Rapid Phenotypic Switching in a Natural Isolate of *Escherichia coli.*

A thesis presented in partial fulfilment of the

requirements for the degree of

Master of Science

in

Genetics

at Massey University, Albany

New Zealand

Stella Margaret Pearless

2022

Abstract

The survivability of any given bacterial population is dependent on its genetic and phenotypic makeup. When cells replicate they usually produce genetically identical daughter cells through a process called binary fission. Although these daughter cells may remain genetically identical and form isogenic populations, bacteria also possess the ability to alter their phenotype independently of other cells in their population. This can result in subpopulations of phenotypically variable cells forming within a larger population, rendering them phenotypically heterogeneous. Phenotypic heterogeneity can arise from multiple molecular mechanisms that cause either genetic or epigenetic changes. Genetic mechanisms include site-specific DNA inversion, slipped-strand mispairing and homologous recombination. Epigenetic changes involve modifications to DNA at the structural level, and in bacteria most commonly refer to methylation. Heterogeneity can provide evolutionary advantages through processes like bet-hedging or the division of labour. A well documented example of evolutionarily advantageous phenotypic heterogeneity is the formation of persister cells within bacterial strains, a leading cause of antibiotic treatment failure.

In this study, we have identified an *Escherichia coli* natural isolate strain (SC375) that is able to rapidly switch between two phenotypes. The phenotypic heterogeneity demonstrated in this strain results in varying colony morphologies influenced by their cellular composition. We initially proposed a DNA inversion mechanism for this switching but subsequently confirmed that all cells remain isogenic regardless of cell phenotype. Through RNA-sequencing we identified three virulence genes that were differentially regulated in both phenotypes, suggestive of an epigenetic regulatory mechanism. We then show, using reporter assays, that two of these genes are expressed in variable levels across subpopulations. We suggest that the rapid phenotypic reversibility of this strain is a possible indicator of epigenetic memory.

Acknowledgements

First and foremost I would like to thank my supervisor Dr. Olin Silander for his support, his trust and for never letting me take myself too seriously. I am grateful for the opportunity to have worked in his lab and for everything I have learnt along the way. Participating in this project has allowed me to discover a passion for an area of science that I didn't know I had and has given me confidence in my own abilities as a scientist. I will always be thankful for the time he has taken to teach me.

Thank you to all the members of the Freed-Silander lab group that I had the pleasure of working with over the last two years. Bhargava for his contributions to my research. Markéta for taking me under her wing and patiently answering every single one of my questions. Georgia for her endless support, friendship, our treat runs and her ability to match my energy no matter how I showed up to the lab.

To Emily, for being my lab partner, best friend and for understanding me in a way that no one else can. Her encouragement and lack of judgement made me look forward to coming into uni everyday, without her I would have been lost.

Finally, to the other important people in my life. Thank you to my parents for making me believe I am capable of anything I set my mind to; there is no way I would have been able to finish the last two years without their endless support. To Hannah for being proud of everything I do and for always setting an example that I would be proud to follow. And to James for always believing in me even when I didn’t believe in myself.

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction

1.1 Bacterial populations exhibit phenotypic variation

The survival of any given cell within a bacterial population is dependent on its genotype and phenotype. Bacteria replicate asexually via binary fission, usually producing two genetically identical daughter cells; continued replication of these daughter cells results in isogenic populations (Narra & Ochman, 2006). The development of single-cell analysis methods has allowed for the quantification of variable phenotypes within such isogenic populations (Buettner et al., 2015; Dey et al., 2015), leading to the discovery that cells are able to differentiate their phenotype independently of each other. Populations that contain variable subpopulations of phenotypically different cells are said to be phenotypically heterogeneous. Heterogeneity can be created through genetic, epigenetic or stochastic (random) means.

Random fluctuations in the synthesis, degradation, and interactions of gene products results in variation in the expression of any given gene across cells. This variation is often termed gene expression noise (Davidson & Surette, 2008; Evans et al., 2018; Patange et al., 2018; Vujovic et al., 2019). Variation in the expression of certain transcripts or proteins can lead to the formation of small subpopulations within a larger population, with each subpopulation being functionally and phenotypically unique.

In contrast to gene expression noise, cells also possess the ability to regulate the expression of their genes in a controlled, responsive manner, independently of the other cells in their population (Jolly et al., 2018). This phenomenon is termed *phenotypic plasticity* (Jolly et al., 2018). Phenotypic plasticity is particularly beneficial for bacterial populations that grow in unstable environments (Binder et al., 2017; Gasperotti et al., 2020; van der Woude, 2011), allowing them to adapt quickly to any changes, for example: to increase survivability under conditions of stress such as heat or antibiotic treatment (Healey et al., 2016; Schreiber et al., 2016). Phenotypic plasticity may also act in combination with stochastic differentiation to ensure that at least some individuals survive sudden environmental stressors, preventing the whole population from extinction.

Epigenetic regulation describes stable changes in gene expression due to modifications of the DNA structure, without any alterations in the genomic sequence (Adhikari & Curtis, 2016; Collier, 2009; Grabiec & Potempa, 2018; Žgur-Bertok, 2007). Changes to DNA modifications can influence gene expression by either inhibiting or facilitating the recruitment of transcription proteins (Trzilova & Tamayo, 2021). However, there are a wide range of other epigenetic mechanisms leading to the heritability of phenotypes without genetic changes, such as through positive feedback loops, the slow dilution of membrane proteins (e.g. transporters), or the inheritance of activated protein products (Azzaz & Fantini, 2022; Harvey et al., 2018; Smits et al., 2006).

In addition to epigenetic mechanisms that affect phenotypes, there are genetic mechanisms that can result in rapid phenotypic switching in bacterial cells. One example is phase variation. Phase variation is a reversible, method of gene regulation that is characterised by its high frequency and “all or nothing” mode of action (ON/OFF switching) (Bayliss et al., 2019; van den Broek et al., 2005; van der Woude, 2011). Phase variation involves the alteration of very specific, predetermined sites on the genome (van der Woude & Bäumler, 2004). A defining characteristic of phase variation is its reversibility, which distinguishes it from other mutational processes (Henderson et al., 1999; Silverman & Simon, 1980; van der Woude & Bäumler, 2004). Phase variation can be regulated through a number of genetic mechanisms including site-specific inversion, homologous recombination and slipped strand mispairing. The ON/OFF switching mechanism of phase variation, can create bimodal gene expression where a promoter is either fully active or not active at all (van der Woude & Bäumler, 2004). Bimodal expression can lead to the formation of two exclusive phenotypes within a population and in extreme cases, leads to bifurcation into two distinct subpopulations (García-Pastor et al., 2019; Sánchez-Romero & Casadesús, 2018). The bifurcation process demonstrates bistability (rather than bimodality, which can be unstable), and reflects the ability of cells to form extreme and stable phenotypes that are unable to exist in intermediate forms (Roberfroid et al., 2016).

1.2 Evolutionary advantages of Phenotypic Variation.

Creating variability can be a costly process that requires bacterial populations to invest in cells with completely different, and sometimes contrasting, specialisations (Lowery et al., 2017). While this investment may appear to be wasteful from the perspective of single cells, it is offset by the potential advantages provided by the resulting phenotypic diversification at a population level. Stochastic differentiation into subpopulations can provide a survival advantage for the population through two evolutionary approaches: the division of labour and bet-hedging **(Fig. 1.1)** (Sánchez-Romero & Casadesús, 2018). The high genetic relatedness between subpopulations acts as a guard against "cheating", in which some cells may opt out of paying the cost of differentiation while still reaping the benefits of the differentiation of other cells in the population (Ackermann et al., 2008).

Bet hedging is a risk-spreading strategy, where generating heterogeneity is used to guard against environmental changes before they are encountered (Carey & Goulian, 2019). The advantages of this strategy are especially relevant for bacteria that live in volatile environments (García-Pastor et al., 2019; Harms et al., 2016). For example, if a bacterial population consisting of isogenic cells produces two distinct phenotypes, and each phenotype has high fitness in at least one environment (but perhaps extremely low fitness in another environment), it ensures that at least some individuals will survive and thrive when there are dramatic environmental changes. This considerably reduces the risk of the whole population being eradicated. Often phenotypes that show the greatest fitness upon exposure to stresses are maladapted to their typical environment, so heterogeneity is only beneficial if the environmental conditions are volatile (Carey & Goulian, 2019). This makes bet-hedging a trade-off between current fitness and future survival.

One bet-hedging strategy is well demonstrated by *Mycobacterium tuberculosis (M. tuberculosis)*, the pathogenic bacteria responsible for the infectious disease tuberculosis (Chisholm & Tanaka, 2016; Davis & Isberg, 2016). *M. tuberculosis* populations are capable of forming small subpopulations of persister cells, which exhibit multidrug tolerance by entering a physiologically dormant state (Harms et al., 2016; Michiels et al., 2016). Dormancy allows persisters to withstand high exposure to antibiotics, but results in a trade off between growth and survival (Wilmaerts et al., 2019). While their lack of proliferative

activity is detrimental to population fitness in their typical environment, it is beneficial upon exposure to antibiotic stressors. It is believed that the presence of persisters are responsible for the characteristic latent infections and long antibiotic treatment required of tuberculosis (Fisher et al., 2017; Hill & Helaine, 2019).

In contrast to bet hedging, phenotypic differentiation resulting in the division of labour is characterised by heterogeneous subpopulations each performing specialised tasks and cooperatively functioning to increase the fitness of the overall population (West & Cooper, 2016). Often subpopulations of these heterogeneous cells sacrifice their own individual fitness to improve the overall fitness of the population (Ackermann et al., 2008). Populations with this dynamic demonstrate a greater combined fitness than if a single cell were to perform all tasks on its own (García-Pastor et al., 2019; Weigel & Dersch, 2018; Zhang et al., 2016). A division of labour approach is particularly useful when a single cell is unable to conduct multiple, necessary functions concurrently (Tsai & Coombes, 2019; Weigel & Dersch, 2018; Zhang et al., 2016). If this division occurs in metabolic genes, it can also have the added benefit of increasing the range of metabolites able to be utilised by a single population (Weigel & Dersch, 2018).

Labour is either divided in a unidirectional or bidirectional manner. When two tasks are divided among individuals in a way that provides a reciprocated fitness benefit for both subpopulations the division of labour is bidirectional. In contrast, a unidirectional division results in one subpopulation performing a task that benefits the whole population but is detrimental to its own fitness (Balaban & Liu, 2019; Giri et al., 2019). Unidirectional division often occurs during the production of a “public good”, which can be a costly process (Becker et al., 2018). If the entire population was forced to invest in the production of this good, growth would be reduced. When only a small subset of the population makes such an investment, the whole population is able to reap the benefits of this good, without a personal cost to each individual, allowing for faster growth at a population wide level (Kramer et al., 2018).

Unidirectional division of labour is demonstrated by the antibiotic producing bacteria Streptomyces coelicolor (S. coelicolor). Antibiotic production is an example of a costly process that benefits a population's survival, by inhibiting the growth of proximal competitors

(Chater et al., 2010). Wild type S. coelicolor show low levels of antibiotic production and rapid growth, while mutant individuals that are able to produce high quantities of antibiotic lose their ability to proliferate as rapidly. Previous work has shown that the limitation of antibiotic production to a small subpopulation of cells allows S. coelicolor to maximise both the amount and variety of antibiotic production in a way that reduces the cost of biosynthesis (Zhang et al., 2020).

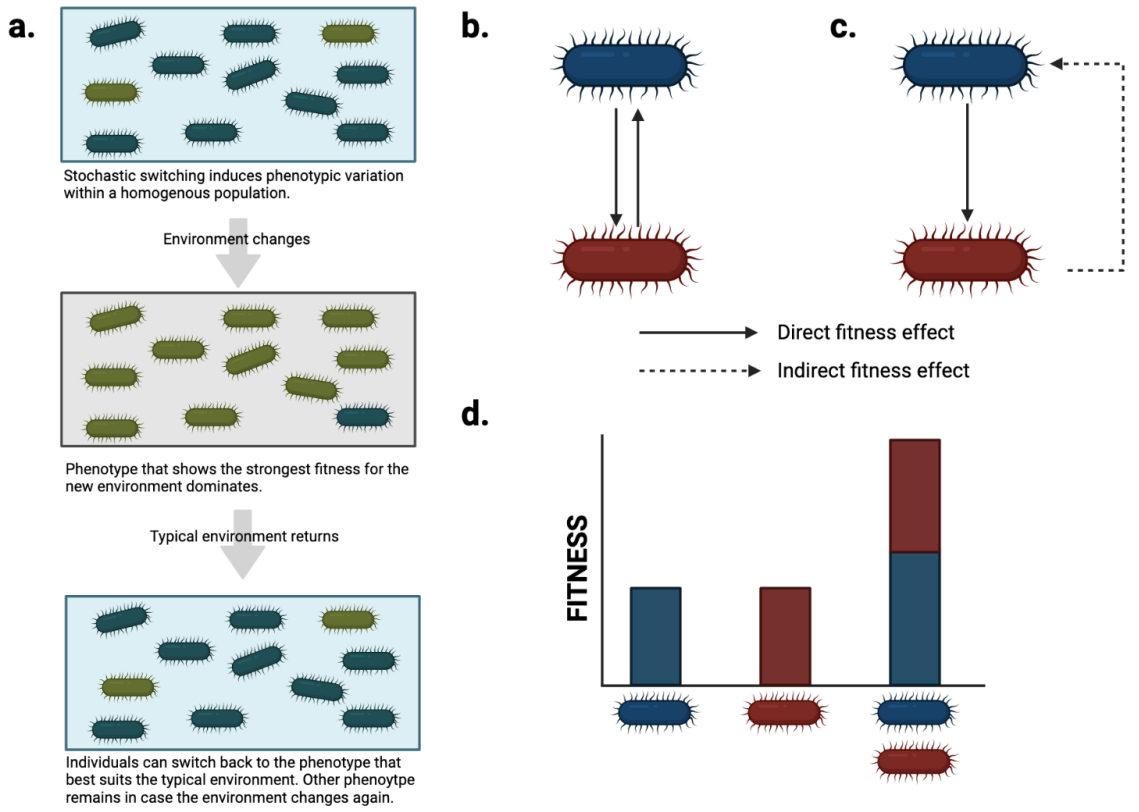


Figure 1.1 Phenotypic variation within a clonal population provides evolutionary advantages through bet-hedging or the division of labour. (a) a schematic demonstrating the process of bet-hedging. Having multiple phenotypes in a homogenous population can prepare a population for environmental changes, preventing the population from growing extinct. **(b)** bidirectional division of labour, where tasks are divided in a way that is mutually beneficial. **(c)** unidirectional division of labour, one individual carries out a task for the rest of the population increasing the fitness of the population as a whole and therefore providing an indirect fitness effect on the individual. **(d)** fitness of the population when two phenotypes are present is higher than if the population had only one phenotype.

1.3 The Proximate Causes of Phenotypic Variation are Numerous

At the molecular level, phenotypic variation in bacteria is regulated by both genetic and epigenetic mechanisms including: site-specific DNA inversion, slipped strand mispairing, homologous recombination, copy-number variation, methylation, or transcription factor activation (and associated feedback loops) (Bikard & Marraffini, 2012; Davidson & Surette, 2008; Van der Woude & Bäumler, 2004). Some of these processes are more prevalent in certain bacteria compared to others. For example, inversion is common in *Escherichia coli* (*E. coli*) (Loiko et al., 2017) while slipped strand mispairing and homologous recombination are common in the *Neisseria* genus (Callaghan et al., 2006; Hamilton & Dillard, 2006; Rotman & Seifert, 2014; Sadarangani et al., 2016; Zelewska et al., 2016). Below I discuss four of these phenotypic variability-generating mechanisms in more detail. Three of these are genetic mechanisms that vary on very short time scales; the fourth is an epigenetic mechanism, DNA methylation, whose rate of variation is not well-established.

1.3.1 Site Specific DNA Inversion

Site specific recombination is a process that involves DNA excision at two defined sites and their subsequent re-annealing to new homologous DNA strands (Olorunniji & Stark, 2010). It involves the actions of a specific recombinase enzyme that identifies pertinent sites for exchange (Olorunniji & Stark, 2010). In the context of bacterial phenotypic variation this refers to the rapid reversible inversion of short sequences, called invertible regions, catalysed by invertase enzymes (Emerson et al., 2009; Goldberg et al., 2014). Invertible regions are flanked by invertible repeats which act as invertase target sites (Emerson et al., 2009; Goldberg et al., 2014; Schwan, 2011). Typically, promoters for target genes are contained inside these invertible regions and their orientation mediates the expression of the downstream gene operon (Jiang et al., 2019). When the promoter is oriented towards the operon, transcription is enabled; when the promoter is directed away from the operon, transcription is inhibited (Coyne et al., 2003; Jiang et al., 2019). Site specific DNA inversion such as this is a genetic mechanism used to regulate phase variation in bacteria.

One of the best studied examples of inversion-mediated phase variation in bacteria is the regulation of the *fimA* and *fimH* genes that encode Type 1 fimbriae in uropathogenic *E. coli*

(**Fig. 1.2**). Switching of *fim* genes allows bacteria to alternate between piliated (ON phase) and non-piliated (OFF-phase) forms (Schwan, 2011). Phase variation within this region involves the inversion of *fimS,* a 314 base pair invertible region containing the *fimA* promoter (Hinde et al., 2005). This inversion is controlled by two regulatory genes *fimE* and *fimB*, located upstream of *fimS,* whose gene products are site specific recombinases (Emerson et al., 2009). These homologous proteins bind to inverted repeats that flank *fimS* and mediate ON/OFF switching (Zhang et al., 2016). FimB coordinates both ON to OFF and OFF to ON switching with equal affinities; FimE appears to be biased towards ON to OFF switching, demonstrating a lower compatibility with regulatory sites after inversion (Schwan, 2011; Zhang et al., 2016). Each of these recombinases are under the control of their own promoter allowing the rate of switching to be regulated by the expression of each protein independently of the other (Dorman & Bogue, 2016). When expressed in equal amounts, switching from the ON phase to OFF phase occurs more frequently due to the directional bias of fimE (Schwan et al., 2018). FimE mediated switching occurs at a rate of 0.3 per cell per generation, compared to FimB mediated switching which occurs at a rate of 10^{-3} to 10^{-4} per cell per generation (Blomfield et al., 1991; Gally et al., 1993; Schwan, 2011).

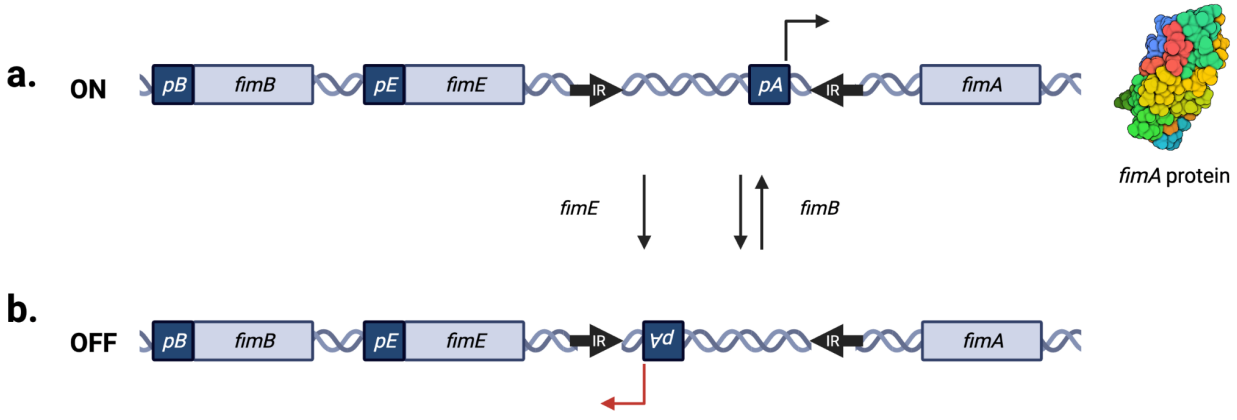


Figure 1.2 Expression of the fimA gene is regulated by site specific inversion of the fimA promoter. pB, pE and pA are the promoters for fimB, fimE and fimA respectively. IR represents inverted repeats. **(a)** in the ON phase pA is oriented towards the fimA gene - fimA is expressed. **(b)** in the OFF phase pA is oriented away from fimA - fimA is not expressed. Inversion is mediated by the invertases fimE and fimB whose genes are located upstream of fimA.

1.3.2 Slipped-Strand Mispairing

Simple sequence repeats (SSRs), or microsatellites, are genomic regions that contain tandem repeat sequences, usually between one and six nucleotides in length (Clayton et al., 2017; Fazekas et al., 2010; Wöhrmann & Weising, 2011). The reversible stochastic expansion or reduction of microsatellites results in variable expression and downstream changes in cell phenotypes (Alamro et al., 2014). SSRs are hypermutable regions prone to insertion or deletion by slipped-strand mispairing, at a rate of 10^{-2} to 10^{-5} per generation (Zhou et al., 2014) **(Fig. 1.3)**. During DNA replication the DNA polymerase and newly generated DNA strand form a complex and can detach from the template strand in error (Sehn, 2015). When dissociation and reattachment happens in a repeat region, the polymerase may reattach a few base pairs upstream or downstream of the original detachment site, causing a misalignment of complementary DNA sequences (Farnoud et al., 2016; Safi et al., 2019). Misalignment can result in the gain or loss of repeat nucleotides, which may lead to frameshift mutations, causing premature stop codons and truncated or non-functional proteins. Frameshifts may also change crucial spacing in promoter regions, altering transcriptional activity (Green et al., 2019; van der Woude, 2011).

Slipped strand mispairing is common in bacteria of the *Neisseria* genus, which live inside animal hosts and closely associate with their mucosal surfaces (Callaghan et al., 2006). The two pathogenic bacterial species within this genus, *N. gonorrhoeae* and *N. meningitidis*, are both able to undergo phase variable switching of proteins on their cell surfaces (Tauseef et al., 2013). The opacity-associated (*opa*) adhesion proteins present on the outer cell membrane of both species are an example of proteins that are under phase variable control. *Opa* proteins mediate interaction with host tissues, contribute to bacterial virulence and regulate the immune reaction of host cells (Alamro et al., 2014; Callaghan et al., 2006; Sadarangani et al., 2016). *N. gonorrhoeae* and *N. meningitidis* have 11-12 and 3-4 *opa* loci, respectively, which can be variably expressed through the mechanisms of slipped-strand mispairing (Rotman & Seifert, 2014). The *opa* gene reading frame contains the pentameric repeat 5'-CTCTT-3' that is subject to expansion or reduction by slipped-strand mispairing, causing a downstream frameshift mutation that can alter the position of the ATG initiation codon within the *opa* reading frame (Sadarangani et al., 2011, 2016; Wisniewski-Dyé & Vial,

2008). The number of genes expressed, and therefore the number of proteins present on the cell surface, varies between each individual cell and is directly dependent on the number of repeats in the *opa* leader sequence. Because of this, *opa* genes have many ON forms and a single OFF form where a frameshift mutation causes the expression of a truncated, non-functional protein (Wisniewski-Dyé & Vial, 2008). Phase variable control of *opa* proteins is an important mechanism that allows bacteria within the *Neisseria* genus to adapt to the different host environments that they encounter during infection (Wisniewski-Dyé & Vial, 2008).

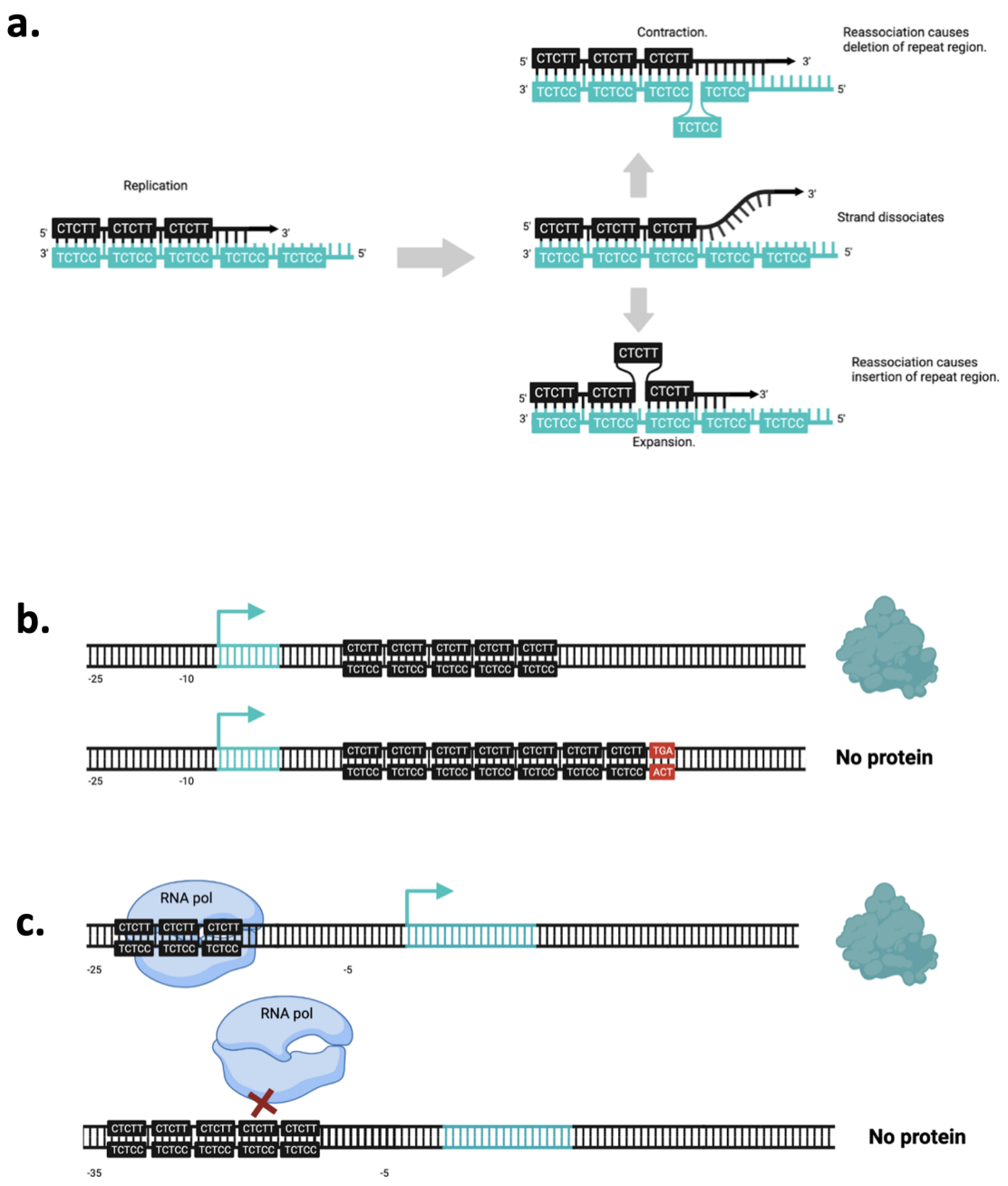


Figure 1.3 Slipped strand mispairing at microsatellites can cause expansion or contraction of repeat sequences. (a) Schematic of the general mechanism of slipped strand mispairing. Detachment of the replicating strand from the template strand can cause insertion or deletion of repeating sequences when the polymerase-DNA complex reattaches incorrectly. **(b)** insertion of repeats within a reading frame leads to downstream frameshift mutations and induces a premature stop codon. **(c)** insertion of repeats within a promoter region can prevent the binding of transcription factors or transcriptional regulator proteins stopping gene expression. Figure adapted from (Hansson, 2018).

1.3.3 Homologous Recombination

In prokaryotic organisms homologous recombination describes the unidirectional transfer of DNA between two genetic sequences of high homology at equivalent genomic loci. General homologous recombination in prokaryotes is equivalent to sex in eukaryotes (Vos, 2009) with the exception that exchange is not reciprocal, analogous to gene conversion (Bobay, 2020). As with the mechanisms above, recombination provides a method for rapidly generating allelic variation (Narra & Ochman, 2006). The process requires long homologous sequences, of at least 50 base pairs and the recruitment of general homologous enzymes - specifically the recombinase *RecA* (Criss et al., 2010; Piazza & Heyer, 2019). Because recombination events can occur at any location along these sequence lengths, a variety of new genetic combinations can be produced, wherein genetic material from the recipient site is displaced with the corresponding sequence from the donor site. This results in a loss of the original sequence in the target site (Paulsson et al., 2017). This mechanism of generating genotypic and phenotypic variation is exemplified in the pathogenic bacteria *Neisseria gonorrhoeae* (*N. gonorrhoeae)*.

