Opinion in Microbiology* 2(2):208-213.
302. Mohanty BK & Kushner SR (1999) Analysis of the function of *Escherichia coli* poly (A) polymerase I in RNA metabolism. *Molecular microbiology* 34(5):1094-1108.
303. Coburn GA & Mackie GA (1996) Differential sensitivities of portions of the mRNA for ribosomal protein S20 to 3′ -exonucleases dependent on oligoadenylation and RNA secondary structure. *Journal of Biological Chemistry* 271(26):15776-15781.
304. Coburn GA & Mackie GA (1998) Reconstitution of the degradation of the mRNA for ribosomal protein S20 with purified enzymes. *Journal of molecular biology* 279(5):1061-1074.
305. Eyre-Walker A & Keightley PD (2007) The distribution of fitness effects of new mutations. *Nature Reviews Genetics* 8(8):610-618.
306. Helling RB, Vargas CN, & Adams J (1987) Evolution of *Escherichia coli* during growth in a constant environment. *Genetics* 116(3):349-358.
307. Rosenzweig RF, Sharp R, Treves DS, & Adams J (1994) Microbial evolution in a simple unstructured environment: genetic differentiation in *Escherichia coli*. *Genetics* 137(4):903-917.
308. Colwell CA (1946) Small colony variants of *Escherichia coli*. *Journal of bacteriology* 52(4):417.
309. Proctor RA, van Langevelde P, Kristjansson M, Maslow JN, & Arbeit RD (1995) Persistent and relapsing infections associated with small colony variants of *Staphylococcus aureus*. *Clinical infectious diseases* 20(1):95-102.
310. Proctor RA, Von Eiff C, Kahl BC, Becker K, McNamara P, Herrmann M, & Peters G (2006) Small colony variants: a pathogenic form of bacteria that facilitates persistent and recurrent infections. *Nature Reviews Microbiology* 4(4):295-305.
311. Seifert H, von Eiff C, & Fätkenheuer G (1999) Fatal case due to methicillin-resistant *Staphylococcus aureus* small colony variants in an AIDS patient. *Emerging infectious diseases* 5(3):450.
312. Swingle EL (1935) Studies on small colony variants of *Staphylococcus aureus*. *Journal of bacteriology* 29(5):467.
313. Acar J, Goldstein F, & Lagrange P (1978) Human infections caused by thiamine-or menadione-requiring *Staphylococcus aureus*. *Journal of clinical microbiology* 8(2):142-147.
314. Kahl B, Herrmann M, Everding AS, Koch HG, Becker K, Harms E, Proctor RA, & Peters G (1998) Persistent infection with small colony variant strains of *Staphylococcus aureus* in patients with cystic fibrosis. *Journal of Infectious Diseases* 177(4):1023-1029.
315. Häußler S, Tümmler B, Weißbrodt H, Rohde M, & Steinmetz I (1999) Small colony variants of *Pseudomonas aeruginosa* in cystic fibrosis. *Clinical infectious diseases* 29(3):621-625.
316. Proctor RA, Kahl B, von Eiff C, Vaudaux PE, Lew DP, & Peters G (1998) Staphylococcal small colony variants have novel mechanisms for antibiotic resistance. *Clinical infectious diseases* 27(Supplement 1):S68-S74.
317. Puentes-Téllez PE & van Elsas JD (2014) Sympatric metabolic diversification of experimentally evolved *Escherichia coli* in a complex environment. *Antonie van Leeuwenhoek* 106(3):565-576.
318. Oliveros JC (2007) VENNY. An interactive tool for comparing lists with Venn Diagrams.

319. Le Gac M, Brazas MD, Bertrand M, Tyerman JG, Spencer CC, Hancock RE, & Doebeli M (2008) Metabolic changes associated with adaptive diversification in *Escherichia coli*. *Genetics* 178(2):1049-1060.
320. Le Gac M, Plucain J, Hindré T, Lenski RE, & Schneider D (2012) Ecological and evolutionary dynamics of coexisting lineages during a long-term experiment with *Escherichia coli*. *Proceedings of the National Academy of Sciences* 109(24):9487-9492.
321. Turner PE, Souza V, & Lenski RE (1996) Tests of ecological mechanisms promoting the stable coexistence of two bacterial genotypes. *Ecology*:2119-2129.
322. Odling-Smee FJ, Laland KN, & Feldman MW (1996) Niche construction. *American Naturalist*:641-648.
323. Thauer R, Zinkhan D, & Spormann A (1989) Biochemistry of acetate catabolism in anaerobic chemotrophic bacteria. *Annual Reviews in Microbiology* 43(1):43-67.
324. Hood RD, Singh P, Hsu F, Güvener T, Carl MA, Trinidad RR, Silverman JM, Ohlson BB, Hicks KG, & Plemel RL (2010) A type VI secretion system of *Pseudomonas aeruginosa* targets a toxin to bacteria. *Cell Host & Microbe* 7(1):25-37.
325. Lemonnier M, Levin BR, Romeo T, Garner K, Baquero M-R, Mercante J, Lemichez E, Baquero F, & Blázquez J (2008) The evolution of contact-dependent inhibition in non-growing populations of *Escherichia coli*. *Proceedings of the Royal Society B: Biological Sciences* 275(1630):3-10.