N. gonorrhoeae possesses a programmed homologous recombination system that allows for variability in the expression of the gene encoding the pilin fibrous protein **(Fig. 1.4)** (Prister et al., 2020). Pilin forms an essential subunit of Type IV pili, a cell-surface protein polymer that has critical roles in epithelial, mucosal and immune cell adhesion, as well as cell motility, two functions that facilitate *N. gonorrhoeae* virulence (Duffin & Barber, 2016; Hill et al., 2016; Rotman et al., 2016). The pilin protein is encoded by *pil* genes that occupy multiple genomic loci. An expression locus, *pilE,* contains the sequence for the functional *pilE* pilin protein; multiple storage loci, *pilS*, contain variations of this sequence that code for truncated non-functional proteins (Cahoon & Seifert, 2013; Hill et al., 2016). These unexpressed “storage” loci contain variable genetic material interjected with conserved DNA sequences that function to facilitate the exchange of genetic material (Davies et al., 2014; Foley, 2015). To create variant *pilE* alleles, these loci undergo unidirectional transfer of genetic material from *pilS* to *pilE*, mediated by *RecA* (Davies et al., 2014; Duffin & Barber, 2016). As a result of this transfer, the corresponding *pilE* sequence is displaced from the genome, while the

pilS locus remains unchanged (Foley, 2015). Unidirectional gene recombination causes the expression of unique and variable *pilE* polypeptides.

The final type IV pilin proteins are composed predominantly of large repeating units of *pilE* (Cehovin et al., 2010; Rotman et al., 2016). The capacity for large amounts of variance in the *pilE* gene means that a multitude of *pilE* protein products are able to be generated, each with potentially different structures and assorted biochemical compositions (Wachter & Hill, 2016). Type IV pili are located on the outer cell surface of pathogenic *N. gonorrhoeae* cells making them the target of host immune cells (Voter et al., 2020). Variation of these pili, through gene conversion at the *pilE* and *pilS* loci, prevents immune detection, facilitating the evasion of host immune responses and limits cross-protection between hosts (Cehovin et al., 2010; Obergfell & Seifert, 2016).

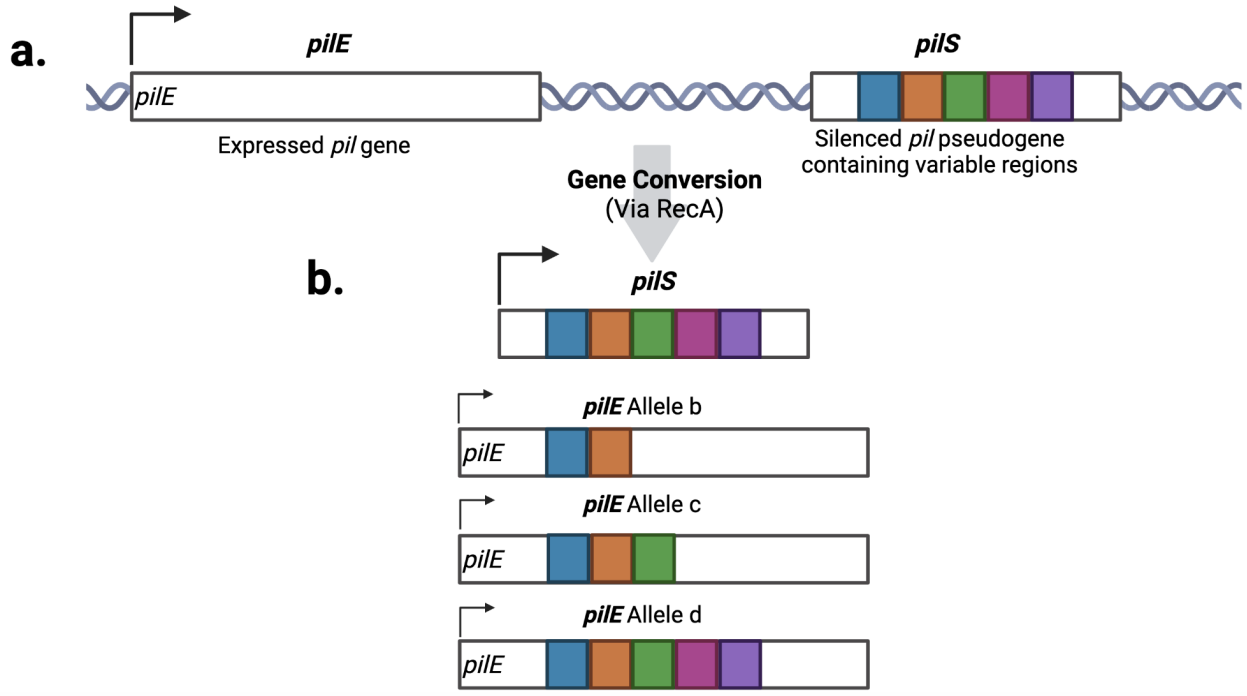


Figure 1.4 Genetic variation in the pilE gene is generated by site specific, unidirectional recombination (gene conversion) between the pilE expression loci and the pilS storage loci. (a) pili expression loci with the pilE promoter and the pilS silenced gene with variable regions. Coloured squares represent variable regions. **(b)** homologous recombination events mediated by RecA result in variable pilE alleles, causing the expression of different proteins. pilE replaces its original sequences with the new variable regions from pilS, pilS remains unchanged.

1.3.4 Methylation

Finally, phenotypic variation can be rapidly generated through epigenetic mechanisms. There are a wide range of epigenetic mechanisms, but they are unified in their ability to affect phenotype in a heritable manner without changing the sequence of the DNA. A common mechanism of epigenetic regulation in eukaryotic organisms is changes to chromatin structure, often through covalent modifications (Adhikari & Curtis, 2016; Savidge, 2016). In bacteria there is no chromatin, but covalent modifications of DNA can affect regulation, for example, methylation (Beaulaurier et al., 2019). DNA methylation regulates gene expression by interfering with the interactions of binding proteins with their target DNA sequence, in either an inhibitory or facilitatory manner (Sánchez-Romero & Casadesús, 2020). In *E. coli,* DNA can be methylated by the enzyme deoxyadenosine methyltransferase (DAM) (Beaulaurier et al., 2019). DAM recognizes the sequence 5'-GATC-3' and has been identified to have roles in a wide range of cellular processes (Davies et al., 2014; van der Woude, 2017), such as biofilm formation, via the expression of the protein Antigen 43.

Antigen 43 (Ag43) is an antigenic outer membrane protein present in *E. coli* that is encoded by the *agn43* gene (Kumar & Rao, 2013). Ag43 encourages the auto-aggregation of cells and is believed to facilitate microcolony and biofilm formation (Haagmans & van der Woude, 2000; Wallecha et al., 2003). Expression of *agn43* is controlled through the opposing actions of a repressor protein (called OxyR) and DAM, which takes on the role of an activator, aiding the binding of RNA polymerase to encourage gene expression **(Fig. 1.5)** (Broadbent et al., 2010; van der Woude, 2011). Methylation of the target site by DAM physically inhibits the binding of OxyR, suppressing it from performing its repressive action; when OxyR is bound, DAM is unable to access the DNA preventing methylation (Chauhan et al., 2013; Wallecha et al., 2003). Methylation of the OxyR binding site switches expression of the *agn43* gene ON while binding of OxyR switches *agn43* gene expression OFF.

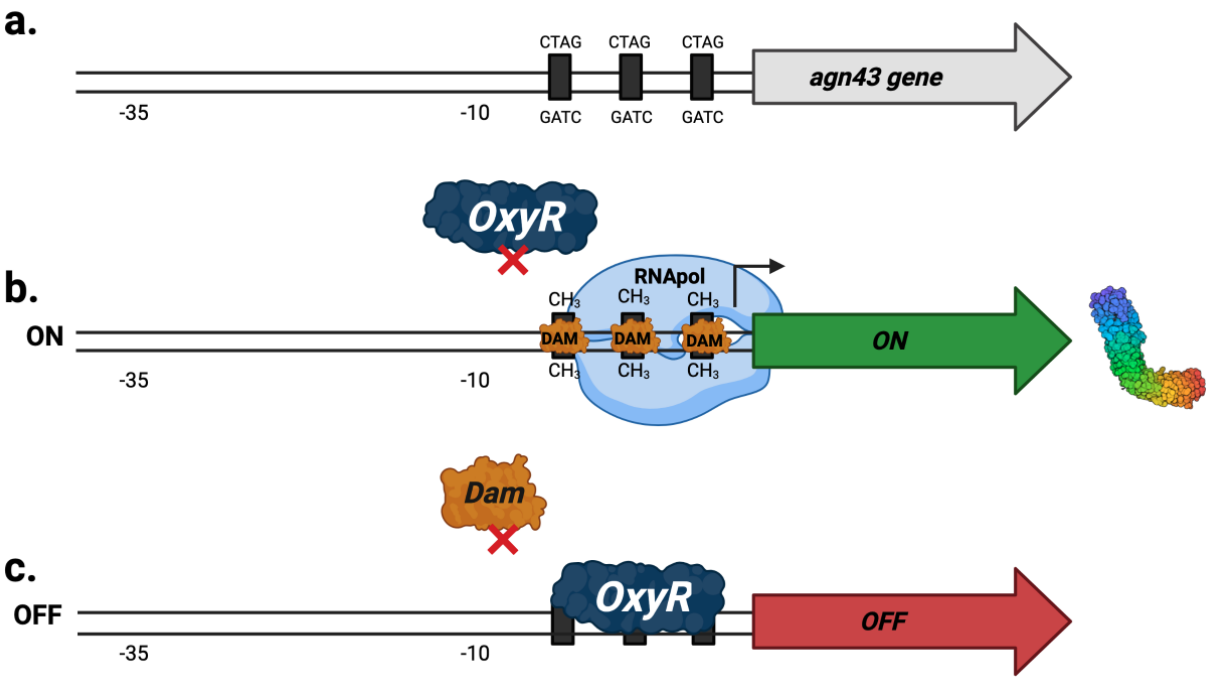


Figure 1.5 Expression of the agn43 gene is regulated by methylation of the promoter site. Black boxes represent deoxyadenosine methyltransferase's (DAM) GATC recognition sites. OxyR is a repressor protein. **(a)** schematic showing the structure of the agn43 gene and its promoter region. **(b)** fully methylated GATC sites by DAM prevents binding of OxyR repressor, but facilitates the binding of RNA polymerase (RNApol): the protein is expressed. **(c)** OxyR repressor binds where GATC sites are not methylated. Binding of OxyR prevents binding methylation of GATC sites by DAM: agn43 gene is not expressed. Adapted from (Seib & Jennings, 2021).

1.4 Antibiotic persister cells and their clinical relevance

A well-known, yet poorly understood example of phenotypic heterogeneity within bacterial populations is the formation of persister cells. Persister cells are genetically identical phenotypic variations of a clonal population that possess temporary multi-drug tolerance (Lewis, 2010; Wood et al., 2013). By transiently switching to a physiologically dormant state, subpopulations of persisters can withstand exposure to high concentrations of antibiotics and are selected for after prolonged antibiotic exposure (Mechler et al., 2015; Van den Bergh et al., 2016). It is believed that entering a dormant state renders antibiotic targets inactive, effectively removing the antibiotics' bactericidal activity (Harms et al., 2016).

In contrast to resistant individuals, persisters are unable to grow during antibiotic treatment (Harms et al., 2016; Maisonneuve & Gerdes, 2014) but are able to re-enter a proliferative state once the stressor is removed, resulting in a population consisting largely of the original susceptible cells (Pu et al., 2017). Populations harbouring persister cells have a characteristic bi-phasic killing curve **(Fig. 1.6b)** (Fauvart et al., 2011). After exposure to stress a rapid initial exponential decline in numbers is followed by a much shallower rate of decline that is indicative of persister cells (Dawan et al., 2020; Harms et al., 2016). The second, much slower, decline is reflective of the rate at which persisters leave their persister state to become susceptible cells, which then die (Barrett et al., 2019).

The discovery of persister cells occurred not long after the discovery of the bactericidal effects of antibiotics (Bigger, 1944; Hobby et al., 1942). There is intense interest in persister cells because antibiotics are used to treat a variety of infections and diseases in both animals and humans. Medicine's reliance on antibiotic treatment has ultimately resulted in the rise of genetically antibiotic resistant strains of bacteria (Gould, 2009; Laxminarayan et al., 2013; Wellington et al., 2013). Persister cells exhibit resistance because of epigenetic changes (Cohen et al., 2013; Levin-Reisman et al., 2017; Mechler et al., 2015; Van den Bergh et al., 2016). In addition to resistance, persisters have been associated with latent infections, such as that of *M. tuberculosis* and *Salmonella enterica* (*S. enterica*) (Hill & Helaine, 2019; Torrey et al., 2016).

Persister cells were first identified in 1944, when J.W Bigger discovered that *Streptococcus pyogenes* (*S. pyogenes*) were not always completely killed when treated with penicillin

(Bigger, 1944). Relative to the time of their initial discovery, little is currently known about the mechanisms that generate persister cells. The most widely recognised mechanism of bacterial persistence is the toxin-antitoxin (TA) system first discovered through the actions of the protein *HipA* (Balaban et al., 2004; Kim & Wood, 2016; Li et al., 2016; Page & Peti, 2016). TA systems are a pair of complementary genes that encode a toxin and its corresponding anti-toxin protein, which acts as an antidote by neutralising toxin activity (Jankevicius et al., 2016). The stable toxin protein interferes with essential cellular processes in such a manner that causes cell dormancy, but its effects are counteracted when the unstable anti-toxin protein is expressed (Hall et al., 2017; Ronneau & Helaine, 2019). During stressful conditions the unstable anti-toxin is degraded, causing the toxin to ultimately send the cell into a dormant state (Page & Peti, 2016). The association between TA systems and bacterial persisters suggests that persistence might be a stress response, where persisters have become dormant to survive the adverse environment.

More recently studies have demonstrated a stochastic, random origin for persistence, suggesting their origin is not a response to environmental change but results from heterogeneity in the population (El Meouche et al., 2016; Harms et al., 2016). Pre-emptive heterogeneity before antibiotic exposure would suggest more of a bet-hedging strategy for persistence rather than a responsive mechanism after the antibiotic is encountered. Further supporting the idea of stochastic persistence, Jiang et al. found an enrichment of invertible regions in known antibiotic resistant genes in *E. coli* (Jiang et al., 2019). The results from the 2019 study suggest that the two possible systems are not mutually exclusive; stochastic generation of persister cells may be through the regulation of TA systems.

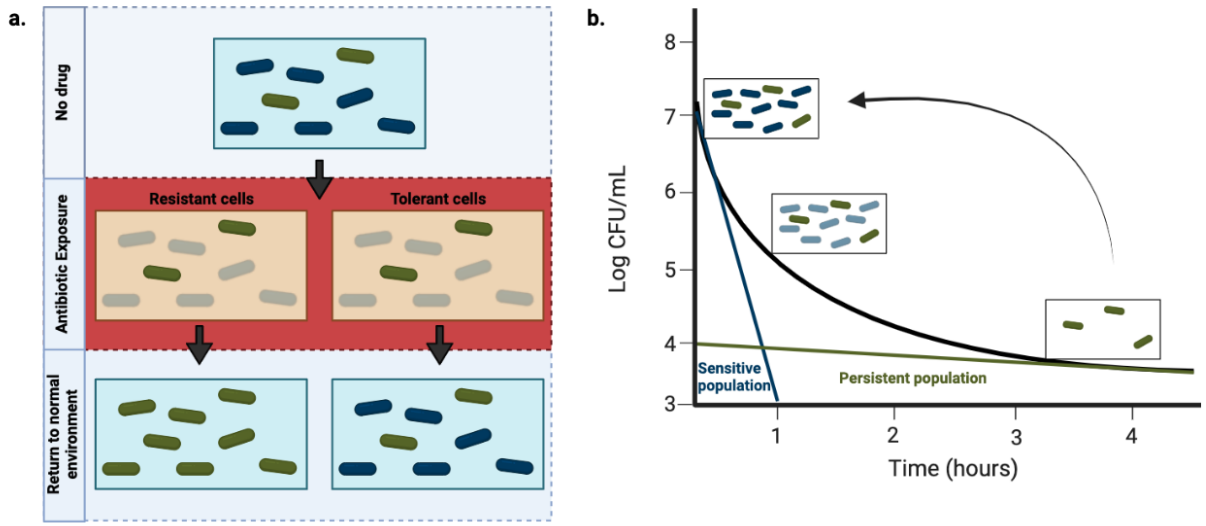


Figure 1.6 Persister cells display multi drug tolerance. (a) The difference between tolerance and resistance. Tolerant cells survive drug exposure by entering a dormant state, then switch back to a proliferative state, regaining susceptibility to antibiotics. Resistant cells grow during antibiotic exposure and produce more identical, resistant cells. **(b)** the biphasic killing curve of populations containing antibiotic persister cells. Figure adapted from (Singh, 2017).

1.5 Research hypothesis and objectives

We first identified a novel, rapid phenotypic switching phenomenon in a natural isolate of *E. coli* (SC375), where an initially isogenic bacterial population formed two extreme phenotypes consisting of “short” and “long” cells, which grew up to 7 times the length of the shortest cells. The two differentiated phenotypes, long and short, suggested a phenotypic switching mechanism. Based on the knowledge of phase variable expression within the *fim* operon of uropathogenic *E. coli*, we first hypothesised that the mechanism behind this switch may be as a result of a similar inversion procedure.

We hypothesise that the phenotypic switching demonstrated by the *E. coli* natural isolate SC375 is under phase variable control regulated by a DNA inversion mechanism.

In order to confirm our hypothesis we will address four objectives:

1. **Confirm the presence of distinct cellular phenotypes within isogenic populations of SC375.**

Cell sizes will be confirmed using both microscopy and flow cytometry. To characterise the dynamics of phenotype switching, we will use time-lapse “movies” of images taken over a 12-24 hour period.

2. **Perform DNA sequencing of short and long phenotypes to pinpoint genetic differences between phenotypes.**

DNA sequencing will be achieved through the use of Oxford Nanopore sequencing to allow identification of structural changes between the two phenotypes.

3. **Compare gene expression levels between the two cellular phenotypes.**

To achieve this, RNA-sequencing will be conducted to identify differences in the expression of any key genes between both cellular phenotypes.

4. **Confirm differences in gene expression between phenotypes using expression reporters.**

To confirm differences in expression and understand the dynamics of gene expression we will construct reporters that allow *in vitro* tracking of gene expression levels.

Chapter 2: Materials and methods

2.1: Strains

Table 2.1. Strains used in this study

Strain	Description	Source
SC375	Natural isolate derived from SC strain collection.	(Ishii et al., 2006)
TOP10	One Shot ™ TOP10 chemically competent *E. coli* cells.	Invitrogen™ Catalogue number: C404010

2.2: Plasmids

Table 2.2 Plasmid constructs used in this study

Plasmid Name	Insert	Backbone	Selectable marker	Source
pUA66	N/A	pUA66 **(Fig.2.1)**	Kanamycin resistance	(Zaslaver et al., 2006)
pfim::GFP	*fim* operon promoter	pUA66	Kanamycin resistance	
pgfc::GFP	*gfc* operon promoter	pUA66	Kanamycin resistance	
paida::GFP	*aidA* gene promoter	pUA66	Kanamycin resistance	
plpp::GFP	*lpp* gene promoter	pUA66	Kanamycin resistance	

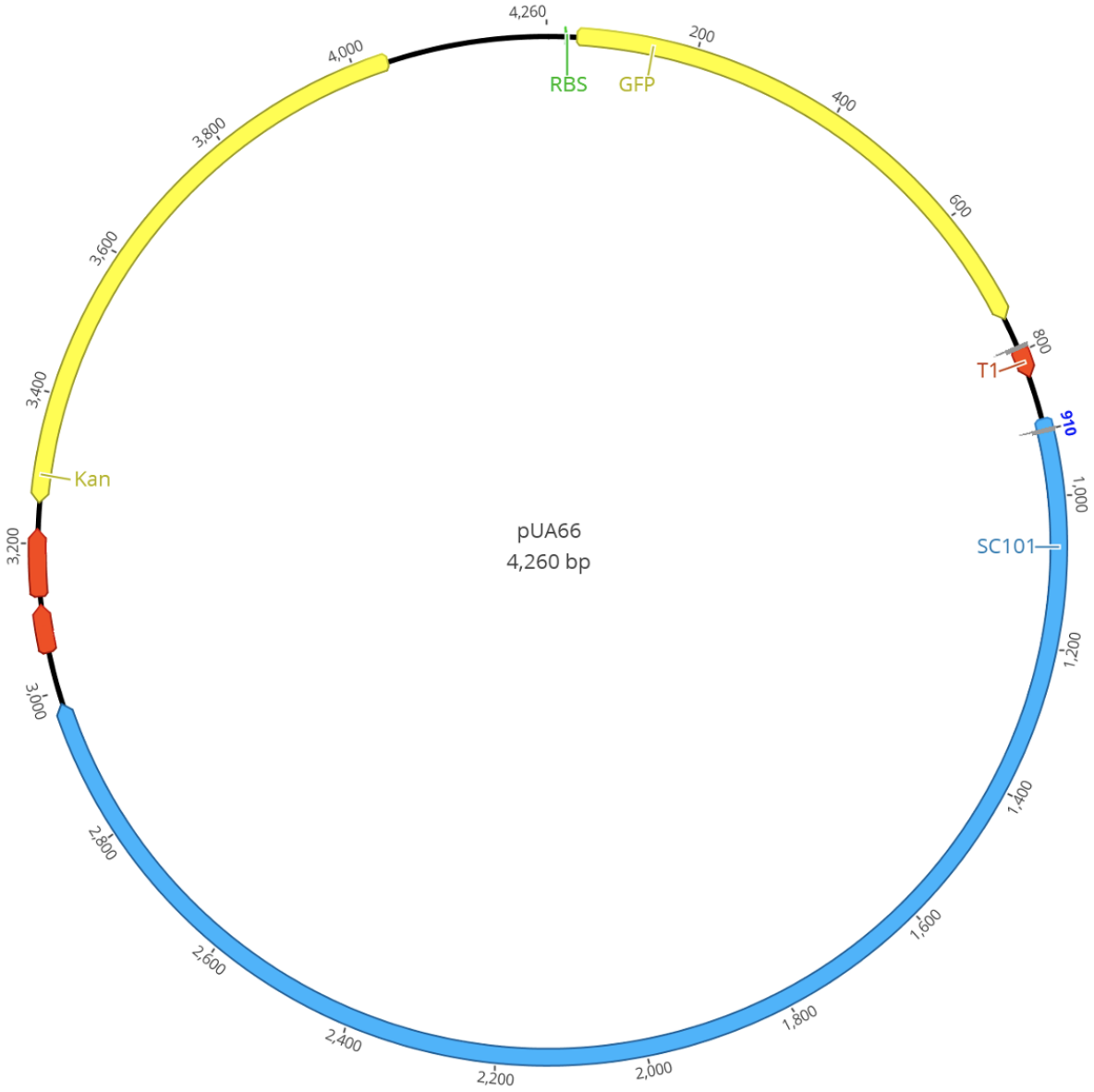


Figure 2.1 Plasmid map of pUA66. This construct contains a GFP reporter downstream of a strong ribosomal binding site and followed by a strong terminator. All promoters were cloned immediately upstream of the ribosomal binding site.

2.3: Media

All media was made with autoclaved components, sterilised at 121^{O}C for 15 minutes (unless specified otherwise), all water used was deionised. All plates were made with 1.5% agar.

Lactose stock solution:

We dissolved 10% Lactose Monohydrate (AJAX FineChem™) measured by weight, in water. The solution was sterilised on a sugars autoclave cycle at 118^{O}C for 20 minutes.

1X magnesium sulphate stock solution:

We dissolved 120.366g/mol of anhydrous magnesium sulphate (Scharlab) in water (measured by weight).

1X calcium chloride stock solution:

We dissolved 147.02g/mol of anhydrous calcium chloride in water (measured by weight).

10X M9 salts stock solution:

We dissolved 112.8g/L of M9 minimal salts in water.

Liquid M9 minimal media with lactose:

For 100mL of media:

10X Lactose	8mL
10X M9 salts stock solution	10mL
1X magnesium sulphate	200μL
1 X calcium chloride	10μL
Water	Up to 100mL

Liquid M9 minimal media with lactose for agar plates:

For 100mL of media:

10X Lactose	16mL
10X M9 salts stock solution	20mL
1X magnesium sulphate	400μL
1 X calcium chloride	20μL
Water	Up to 100mL

10X PBS stock solution (ph 6.8):

For 100mL of 10X PBS:

Disodium phosphate	1.78g
Potassium phosphate	0.24g
Sodium chloride	8g
Potassium chloride	0.2g

1X PBS solution (pH 7.4):

For 100mL:

10X PBS stock solution	10mL
Water	90mL

Liquid Luria Bertani Broth (LB):

We dissolved 25g of Luria Bertani Broth Base (Miller's LB Broth Base™) (Invitrogen) in 1 litre of water.

Liquid Luria Bertani Broth (LB) for agar plates:

For 200mL:

Luria Bertani Broth Base (Miller's LB Broth Base™) (Invitrogen)	10g
Agar	3g
Water	200mL

2.4: Antibiotics

Kan50

We dissolved Kanamycin sulphate (VWR Chemicals) in water for a final concentration of 50μL/100mL. From here onwards Kanamycin sulphate at a concentration of 50μL/100mL, will be referred to as Kan50.

2.5: Storage of bacterial cultures

We streaked bacterial strains on LB agar media and grew them at 37°C overnight for at least 12 hours. Colonies were used to inoculate LB liquid media and grown again at 37°C overnight for at least 12 hours. Strains containing plasmids with a Kanamycin resistance selection gene were grown with the addition of Kan50 at a ratio of 1μL:1mL of total liquid volume. 750μL of this overnight culture was added to 350μL of 30% glycerol. We froze samples in duplicate at -80°C for permanent storage.

2.6: Cell culturing

We grew all bacterial cultures at 37°C in either liquid or solid agar media. Liquid cultures were grown overnight for between 12 to 18 hours in a 15mL falcon tube with shaking at 250 RPM; strains plated on agar media were grown for up to 48 hours in order to produce

colonies large enough to be phenotypically identified visually. We inoculated all overnight cultures from a SC375 glycerol stock frozen at -80°C unless specified. We cultured SC375 cells containing plasmid constructs in the same manner, with the addition of Kan50 at the aforementioned ratio of 1μL:1mL of total liquid volume.

2.7: Colony Selection

We plated 20μL of a 1.0e5-fold dilution of an overnight culture of the SC375 strain onto each of 12 agar plates. Plates were incubated overnight at 37°C. We selected colonies by visual characterisation of their morphology type. For analysis of SC375 cells containing plasmid constructs, colonies were selected based on their ability to fluoresce under blue light, and classification as either fluorescent, non-fluorescent or mixed.

2.8: Flow cytometry of resuspended colonies for cell size and fluorescence

For flow cytometry of colony morphology, we selected and resuspened each three of each colony type in 100μL of 1X PBS each in a single well of a 96 well microplate (resuspended colonies created biological replicates). Each resuspension was further diluted serially up to 1.0e3-fold in 1X PBS with 2% formaldehyde to create technical replicates.

For analysis of SC375 cells containing plasmid constructs, we selected colonies based on their fluorescence ability and resuspended three of each colony type in 500μL of 1X PBS with 2% formaldehyde to create biological replicates. We pipetted 150μL of each resuspended colony solution into individual wells on a 96 well plate to create technical replicates. Technical replicates were diluted serially up to 1.0e2-fold.

The lowest dilutions were used for data collection to avoid overflow. We collected forward and side scatter data as indicators for cell size and complexity, as well as GFP data to measure gene expression. GFP fluorescence was collected using a 488 nm excitation and 517/13nm bandpass filter. We exported the data into Flow Cytometry Standard (FCS) files. We analysed the data in R (version 4.0.3; R Core Team 2021) using the flowCore package (version 1.44.1; Hahne et al., 2009).

2.9: Microscopy

We performed microscopy using a Nikon Eclipse Ti2-E inverted microscope in conjunction with the NIS-Elements imaging software.

2.9.1: Colony microscopy

We photographed colonies on agar plates using a dissecting microscope. To get higher magnification images of single colonies, we used the Nikon Eclipse Ti2-E inverted microscope. Small agar plates of M9 minimal media with 0.8% lactose were made in the wells of a six-well plate. We diluted overnight cultures of SC375 up to 1.0e7-fold, and grew them for 48 hours at 37°C. Colonies were imaged using a 10X phase contrast dry objective along with a phase contrast optical configuration. Images were denoised using the AI denoising module of the NIS-elements software (M. Davis, 2019).

2.9.2: Microscopy

We made microscope slides by cutting out slices of agar from premade plates with a sterile scalpel. Agar slices were put into a 15mL falcon tube and heated until liquid enough to be pipetted. Using a 1mL pipette, we transferred liquid agar into cavities on a double cavity microscope slide, placing a glass coverslip on immediately after transfer. Agar was left until solidified and the glass coverslip was removed. Using the same sterile pipette, we cut 40mm x 40mm squares in the centre of the cavity. We pipetted 1.5μL of liquid culture onto the centre of the slide and spread to the edges using the same pipette tip. Slides were left uncovered to dry for five minutes. We secured glass coverslips over dried pads with high-vacuum silicone grease (Dow Corning®), ensuring that the coverslip fully contacted the agar pad.

We imaged cells using phase contrast microscopy with a 100X oil immersion objective (NA 1.4). Final images were run through the NIS-elements AI denoising software before being exported to 16-bit.tiff files for analysis. For fluorescence imaging an additional 470 nm optical configuration was used and the intensity was adjusted to ensure there were no points of saturation. These images were not run through the denoising software to preserve the integrity of the fluorescence data.

For time lapse microscopy we diluted overnight cultures of SC375 grown for 12 hours in 1 X PBS to create 1.0e1-fold and 1.0e2-fold serial dilutions. We made microscope slides using the 1.0e2-fold dilution. We used an oil immersion 100X phase contrast objective and a phase contrast optical configuration to take images every two minutes over a 24 hour period. We ran final images through the NIS-elements AI denoising software before being exported to 16-bit.tiff files for analysis. Fluorescence data was collected in the same manner as mentioned above. When taking time-lapse images to measure fluorescence, phase contrast and fluorescence images were taken every three minutes, to reduce photobleaching. Once again, fluorescence images were not run through the denoising software to preserve the integrity of the fluorescence data.

For large scale culture microscopy at different time points, we resuspended triplicate biological samples of each colony phenotype (compact, mixed and halo OR fluorescent, non-fluorescent and mixed) in 200µL of M9 minimal media with 0.8% lactose (and when transformed strains were used, Kan50 at a volume of 1µL:1mL). We transferred 20µL of each colony resuspension into 180µL of fresh M9 media and left them to grow at 37°C for 12 hours with shaking at 250 RPM. We made microscope slides as outlined above with 1.5µL samples of each colony resuspension at 0, 2, 4 and 12 hour time points (for a total of 9 microscope slides at each time point). We used an oil immersion 100X phase contrast objective and a phase contrast optical configuration to take images at 5 different slide locations on each microscope slide. Slide locations were selected randomly based on incremental adjustments of XY values. Undiluted, resuspended colonies were used for imaging at 0 hours. Undiluted samples of each growth culture were used for microscopy imaging after 2, 4 and 12 hours of growth. Fluorescence and size data was collected in the same manner as above, with the addition of a 470 nm optical configuration.

2.9.3: Cell segmentation and analysis

We used the cell segmentation software SuperSegger from the Wiggins lab (Stylianidou et al., 2016) to perform cell segmentation.

2.10: Gibson Assemblies

For all reporter assays, we cloned inserted PCR products of the promoter region from each gene-of-interest into a pUA66 plasmid backbone using a Gibson Assembly method (Gibson, 2009) and NEB™ reagents.

2.10.1: Primer design.

We designed primers using *primer3* in Geneious version 9.1.8. Primer details are shown in **table 2.3** below.

Table 2.3. Final PCR primer pair sequences.

Gene	Primer Name	Primer Sequence	%GC content
fim	F3 fim fwd U590	5'-TTTCGTCTTCACCTCGAGGAGGTGGCCATTCTTCTCAGG–3'	53.8%
fim	F3 fim rev D115	5'-TAAATCTAGAGGATCCCGCAGATGCATTTAACCCGCC-3'	48.6%
gfc	F3 gfcA fwd U581	5'-TTTCGTCTTCACCTCGAACGGTCGTAAAGAAGTTCAGGTC–3'	47.5%
gfc	F3 gfcA rev D129	5'-TAAATCTAGAGGATCCCCGGTAGTGGTCGTTGTGGTG–3'	51.4%
aidA	F3 aidA fwd U1850	5'-TTTCGTCTTCACCTCGATATTCGCCATACTGTACGTTATACC–3'	42.9%
aidA	F3 aidA rev D250	5'-TAAATCTAGAGGATCCCACTTCCCCTTTAAAGTGAACGGTC–3'	43.9%

2.10.2: PCR amplification of promoter regions.

We cloned promoter regions of interest by PCR using the HiFi polymerase enzyme reaction. We determined annealing temperatures for primer pairs using the NEB TM Calculator (https://tmcalculator.neb.com/#!/main); resulting annealing temperatures are shown in **Table 2.4**. We set up reaction mixtures as per **Table 2.5**. PCR reactions for promoters *gfc*, *fim*, and

aidA-I were run for 25, 25 and 30 cycles respectively, with amplification time for all three promoter regions set to two minutes.

Table 2.4. Annealing temperatures of primers as determined by the NEB TM Calculator.

Primer pair	Annealing Temperature (°C)
Forward: F3 fim fwd U590 Reverse: F3 fim rev D115	63
Forward: F3 gcfa fwd U581 Reverse: F3 gcfa rev D129	65
Forward: F3 aidA fwd U1850 Reverse: F3 aidA rev D250	65
pUA66 vector primers	55

Table 2.5. PCR reaction mixture composition for insert amplification.

Reagents	Volume for 2 reactions (µL)	Volume for 3 reactions (µL)
ddH_20	65	97.5
5 x HF buffer	20	30
Forward Primer	5	7.5
Reverse Primer	5	7.5
DNTP mix	2	3
Phusion DNA polymerase	1	1.5
	Aliquot 49µL	Aliquot 49µL
+ Template DNA	1	1

2.10.3: Quality check of PCR products

To check for successful PCR amplification we mixed 5µL of PCR reaction with 1µL of 6X loading dye (Thermofisher) and loaded the mixture on a 1% agarose gel alongside a 1KB DNA ladder. We ran the gel for 30 minutes, with the voltage set to 60mV, using a 1X TAE buffer.

1% agarose gel

We dissolved 0.5g of agarose in 50µL of 1X TAE buffer and, once slightly cooled, we added 5µL of SYBR™ Safe DNA Gel Stain. We poured the gel into a casting tray and left it to set for one hour.

2.10.4: Determining concentration of PCR reactions

Once we confirmed that amplified inserts were the correct length, we measured the concentration of PCR products by mixing 5µL of reaction with 195µL of buffer and using an

Invitrogen Qubit Fluorometer and the Double Strandard Broad Range reagents. Concentrations were measured in ng/μL.

2.10.5: Gibson Assemblies

We performed Gibson Assemblies to ligate PCR promoter products into our pUA66 plasmid backbone, as per manufacturer's instructions (https://nebiolabs.co.nz/protocols/2014/11/26/nebuilder-hifi-dna-assembly-reaction-protocol). The required insert volume was calculated using the NEB ligation calculator: http://nebiocalculator.neb.com/#!/ligation, each reaction had an insert: plasmid ratio of 2:1. Reaction mixtures were set up as per **table 2.6.**

Table 2.6 Gibson Assembly reaction volumes.

Reaction volumes (μL)	
Recommended DNA molar ratio	Vector:Insert 1:2
NEBuilder HiFi DNA Assembly Master Mix	10
Vector backbone	2
Insert fragments	As determined by NEB ligation calculator
Deionised H_20	10 - insert volume
Total volume	20

2.11: Bacterial transformation

We first transformed Top10 cells with plasmid constructs. We then extracted plasmids from transformed Top10 cells and electroporated them into electrocompetent SC375 cells for further downstream analysis.

2.11.1: Making electrocompetent bacteria cells

We made overnight cultures of Top10 cells by inoculating 3mL of LB, in a 15mL falcon tube with a single Top10 colony. Inoculated LB was left to grow overnight at 37°C, with shaking at 250 RPM. 100mL of LB in a one litre flask was left overnight at the same conditions, to aerate the LB. 1mL of the overnight culture was transferred to the 100mL of pre-warmed and pre-shaken LB and left it to grow to an OD of 0.45-0.5 at 37°C, shaking at 250 RPM. Once the OD was reached, we immediately put the culture on ice, and swirled it in an ice slurry for 10 minutes. After swirling, the flask was kept on ice and left to incubate in the fridge for one hour. We then split the culture equally into two pre-chilled 50mL falcon tubes and centrifuged both tubes for 10 minutes at 4°C at 4200 x g. After centrifugation we poured off the supernatant, and resuspended the pellet in pre-chilled 10% glycerol, then replaced it in ice and refrigerated for 10 minutes. We repeated the centrifugation and ice incubation steps with resuspended pellet two more times, with the exception of the centrifuge now being spun at 4500 x g, and supernatant now being removed with a pipette. After the third round of centrifugation, we resuspended the pellet in the supernatant left behind after removal with a pipette. We pooled all cells together and transferred aliquots of 70μL to pre-frozen 1.5mL tubes. We stored electrocompetent Top10 cells at -80°C. This process was repeated with the SC375 strain to make electrocompetent SC375 cells.

2.11.2: Electroporation of electrocompetent cells

We transformed electrocompetent cells with assembled pUA66 plasmids by electroporation. We mixed 2μL of plasmid with 70μL of thawed electrocompetent top10 cells. We electroplated cell-plasmid mixtures at 1850mV, 500μL of warmed SOS-media was added immediately after electroporation. We transferred mixtures into 5mL microcentrifuge tubes and left to recover for one hour at 37°C, shaking at 250 RPM. After one hour, 200μL of 1.0e1-fold, 1.0e2-fold and two measures of undiluted cells were plated on pre-warmed LB plates containing Kan50, to ensure the cells retained the plasmids. We electroporated electrocompetent cells without plasmid constructs and plated them as a negative control. Plates were left to grow at 37°C overnight. Successful transformation was deemed to be colony growth on plates, alongside no growth of the negative control.

2.11.3: Plasmid extraction

We extracted plasmids were from electrocompetent cells following protocol Agilent StrataPrep Plasmid Miniprep Kit as outlined in: https://www.agilent.com/cs/library/usermanuals/public/400761.pdf.

2.11.4: Confirming gibson assembly ligation

We performed PCR of transformed electrocompetent cells to confirm correct insert ligation of gibson assemblies. We used a single colony to inoculate 3mL of LB with the addition of 3μL of Kan50. The inoculated culture was then grown overnight at 37°C with shaking at 250 RPM. We mixed 5μL of overnight culture with 95μL of water and incubated at 95°C for five minutes, to be used as template DNA. We set up PCR reactions as outlined in **Table 2.7**. We ran PCR products on a 1% agarose gel to confirm expected sequence length. We performed an ExoSAP clean up on PCR products to prepare them for sequencing (outlined in **Table 2.8**). Samples were sent to Macrogen, Inc. to be sequenced.

Table 2.7 PCR reaction mixtures for insert sequencing preparation.

Reagent	Concentration	Volume per 1 reaction (μL)	Volume per 4 reactions (μL)
ddh_20		3	12
ThermoScientific DreamTag Hot Start PCR Master Mix	2X	5	20
Primer F	10μM	0.5	2
Primer R	10μM	0.5	2
			Aliquot 9μL
Add template DNA		1	

Table 2.8 ExoSAP Reagent setup

Reagents	Volume per reaction (µL)	Volume per 4 reactions (µL)
CutSmart buffer	2	8
Exo 1 buffer	2	8
Exo 1 enzyme	1.25	5
rSAP enzyme	2.5	10
Final Volume	7.75	31
Aliquot per reaction	7.75	7.75
Final volume per reaction (incl PCR products)	20.25	20.25

2.12: Genome Sequencing

2.12.1: DNA extraction

We performed DNA extraction following the promega protocol 3.G “Isolating genomic DNA from Gram Positive and Gram Negative Bacteria" as outlined in: https://worldwide.promega.com/-/media/files/resources/protocols/technical-manuals/0/wizard-genomic-dna-purification-kit-protocol.pdf?la=en, with four minor changes: Protein Precipitation Solution was pre chilled at 3°C before use; steps 12 & 13 and 16 & 17 were repeated twice each; sterile water was used for the rehydration of DNA in place of DNA Rehydrate Solution; and DNA was left to rehydrate overnight at room temperature.

2.12.2: Quality check of extracted DNA

We checked DNA quality (fragmentation) by mixing 5µL of extracted DNA with 1µL of 6X loading dye (Thermofisher) and loaded it on a 0.7% agarose gel alongside a 1 Kbp DNA ladder. We ran the gel for one hour, with the voltage set to 120mV, using a 1X TAE buffer.

0.7% agarose gel

0.35g of agarose was dissolved in 50µL of 1X TAE buffer and, once slightly cooled, 5µL of SYBR™ Safe DNA Gel Stain was added. We poured the gel into a casting tray and left it to set for one hour.

1X TAE Buffer

For 1 litre:

Tris-Acetate-EDTA 50X solution (Fisher BioReagents)	20mL
Water	980mL

2.12.3: Nanopore sequencing of extracted genomic DNA

We performed sequencing using the Oxford Nanopore MinION following the Rapid Barcoding Sequencing procedure as outlined in: https://community.nanoporetech.com/protocols/rapid-barcoding-sequencing-sqk-rbk004/checklist_example.pdf, skipping the optional concentration steps. The nanopore was set to run for 72 hours. The data was basecalled using Guppy 4.0.

2.12.4: Genome assembly and alignment

Flye (version 2.7.1) was used to create de novo genome assemblies from the nanopore .fastq output data. The assemblies were polished with racon (Vaser et al., 2017), followed by medaka tools (Huang et al., 2020). Consensus sequences were aligned using Mauve for visualisation (Darling et al., 2004).

2.13: RNA-sequencing

2.13.1: RNA extraction

We quenched exponential phase cultures using a 5% phenol in ethanol solution at a volume of 50% of the total culture in each tube. Cultures were incubated on ice for 15 minutes, then centrifuged at 4°C for seven minutes at 7000 x g. Following centrifugation, we discarded the supernatant and re-dispersed pellets in RNAlater solution and stored them in a refrigerator overnight at 4°C. The following day we processed samples using the standard protocol from

Zymo-quick RNA miniprep with the addition of a lysozyme digestion step. We prepared the lysozyme solution using TE buffer with a concentration of 2mg/mL. We incubated the bacterial pellet in RNAlater with the lysozyme solution for approximately five minutes before proceeding with the quickRNA miniprep protocol.

2.13.2: RNA-sequencing

We sent extracted RNA to Custom Science Ltd., where samples were sequenced by Illumina's stranded RNA sequencing, which also included rRNA depletion.

2.13.3: Analysis of sequencing data

We mapped RNA Illumina sequencing results using bwa mem (Li & Durbin, 2009), counted with feature count (Liao et al., 2014) and analysed differential gene expression with DESeq2 (Love et al., 2014).

Chapter 3: Results

3.1: E. coli SC375 exhibits heterogeneous colony morphologies

SC375 is a natural isolate of *E. coli* that was collected during routine sampling from the foreshore of the St Louis River, which is the outflow of Lake Superior at Duluth, Minnesota. During plating of this strain on minimal media (M9) agar plates, we found that two distinct colony morphologies were present. We first checked for possible contamination of the culture by restreaking three times to a single colony; each time, after regrowth of the cells in liquid media, we observed the same two distinct phenotypes. To understand the genesis of this phenomenon, we first tested whether this phenotype switching was dependent on nutrient source. We found that the two morphologies were much less distinct on rich media (LB) and the frequency of one colony type was considerably less on minimal glucose media and more on minimal lactose media. Thus to examine the colony morphologies in more detail, we plated dilutions of overnight cultures of SC375 onto 0.8% minimal lactose plates, incubated overnight, and photographed them using a dissecting microscope.

As shown in **Fig. 3.1a** we were able to visualise the growth of distinct colony types, confirming that SC375 is a phenotypically heterogeneous strain. Cells grew into either small, circular compact colonies with well-defined rounded edges, or larger colonies with irregular edges that formed a translucent halo around a more concentrated centre. We named these colony types after their dominant feature and therefore going forward, refer to them as “compact” and “halo” colonies respectively. In addition to compact and halo colonies we discovered a third intermediate colony phenotype that we could not clearly classify as either halo or compact. These colonies had a compact-like centre with ill-defined irregular outer edges and demonstrated a smaller halo-like semi-transparent ring around a well defined centre; we therefore referred to these as “mixed” colonies. The compact phenotype appeared to be the prevailing morphology, while halo and mixed colonies were less common. This, together with previous results from SC375 grown on LB agar plates (**Fig. 5.1**) in which we observed exclusively compact colonies suggested that, in the growth environments tested, compact colonies were the default phenotype, while halo colonies were the anomaly.

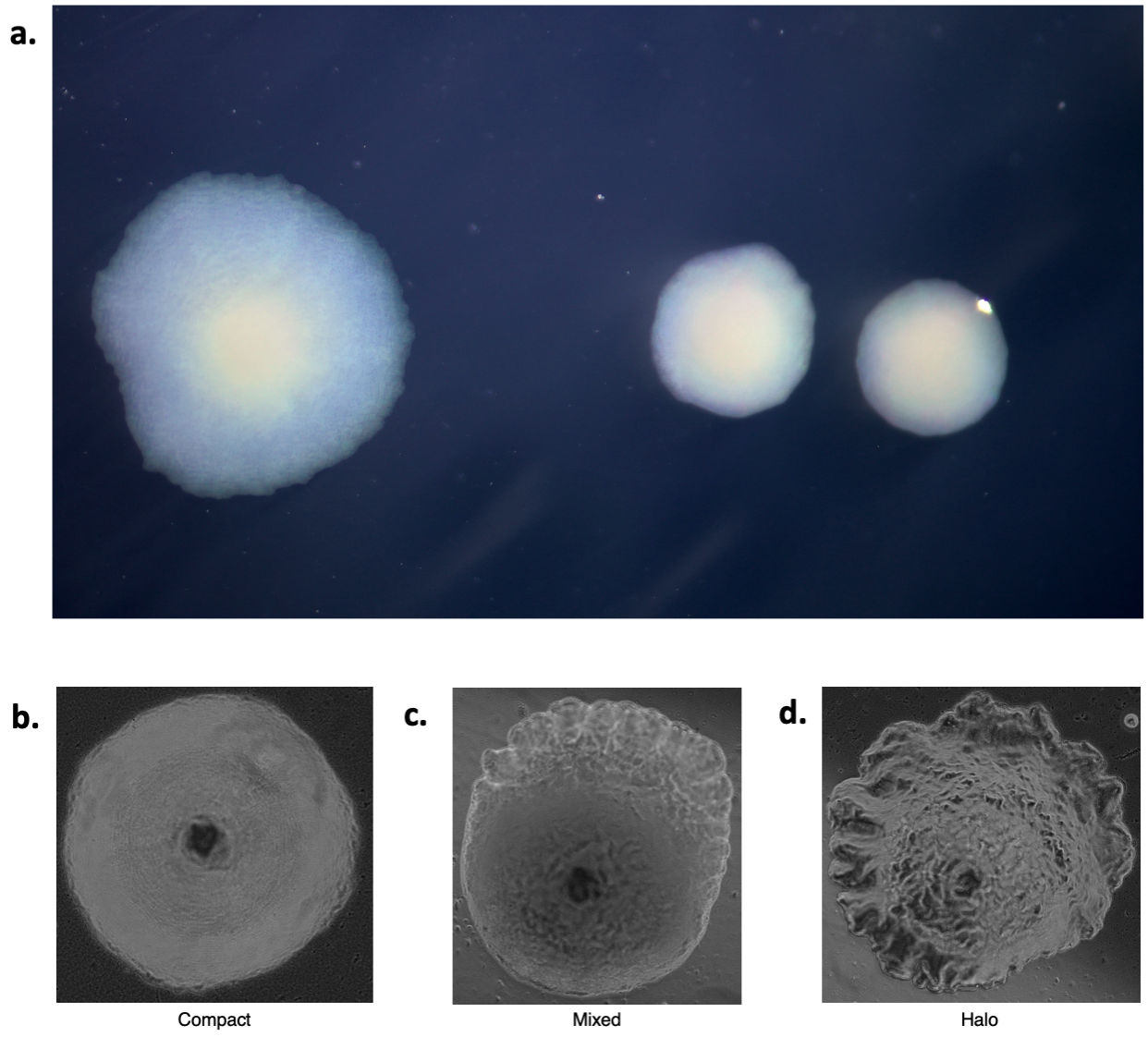


Figure 3.1 The SC375 natural *E. coli* isolate forms phenotypically variable colonies when grown on M9 minimal media. (a) Colony growth of SC375 cells on M9 agar plates with 0.8% lactose, photographed using a dissecting microscope. Cells form either (from left to right) halo colonies, mixed colonies or compact colonies. Individual colonies photographed at 10X magnification **(b)** compact colony **(c)** mixed colony **(d)** halo colony.

In order to better visualise the morphological differences between the compact, mixed and halo colony types, we further analysed plate growth using a light microscope. Colonies were again grown in minimal lactose media and photographed at a magnification of 10X using a Nikon Ti-2 inverted microscope (*Chapter 2.9.1: Colony microscopy*). Along with their characteristic irregular edges, we observed that halo colonies (**Fig. 3.1d**) appeared to have a wrinkly surface very similar to that of the Wrinkly Spreader colonies of *Pseudomonas*

fluorescens SBW25 (J. H. Green et al., 2011; Spiers et al., 2002). Wrinkles also appeared to be larger towards the edges and smaller within the colony centre. Compact colonies (**Fig. 3.1b**) had a smooth, consistent surface, complementary to their distinctive edges. The intermediate mixed phenotype (**Fig. 3.1c**), clearly exhibited characteristics of both colony types, having both wrinkled and smooth colony surfaces.

3.2: Individual cell phenotypes differ between colony morphologies

After confirming that SC375 cells were able to form differential colony morphologies, we hypothesised that individual cells would also display different phenotypes, and that colony morphology is directly influenced by these phenotypes. To investigate this, we used flow cytometry to determine the approximate cellular morphologies comprising each colony type. Flow cytometry uses light scattering to determine cell complexity and relative cell volume (Romano et al., 2003). Cells of different sizes can be differentiated based on their forward and side scatter which reflect their granularity and size respectively (Inglis et al., 2008).

We grew colonies on M9 lactose plates (*Chapter 2.6: Cell culturing*) and selected the colonies used for downstream analysis based on their characterisation as either a compact or halo colony. We collected flow cytometry data for a total of 21 colonies, and identified two different scatter patterns (**Fig. 3.2**). Cells derived from compact colonies consisted of a single cluster of cells with similar size and complexity (**Fig. 3.2a**). Cells derived from halo colonies exhibited two clusters that varied the most by their side scatter values, which was suggestive of two subpopulations of different sized cells (**Fig. 3.2b**). These results suggested that compact colonies are composed of smaller cells, while halo colonies consist of a mixture of both large and small cells.

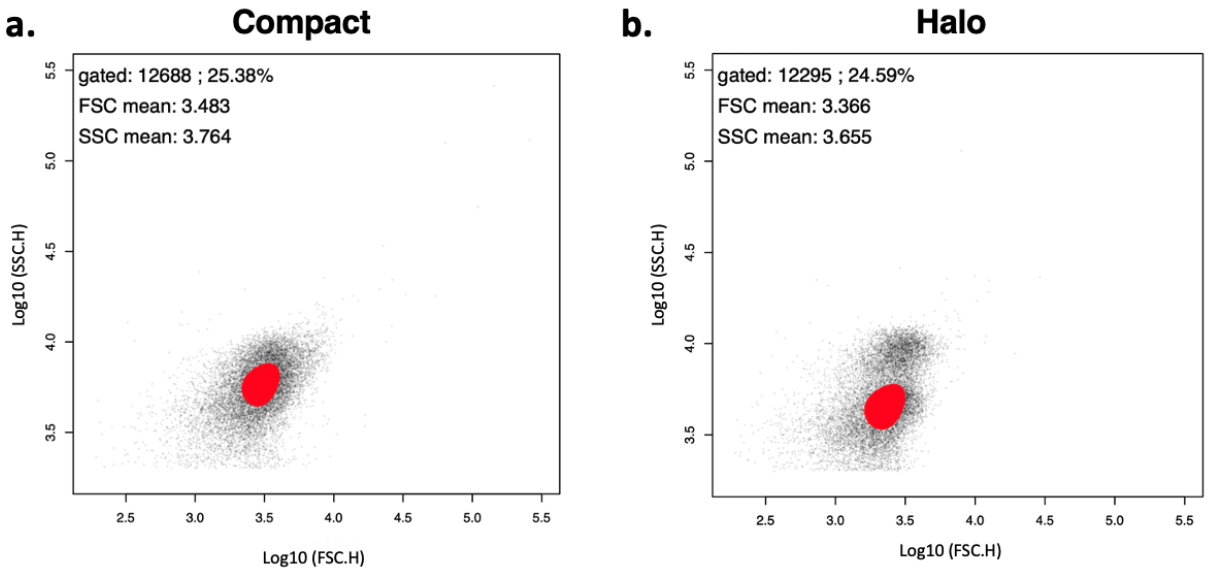


Figure 3.2. Differential colony morphologies are reflective of the phenotypes of cells within colonies. Flow cytometry was used to measure granularity (FSC) and size (SSC) and the results were plotted as cell density maps. **(a)** cells within compact colonies share a single phenotype, as shown by the single mass of cell densities. **(b)** halo colonies consist of two major cellular phenotypes as demonstrated through the two visible groupings. The red regions indicate the area of the highest density of cells.

We used microscopy to gain a better understanding of the distribution of individual cellular phenotypes within each colony morphology. We selected compact, halo and mixed colonies in triplicate and resuspended each colony in 1X PBS. Undiluted resuspensions were used to make microscope slides as outlined in the (*2.9.2: Microscopy)*. We took phase contrast images using a 100X oil objective (**Figs. 3.3 a**, **b** and **c**).

By initial visual analysis, we confirmed that compact colonies consisted of all short cells, while mixed and halo colonies had both long and short cells. Interestingly, distinguishing between mixed and halo colonies proved very difficult.

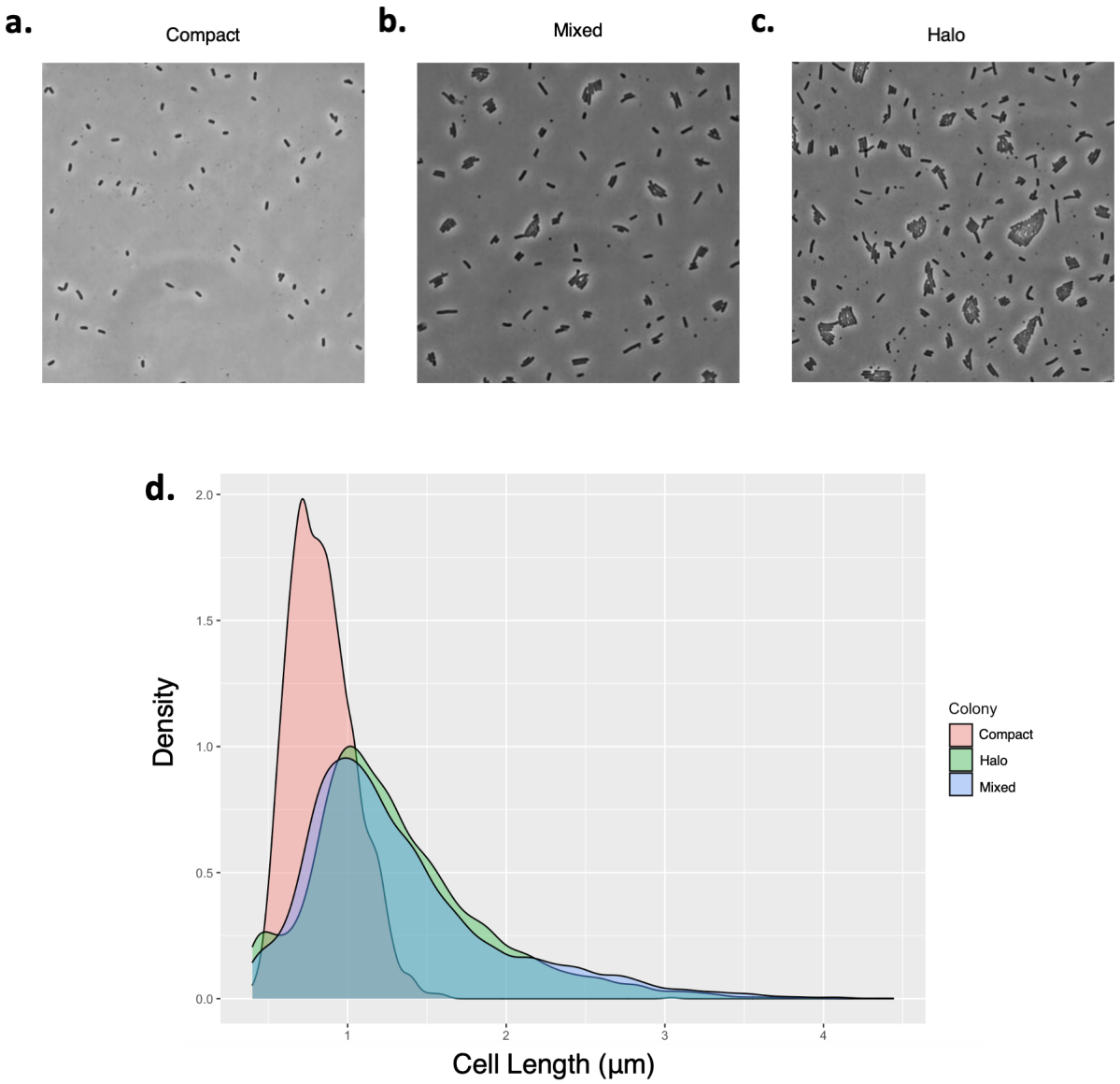


Figure 3.3 The presence of long cells in colonies determines colony morphology. Sample phase contrast images of undiluted resuspended colonies, taken at a magnification of 100X showing a **(a)** compact colony **(b)** mixed colony **(c)** halo colony. **(d)** Distribution of cell length by colony type. The distributions appear left-censored because segmented regions less than 0.4µm in length were excluded as likely dead or non-replicating cells.