326. Covert AW, Lenski RE, Wilke CO, & Ofria C (2013) Experiments on the role of deleterious mutations as stepping stones in adaptive evolution. *Proceedings of the National Academy of Sciences* 110(34):3171-3178.
327. Cowperthwaite MC, Bull J, & Meyers LA (2006) From bad to good: Fitness reversals and the ascent of deleterious mutations. *PLoS computational biology* 2(10):e141.
328. Silander OK, Tenaillon O, & Chao L (2007) Understanding the evolutionary fate of finite populations: the dynamics of mutational effects. *PLoS biology* 5(4):e94.
329. Novak M, Pfeiffer T, Lenski RE, Sauer U, & Bonhoeffer S (2006) Experimental tests for an evolutionary trade-off between growth rate and yield in *Escherichia coli*. *The American Naturalist* 168(2):242-251.
330. Dauvillée D, Kinderf IS, Li Z, Kosar-Hashemi B, Samuel MS, Rampling L, Ball S, & Morell MK (2005) Role of the *Escherichia coli glgX* gene in glycogen metabolism. *Journal of Bacteriology* 187(4):1465-1473.
331. Hunter S, Jones P, Mitchell A, Apweiler R, Attwood TK, Bateman A, Bernard T, Binns D, Bork P, & Burge S (2011) InterPro in 2011: new developments in the family and domain prediction database. *Nucleic acids research*:gkr948.
332. Baquero F & Lemonnier M (2009) Generational coexistence and ancestor's inhibition in bacterial populations. *FEMS microbiology reviews* 33(5):958-967.
333. Atlung T, Nielsen A, & Hansen FG (1989) Isolation, characterization, and nucleotide sequence of appY, a regulatory gene for growth-phase-dependent gene expression in *Escherichia coli. Journal of bacteriology* 171(3):1683-1691.
334. Atlung T, Sund S, Olesen K, & Brøndsted L (1996) The histone-like protein H-NS acts as a transcriptional repressor for expression of the anaerobic and growth phase activator AppY of *Escherichia coli*. *Journal of Bacteriology* 178(12):3418-3425.
335. Nesbit A, Fleischhacker A, Teter S, & Kiley P (2012) ArcA and AppY antagonize IscR repression of hydrogenase-1 expression under anaerobic conditions, revealing a novel mode of O2 regulation of gene expression in *Escherichia coli*. *Journal of bacteriology* 194(24):6892-6899.

336. Pinske C, McDowall JS, Sargent F, & Sawers RG (2012) Analysis of hydrogenase 1 levels reveals an intimate link between carbon and hydrogen metabolism in Escherichia coli K-12. *Microbiology* 158(Pt 3):856-868.
337. Atlung T, Knudsen K, Heerfordt L, & Brøndsted L (1997) Effects of sigmaS and the transcriptional activator AppY on induction of the *Escherichia coli hya* and *cbdAB-appA* operons in response to carbon and phosphate starvation. *Journal of bacteriology* 179(7):2141-2146.
338. Loui C, Chang AC, & Lu S (2009) Role of the ArcAB two-component system in the resistance of *Escherichia coli* to reactive oxygen stress. *Bio Medical Central Microbiology* 9(1):183.
339. Iuchi S & Lin E (1988) *arcA* (dye), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways. *Proceedings of the National Academy of Sciences* 85(6):1888-1892.
340. Bidart GN, Ruiz JA, de Almeida A, Méndez BS, & Nikel PI (2012) Manipulation of the anoxic metabolism in *Escherichia coli* by ArcB deletion variants in the ArcBA two-component system. *Applied and environmental microbiology* 78(24):8784-8794.
341. Taylor BL & Zhulin IB (1999) PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews* 63(2):479-506.
342. Stover C, Pham X, Erwin A, Mizoguchi S, Warrener P, Hickey M, Brinkman F, Hufnagle W, Kowalik D, & Lagrou M (2000) Complete genome sequence of *Pseudomonas aeruginosa* PAO1, an opportunistic pathogen. *Nature* 406(6799):959-964.
343. Mathee K, Narasimhan G, Valdes C, Qiu X, Matewish JM, Koehrsen M, Rokas A, Yandava CN, Engels R, & Zeng E (2008) Dynamics of *Pseudomonas aeruginosa* genome evolution. *Proceedings of the National Academy of Sciences* 105(8):3100-3105.
344. Dragosits M & Mattanovich D (2013) Adaptive laboratory evolution-principles and applications for biotechnology. *Microbial Cell Factories* 12:64.
345. Yazdani SS & Gonzalez R (2007) Anaerobic fermentation of glycerol: a path to economic viability for the biofuels industry. *Current Opinion in Biotechnology* 18(3):213-219.
346. Lin Y & Tanaka S (2006) Ethanol fermentation from biomass resources: current state and prospects. *Applied microbiology and biotechnology* 69(6):627-642.
347. Fischer CR, Klein-Marcuschamer D, & Stephanopoulos G (2008) Selection and optimization of microbial hosts for biofuels production. *Metabolic engineering* 10(6):295-304.
348. MacLean RC, Bell G, & Rainey PB (2004) The evolution of a pleiotropic fitness tradeoff in *Pseudomonas fluorescens*. *Proceedings of the National Academy of Sciences of the United States of America* 101(21):8072-8077.