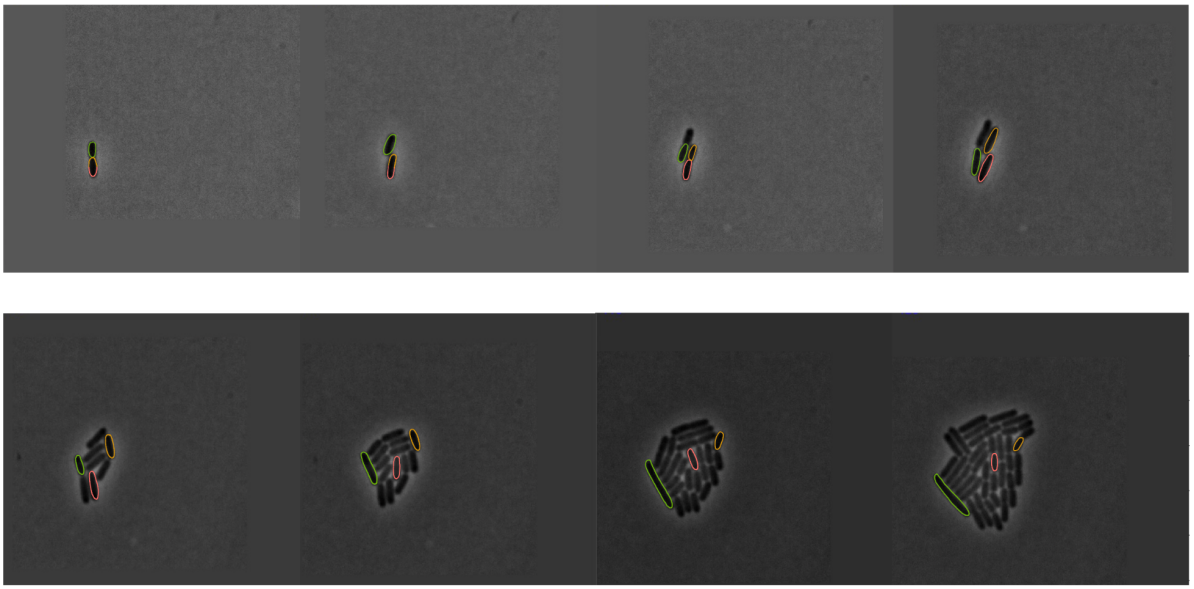
To analyse the data more quantitatively, we segmented individual cells from the phase contrast images using the cell segmentation software *SuperSegger* (Stylianidou et al., 2016), and compared the distribution of cell lengths in each colony type (**Fig. 3.3d**). All three colonies showed unimodal distributions of cell length with long tails. The compact colonies

had uniformly short cells. This is apparent as a tall, narrow peak with cell sizes concentrated between 0.4 and 1.3µm. Cell lengths were more variable within mixed and halo colonies; most notable is the presence of the right tail of the distribution showing long cells present at small quantities. This contrasts with our expectations from the previous flow cytometry data, which suggested mixed and halo colonies had a bimodal distribution of cell length. Nevertheless the microscopy data further supported the hypothesis that the presence of even a small fraction of longer cells within colonies was enough to alter colony phenotype.

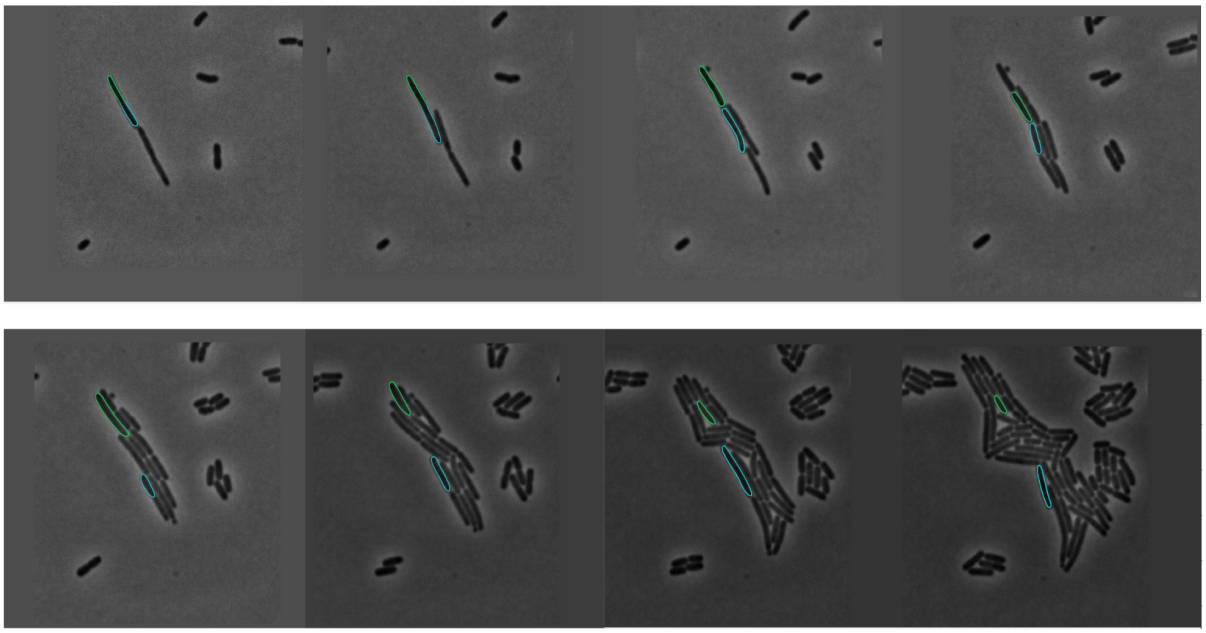
3.3: Cells switch rapidly switch between long- and short-cell phenotypes

To characterise the phenotypic switching between short- and long-cell phenotypes, we performed time-lapse microscopy, to monitor the length and growth of cells over time. We created time-lapse movies of cell growth, by taking sequential images every two minutes for a period of up to 24 hours. Time-lapse microscopy allowed us to visualise cell division and track the rate of phenotype switching (**Figs. 3.4 a**, **b** and **c**). We observed phenotypic switching in both directions (i.e. long to short cells and the reverse). As expected, we found that it was more common for long cells to switch to short (**Fig. 3.4 b** and **c**).

a.



b.



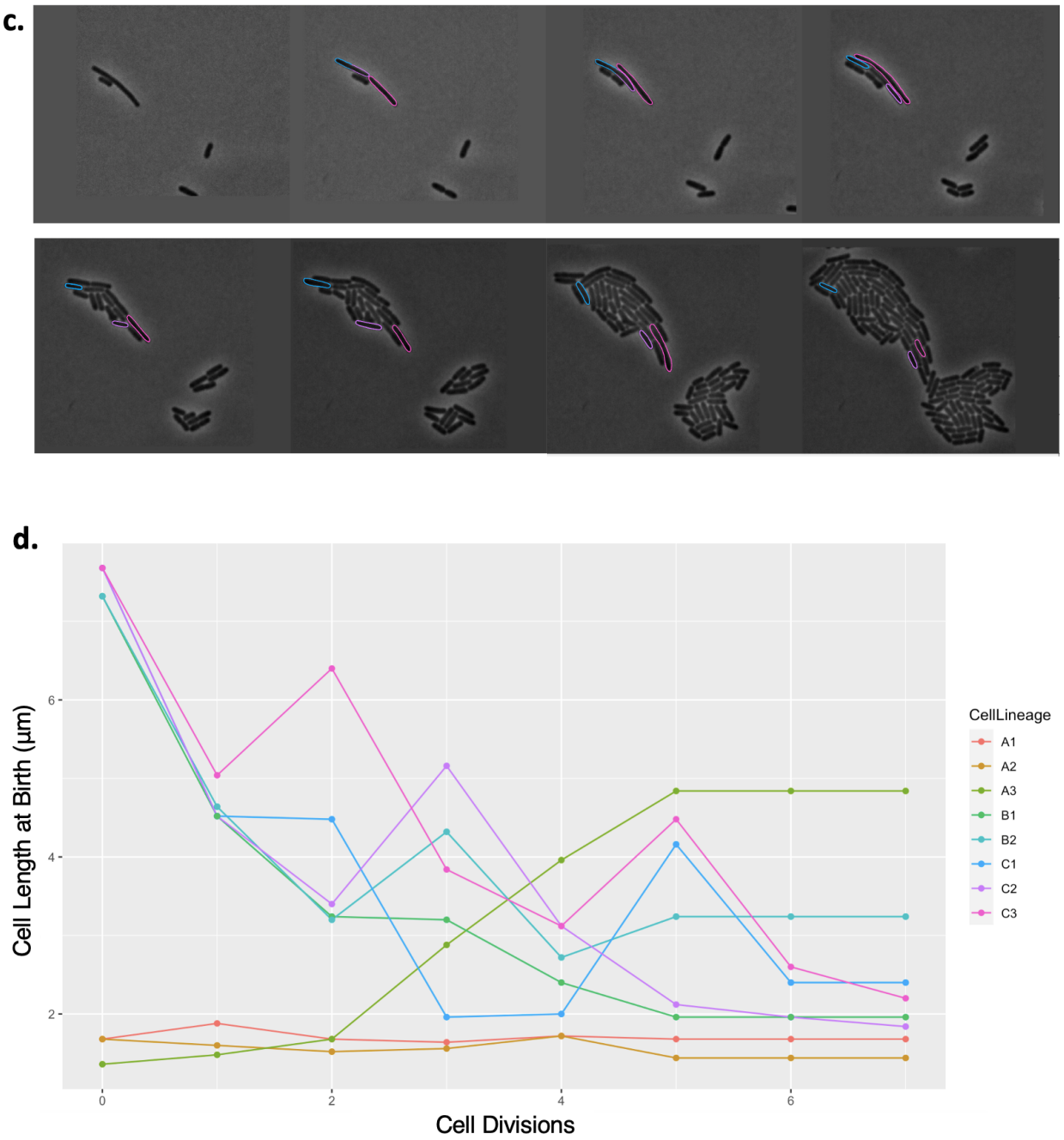


Figure 3.4 SC375 cells rapidly switch between short and long phenotypes. Time-lapse microscopy of cell growth over time, showing changes in the length of cell progeny of an initially **(a)** short cell **(b)** long cell **(c)** long cell. **(d)** Cell length at birth of cell progeny over a period of generations. The name of each cell line refers to the corresponding time-lapse image of its cell growth in figures **a,b** and **c.** For clarity the division times for all cells has been aligned, so in some cases there are an unequal number of time steps between each cell division. Lineage A3 switches from a short to long phenotype, doubling in length at birth (and reaching a stable plateau); several other lineages half their size at birth.

We quantified cell switching characteristics by tracking the cell length at birth of each cell's progeny over a period of cell divisions (**Fig. 3.4d**). We found that short cells exhibited extremely stable phenotypes, with cell lengths at birth varying by less than 10%. We identified few long cells that remained long, and found that when they transitioned to short cells, this transition was extremely unstable, with length at birth both decreasing and increasing during the transition, sometimes by more than 25%. This variability appeared to stabilise once cells fully transitioned to short cells. Finally, in the time-lapse movies we analysed, we found only a single cell that switched from short to long; however, this transition appeared to be relatively stable (A3 in **Fig. 3.4**). By establishing a pattern of cell switching, we were also able to confirm that long cells are still viable and able to reproduce, ruling out cell division errors as a cause for the formation of long cells.

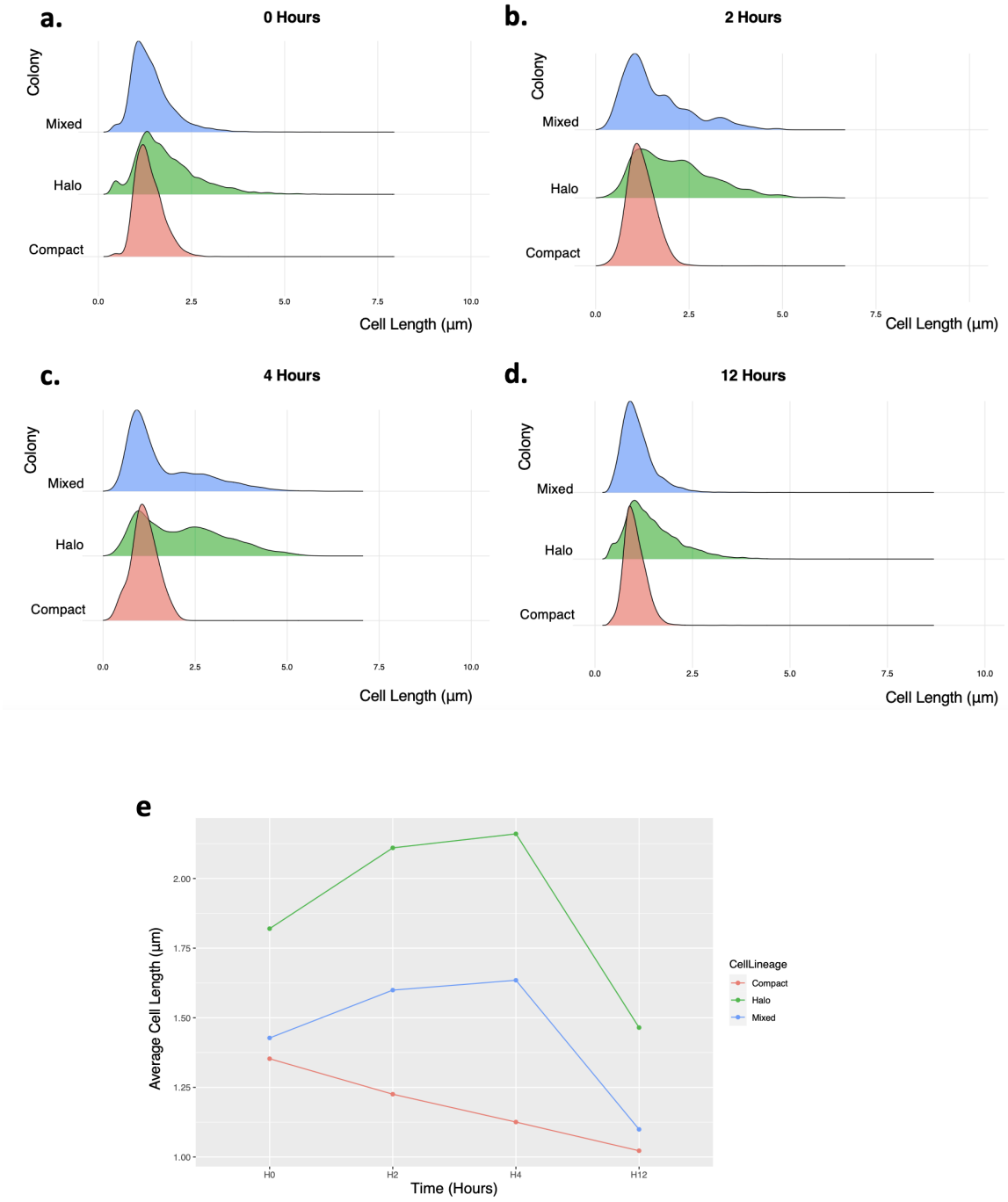


Figure 3.5 Long cells show a more dynamic pattern of switching on a colony wide level in comparison to short cells. Distribution of cell lengths in each colony type after sampling of liquid cultures grown for **(a)** 0 hours **(b)** 2 hours **(c)** 4 hours **(d)** 12 hours. **(e)** Average cell length of all cells in each colony over different time points.

Analysis of time-lapse imaging is limited by the cell segmentation step and tracking across frames, as the proximity of cells makes segmentation difficult. In addition, there are few independent data points when observing microcolony growth, as most cells arise from a single progenitor cell. In order to sample a more independent set of cells and quantitatively characterise cell size across the population, we tracked the frequency of cell phenotypes in populations by sampling overnight cultures at varying time points for microscopic analysis. We resuspended triplicate samples of each colony type (compact, mixed and halo) in M9 minimal media and left them to grow for 12 hours. We then diluted these into fresh media (*Chapter 2.9.2: Microscopy*) and sampled them at 0 hours (resuspended colonies), 2 hours, 4 hours, and 12 hours. We used undiluted samples to take phase contrast images at each time point and segmented the images to measure variation in cell lengths over time.

At hour 0 we expected compact colonies to contain primarily short cells, and mixed and halo colonies to have both short and longer cells. We aimed to use the cell length distribution patterns of each colony type over time, to follow the switching dynamics of each cellular phenotype. Due to the less frequent switching from the short to long phenotype, we expected that compact colonies would remain composed primarily of short cells, with a few outliers that demonstrated switching from short to long. We expected that halo and mixed colonies would show bimodal distributions of cell sizes, representing the two cell phenotypes.

At hour 0 (**Fig. 3.5a**), compact and mixed colonies showed unimodal right skewed distributions very similar to those shown in **Fig. 3.3d**. The modal cell lengths for all three colony types was similar, approximately 1.2µm. Mixed colonies exhibited a slightly longer tail; however, halo colonies had a long tail extending to just over 4µm. This further reinforced our previous findings that compact colonies were made up of only short cells while halo and mixed colonies had both cell types. The mean cell length (**Fig. 3.5e**) of halo colonies at 0 hours was 1.82µm; in mixed colonies the mean cell length was 1.43µm - not much higher than the 1.35µm average of compact colonies.

At hour 2 (**Fig. 3.5b**) the modal cell length in compact colonies became slightly shorter, with very few cells being longer than 2µm. This remained consistent after 4 hours (**Fig. 3.5c**), with cell lengths varying from 0.5µm to 2µm. Finally, after 12 hours (**Fig. 3.5d**) the cell length distribution narrowed further to show a modal length of 1µm. Overall, mean cell length over

time (**Fig. 3.5e**) decreased from 1.35µm at hour 0, to 1µm after 12 hours. Along with the narrowing of distribution peaks over time, the downward trend suggested that short cells tend to stay short over time and are even able to get shorter. However, this result is surprising in that bacterial cells usually become longer when they enter exponential growth.

After 2 hours (**Fig. 3.5b**), halo colonies formed a bimodal peak that skewed considerably further to the right. The longest cells reached a length of 5µm or more. At 4 hours (**Fig. 3.5c**) cell length was clearly bimodal, with peaks at 1µm and 2.4 µm. After 12 hours (**Fig. 3.5d**) the distribution once again became unimodal, although surprisingly with a shorter right tail than at hour 0. Mean cell lengths at each time point clearly indicated that cells got longer after 2 and 4 hours of growth, and then fell to 1.46µm - again, lower than their initial average length of 1.82µm at 0 hours.

Mixed colonies followed a pattern that reflected an intermediate between compact and halo colonies. A bimodal distribution was evident at 4 hours (**Fig. 3.5c**) although, relative to the halo colonies, the majority of cells were of shorter length. Mean cell lengths followed the same pattern as halo colonies, reaching less extreme values; they increased at 2 and 4 hours and decreased again at 12 hours. Average cell lengths at 0 hours was 1.43µm, only 0.08µm longer than that of compact colonies. After 12 hours the average dropped down to 1.1µm, which was 0.04µm longer than the compact cell average. In comparison, the average cell length at 12 hours of halo colonies was 1.46µm - 0.1µm longer than the starting value for compact colonies. Together, the cytometry, time-lapse microscopy, and static microscopy indicated that cell morphology changed rapidly, and most clearly, that long-cell phenotypes could develop very quickly. We next sought to discover whether there was a specific molecular mechanism leading to the differentiation of these two phenotypes.

3.4: Long- and short-cell phenotypes do not differ in genomic sequence

We hypothesised that there may be a genetic basis for the phenotypic variation demonstrated by the SC375 strain. Under this hypothesis, after a genetic switch occurred, cells would begin the transition to a different phenotype, and would remain that phenotype until the switch reverted. Thus, even though the switch is instantaneous and discrete, phenotypes themselves are not necessarily discrete (i.e. either long or short) as there is a transition period. To test this possibility, we extracted DNA from each of the two colony phenotypes, performed Oxford Nanopore whole genome sequencing, and assembly (see *2.12: Genome Sequencing*). Alignment of the two genome assemblies suggested a possible indel within the *FocC_1* gene of the *fim* operon, which encodes the chaperon protein FocC. The indel was a four nucleotide deletion of CTCT within a TTCTCTCTTTTT repeat sequence **(Fig. 5.1a)**, which only appeared to be present in the short cell genome assembly. Based on our previous knowledge of the regulation and phase variable control of gene expression in the *fim* operon, we suspected that this indel could have a role in the formation of the two phenotypes. The deletion resulted in the truncation of the functional FocC protein into a non-functional, shorter, protein and an additional hypothetical protein **(Fig. 5.1b)**. In an attempt to confirm the role of this mutation, we aligned individual reads from each population to the assembled genomes. This would allow us to identify changes in the genome structure even if it occurred in only a subset of cells. However, we found no consistent differences in the variants present in reads from short- or long-cell phenotypes. Furthermore, we could not identify even a subset of reads having a different structure or consistently different sequence. This suggested that the indel differentiating the long-cell and short-cell assembly was most likely present due to errors in the assembly process rather than being a true polymorphism. We cannot exclude the hypothesis that long cells differ genetically, but their frequency was not high enough to determine this difference.

3.5: Long- and short-cell phenotypes differ in gene expression profiles

Although we could not determine a consistent genetic change that appeared to be causal of the two phenotypes, we hypothesised that gene expression states might differ between the two phenotypes. In order to measure gene expression, we grew resuspended compact and halo colonies to exponential phase. RNA was extracted from 50 pooled compact and 50 pooled halo colony samples and sequenced using Illumina's stranded RNA sequencing. Sequencing results were analysed using DESeq2 to identify genes that were expressed differentially in either cell type (DESeq2 outputs are shown in **Tables 5.1** and **Table 5.2**).

We identified two groups of genes that were either up- or down-regulated in long cells, compared to short cells. Down-regulated genes are shown in **Fig. 3.6a** and up-regulated genes are shown in **Fig 3.6b**. All changes had significant p-values of less than 2e-5. The down-regulated genes shown all exhibited a log2-fold decrease of at least 2. The up-regulated genes shown in **Fig 3.6b** exhibited a log2 fold increase of at least 2.4.

In order to identify genes of interest for further study, we analysed the relative expression of each gene (**Fig. 3.6 c** and **d**). We hypothesised that variation in expression of genes that are more highly expressed on a relative level were most likely to affect phenotype, even if the expression differences were caused by upstream regulatory changes. Of particular interest in the down-regulated genes was the *fimAICDFGH* gene cluster, which contains all the subunits of the Type I fimbriae, and has been found to be under phase-switching control. All seven *fim* genes had significant fold-decreases in gene expression in long cells. The *fim* cluster of genes are located on the same operon, under the control of the same promoter. Analysis of the relative gene expression of negatively regulated genes concluded that the *fimA* gene was expressed at the highest level (**Fig. 3.6c**). Several genes within the *gfc* operon were also expressed at a high level: *gfcB, etp* and *etk*. Due to the high basal gene expression of *gfc* and the role of the *gfc* operon in capsule synthesis and subsequent excretion, we selected the *gfc* operon as a second negatively regulated gene cluster of interest. Finally, we selected *Aida_I* for further analysis because of its high basal expression, its role as an outer membrane autotransporter and its highly up-regulated expression in the long cell phenotype.

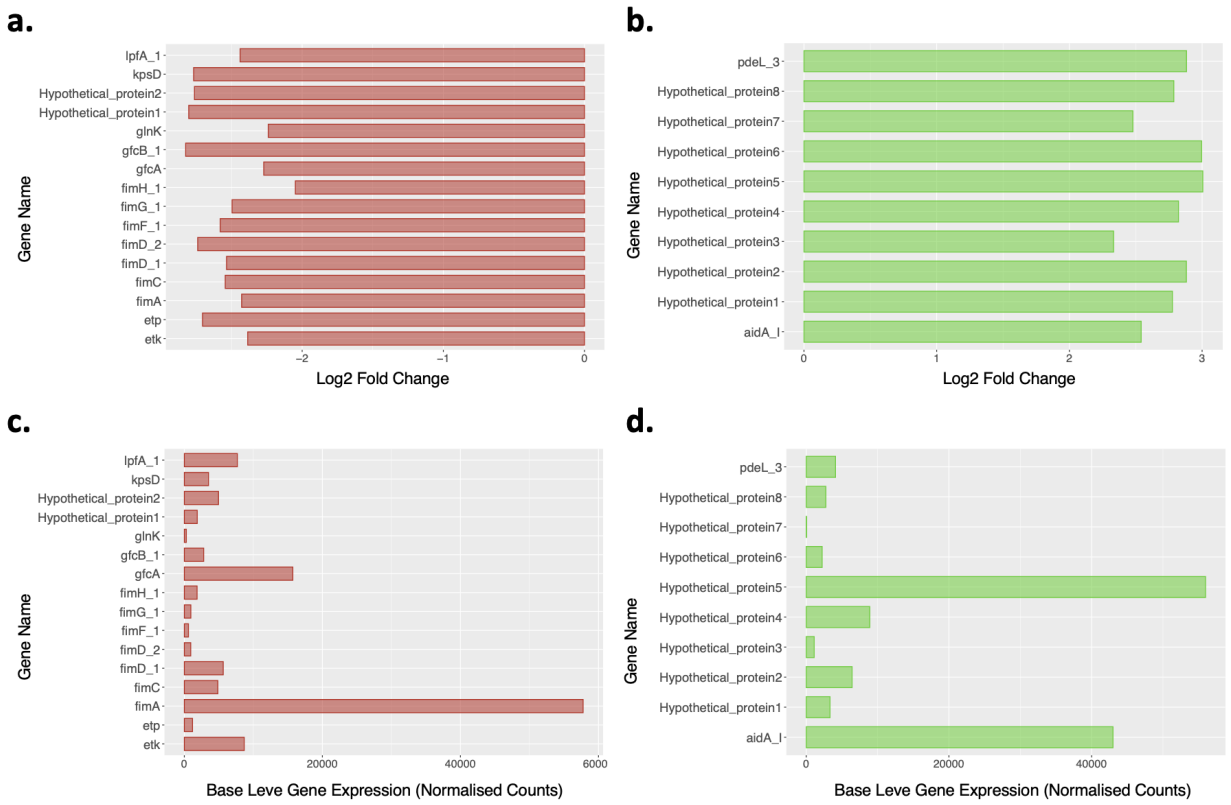


Figure 3.6 SC375 cellular phenotypes demonstrate variation in the expression of three key genes. RNA sequencing analysis of differential cell phenotypes identified clusters of genes positively and negatively regulated in long cells in comparison to short cells. Log2 fold change of expression in **(a)** negatively regulated genes and **(b)** positively regulated genes. Basal gene expression of **(c)** negatively regulated genes and **(d)** positively regulated genes.

3.6: Expression patterns correlate with colony phenotype

To investigate variation in gene expression at a cellular level, we cloned the promoter regions of our three genes of interest upstream of a strong ribosomal binding site and Green Fluorescent Protein (GFP) on a low copy-number plasmid. This allowed us to determine the relative cellular mRNA expression levels by measuring the fluorescence data of each cell.

The promoter regions for the *gfc* operon and the *fim* gene cluster were 710bp and 700bp, respectively. The *fim* promoter region included 105 base pairs from the upstream gene and 115 base pairs from the downstream gene; the *gfc* promoter region included 139 base pairs of the upstream gene and 129 base pairs of the downstream gene. We included part of the upstream and downstream genes as there are known to be regulatory elements in these regions. The analogous promoter region for the *aidA* gene was over 2000 base pairs. This

was difficult to clone into our 4620 base pair vector backbone and we were unable to confirm successful transformation of the paidA::GFP plasmid construct into our SC375 cells. For this reason, the results below are limited to the *gfc* and *fim* promoters.

We looked at how colony fluorescence (indicative of *fim* or *gfc* operon expression) correlated with morphology. We imaged plated colonies using both a phase contrast and a 470nm optical configuration. Because our RNA-seq data indicated that both the *fim* operon and *gfc* operon were upregulated in short cells, we expected short cells to have greater fluorescence values than long cells. Therefore compact colonies would be expected to have a greater fluorescence than halo and mixed colonies.

Colonies containing pfim::GFP are shown in **Fig. 3.7.** Phase contrast images of colonies containing either the pfim::GFP or the pgfc::GFP (**Fig. 3.8**) plasmids demonstrated identical characteristics to those in **Fig. 3.1**, confirming the presence of plasmids within cells does not influence cell morphology.

Compact colonies tended to be small with defined edges (**Fig. 3.7a**). The shape and characteristics seen in the phase contrast image of this colony are clearly identifiable in the corresponding fluorescence image. A very large proportion of cells within the compact colony were highly fluorescent. Comparatively, the halo colony had limited fluorescence localised in the central portion of the colony (where the cells are the densest), with a swirling, inconsistent pattern of fluorescence towards the outer edges (**Fig. 3.7b**). In the fluorescence image, colony edges were indistinct. Because halo colonies contained a large proportion of long cells, which give the colony its unusual wrinkly shape, we hypothesise that the lack of fluorescence at the edges of this colony was due to a deposit of long cells at these locations, and the overall “swirled” pattern is due primarily to the long cell morphology and the forces affecting neighbouring cells during colony growth. The intermediate morphology of the mixed colony (**Fig. 3.7c**) was very clear in both the phase contrast and the fluorescence images. Half of the colony had the smooth edges characteristic of compact colonies while the other half had the irregular edges and wrinkled surface, affiliated with halo colonies. The fluorescence image followed the same pattern: half of the colony was brightly fluorescent with defined edges, while the other half had the swirling fluorescence pattern of the halo

colonies. It was easy to distinguish differences in the mixed colony morphology using the fluorescence image alone.

Colonies transformed with the pgfc::GFP plasmid (**Fig. 3.8**) followed a similar pattern as the pfim::GFP colonies in **Fig 3.7**. The compact colony (**Fig. 3.8a**) showed the highest level of fluorescence, corresponding with the hypothesis that compact colonies are composed of all short cells and that these express the *gfc* operon at high levels. The halo colony **(Fig. 3.8b**) had its characteristic ill-defined borders and wrinkly surface, visible in the phase contrast image, and the newly determined unique fluorescent centre with a swirling pattern towards the outer edges. The mixed colony (**Fig. 3.8c**) of pgfc::GFP transformed cells appeared to have an unusual combination of a mostly wrinkly surface with defined edges on one half of its border. However, the fluorescence patterns were as expected, with the greatest intensity of fluorescence corresponding with compact colony features.

The increased fluorescence of compact colonies of both pgfc::GFP and pfim::GFP transformed cells confirmed that shorter cells have the greatest promoter activity, even for promoters present on multicopy plasmids. Coupled with phase contrast imagery, we were able to use fluorescence microscopy to determine the location of each cell phenotype within halo and mixed colonies; shorter cells appeared to be more common at the centre, while longer cells were localised towards the outer colony edges.

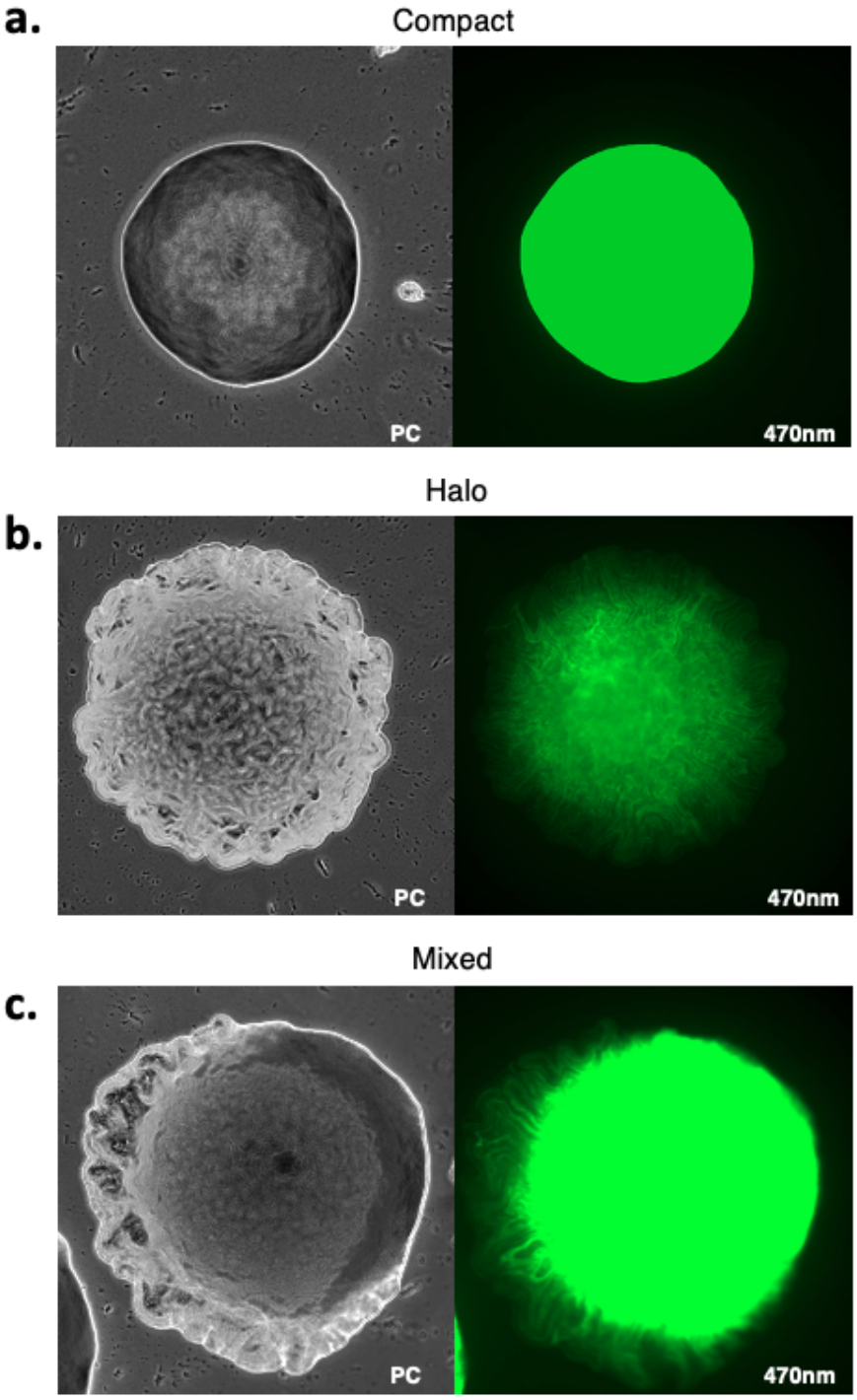


Figure 3.7 The compact colony morphology of pfim::GFP transformed cells showed the greatest fluorescence under 470 nm light, the halo colony morphology was the least fluorescent. 10X phase contrast (PC) and their corresponding 470nm (470nm) fluorescence microscopy images of SC375 transformed with a fim-pUA66 expression vector. All fluorescence images were taken at 100ms exposure at 65% intensity. Images show a **(a)** compact colony **(b)** halo colony **(c)** mixed colony.

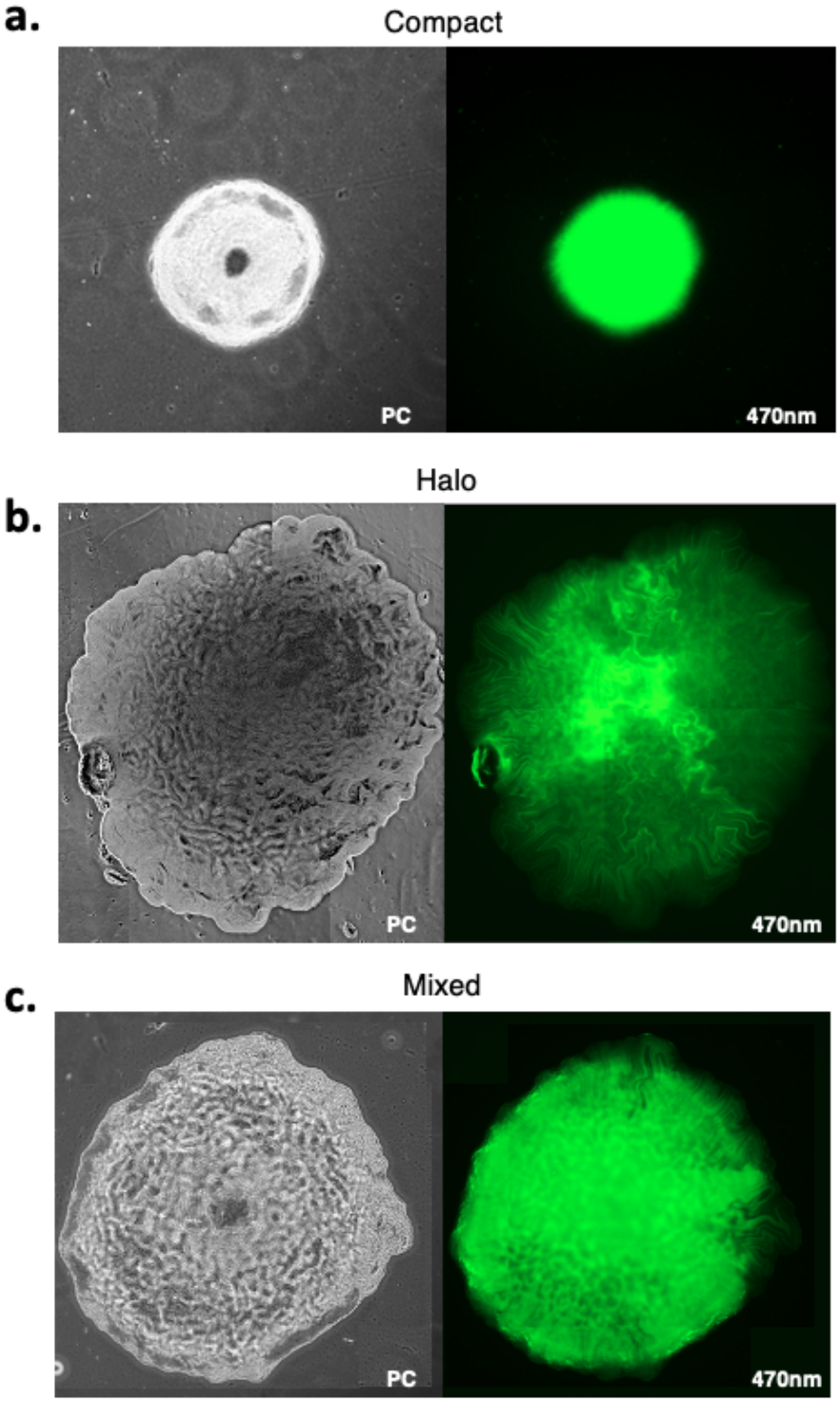


Figure 3.8 Colony fluorescence of pgfc::GFP transformed cells is positively associated with small cell composition. 10X phase contrast (PC) and their corresponding 470nm (470nm) fluorescence microscopy images of SC375 transformed with a fim-pUA66 expression vector. All fluorescence images were taken at 100ms exposure at 65% intensity. Images show a **(a)** compact colony **(b)** halo colony **(c)** mixed colony.

3.7: Short cells exhibit higher activity of the fim and gfc promoters

Following the confirmation of gene expression phenotypes through fluorescence microscopy, we again used flow cytometry as independent confirmation that expression of the *fim* and *gfc* gene clusters were greater in short cell phenotypes. Rather than selecting colonies based on colony (and therefore cell) morphology and testing the fluorescence, following the above results we selected colonies based on their fluorescence and tested differences in cell morphology. We selected triplicates of fluorescent, non-fluorescent and mixed colonies from each transformed strain and collected forward scatter, side scatter and fluorescence data (488 nm excitation, 513/17nm bandpass emission) for each colony. For comparison of fluorescence, we used an empty pUA66 plasmid as a negative control. We predicted that fluorescent colonies would be composed of entirely small cells, and non-fluorescent colonies would show a cell composition pattern similar to that of halo colonies.

Flow cytometry results for pgfc::GFP are shown in **Figure 3.9** and **Figures 5.9** to **5.14**. Cell size (as inferred by side scatter (SSC) data) for fluorescent (**Fig 3.9a**), non-fluorescent (**Fig. 3.9d**) and mixed (**3.9g**) colonies all had the expected cell size patterns: fluorescent colonies showed two different cell size peaks indicating the presence of both long and short phenotypes; however, considerably fewer cells were in the high-SSC cluster, indicating long cells were not present in large numbers. Colony fluorescence of the compact morphology (**Fig. 3.9b**) showed two fluorescence peaks at values of GFP 3-4 and 4-5 log10 of GFP. In addition, almost all high-SSC cells had low fluorescence. We found that the fluorescent cells within fluorescent colonies (**Fig. 3.9c**) all had low SSC, with a value of 3.0-3.5 (log10 arbitrary units).

Non-fluorescent colonies (**Fig. 3.9d**) showed one large cluster of cell size at a high SSC value, centred at 3.7, indicative of the longer cellular phenotype. A small smear below 3.5 suggested that short cells are present, but in very low numbers. Colony fluorescence (**Fig. 3.9e**) also showed two peaks, each of a similar size, at 2.0-3.0 and 3.0-4.0. This suggested that some cells may have recently had high fluorescence, but this fluorescence was diluted as the cells became longer; other cells have remained long and therefore filtered the majority of the GFP out. Comparison of cell size and fluorescence (**Fig. 3.9f**), showed two adjacent

clusters of cells with fluorescence values of between 2.0-3.0 and 3.0-4.0. Both fluorescence clusters had SSC values of between 3.5 and 4.0.

Mixed colonies showed the expected two cluster pattern of cell size, however with a greater density of high-SSC cells. Mixed colonies (**Fig. 3.9h**) showed an inverse pattern to that of non-fluorescent colonies, with the second of the two peaks being larger. This is suggestive of a larger quantity of fluorescent cells within mixed colonies than in non-fluorescent colonies. Cell size-fluorescence data (**Fig. 3.9i**) for mixed colonies is comparatively intermediate. Two clusters distinguishable by the y side scatter axis, differentiate in their log10 GFP values. The cluster of lower SSC cells have higher log10 GFP values of between 4.0-4.5. The higher SSC cluster of cells have lower log10 GFP values of below 3.5.

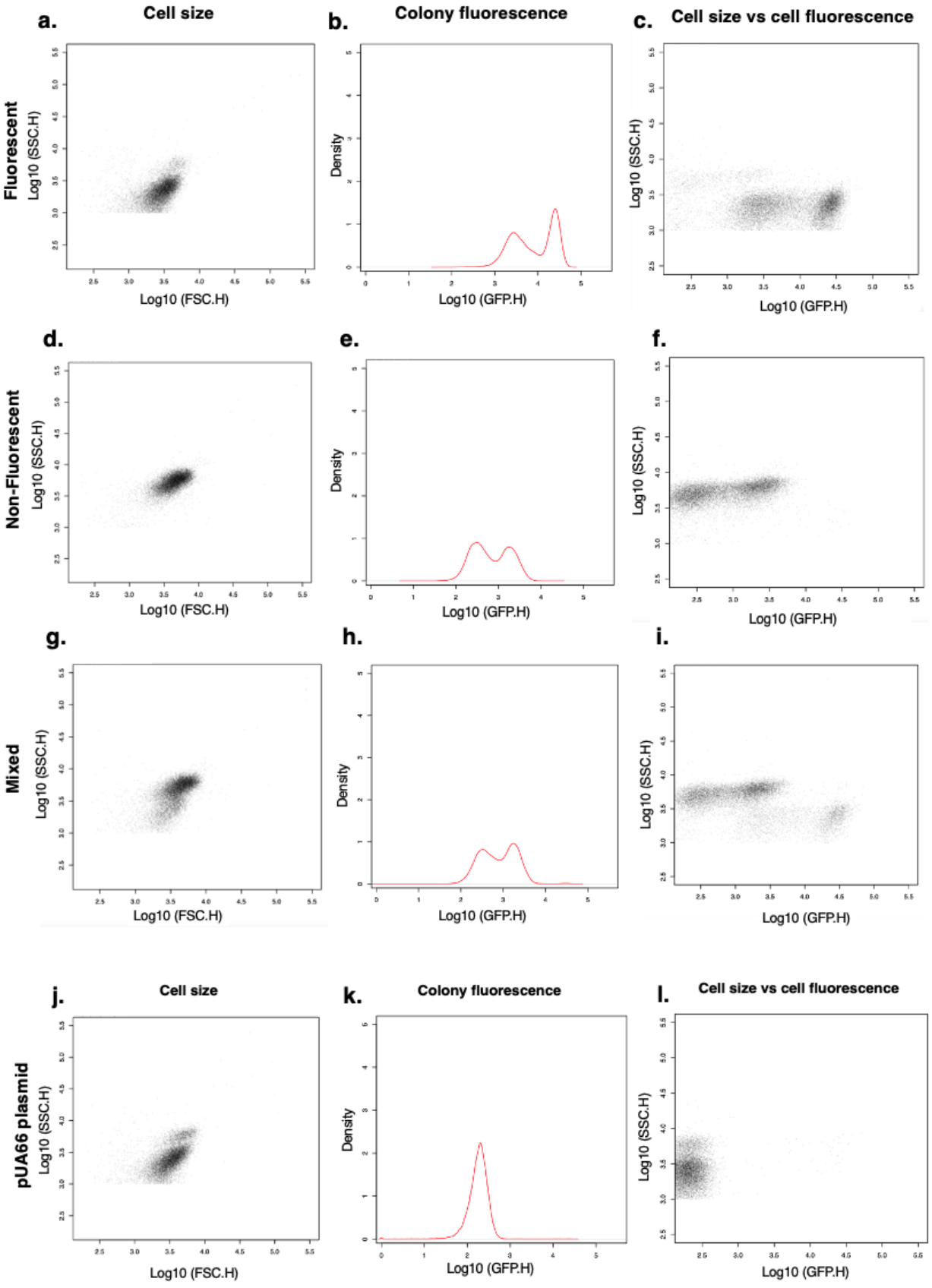

a.
Cell size
b.
Colony fluorescence
c.
Cell size vs cell fluorescence
Fluorescent
Log10 (SSC.H)
Log10 (FSC.H)
Density
Log10 (GFP.H)
d.
e.
f.
Non-Fluorescent
g.
h.
i.
Mixed
j.
Cell size
k.
Colony fluorescence
l.
Cell size vs cell fluorescence
pUA66 plasmid

Figure 3.9 SC375 cells transformed with the pgfc::GFP plasmid show greater fluorescence in the short cell phenotype than in the long cell phenotype. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within a fluorescent, non-fluorescent and mixed colony. **(a)** Cell size **(b)** colony fluorescence and **(c)** cell fluorescence of cells within a fluorescent colony. **(d)** Cell size **(e)** colony fluorescence and **(f)** cell fluorescence of cells within a non-fluorescent colony. **(g)** Cell size **(h)** colony fluorescence and **(i)** cell fluorescence of cells within a mixed colony. **(j)** Cell size **(b)** colony fluorescence and **(c)** cell fluorescence of cells within the control colony. Coupled with the results of our negative control, we determined that this lower fluorescence value was due to cell autofluorescence from exposure to the 470 nm light used to measure GFP.

Flow cytometry results from the pfim::GFP transformed strain are shown in **Figure 3.10.** Cell size results for fluorescent (**Fig. 3.10a**), non-fluorescent (**Fig. 3.10d**) and mixed colonies (**Fig. 3.10g**) all showed two clusters suggesting the presence of both long and short cells. Although these results indicate the presence of both cellular phenotypes in all three colony types, non-fluorescent colonies exhibited a more dense cluster of cells with a larger side scatter value (i.e. larger cells). This suggested non-fluorescent colonies have a larger proportion of longer cells. Fluorescence patterns for all three colony types varied unexpectedly. Fluorescent colonies (**Fig. 3.10b**) showed two peaks with the high fluorescence peak approximately two logs more fluorescent (1.0e2-fold). The colonies we identified as non-fluorescent (**Fig. 3.10e**) had a single fluorescence peak at 2.3 on a log10 scale. Mixed colonies (**Fig. 3.10h**) had two peaks of fluorescence, similar to the peaks demonstrated by the colonies we identified as fluorescent. However, the first peak had more than double the number of cells as the second peak.

We also examined fluorescence data in direct relation to cell size. As expected, both fluorescent (**Fig. 3.10c**) and mixed colonies (**Fig. 3.10i**) showed two clusters of fluorescence. Notably, a large fraction of the highly fluorescent cells had high SSC values, which we infer as being long. In comparison, non-fluorescent colonies (**Fig. 3.10f**) contained a single cluster of fluorescence, however in this cluster there were two positions of cell density with varying cell size values. In non-fluorescent colonies, it appeared that cell size did not influence fluorescence.

While each of the three colony morphologies of pfim::GFP transformed cells showed the expected variation in fluorescence data, the composition of each cell type was not as expected. All colonies had similar bimodal distributions of SSC values. In addition, we found

that fluorescent colonies in fact had a higher proportion of high-SSC cells than non-fluorescent colonies

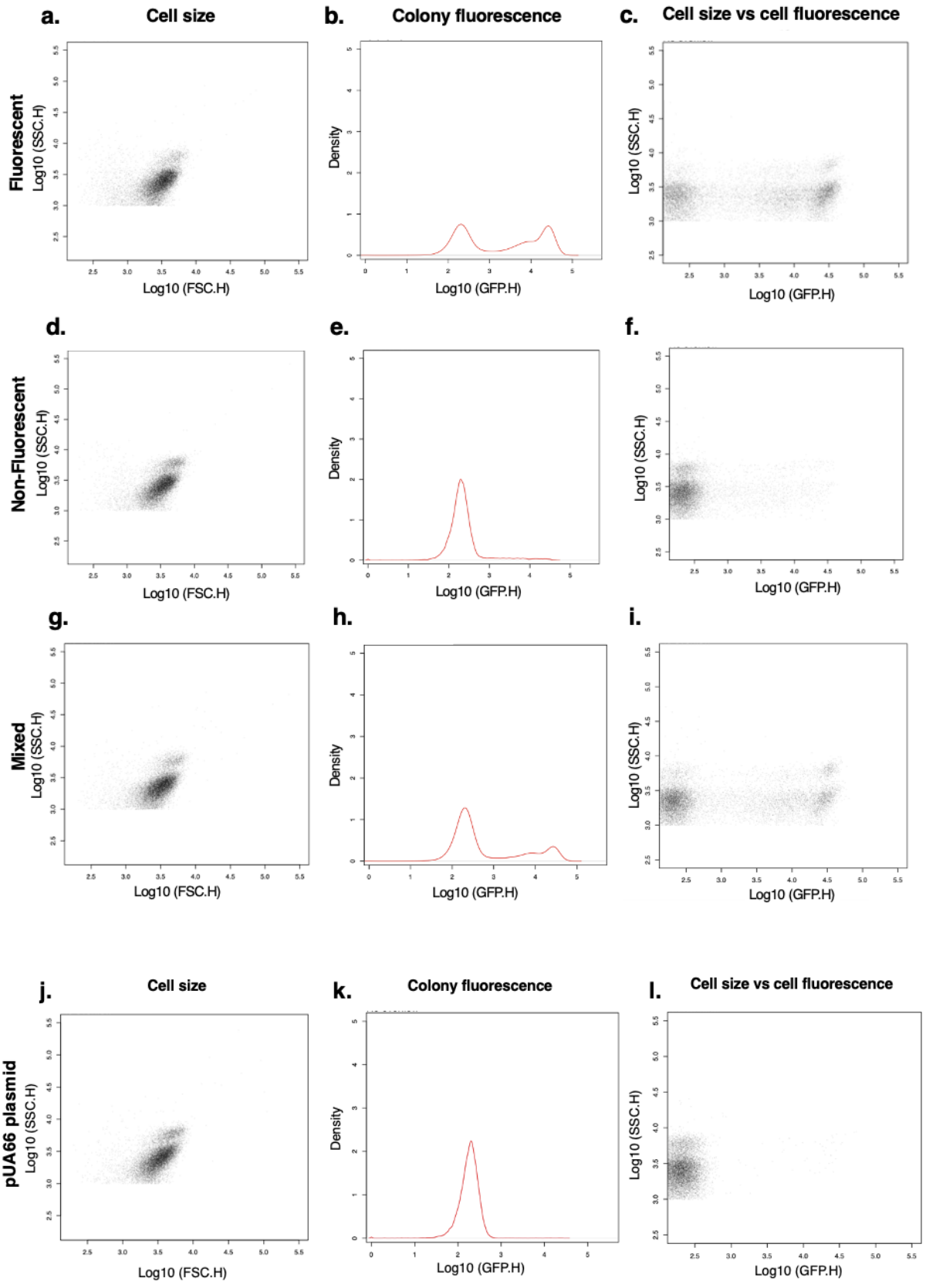


Figure 3.10 Colonies grown from cells transformed with the pfim::GFP plasmid show no relationship between cell length and fluorescence. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) of the cells within fluorescent, non-fluorescent and mixed colonies. Cell size **(b)** colony fluorescence and **(c)** cell fluorescence of cells within a fluorescent colony. **(d)** Cell size **(e)** colony fluorescence and **(f)** cell fluorescence of cells within a non-fluorescent colony. **(g)** Cell size **(h)** colony fluorescence and **(i)** cell fluorescence of cells within a mixed colony. **(j)** Cell size **(b)** colony fluorescence and **(c)** cell fluorescence of cells within the control colony. Coupled with the results of our negative control, we determined that this lower fluorescence value was due to cell autofluorescence from exposure to the 470 nm light used to measure GFP.

3.8: Fluorescence and cell lengths change over time

Up until this point, we have only conducted fluorescence analysis of our transformed strains within colonies - a single snapshot in time at stationary phase. To get a more thorough picture of the changes in gene expression over time, we used flow cytometry to perform a large-scale analysis of length and fluorescence over multiple time points.

We inoculated liquid cultures from a glycerol stock of transformed strains, and grew them for either 4 hours or 12 hours, and collected cytometry data. At 4 hours, we expected most of our cells to be in exponential phase. pfim::GFP clearly changed between 4 hours and 12 hours of growth. At both time points, two subpopulations of cells, differing in their cell length (SSC), were present within the larger population. After both 4 hours (**Fig. 3.11a**) and 12 hours (**Fig. 3.11d**) shorter cells were the dominant phenotype.

After 4 hours of growth the negative control (**Fig. 3.11c**) had fluorescence values below 2.75 (i.e. background cellular auto-fluorescence), so fluorescence values above this point were deemed significant, this value persisted until 12 hours (**Fig. 3.11f**). We observed the greatest fluorescence values for pfim::GFP transformed cells at 4 hours; a smear-like pattern of fluorescence values suggested a large range of fluorescence values rather than two significant extremes. There was no obvious distinctive relationship between fluorescence values and side scatter values, with long and short cells equally lacking in fluorescence at either time point. After 12 hours the already minimal fluorescence decreased, despite an increase in short cells within the population (**Fig. 3.11d**).

In contrast, the fluorescence of pgfc::GFP transformed cells remained stable over time; short and long cells were present at both time points. After 4 hours (**Fig. 3.11b**), there was an equal proportion of long and short cells within the population; there were two concentrations

of data on our SSC x GFP plot (**Fig. 3.11b**). Each cluster differentiated on both their SSC and GFP values, indicating the presence of cells that vary in cell length and fluorescence. Both fluorescence values were greater than the 2.75 (log10 of GFP) of our negative control, so we could rule out solely autofluorescence. Short cells had a much greater fluorescence of 4.5 compared to the 3.5 of long cells. Again, we view this as most likely to be residual fluorescence from transitioning from short to long, or as leakiness of the promoter. After 12 hours, short cells became the dominant phenotype, although some long cells remained present (**Fig. 3.11e**). Short cells also continued to show high fluorescence values of 4.5 (**Fig. 3.11e**). The long cells that remained in the population were still fluorescent.

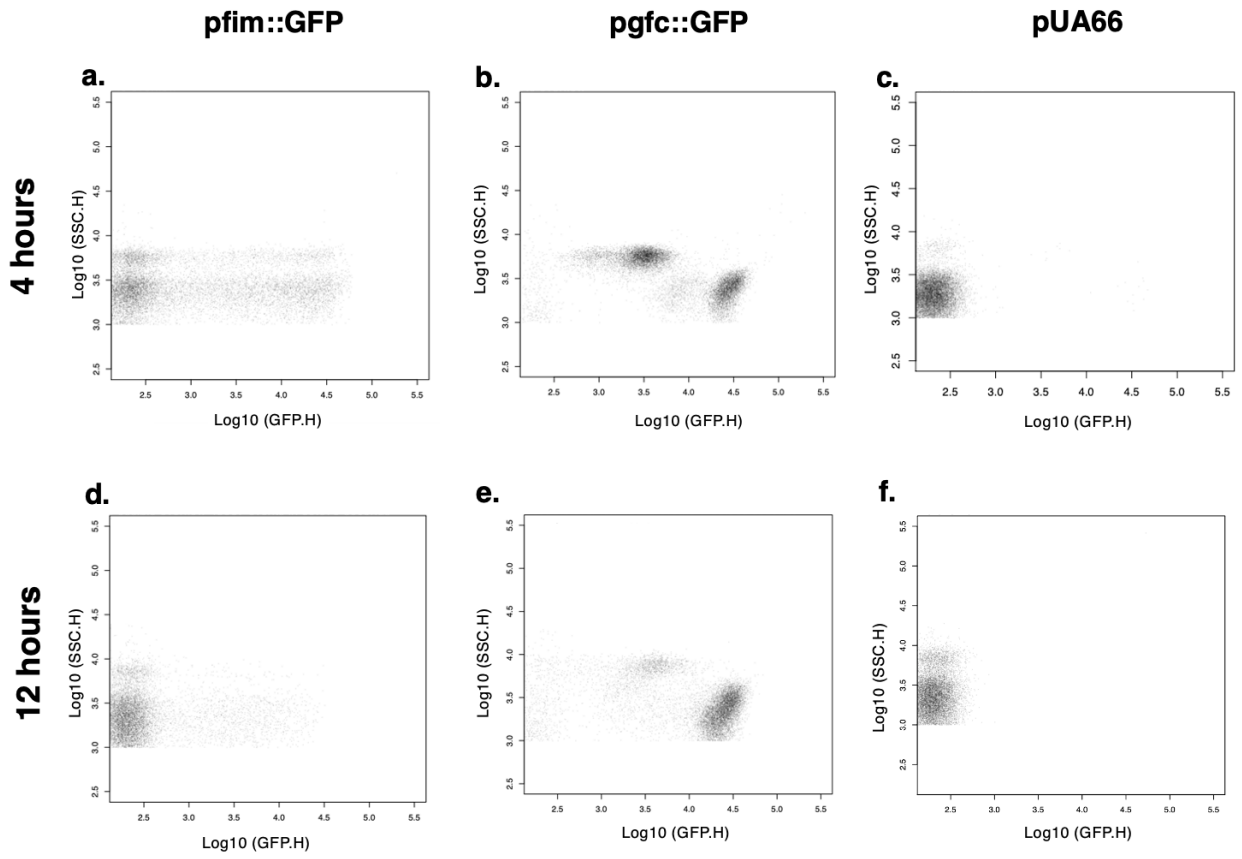


Figure 3.11 The fluorescence of SC375 transformed strains changes over time. Liquid cultures inoculated with glycerol stocks of pfim:GFP, pgfc:GFP and pUA66 transformed cells were grown for 12 hours. We performed flow cytometry to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) of each culture after 4 and 12 hours of growth. The relationship between cell size and fluorescence is plotted as SSC and GFP values at 4 hours **(a)** pfim::GFP **(b)** pgfc::GFP **(c)** pUA66 and 12 hours **(d)** pfim::GFP **(e)** pgfc::GFP **(f)** pUA66.

3.9: Changes in morphology correlate with changes in gene expression

For a more accurate measure of the relationship between cell fluorescence and cell length over time, we repeated the large scale culture microscopy procedure outlined in chapter *2.9.2: Microscopy* with our GFP reporter strains. Performing microscopy in this manner allowed us to extract data from individual cells within a whole population. We resuspended three of each colony type (fluorescent, mixed and non-fluorescent) for each strain and left them to grow for 12 hours. Samples were taken at 0 hours (resuspended colonies), 2 hours, 4 hours and 12 hours and photographed microscopically. Untransformed cells acted as a control.

To identify any correlation between cell length and fluorescence we created 2d density plots and scatter plots from the segmented cell data (**Fig. 3.12** and **Fig. 3.13**). To rule out autofluorescence, we used untransformed cells as a measure of autofluorescence (**Fig. 3.14**). While fluorescence levels remained consistently under 2.25 (log10 of GPF) at 0 hours, 2 hours and 12 hours, at 4 hours mean autofluorescence reached values of up to 2.5. This was significant because both transformed strains reached peak fluorescence values after 4 hours. For downstream analysis of transformed strains we only considered fluorescence values above 2.25 to be significant at 0, 2 and 12 hours. At 4 hours our marker for significant fluorescence were values of 2.5 and above.

The distribution of long and short cells within pfim::GFP transformed populations (**Fig. 3.12**) remained relatively consistent from 0 hours to 4 hours. At 0 hours (**Fig. 3.12a**), there were two main clusters of cells differing by both their length and mean fluorescence. These clusters showed the inverse relationship between length and fluorescence than we expected; longer cells had a greater mean fluorescence, and shorter cells had insignificant fluorescence values. The scatter plot at hour 0 (**Fig. 3.12b**) confirmed this relationship, showing a statistically significant but weak positive correlation of 0.073 between cell length and fluorescence. This correlation decreased after 2 hours (**Fig. 3.12d**). The two main subpopulations of cells remained, but there was a greater proportion of shorter non-fluorescent cells than longer fluorescent cells (**Fig.3.12c**). After 2 hours cells within both clusters became longer; although there was still a clear distinction between subpopulations of different sizes.

After 4 hours of growth there was no statistically significant correlation between length and mean fluorescence (**Fig. 3.12f**). There was a greater proportion of longer cells at 4 hours and the cells within this subpopulation appeared to be longer than after 2 hours of growth (**Fig. 3.12e**). Due to the increase in autofluorescence after 4 hours, the fluorescence values of both main clusters of cells appeared to be insignificant. However, at 0, 2 and 4 hours there was an additional very small cluster of short cells that were highly fluorescent (with values of 3.0-3.25 log10 of GPF and cell lengths of between 1.0 and 2.0µm). Although limited in number, their presence did confirm it was possible for small cells to be and remain fluorescent over these time intervals, even when autofluorescence increases.

A very weak negative correlation between cell length and fluorescence was present after 12 hours (**Fig. 3.12h**). The proportion of long and short cells within the population changed drastically after 12 hours. While there were two clusters of cells present, these clusters differed the most by their length. The first cluster contained cells between 0.4 and 1µm, with the second cluster containing cells between 1 and 1.4µm. Regardless of length, after 12 hours a large proportion of cells had significant mean fluorescence values, although the subpopulation of shorter cells showed marginally higher mean fluorescence. Two subpopulations of cell size remained present at all time points; after 12 hours, the greatest proportion of cells were less than 2.5µm long.

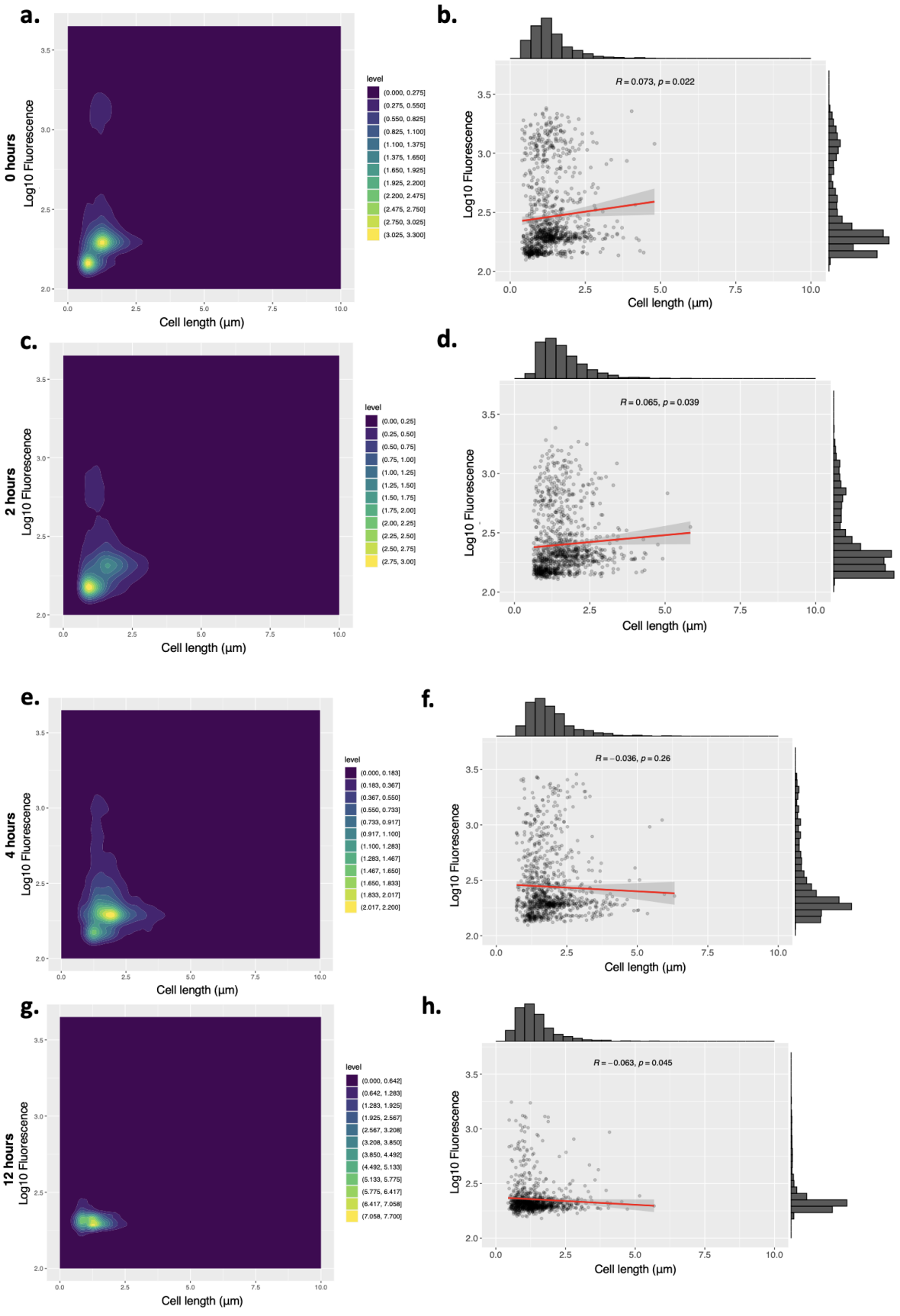

a.
b.
c.
d.
e.
f.
g.
h.
0 hours
2 hours
4 hours
12 hours
Log10 Fluorescence
Cell length (μm)
level
R = 0.073, p = 0.022
R = 0.065, p = 0.039
R = −0.036, p = 0.26
R = −0.063, p = 0.045

Figure 3.12 Changes in cell length were not strongly correlated with fluorescence changes in pfim::GFP transformed cells. Liquid cultures of resuspended colonies were grown for 12 hours, samples were taken at 0, 2, 4 and 12 hours and photographed microscopically. Segmented cell data for each time point was plotted as 2d density plots at **(a)** 0 hours, **(c)** 2 hours, **(e)** 4 hours and **(g)** 12 hours, and scatter plots at **(b)** 0 hours, **(d)** 2 hours, **(f)** 4 hours and **(h)** 12 hours. Scatter plots were constructed using a random sample of 1000 cells from each data set of pooled colony values.

At 0 hours, the population of pgfc::GFP transformed SC375 cells showed no correlation between cell length and fluorescence (**Fig. 3.13b**). A very large proportion of the population had significant mean fluorescence of over 2.25. With the exception of a small cluster of cells between 0.8 and 1.8µm in length (that had the greatest fluorescence values of 3.1 to 3.6) there was no evident differentiation into subpopulations (**Fig. 3.13a**). Cells up to 3.2 µm long were significantly fluorescent, and insignificant values were observed in cells as small as 0.4µm long.

After 2 hours of growth, two subpopulations of cells that differed the most in their length were evident (**Fig. 3.13c**). The most fluorescent subpopulation contained smaller cells of less than 2.5µm long, while the less fluorescent subpopulation contained cells that ranged from 0.4 to 5.0µm. Even the most lowly fluorescent cells had values above 2.25, indicating that they were significant and not due to autofluorescence. These populations remained at further time points. After 4 hours, there appeared to be a population bifurcation. The largest proportion of cells were short and highly fluorescent (**Fig. 3.13e**). The less fluorescent cluster of cells lost their ability to fluoresce, with no significant values above 2.5; cells within this cluster still ranged from 0.4 to 5.0µm long.

After 12 hours, both subpopulations of cells switched shorter, while still differing by their length. Highly fluorescent cells were between 0.4 and 1.4µm long. The less fluorescent subpopulation of cells were between 1.0 and 2.7µm long - shorter than their values at 4 hours.

We observed a negative correlation between cell length and fluorescence, with a statistically correlation coefficient of -0.35 at 2 hours (**Fig. 3.13d**). This negative correlation remained throughout the rest of the experimental procedure, with statistically significant correlation coefficients of -0.42 at 4 hours (**Fig. 3.13f**) and -0.32 at 12 hours (**Fig. 3.13h**).

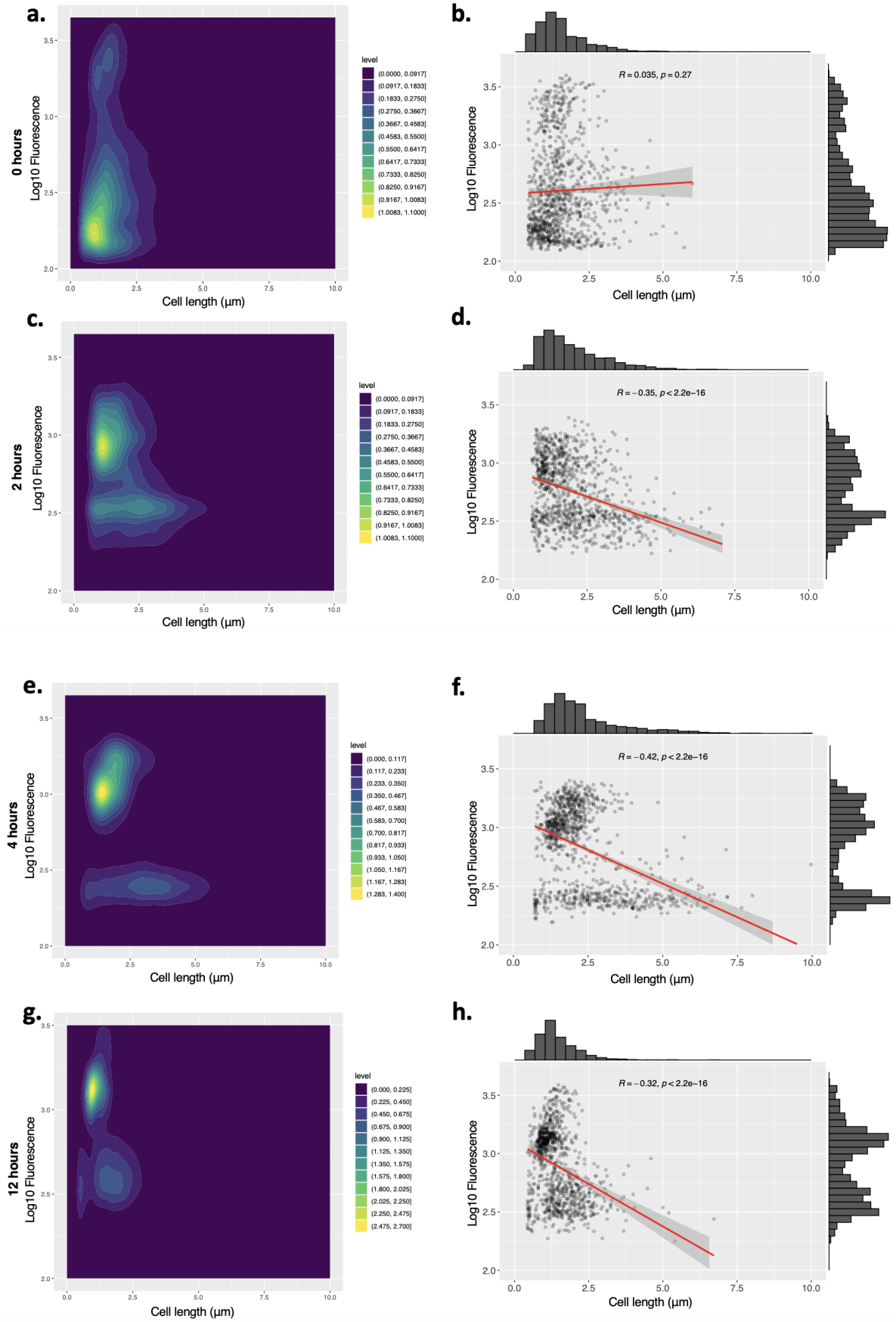

a.
b.
c.
d.
e.
f.
g.
h.
0 hours
2 hours
4 hours
12 hours
Log10 Fluorescence
Cell length (μm)
level
R = 0.035, p = 0.27
R = −0.35, p < 2.2e−16
R = −0.42, p < 2.2e−16
R = −0.32, p < 2.2e−16

Figure 3.13 SC375 cells transformed with the pgfc::GFP expression plasmid show a negative correlation between cell length and cell fluorescence. Liquid cultures of resuspended colonies were grown for 12 hours, samples were taken at 0, 2, 4 and 12 hours and photographed microscopically. Segmented cell data for each time point was plotted as 2d density plots at **(a)** 0 hours, **(c)** 2 hours, **(e)** 4 hours and **(g)** 12 hours, and scatter plots at **(b)** 0 hours, **(d)** 2 hours, **(f)** 4 hours and **(h)** 12 hours. Scatter plots were constructed using a random sample of 1000 cells from each data set of pooled colony values.

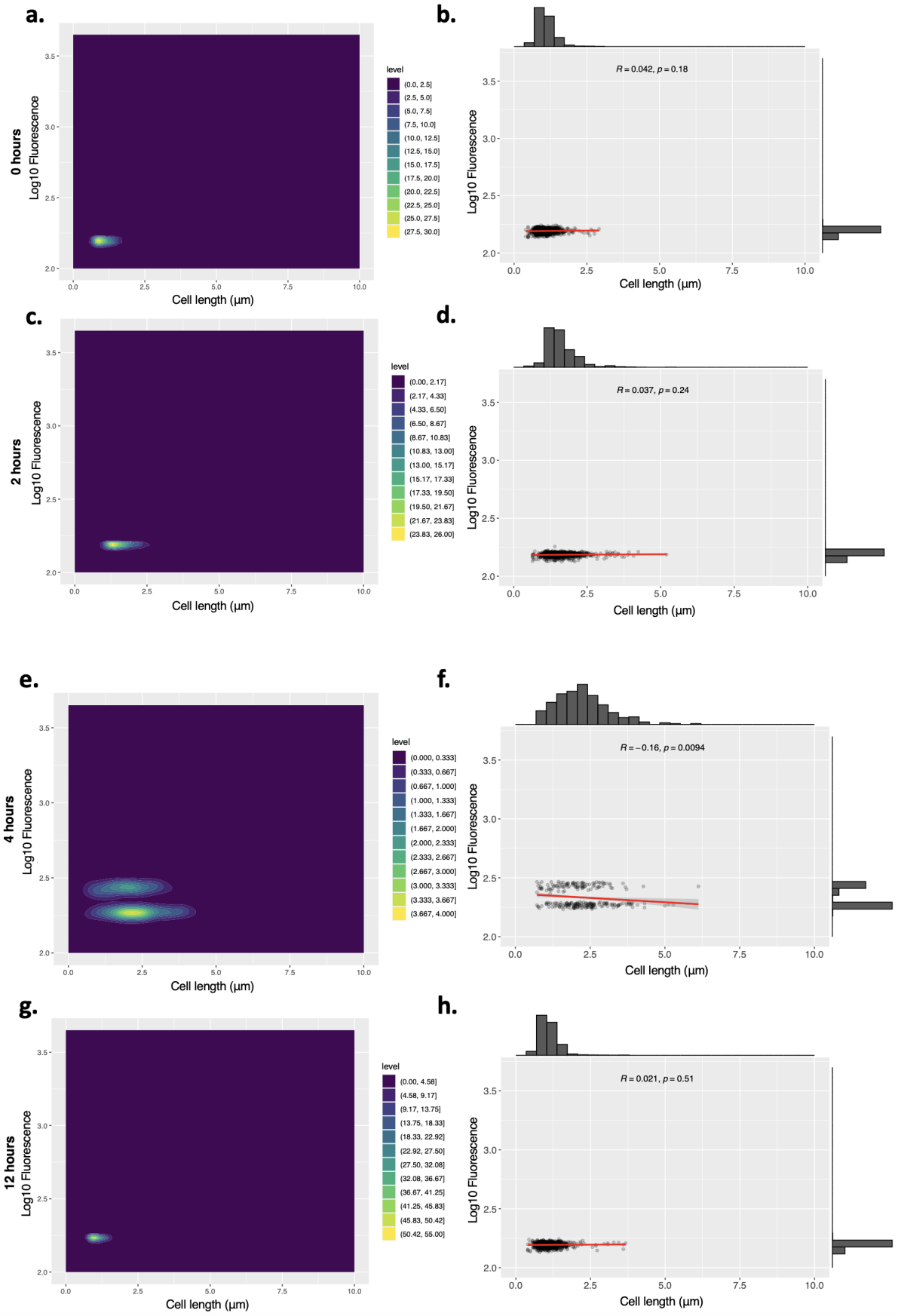

a.
b.
c.
d.
e.
f.
g.
h.
0 hours
2 hours
4 hours
12 hours
Log10 Fluorescence
Cell length (μm)
level
(0.0, 2.5]
(2.5, 5.0]
(5.0, 7.5]
(7.5, 10.0]
(10.0, 12.5]
(12.5, 15.0]
(15.0, 17.5]
(17.5, 20.0]
(20.0, 22.5]
(22.5, 25.0]
(25.0, 27.5]
(27.5, 30.0]
R = 0.042, p = 0.18
(0.00, 2.17]
(2.17, 4.33]
(4.33, 6.50]
(6.50, 8.67]
(8.67, 10.83]
(10.83, 13.00]
(13.00, 15.17]
(15.17, 17.33]
(17.33, 19.50]
(19.50, 21.67]
(21.67, 23.83]
(23.83, 26.00]
R = 0.037, p = 0.24
(0.000, 0.333]
(0.333, 0.667]
(0.667, 1.000]
(1.000, 1.333]
(1.333, 1.667]
(1.667, 2.000]
(2.000, 2.333]
(2.333, 2.667]
(2.667, 3.000]
(3.000, 3.333]
(3.333, 3.667]
(3.667, 4.000]
R = −0.16, p = 0.0094
(0.00, 4.58]
(4.58, 9.17]
(9.17, 13.75]
(13.75, 18.33]
(18.33, 22.92]
(22.92, 27.50]
(27.50, 32.08]
(32.08, 36.67]
(36.67, 41.25]
(41.25, 45.83]
(45.83, 50.42]
(50.42, 55.00]
R = 0.021, p = 0.51

Figure 3.14 Untransformed cells have increased levels of autofluorescence after 4 hours of growth. Liquid cultures of resuspended colonies were grown for 12 hours, samples were taken at 0, 2, 4 and 12 hours and photographed microscopically. Segmented cell data for each time point was plotted as 2d density plots at **(a)** 0 hours, **(c)** 2 hours, **(e)** 4 hours and **(g)** 12 hours, and scatter plots at **(b)** 0 hours, **(d)** 2 hours, **(f)** 4 hours and **(h)** 12 hours. Scatter plots were constructed using a random sample of 1000 cells from each data set of pooled colony values.

Chapter 4: Discussion

4.1: Identification of a phenotypically heterogeneous E. coli natural isolate

Bacterial populations possess the ability to alter their phenotype independently of each other, forming subpopulations of physically and functionally different cells. When a population contains varying cellular phenotypes it is considered to be phenotypically heterogeneous. Phenotypic heterogeneity can be regulated through a number of molecular mechanisms, either genetic or epigenetic, or it can occur randomly as a result of mutations. Intentionally generating heterogeneity can improve the ability of a population to adapt and evolve, increasing its overall fitness.

In bacteria, genes that display phenotypic heterogeneity often play important roles in virulence and host cell invasion (van der Woude, 2011; van der Woude & Bäumler, 2004). The most common example of phenotypic variation in bacteria is the regulation of fimbriae formation in *Uropathogenic E.coli* (Holden et al., 2007). However, phenotypic heterogeneity is exemplified in the colonisation process of *Pseudomonas fluorescens (P. fluorescens*) (Sánchez-Contreras et al., 2002; van den Broek et al., 2005). *P. fluorescens* is a gram-negative, rod shaped bacteria that forms varying colony morphologies made up of phenotypically different cells (Sánchez-Contreras et al., 2002). It has been proposed that a *P. fluorescens* population with different subpopulations is more successful when root colonisation becomes competitive (Martínez-Granero et al., 2005, 2006). This provides an alternative motivation for variable gene expression in gram-negative bacteria that is not associated with host cell immune evasion.

We have identified a natural isolate of *E. coli*, called SC375, that forms non-uniform colonies when grown on M9 minimal media. Two extreme colony morphologies that we have denoted as either compact or halo are distinguishable by size and shape, along with an intermediate mixed colony that combines features of the two (**Fig. 3.1**). Using flow cytometry we were able to determine that the different colony morphologies were caused by the presence of different cellular phenotypes (**Fig. 3.2**). Cells varied in their length, either remaining short or

growing up to seven times the length of the smallest cell. Halo and mixed colonies contained a mix of short and long cells while compact colonies had exclusively short cells.

When grown on LB agar, the SC375 strain formed exclusively compact colonies (**Fig. 5.1**). LB media has an abundance of nutrients that provide optimal conditions for cell growth. In contrast, minimal M9 media contains only salts and nitrogen and requires supplementation with sugars, to allow for the appropriate conditions for colony formation. Due to the limited nutrient availability of M9 media, we propose that differential colony formation in these conditions is a result of an attempt to survive in a state of stress. SC375 cells could benefit from introducing subpopulations of phenotypically heterogeneous cells in a similar manner to the *P. fluorescens* strain. Cells with different phenotypes could potentially carry out different functions which benefit the population as a division of labour. Because only compact colonies are formed when SC375 is grown on LB media, we determined that halo colonies were a deviation from regular cell growth. Therefore long cells would be a result of any stress induced phenotypic heterogeneity, and provide a survival advantage in non-preferred conditions.

We initially discovered these varying colony morphologies during a large-scale replica plating procedure. Multiple natural isolate strains were grown on a variety of growth media with different carbon sources. The different morphologies were immediately obvious when grown on M9 supplemented with lactose. It is widely acknowledged that *E. coli* shows the best growth in glucose (Bren et al., 2016; Law et al., 2002). Lactose utilisation requires the induction of a new metabolic pathway, an action that is only possible when glucose is present at very low volumes (Lewis, 2013; Reznikoff, 1992). For growth on M9 supplemented with lactose, this metabolic pathway would be required to work highly efficiently. The involvement of metabolic genes provides further potential for a possible division of labour motivation for heterogeneity in the SC375 strain. It is common for different subpopulations of cells to form phenotypes that are more efficient at using alternative metabolites in an attempt to better the survival of the population as a whole (Tsai & Coombes, 2019; Weigel & Dersch, 2018).

4.2: Switching between the two cellular phenotypes is reversible

On observation of cell growth recorded by time-lapse videos we were able to see that phenotypes can switch from one to the other in either direction. This was obvious in time-lapse movies, of which still images are shown in **Figure 3.5** and **Figure 5.2**. By tracking the length of cell progeny over time we were also able to determine that this switching happens rapidly on a manner of generations. The asexual nature of cell division in bacteria, means that daughter cells are genetically identical to their parent cell. If switching was only possible in one direction we would expect all daughter cells to be the same length as their parent cell at birth. Because cells elongate during division (Egan & Vollmer, 2013; Rothfield et al., 1999), recording cell length at birth was crucial to ensure that we got an accurate measure of length. Recording at any other time point could have potentially reflected typical cell cycle lengthening.

Long cells remained viable even when they reached lengths of over 4μm (**Fig. 3.4d**). Often when cells reach extreme lengths, it is as a result of errors during cell division. We showed that long cells are still able to grow and divide, regardless of how long they are at birth. This confirms that the heterogeneity in cell length of the SC375 strain is intentional, and therefore likely has an evolutionary advantage.

Interestingly, when we analysed switching at the population level, it appeared that switching occurred more frequently from long to short. Because compact colonies only contained short cells we were able to use cell length densities over time to infer large-scale patterns of short cell switching. Although the vast majority of cells within these colonies were short and remained short over time, there are outliers present and visible in **Figure 3.5** and **Figure 5.3**. This leads us to believe that switching is possible in the short to long direction, however it is rare.

Regardless of starting cell length, all cells switched back to short after a period of 12 hours. Average cell length for each colony type at 12 hours, reached a value of below their initial average starting lengths (**Fig. 3.5e**). Along with the narrowing of distribution peaks over time, the downward trend of average cell length in compact colonies suggested that short cells tend to stay short over time and are even able to get shorter. Long cells, exhibited by the cell length patterns of mixed and halo colonies, showed an upward trend before getting

significantly shorter between 4 and 12 hours. Based on the cell distribution patterns of halo and mixed colonies in **Figures 3.5a**, **b**, **c** and **d** coupled with the average cell length patterns, it seemed that the most likely trajectory for long cells was to get longer during exponential growth phases, before switching short at stationary phase.

4.3: SC375 phenotypes show variation in the expression of three virulence genes

Genome sequencing of both phenotypes confirmed that the SC375 strain was isogenic. Initially we had proposed that there was a genetic cause for the phenotypic variation, more specifically a DNA inversion similar to that of the *fim* operon (Emerson et al., 2009; Hinde et al., 2005; Schwan, 2011). Our nanopore sequencing results disproved this, showing no variation in sequence regardless of phenotype.

It seemed likely that if alterations in the genome were not causative of phenotypic heterogeneity, then regulation of phenotype formation was at the epigenome level. Epigenetic regulation affects gene expression at the transcription step (Jaenisch & Bird, 2003). In eukaryotes the most common epigenetic modification is methylation, which can influence the recruitment of repressors or activators to gene regulatory regions (Sánchez-Romero & Casadesús, 2020). For this reason, the best measure of epigenetic regulation is gene expression. When genes are expressed, DNA is directly translated into RNA by polymerase enzymes, and this RNA is used as a template to make proteins. The sequential procedure of gene expression means that we can use the corresponding RNA levels of any given gene to quantify the rate at which it is expressed. By performing RNA sequencing analysis we were able to identify two groups of genetic loci that were expressed differentially in long cells compared to the short cells.

Of the 10 genetic loci that were considered to have higher levels of gene expression in the long cell phenotype, only two were associated with known protein encoding genes (**Fig. 3.6d** and **Table 5.1**). Both genetic loci showed significant p-values confirming that variation in their expression were due to phenotypic changes and not just by random chance. These two genes were *AidA-I* and *pdeL_3.* To identify the gene best suited for downstream analysis, we looked at the baseMean values of each gene. The baseMean values refer to the average of the normalised count values, or the average amount of RNA present in both sample types for

that gene. *AidA-I* was expressed at a rate 10.5 times greater than pdeL_3. We concluded that because *AidA-I* is basally expressed at significantly higher values, any variation in its expression would have a greater impact than genes that were more lowly expressed.

The *AidA-I* gene encodes AIDA-I, *e.coli* adhesion involved in diffusion adherence, an outer membrane autotransporter protein responsible for attachment to host epithelial cells (Ngeleka et al., 2003). Because adhesion is an important step in infectious disease development, the AIDA-I protein plays a crucial role in *E. coli* virulence (Benz & Schmidt, 1992; Charbonneau et al., 2006; Zhao et al., 2009). A greater expression of AIDA-I in long cells would suggest that the longer phenotype is more efficient in adhesion than the short phenotype.

A larger quantity of genetic loci identified to be negatively regulated in long cells are associated with known protein encoding genes (**Fig. 3.6c** and **Table 5.2**). Out of the 16 loci identified, only two were considered to encode hypothetical proteins. Because the other 14 loci had significant p-values we once again looked for genes that had the greatest basal expression. The *fimA* gene had the highest basal expression of all 16 loci, with more than three times the RNA quantity than the next closest gene. Because the *fimA* gene is part of a large cluster of genes within the same operon, under the same promoter, selecting *fimA* for downstream analysis allowed us to see how expression of the whole operon affected SC375 phenotype. The *fimA* gene in particular is responsible for the FimA protein, the major pilin subunit (Nagano et al., 2012). Although we have established that the *fim* operon is the most well known example of phase variation in *uropathogenic E. coli*, its regulatory methods involve a DNA inversion upstream of the *fimA* gene promoter. Variation in the expression of the *fimA* gene without this inversion seems unusual. This could reflect either an error in our sequencing and alignment process, or alternatively our RNA-sequencing method.

The second most highly expressed gene was *gfcA*, a gene located within the *gfc* operon. There more *gfc* operon genes: *gfcB*, etp and etk were also identified to be negatively regulated in long cells. Enteropathogenic and enterohemorrhagic strains of *E. coli* are covered in extracellular polysaccharides and lipopolysaccharides that form a protective capsule around each cell (Larson et al., 2021; Sathiyamoorthy et al., 2011)**.** The *gfc* operon is responsible for the production and exportation of the group 4 capsule (Larson et al., 2021).

Group 4 capsules contain cell surface antigens and polypeptides that help pathogenic *E. coli* evade host cells (Whitfield & Roberts, 1999). The *gfcA* gene encodes the shortest operonic gene, a protein with a large component of threonine residues, that is predicted to encode an inner membrane protein (Ghovvati et al., 2018; Larson et al., 2021). Group 4 capsule secretion and *gfcA* expression has been associated with the evasion of human immune cells (Thomassin et al., 2013).

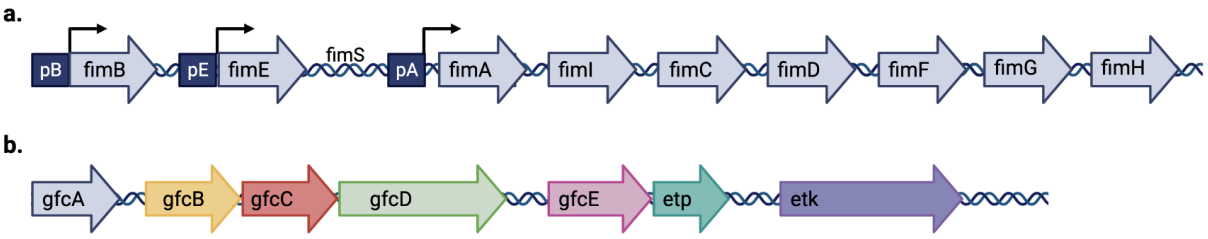


Figure 4.1 Operonic structure of the two gene clusters upregulated in the short cell phenotype. (a) structure of the fim operon (regulation of the fimS invertible region is shown in **Fig. 1.2**). **(b)** structure of the gfcABCD operon.

4.4: Associating fluorescence with cell length in resuspended colonies

We have well established the cellular phenotype composition of each colony morphology. We know that compact colonies consist of predominantly short cells, and mixed and halo colonies contain both long and short cells. Based on our RNA-sequencing data we identified our three genes of interest with varying expression levels in either phenotype. We predicted that using our expression constructs, we would be able to associate expression levels with fluorescence intensity in either phenotype.

We mentioned in *Chapter 3.5* that we were unable to confirm transformation of the paida::GFP plasmid into our electrocompetent SC375 cells (**Figure. 5.5**). The insert length was approximately 2100bp, which was nearly half of the length of our pUA66 backbone. This made it difficult to follow the gibson assembly method of plasmid construction, because large volumes of the PCR product were required for successful ligation. We were able to confirm that all of our promoter region inserts for each gene or gene cluster were successful (**Fig. 5.4b**). Our primers were designed correctly and our PCR reactions were set up in the correct manner. We attributed our errors in transformation to the long insert length. A successfully

constructed paida::GFP plasmid was 6720bp long, as opposed to the 5320bp for pfim::GFP or 5330bp for pgfc::GFP. Further gene expression analysis was conducted using only the pfim::GFP and the pgfc::GFP plasmids.

Fluorescence imaging of transformed colonies showed the physical distributions of short and long cells within each colony morphology. The *gfc* and *fim* gene clusters were both upregulated in short cells, therefore we would expect compact colonies to be fluorescent; mixed and halo colonies should be either not fluorescent at all or only exhibiting low levels of fluorescence. Imaging of the different colony morphologies with phase contrast and 470nm optical configurations partially confirmed our findings (**Fig. 3.7** and **Fig. 3.8**). Compact colonies were very highly fluorescent. Colony characteristics were just as obvious in the fluorescence image as they were in their corresponding phase contrast images. Non-fluorescent colonies showed comparatively low levels of fluorescence and it was not possible to identify colony boundaries. Because we know that the most fluorescent cells are short cells, any areas that lack fluorescence are very likely to be long cells. Therefore we could see that halo colonies retained a low proportion of short cells within the colony centre, while containing a large amount of long cells towards the outer edges. A deposit of long cells around the edges, correlates with the translucent halo, concluding that long cells are responsible for this morphology characteristic.

For comparison of fluorescence and cell length at the single cell level, we used flow cytometry to analyse each colony. The pgfc::GFP transformed strain confirmed our expected pattern, with short cells fluorescing highly, and long cells only showing low levels of fluorescence (**Fig. 3.9**). We also noted that, in comparison to results from the empty pUA66 plasmid that we used as a negative control, long cells showed low levels of fluorescence rather than no fluorescence at all. In contrast, the pfim::GFP transformed strain showed no association between cell length and fluorescence (**Fig. 3.10**). Both long and short cells were equally as fluorescent as each other in every colony type, regardless of whether the colony was fluorescent, mixed or non-fluorescent. The proportion of long cells within a fluorescent colony of the pfim::GFP transformed strain was minimal, however it appeared that all long cells present had significant fluorescence values (**Fig. 3.10c**).

It is worth noting that once colonies have formed on agar plates, most of the cells within them are in stationary phase. Stationary phase describes a non-proliferative state where cells are still metabolically active but not growing (Nyström, 2004; Zambrano & Kolter, 1996), as a result of lessened access to nutrients. During this phase, cells alter their gene expression to ensure that resources are put into promoting cell survival (Jaishankar & Srivastava, 2017). Because the *fim* operon is responsible for proteins involved in cell surface fimbriae, it is possible that at stationary phase its expression is altered in favour of metabolic genes. Therefore to get a better understanding of how gene expression changes within each cellular phenotype, we need to investigate how it changes over time, through all growth phases.

4.5 Gene expression is associated with cell length changes

If changes in gene expression were responsible for the phenotypic switching we observed in the SC375 strain, we would expect any changes in cell length over time to directly correspond with changes in gene expression. In this study we analysed the relationship between the two by tracking growth and fluorescence of our transformed strains over time.

Our initial approach was to perform flow cytometry on glycerol inoculated cultures that had been left to grow for 4 and 12 hours. We suspected that after 4 hours cells would be in exponential phase, a point of rapid cell growth and division. This was also the time point that cells were extracted for RNA sequencing. Based on these two factors, we expected to see a strong association between cell length and the expression of both the *fim* and *gfc* operons. After 12 hours, we would expect cells to reach stationary phase, and therefore show expression and length patterns similar to the resuspended colonies.

Once again, expression of the *gfc* genes was as expected, with long cells showing lower GFP values than short cells. After 4 hours of growth there was a fairly equal proportion of long and short cells within the population (**Fig. 3.11b**). Cell clusters differed on both their x and y values, making it easy to visualise the association between size and fluorescence of pgfc::GFP transformed cells. After 12 hours the association between length and fluorescence remained (**Fig. 3.11e**), however, there was a much greater proportion of

shorter, more fluorescent, cells within the population than longer, less fluorescent, cells (**Fig. 3.11e**).

After both 4 and 12 hours of growth, there was no association between cell length and fluorescence in pfim::GFP transformed colonies (**Fig. 3.11a**). Both long and short cells demonstrated the ability to fluoresce at significant values. The strain did however, have varying proportions of long and short cells at each time point. After 4 hours (**Fig. 3.11a**) long cells were present, although not in equal amounts to small cells like we saw in pgfc::GFP transformed cells. After 12 hours (**Fig. 3.11d**) their volume decreased even more, showing a much greater proportion of short cells making up the population. This change also seemed to correspond with a change in the GFP values of each population. After 12 hours of growth fluorescence decreased quite substantially, with minimal significant fluorescence values. This directly opposes what we would have expected if fluorescence was related to cell length.

We were able to more directly associate cell length with fluorescence on a single cell level using microscopic analysis. Because we can gather data from each individual cell within a microscopy image, we are able to confidently discern values for cell length and the mean fluorescence of each cell. In our attempt to make this processing high throughput, we grew cells over time in liquid culture and took images of samples removed at 0, 2, 4 and 12 hours. We discovered that the rate limiting step of the superSegger segmentation process is the linking of cells between each image. If we remove this step, it decreases the procedure from 16 hours to a matter of minutes. Performing microscopy in this way allowed us to gather data from thousands of individual cells to give a broad understanding of the changes in each transformed strain at a population wide level.

At nearly every time point, expression of the *gfc* operon was negatively correlated with cell length (**Fig. 3.13**). Subpopulations of cells that varied in their mean fluorescence values were visible at all time points, and after 2 hours (**Fig. 3.13c**) these subpopulations also began to vary in their cell length. Subpopulations remained at 12 hours, even when cells were likely in the stationary phase. At hour 0, there is a very (insignificantly) weak positive correlation between cell length and mean fluorescence (**Fig. 3.13b**). Although it is clear that at this time, most fluorescent cells are short, there are some longer cells that have significant

fluorescence levels. We propose that this is due to an increase in the expression of the *gfc* operon that occurs before cells switch to short.

No strong correlation, either positive or negative, was observed between cell length and gene expression in pfim::GFP transformed cells. At 0, 2 and 4 hours populations contain subpopulations that differ by cell length and fluorescence. As established in our flow cytometry data, these subpopulations have an opposite relationship between length and fluorescence to what we expected based on our RNA sequencing data. With the exception of a small cluster of highly fluorescent short cells, short cells remained predominantly less fluorescent than long cells. After 4 and 12 hours the difference between their two values decreased.

Peak *fim* gene expression was reached at 4 hours (**Fig. 3.12 e** and **f, Fig. 5.12**). At 12 hours, the length of the longest cells within the population had decreased significantly. We noted in time-lapse movies of cells transformed with the pfim::GFP strain, fluorescence intensity of long cells increased just before they switched to short. We attribute this to an increase in the expression of the genes within the *fim* operon. We suspected that, like the *gfc* operon, an increase in *fim* expression occurred before long cells made the switch to short cells.

The *fim* operon contains nine expressed genes. Of these nine genes, we identified six that were downregulated in long cells, comparative to their expression in short cells. Of these six *fim* genes, *fimA* was expressed at the highest basal level. It showed a Log2 fold change of -2.43, indicating that its expression in long cells was about 20% of that in short cells. The expression of the FimA protein is regulated through the action of two other proteins encoded within the fim operon, FimB and FimE (Holden et al., 2007; Schwan, 2011; Zhang et al., 2016). FimB and FimE are both invertases that control the inversion of an invertible region upstream of *fimA* called *fimS* (Emerson et al., 2009; Hinde et al., 2005). The FimB and FimE proteins are highly basic, a characteristic of many DNA binding proteins (Schwan, 2011). It is possible that when we inserted the *fim* promoter region, controlling the expression of *fimA*, into our pUA66 backbone it affected the ability of FimB and FimE to bind to the regulatory regions within *fimS*. This could result in *fimS* being fixed in the ON orientation, which would mean that it is not subjected to the same regulatory mechanisms as genomic *fimA.*

We confirmed that the PCR product of our *fim* promoter sequence shared 100% sequence identity with the SC375 genomic *fim* operon (**Fig. 5.7**). Therefore we can conclude that mutations are not the cause of irregular *fim* expression.

In comparison, little is known about the regulation of the expression of the *gfc* operon, but it has been linked to the production and exportation of the group 4 capsule (Larson et al., 2021)**.** The most important role of the cell surface antigens and polypeptides within group 4 capsules is their ability to help pathogenic *E.coli* evade host cells (Whitfield & Roberts, 1999). Because capsules act to envelop bacteria, we propose that any variation in the gene expression of the proteins that contribute to their formation, could result in dramatic changes to the cell surface.

4.6 The importance of understanding the fundamental mechanisms of phenotypic variation

The generation of phenotypic heterogeneity is important for bacterial survival. It plays a fundamental role in adaptation and evolution, which is particularly important for populations growing in unstable conditions. A broad fundamental understanding of the processes and mechanisms in which bacteria generate and retain phenotypic heterogeneity is beneficial in many ways. Knowledge of these bacterial processes can be translated to areas of biotechnology including identifying vaccine targets and combatting resilient bacterial infections.

Uropathogenic E. *coli* is responsible for a wide range of diseases including urinary tract infections (UTIs) and intestinal diseases (Marrs et al., 2005). Since the discovery of antibiotics in the 1940's (Bigger, 1944; Hobby et al., 1942) they have been the primary mode of treatment for bacterial infections in humans and animals. Along with this new "miracle" drug, came the rise in antibiotic persisters (Bigger, 1944), a small subpopulation of cells that exhibit temporary multidrug tolerance (Harms et al., 2016; Michiels et al., 2016). The presence of persisters within a bacterial strain have been linked to latent infections that can prove costly to individuals and populations (Hill & Helaine, 2019; Torrey et al., 2016).

Persister formation is best described as a harbouring of phenotypically heterogeneous cells within a much larger population. The smaller subpopulation of persister cells demonstrates a

bet-hedging strategy of heterogeneity that aids in population survival only after exposure to antibiotics. It is not yet well understood how these persister populations form, although a recent study has identified invertible DNA regions within persistence genes (Jiang et al., 2019), suggesting a possible link to phase variation. However, it is widely believed that persister cells are genetically identical (Kaldalu et al., 2020; Lewis, 2007; Willenborg et al., 2014) which would mean persistence is mediated by epigenetic regulation.

The generation of persister cells is a reversible process (Andersson, 2003; Kaldalu et al., 2020; Sundqvist, 2014). Tolerant persister cells temporarily enter a state of dormancy and are able to return to a proliferative state when the antibiotic is removed (Lewis, 2010; Wood et al., 2013). Despite the removal of the environmental stressor, a subpopulation of these persister cells tends to remain. A 2018 study proposed that this was due to bacterial memory, with populations showing they are capable of retaining this ability long term (Miyaue et al., 2018). This opposes the theory that persister cells are stochastically generated.

The phenotypic switching mechanism that we have identified in this thesis is also a reversible process. It is not regulated genetically but rather through alterations in gene expression which is an epigenetic process. The most common form of epigenetic regulation in prokaryotic cells is methylation (Sánchez-Romero & Casadesús, 2020), the addition of methyl groups to the DNA structure itself. Methylation acts as a physical barrier that either aids or inhibits the recruitment of transcriptional proteins to the DNA sequence. Interestingly, methylation dependent DNA modifications are heritable, which makes them an important factor in bacterial memory (Casadesús & D'Ari, 2002; Sánchez-Romero & Casadesús, 2020). The rapid nature of the phenotypic switching in the SC375 strain, coupled with the ability of cells to be either long or short regardless of their parent cell phenotype, is suggestive of memory playing an important role in cell length regulation.

All three of the genes that we identified to have differential regulation in each phenotype were virulence genes, characteristic of infectious *E. coli* strains. Of particular interest was the expression of the *gfc* operon - a collection of genes that are responsible for producing and exporting the products in bacterial group 4 capsules. This capsule in particular has been associated with immune evasion in human hosts (Thomassin et al., 2013), playing a crucial part of the strains virulence capabilities. Virulence genes are most highly expressed in

infectious strains, and are typically on the surface of a bacterial cell (Guiney, 1997; Ørskov, 1978). Cell surface proteins such as these are often used as vaccine targets (Sundar & Singh, 2014; Wells et al., 2007). Because genes that show variation in their expression may not be consistently present, or could potentially show alterations in their biochemical structure, they would not be suitable vaccine targets. This could potentially rule out a large number of otherwise favourable protein targets in vaccine development. An understanding of the nature of the variable expression of cell surface proteins could potentially broaden the types of proteins available for use as vaccine targets, aiding vaccine development.

4.7 Conclusions and future perspectives

In this thesis, we have identified a rapid phenotypic switching mechanism in a natural isolate of *E.coli*. This isolate strain has the ability to form either of two extreme colony morphologies, influenced by their cellular composition.

Microscopic analysis allowed us to establish that this switching occurs over a matter of generations. Cells are able to remain short or grow long, and switch between the two phenotypes in either direction. This switching happens on a large scale, causing the population to bifurcate into two subpopulations of either phenotype. Due to the extreme nature of cell lengths, we initially proposed a phase variable-like mechanism for the control of phenotype formation. We had proposed that regulation of such a mechanism would be due to a DNA inversion, similar to that observed in the *fim* operon. Genome sequencing disproved this, confirming that cells remained isogenic regardless of their length.

RNA sequencing analysis identified variations in the expression patterns of three virulence genes. Genes within the *fim* and *gfc* operons, were upregulated in short cells, and down regulated in long cells. A single significant gene, *AidA-I*, which encoded an outer membrane autotransporter protein responsible for attachment to host epithelial cells, was upregulated in long cells. Unfortunately due to the large size of the promoter region that controls the expression of this gene we were unable to continue with downstream analysis.

We were able to confidently identify a correlation between the expression of the *gfc* operon and changes in cell length. This gene was consistently downregulated in long cells and upregulated in short cells at all time points during cell growth. Although we were not able to

see a complete absence of expression in long cells, we attributed this to the rapid switching seen between each phenotype. It is likely that an increase in expression precedes the switch from long to short, a delay that is caused by the build up of protein products.

The methods in this study were not sufficient to determine exactly how expression of the *gfc* operon is regulated. By performing genomic sequencing, we were able to rule out a genetic cause and we have therefore concluded that this gene is regulated through epigenetic modifications. The most common epigenetic modification in bacteria is methylation, so we propose that it is likely methylation that plays a role in *gfc* operonic expression. Methylation analysis can be performed through nanopore sequencing, as the sequencing procedure is able to detect nucleotide modifications.

Because of the rapid nature of switching and the ability of progeny to be long or short regardless of parent cell phenotype, we propose that this switch could reflect bacterial memory. Bacterial memory is often exemplified through heritable DNA methylation patterns; for this reason investigating a potential role of methylation in the expression of our three key genes of interest could confirm our hypothesis. It would also be beneficial to see if the SC375 strain retains its ability to form a long cell phenotype when transformed from growth in M9 media, to LB media. If the strain does retain the ability to form either phenotype, it could be a further indicator of memory.

We discovered the ability of this strain to form different colony morphologies during a replica plating procedure. Although this involved growing *E. coli* on agar plates with multiple nutrient sources, in this thesis we focused on M9 supplemented with lactose. This was the nutrient combination that we found had the most extreme effect on colony morphology. Further experimental work on this strain should involve growth on multiple different carbon sources, and at different temperatures. This could potentially discover a link between metabolism and switching events.

We were unable to continue with the downstream analysis of the expression of the *AidA-I* gene. The length of its gene regulatory region caused problems during the ligation and transformation process. Further studies should involve troubleshooting this step, to include an *AidA-I* plasmid construct.

The findings in this research, help provide a fundamental understanding of how phenotypic heterogeneity is regulated in bacterial cells. We have identified a natural isolate strain that could display bacterial memory in the regulation of virulence genes. Knowledge of how heterogeneity of these genes is generated and maintained in an isogenic strain could provide further insights into how infectious bacteria are able to evade host immune cells and antibiotic treatments.

References

Ackermann, M., Stecher, B., Freed, N. E., Songhet, P., Hardt, W.-D., & Doebeli, M. (2008). Self-destructive cooperation mediated by phenotypic noise. *Nature*, *454*(7207), 987–990. https://doi.org/10.1038/nature07067

Adhikari, S., & Curtis, P. D. (2016). DNA methyltransferases and epigenetic regulation in bacteria. *FEMS Microbiology Reviews*, *40*(5), 575–591. https://doi.org/10.1093/femsre/fuw023

Alamro, M., Bidmos, F. A., Chan, H., Oldfield, N. J., Newton, E., Bai, X., Aidley, J., Care, R., Mattick, C., Turner, D. P. J., Neal, K. R., Ala'Aldeen, D. A. A., Feavers, I., Borrow, R., & Bayliss, C. D. (2014). Phase Variation Mediates Reductions in Expression of Surface Proteins during Persistent Meningococcal Carriage. *Infection and Immunity*, *82*(6), 2472–2484. https://doi.org/10.1128/IAI.01521-14

Andersson, D. I. (2003). Persistence of antibiotic resistant bacteria. *Current Opinion in Microbiology*, *6*(5), 452–456. https://doi.org/10.1016/j.mib.2003.09.001

Azzaz, F., & Fantini, J. (2022). The epigenetic dimension of protein structure. *Biomolecular Concepts*, *13*(1), 55–60. https://doi.org/10.1515/bmc-2022-0006

Balaban, N. Q., & Liu, J. (2019). Evolution Under Antibiotic Treatments: Interplay Between Antibiotic Persistence, Tolerance, and Resistance. In K. Lewis (Ed.), *Persister Cells and Infectious Disease* (pp. 1–17). Springer International Publishing. https://doi.org/10.1007/978-3-030-25241-0_1

Balaban, N. Q., Merrin, J., Chait, R., Kowalik, L., & Leibler, S. (2004). Bacterial Persistence as a Phenotypic Switch. *Science*, *305*(5690), 1622–1625. https://doi.org/10.1126/science.1099390

Barrett, T. C., Mok, W. W. K., Murawski, A. M., & Brynildsen, M. P. (2019). Enhanced antibiotic resistance development from fluoroquinolone persisters after a single exposure to antibiotic. *Nature Communications*, *10*(1), 1177. https://doi.org/10.1038/s41467-019-09058-4

Bayliss, C. D., Fallaize, C., Howitt, R., & Tretyakov, M. V. (2019). Mutation and Selection in Bacteria: Modelling and Calibration. *Bulletin of Mathematical Biology*, *81*(3), 639–675. https://doi.org/10.1007/s11538-018-0529-9

Beaulaurier, J., Schadt, E. E., & Fang, G. (2019). Deciphering bacterial epigenomes using modern sequencing technologies. *Nature Reviews Genetics*, *20*(3), 157–172. https://doi.org/10.1038/s41576-018-0081-3

Becker, F., Wienand, K., Lechner, M., Frey, E., & Jung, H. (2018). Interactions mediated by a public good transiently increase cooperativity in growing Pseudomonas putida metapopulations. *Scientific Reports*, *8*(1), 4093. https://doi.org/10.1038/s41598-018-22306-9

Benz, I., & Schmidt, M. A. (1992). AIDA-I, the adhesin involved in diffuse adherence of the diarrheagenic Escherichia coli strain 2787 (O126:H27), is synthesised via a precursor molecule. *Molecular Microbiology*, *6*(11), 1539–1546. https://doi.org/10.1111/j.1365-2958.1992.tb00875.x

Bigger, JosephW. (1944). TREATMENT OF STAPHYLOCOCCAL INFECTIONS WITH PENICILLIN BY INTERMITTENT STERILISATION. *The Lancet*, *244*(6320), 497–500. https://doi.org/10.1016/S0140-6736(00)74210-3

Bikard, D., & Marraffini, L. A. (2012). Innate and adaptive immunity in bacteria: Mechanisms of programmed genetic variation to fight bacteriophages. *Current Opinion in Immunology*, *24*(1), 15–20. https://doi.org/10.1016/j.coi.2011.10.005

Binder, D., Drepper, T., Jaeger, K.-E., Delvigne, F., Wiechert, W., Kohlheyer, D., & Grünberger, A. (2017). Homogenising bacterial cell factories: Analysis and engineering of phenotypic heterogeneity. *Metabolic Engineering*, *42*, 145–156. https://doi.org/10.1016/j.ymben.2017.06.009

BioRender. (n.d.). Retrieved March 30, 2022, from https://biorender.com/

Blomfield, I. C., McClain, M. S., Princ, J. A., Calie, P. J., & Eisenstein, B. I. (1991). Type 1 fimbriation and fimE mutants of Escherichia coli K-12. *Journal of Bacteriology*, *173*(17), 5298–5307. https://doi.org/10.1128/jb.173.17.5298-5307.1991

Bobay, L.-M. (2020). CoreSimul: A forward-in-time simulator of genome evolution for prokaryotes modelling homologous recombination. *BMC Bioinformatics*, *21*(1), 264. https://doi.org/10.1186/s12859-020-03619-x

Bren, A., Park, J. O., Towbin, B. D., Dekel, E., Rabinowitz, J. D., & Alon, U. (2016). Glucose becomes one of the worst carbon sources for E.coli on poor nitrogen sources due to suboptimal levels of cAMP. *Scientific Reports*, *6*(1), 24834.

https://doi.org/10.1038/srep24834

Broadbent, S. E., Davies, M. R., & Van Der Woude, M. W. (2010). Phase variation controls expression of Salmonella lipopolysaccharide modification genes by a DNA methylation-dependent mechanism: Phase variable modification of Salmonella O-antigen. *Molecular Microbiology*, *77*(2), 337–353. https://doi.org/10.1111/j.1365-2958.2010.07203.x

Buettner, F., Natarajan, K. N., Casale, F. P., Proserpio, V., Scialdone, A., Theis, F. J., Teichmann, S. A., Marioni, J. C., & Stegle, O. (2015). Computational analysis of cell-to-cell heterogeneity in single-cell RNA-sequencing data reveals hidden subpopulations of cells. *Nature Biotechnology*, *33*(2), 155–160. https://doi.org/10.1038/nbt.3102

Cahoon, L. A., & Seifert, H. S. (2013). Transcription of a cis-acting, Noncoding, Small RNA Is Required for Pilin Antigenic Variation in Neisseria gonorrhoeae. *PLoS Pathogens*, *9*(1), e1003074. https://doi.org/10.1371/journal.ppat.1003074

Callaghan, M. J., Jolley, K. A., & Maiden, M. C. J. (2006). Opacity-Associated Adhesin Repertoire in Hyper Invasive *Neisseria meningitidis*. *Infection and Immunity*, *74*(9), 5085–5094. https://doi.org/10.1128/IAI.00293-06

Carey, J. N., & Goulian, M. (2019). A bacterial signaling system regulates noise to enable bet hedging. *Current Genetics*, *65*(1), 65–70. https://doi.org/10.1007/s00294-018-0856-2

Casadesús, J., & D'Ari, R. (2002). Memory in bacteria and phage: Review articles. *BioEssays*, *24*(6), 512–518. https://doi.org/10.1002/bies.10102

Cehovin, A., Winterbotham, M., Lucidarme, J., Borrow, R., Tang, C. M., Exley, R. M., & Pelicic, V. (2010). Sequence conservation of pilus subunits in Neisseria meningitidis. *Vaccine*, *28*(30), 4817–4826. https://doi.org/10.1016/j.vaccine.2010.04.065

Charbonneau, M.-È., Berthiaume, F., & Mourez, M. (2006). Proteolytic Processing Is Not Essential for Multiple Functions of the Escherichia coli Autotransporter Adhesin Involved in Diffuse Adherence (AIDA-I). *Journal of Bacteriology*, *188*(24), 8504–8512. https://doi.org/10.1128/JB.00864-06

Chater, K. F., Biró, S., Lee, K. J., Palmer, T., & Schrempf, H. (2010). The complex extracellular biology of *Streptomyces*. *FEMS Microbiology Reviews*, *34*(2), 171–198. https://doi.org/10.1111/j.1574-6976.2009.00206.x

Chauhan, A., Sakamoto, C., Ghigo, J.-M., & Beloin, C. (2013). Did I Pick the Right Colony? Pitfalls in the Study of Regulation of the Phase Variable Antigen 43 Adhesin. *PLoS ONE*, *8*(9), e73568. https://doi.org/10.1371/journal.pone.0073568

Chisholm, R. H., & Tanaka, M. M. (2016). The emergence of latent infection in the early evolution of *Mycobacterium tuberculosis*. *Proceedings of the Royal Society B: Biological Sciences*, *283*(1831), 20160499. https://doi.org/10.1098/rspb.2016.0499

Clayton, W., Eaton, C. J., Dupont, P.-Y., Gillanders, T., Cameron, N., Saikia, S., & Scott, B. (2017). Analysis of simple sequence repeat (SSR) structure and sequence within Epichloë endophyte genomes reveals impacts on gene structure and insights into ancestral hybridization events. *PLOS ONE*, *12*(9), e0183748. https://doi.org/10.1371/journal.pone.0183748

Cohen, N. R., Lobritz, M. A., & Collins, J. J. (2013). Microbial Persistence and the Road to Drug Resistance. *Cell Host & Microbe*, *13*(6), 632–642. https://doi.org/10.1016/j.chom.2013.05.009

Collier, J. (2009). Epigenetic regulation of the bacterial cell cycle. *Current Opinion in Microbiology*, *12*(6), 722–729. https://doi.org/10.1016/j.mib.2009.08.005

Coyne, M. J., Weinacht, K. G., Krinos, C. M., & Comstock, L. E. (2003). Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proceedings of the National Academy of Sciences*, *100*(18), 10446–10451. https://doi.org/10.1073/pnas.1832655100

Criss, A. K., Bonney, K. M., Chang, R. A., Duffin, P. M., LeCuyer, B. E., & Seifert, H. S. (2010). Mismatch Correction Modulates Mutation Frequency and Pilus Phase and Antigenic Variation in *Neisseria gonorrhoeae*. *Journal of Bacteriology*, *192*(1), 316–325. https://doi.org/10.1128/JB.01228-09

Darling, A. C. E., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. *Genome Research*, *14*(7), 1394–1403. https://doi.org/10.1101/gr.2289704

Davidson, C. J., & Surette, M. G. (2008). Individuality in Bacteria. *Annual Review of Genetics*, *42*(1), 253–268. https://doi.org/10.1146/annurev.genet.42.110807.091601

Davies, J. K., Harrison, P. F., Lin, Y.-H., Bartley, S., Khoo, C. A., Seemann, T., Ryan, C. S., Kahler, C. M., & Hill, S. A. (2014). The Use of High-Throughput DNA Sequencing in

the Investigation of Antigenic Variation: Application to Neisseria Species. *PLoS ONE*, *9*(1), e86704. https://doi.org/10.1371/journal.pone.0086704

Davis, K. M., & Isberg, R. R. (2016). Defining heterogeneity within bacterial populations via single cell approaches. *BioEssays*, *38*(8), 782–790. https://doi.org/10.1002/bies.201500121

Dawan, J., Wei, S., & Ahn, J. (2020). Role of antibiotic stress in phenotypic switching to persister cells of antibiotic-resistant Staphylococcus aureus. *Annals of Microbiology*, *70*(1), 1. https://doi.org/10.1186/s13213-020-01552-1

Dey, S. S., Kester, L., Spanjaard, B., Bienko, M., & van Oudenaarden, A. (2015). Integrated genome and transcriptome sequencing of the same cell. *Nature Biotechnology*, *33*(3), 285–289. https://doi.org/10.1038/nbt.3129

Dorman, C. J., & Bogue, M. M. (2016). The Interplay between DNA Topology and Accessory Factors in Site-Specific Recombination in Bacteria and their Bacteriophages. *Science Progress*, *99*(4), 420–437. https://doi.org/10.3184/003685016X14811202974921

Duffin, P. M., & Barber, D. A. (2016). DprA is required for natural transformation and affects pilin variation in Neisseria gonorrhoeae. *Microbiology*, *162*(9), 1620–1628. https://doi.org/10.1099/mic.0.000343

Egan, A. J. F., & Vollmer, W. (2013). The physiology of bacterial cell division. *Annals of the New York Academy of Sciences*, *1277*(1), 8–28. https://doi.org/10.1111/j.1749-6632.2012.06818.x

El Meouche, I., Siu, Y., & Dunlop, M. J. (2016). Stochastic expression of a multiple antibiotic resistance activator confers transient resistance in single cells. *Scientific Reports*, *6*(1), 19538. https://doi.org/10.1038/srep19538

Emerson, J. E., Reynolds, C. B., Fagan, R. P., Shaw, H. A., Goulding, D., & Fairweather, N. F. (2009). A novel genetic switch controls phase variable expression of CwpV, a *Clostridium difficile* cell wall protein. *Molecular Microbiology*, *74*(3), 541–556. https://doi.org/10.1111/j.1365-2958.2009.06812.x

Evans, C. R., Fan, Y., Weiss, K., & Ling, J. (2018). Errors during Gene Expression: Single-Cell Heterogeneity, Stress Resistance, and Microbe-Host Interactions. *MBio*, *9*(4). https://doi.org/10.1128/mBio.01018-18

Farnoud, F., Schwartz, M., & Bruck, J. (2016). The Capacity of String-Duplication Systems.

IEEE Transactions on Information Theory, *62*(2), 811–824. https://doi.org/10.1109/TIT.2015.2505735

Fauvart, M., De Groote, V. N., & Michiels, J. (2011). Role of persister cells in chronic infections: Clinical relevance and perspectives on anti-persister therapies. *Journal of Medical Microbiology*, *60*(6), 699–709. https://doi.org/10.1099/jmm.0.030932-0

Fazekas, A. J., Steeves, R., & Newmaster, S. G. (2010). Improving sequencing quality from PCR products containing long mononucleotide repeats. *BioTechniques*, *48*(4), 277–285. https://doi.org/10.2144/000113369

Fisher, R. A., Gollan, B., & Helaine, S. (2017). Persistent bacterial infections and persister cells. *Nature Reviews Microbiology*, *15*(8), 453–464. https://doi.org/10.1038/nrmicro.2017.42

Foley, J. (2015). Mini-review: Strategies for Variation and Evolution of Bacterial Antigens. *Computational and Structural Biotechnology Journal*, *13*, 407–416. https://doi.org/10.1016/j.csbj.2015.07.002

Gally, D. L., Bogan, J. A., Eisenstein, B. I., & Blomfield, I. C. (1993). Environmental regulation of the fim switch controlling type 1 fimbrial phase variation in Escherichia coli K-12: Effects of temperature and media. *Journal of Bacteriology*, *175*(19), 6186–6193. https://doi.org/10.1128/jb.175.19.6186-6193.1993

García-Pastor, L., Puerta-Fernández, E., & Casadesús, J. (2019). Bistability and phase variation in Salmonella enterica. *Biochimica et Biophysica Acta (BBA) - Gene Regulatory Mechanisms*, *1862*(7), 752–758. https://doi.org/10.1016/j.bbagrm.2018.01.003

Gasperotti, A., Brameyer, S., Fabiani, F., & Jung, K. (2020). Phenotypic heterogeneity of microbial populations under nutrient limitation. *Current Opinion in Biotechnology*, *62*, 160–167. https://doi.org/10.1016/j.copbio.2019.09.016

Ghovvati, S., Pezeshkian, Z., & Mirhoseini, S. Z. (2018). In silico analysis of different signal peptides to discover a panel of appropriate signal peptides for secretory production of Interferon-beta 1b in Escherichia coli. *Acta Biochimica Polonica*, *65*(4), 521–534. https://doi.org/10.18388/abp.2018_2351

Gibson, D. (2009). One-step enzymatic assembly of DNA molecules up to several hundred kilobases in size. *Protocol Exchange*. https://doi.org/10.1038/nprot.2009.77

Giri, S., Waschina, S., Kaleta, C., & Kost, C. (2019). Defining Division of Labor in Microbial Communities. *Journal of Molecular Biology*, *431*(23), 4712–4731. https://doi.org/10.1016/j.jmb.2019.06.023

Goldberg, A., Fridman, O., Ronin, I., & Balaban, N. Q. (2014). Systematic identification and quantification of phase variation in commensal and pathogenic Escherichia coli. *Genome Medicine*, *6*(11), 112. https://doi.org/10.1186/s13073-014-0112-4

Gould, I. M. (2009). Antibiotic resistance: The perfect storm. *International Journal of Antimicrobial Agents*, *34*, S2–S5. https://doi.org/10.1016/S0924-8579(09)70549-7

Grabiec, A. M., & Potempa, J. (2018). Epigenetic regulation in bacterial infections: Targeting histone deacetylases. *Critical Reviews in Microbiology*, *44*(3), 336–350. https://doi.org/10.1080/1040841X.2017.1373063

Green, J. H., Koza, A., Moshynets, O., Pajor, R., Ritchie, M. R., & Spiers, A. J. (2011). Evolution in a test tube: Rise of the Wrinkly Spreaders. *Journal of Biological Education*, *45*(1), 54–59. https://doi.org/10.1080/00219266.2011.537842

Green, L. R., Haigh, R. D., & Bayliss, C. D. (2019). Determination of Repeat Number and Expression States of Phase-Variable Loci Through Next Generation Sequencing and Bioinformatic Analysis. In K. L. Seib & I. R. Peak (Eds.), *Neisseria meningitidis* (Vol. 1969, pp. 83–92). Springer New York. https://doi.org/10.1007/978-1-4939-9202-7_5

Guiney, D. G. (1997). Regulation of bacterial virulence gene expression by the host environment. *Journal of Clinical Investigation*, *99*(4), 565–569. https://doi.org/10.1172/JCI119196

Haagmans, W., & van der Woude, M. (2000). Phase variation of Ag43 in Escherichia coli: Dam-dependent methylation abrogates OxyR binding and OxyR-mediated repression of transcription. *Molecular Microbiology*, *35*(4), 877–887. https://doi.org/10.1046/j.1365-2958.2000.01762.x

Hall, A. M., Gollan, B., & Helaine, S. (2017). Toxin–antitoxin systems: Reversible toxicity. *Current Opinion in Microbiology*, *36*, 102–110. https://doi.org/10.1016/j.mib.2017.02.003

Hamilton, H. L., & Dillard, J. P. (2006). Natural transformation of *Neisseria gonorrhoeae*: From DNA donation to homologous recombination: Natural transformation of *Neisseria gonorrhoeae*. *Molecular Microbiology*, *59*(2), 376–385.

https://doi.org/10.1111/j.1365-2958.2005.04964.x

Hansson, O. (2018). *Development of computer software to characterise and simulate molecular biology processes used in forensic DNA profiling assays*. https://doi.org/10.13140/RG.2.2.19317.22245

Harms, A., Maisonneuve, E., & Gerdes, K. (2016). Mechanisms of bacterial persistence during stress and antibiotic exposure. *Science*, *354*(6318), aaf4268. https://doi.org/10.1126/science.aaf4268

Harvey, Z. H., Chen, Y., & Jarosz, D. F. (2018). Protein-Based Inheritance: Epigenetics beyond the Chromosome. *Molecular Cell*, *69*(2), 195–202. https://doi.org/10.1016/j.molcel.2017.10.030

Healey, D., Axelrod, K., & Gore, J. (2016). Negative frequency-dependent interactions can underlie phenotypic heterogeneity in a clonal microbial population. *Molecular Systems Biology*, *12*(8), 877. https://doi.org/10.15252/msb.20167033

Henderson, I. R., Owen, P., & Nataro, J. P. (1999). Molecular switches—The ON and OFF of bacterial phase variation. *Molecular Microbiology*, *33*(5), 919–932. https://doi.org/10.1046/j.1365-2958.1999.01555.x

Hill, P. W. S., & Helaine, S. (2019). Antibiotic Persisters and Relapsing Salmonella enterica Infections. In K. Lewis (Ed.), *Persister Cells and Infectious Disease* (pp. 19–38). Springer International Publishing. https://doi.org/10.1007/978-3-030-25241-0_2

Hill, S. A., Masters, T. L., & Wachter, J. (2016). Gonorrhoea – an evolving disease of the new millennium. *Microbial Cell*, *3*(9), 371–389. https://doi.org/10.15698/mic2016.09.524

Hinde, P., Deighan, P., & Dorman, C. J. (2005). Characterization of the Detachable Rho-Dependent Transcription Terminator of the *fimE* Gene in *Escherichia coli* K-12. *Journal of Bacteriology*, *187*(24), 8256–8266. https://doi.org/10.1128/JB.187.24.8256-8266.2005

Hobby, G. L., Meyer, K., & Chaffee, E. (1942). Observations on the Mechanism of Action of Penicillin. *Experimental Biology and Medicine*, *50*(2), 281–285. https://doi.org/10.3181/00379727-50-13773

Holden, N., Totsika, M., Dixon, L., Catherwood, K., & Gally, D. L. (2007). Regulation of P-Fimbrial Phase Variation Frequencies in *Escherichia coli* CFT073. *Infection and Immunity*, *75*(7), 3325–3334. https://doi.org/10.1128/IAI.01989-06

Huang, Y.-T., Liu, P.-Y., & Shih, P.-W. (2020). *High-Quality Genomes of Nanopore Sequencing by Homologous Polishing* [Preprint]. Bioinformatics. https://doi.org/10.1101/2020.09.19.304949

Inglis, D. W., Davis, J. A., Zieziulewicz, T. J., Lawrence, D. A., Austin, R. H., & Sturm, J. C. (2008). Determining blood cell size using microfluidic hydrodynamics. *Journal of Immunological Methods*, *329*(1–2), 151–156. https://doi.org/10.1016/j.jim.2007.10.004

Ishii, S., Ksoll, W. B., Hicks, R. E., & Sadowsky, M. J. (2006). Presence and Growth of Naturalised *Escherichia coli* in Temperate Soils from Lake Superior Watersheds. *Applied and Environmental Microbiology*, *72*(1), 612–621. https://doi.org/10.1128/AEM.72.1.612-621.2006

Jaenisch, R., & Bird, A. (2003). Epigenetic regulation of gene expression: How the genome integrates intrinsic and environmental signals. *Nature Genetics*, *33*(3), 245–254. https://doi.org/10.1038/ng1089

Jaishankar, J., & Srivastava, P. (2017). Molecular Basis of Stationary Phase Survival and Applications. *Frontiers in Microbiology*, *8*, 2000. https://doi.org/10.3389/fmicb.2017.02000

Jankevicius, G., Ariza, A., Ahel, M., & Ahel, I. (2016). The Toxin-Antitoxin System DarTG Catalyzes Reversible ADP-Ribosylation of DNA. *Molecular Cell*, *64*(6), 1109–1116. https://doi.org/10.1016/j.molcel.2016.11.014

Jiang, X., Hall, A. B., Arthur, T. D., Plichta, D. R., Covington, C. T., Poyet, M., Crothers, J., Moses, P. L., Tolonen, A. C., Vlamakis, H., Alm, E. J., & Xavier, R. J. (2019). Invertible promoters mediate bacterial phase variation, antibiotic resistance, and host adaptation in the gut. *Science*, *363*(6423), 181–187. https://doi.org/10.1126/science.aau5238

Jolly, M. K., Kulkarni, P., Weninger, K., Orban, J., & Levine, H. (2018). Phenotypic Plasticity, Bet-Hedging, and Androgen Independence in Prostate Cancer: Role of Non-Genetic Heterogeneity. *Frontiers in Oncology*, *8*, 50. https://doi.org/10.3389/fonc.2018.00050

Kaldalu, N., Hauryliuk, V., Turnbull, K. J., La Mensa, A., Putrinš, M., & Tenson, T. (2020). *In Vitro* Studies of Persister Cells. *Microbiology and Molecular Biology Reviews*, *84*(4), e00070-20. https://doi.org/10.1128/MMBR.00070-20

Kim, J.-S., & Wood, T. K. (2016). Persistent Persister Misperceptions. *Frontiers in Microbiology*, *07*. https://doi.org/10.3389/fmicb.2016.02134

Kramer, J., López Carrasco, M. Á., & Kümmerli, R. (2018). *Positive linkage between bacterial social traits reveals that homogeneous rather than specialized behavioral repertoires prevail in natural* Pseudomonas *communities* [Preprint]. Microbiology. https://doi.org/10.1101/477471

Kumar, R., & Rao, D. N. (2013). Role of DNA methyltransferases in epigenetic regulation in bacteria. *Sub-Cellular Biochemistry*, *61*, 81–102. https://doi.org/10.1007/978-94-007-4525-4_4

Larson, M. R., Biddle, K., Gorman, A., Boutom, S., Rosenshine, I., & Saper, M. A. (2021). Escherichia coli O127 group 4 capsule proteins assemble at the outer membrane. *PLOS ONE*, *16*(11), e0259900. https://doi.org/10.1371/journal.pone.0259900

Law, J., Lee, S., Tseng, A., Tsui, K. W., & Yu, N. (2002). *The Role of Glycerol and Isopropyl Thiogalactoside in Escherichia coli Growth and Lactose Induction of β-Galactosidase*. *2*, 6.

Laxminarayan, R., Duse, A., Wattal, C., Zaidi, A. K. M., Wertheim, H. F. L., Sumpradit, N., Vlieghe, E., Hara, G. L., Gould, I. M., Goossens, H., Greko, C., So, A. D., Bigdeli, M., Tomson, G., Woodhouse, W., Ombaka, E., Peralta, A. Q., Qamar, F. N., Mir, F., … Cars, O. (2013). Antibiotic resistance—The need for global solutions. *The Lancet Infectious Diseases*, *13*(12), 1057–1098. https://doi.org/10.1016/S1473-3099(13)70318-9

Levin-Reisman, I., Ronin, I., Gefen, O., Braniss, I., Shoresh, N., & Balaban, N. Q. (2017). Antibiotic tolerance facilitates the evolution of resistance. *Science*, *355*(6327), 826–830. https://doi.org/10.1126/science.aaj2191

Lewis, K. (2007). Persister cells, dormancy and infectious disease. *Nature Reviews Microbiology*, *5*(1), 48–56. https://doi.org/10.1038/nrmicro1557

Lewis, K. (2010). Persister Cells. *Annual Review of Microbiology*, *64*(1), 357–372. https://doi.org/10.1146/annurev.micro.112408.134306

Lewis, M. (2013). Allostery and the lac Operon. *Journal of Molecular Biology*, *425*(13), 2309–2316. https://doi.org/10.1016/j.jmb.2013.03.003

Li, H., & Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler

transform. *Bioinformatics*, *25*(14), 1754–1760. https://doi.org/10.1093/bioinformatics/btp324

Li, T., Yin, N., Liu, H., Pei, J., & Lai, L. (2016). Novel Inhibitors of Toxin HipA Reduce Multidrug Tolerant Persisters. *ACS Medicinal Chemistry Letters*, *7*(5), 449–453. https://doi.org/10.1021/acsmedchemlett.5b00420

Liao, Y., Smyth, G. K., & Shi, W. (2014). featureCounts: An efficient general purpose program for assigning sequence reads to genomic features. *Bioinformatics*, *30*(7), 923–930. https://doi.org/10.1093/bioinformatics/btt656

Loiko, N. G., Lobanov, K. V., Nikolaev, Yu. A., Kozlova, A. N., & El'-Registan, G. I. (2017). Regulation of phase variation in type I pili formation in Escherichia coli: Role of alkylresorcinols, microbial autoregulators. *Microbiology*, *86*(5), 560–570. https://doi.org/10.1134/S0026261717050149

Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12), 550. https://doi.org/10.1186/s13059-014-0550-8

Lowery, N. V., McNally, L., Ratcliff, W. C., & Brown, S. P. (2017). Division of Labor, Bet Hedging, and the Evolution of Mixed Biofilm Investment Strategies. *MBio*. https://doi.org/10.1128/mBio.00672-17

Maisonneuve, E., & Gerdes, K. (2014). Molecular Mechanisms Underlying Bacterial Persisters. *Cell*, *157*(3), 539–548. https://doi.org/10.1016/j.cell.2014.02.050

Marrs, C. F., Zhang, L., & Foxman, B. (2005). *Escherichia coli* mediated urinary tract infections: Are there distinct uropathogenic *E. coli* (UPEC) pathotypes? *FEMS Microbiology Letters*, *252*(2), 183–190. https://doi.org/10.1016/j.femsle.2005.08.028

Martínez-Granero, F., Capdevila, S., Sánchez-Contreras, M., Martín, M., & Rivilla, R. (2005). Two site-specific recombinases are implicated in phenotypic variation and competitive rhizosphere colonization in Pseudomonas fluorescens. *Microbiology*, *151*(3), 975–983. https://doi.org/10.1099/mic.0.27583-0

Martínez-Granero, F., Rivilla, R., & Martín, M. (2006). Rhizosphere Selection of Highly Motile Phenotypic Variants of *Pseudomonas fluorescens* with Enhanced Competitive Colonization Ability. *Applied and Environmental Microbiology*, *72*(5), 3429–3434. https://doi.org/10.1128/AEM.72.5.3429-3434.2006

Mechler, L., Herbig, A., Paprotka, K., Fraunholz, M., Nieselt, K., & Bertram, R. (2015). A Novel Point Mutation Promotes Growth Phase-Dependent Daptomycin Tolerance in Staphylococcus aureus. *Antimicrobial Agents and Chemotherapy*, *59*(9), 5366–5376. https://doi.org/10.1128/AAC.00643-15

Michiels, J. E., Van den Bergh, B., Verstraeten, N., & Michiels, J. (2016). Molecular mechanisms and clinical implications of bacterial persistence. *Drug Resistance Updates*, *29*, 76–89. https://doi.org/10.1016/j.drup.2016.10.002

Miyaue, S., Suzuki, E., Komiyama, Y., Kondo, Y., Morikawa, M., & Maeda, S. (2018). Bacterial Memory of Persisters: Bacterial Persister Cells Can Retain Their Phenotype for Days or Weeks After Withdrawal From Colony–Biofilm Culture. *Frontiers in Microbiology*, *9*. https://www.frontiersin.org/article/10.3389/fmicb.2018.01396

Nagano, K., Abiko, Y., Yoshida, Y., & Yoshimura, F. (2012). Porphyromonas gingivalis FimA fimbriae: Roles of the fim gene cluster in the fimbrial assembly and antigenic heterogeneity among fimA genotypes. *Journal of Oral Biosciences*, *54*(3), 160–163. https://doi.org/10.1016/j.job.2012.07.002

Narra, H. P., & Ochman, H. (2006). Of What Use Is Sex to Bacteria? *Current Biology*, *16*(17), R705–R710. https://doi.org/10.1016/j.cub.2006.08.024

Ngeleka, M., Pritchard, J., Appleyard, G., Middleton, D. M., & Fairbrother, J. M. (2003). Isolation and Association of Escherichia Coli AIDA-I/STb, Rather than EAST1 Pathotype, with Diarrhea in Piglets and Antibiotic Sensitivity of Isolates. *Journal of Veterinary Diagnostic Investigation*, *15*(3), 242–252. https://doi.org/10.1177/104063870301500305

Nyström, T. (2004). Stationary-Phase Physiology. *Annual Review of Microbiology*, *58*(1), 161–181. https://doi.org/10.1146/annurev.micro.58.030603.123818

Obergfell, K. P., & Seifert, H. S. (2016). The Pilin N-terminal Domain Maintains Neisseria gonorrhoeae Transformation Competence during Pilus Phase Variation. *PLOS Genetics*, *12*(5), e1006069. https://doi.org/10.1371/journal.pgen.1006069

Olorunniji, F. J., & Stark, W. M. (2010). Catalysis of site-specific recombination by Tn *3* resolvase. *Biochemical Society Transactions*, *38*(2), 417–421. https://doi.org/10.1042/BST0380417

Ørskov, F. (1978). Virulence Factors of the Bacterial Cell Surface. *The Journal of Infectious*

Diseases, *137*(5), 630–633.

Page, R., & Peti, W. (2016). Toxin-antitoxin systems in bacterial growth arrest and persistence. *Nature Chemical Biology*, *12*(4), 208–214. https://doi.org/10.1038/nchembio.2044

Patange, O., Schwall, C., Jones, M., Villava, C., Griffith, D. A., Phillips, A., & Locke, J. C. W. (2018). Escherichia coli can survive stress by noisy growth modulation. *Nature Communications*, *9*(1), 5333. https://doi.org/10.1038/s41467-018-07702-z

Paulsson, J., El Karoui, M., Lindell, M., & Hughes, D. (2017). The processive kinetics of gene conversion in bacteria: Processivity of gene conversion. *Molecular Microbiology*, *104*(5), 752–760. https://doi.org/10.1111/mmi.13661

Piazza, A., & Heyer, W.-D. (2019). Homologous Recombination and the Formation of Complex Genomic Rearrangements. *Trends in Cell Biology*, *29*(2), 135–149. https://doi.org/10.1016/j.tcb.2018.10.006

Prister, L. L., Yin, S., Cahoon, L. A., & Seifert, H. S. (2020). Altering the *Neisseria gonorrhoeae pilE* Guanine Quadruplex Loop Bases Affects Pilin Antigenic Variation. *Biochemistry*, *59*(10), 1104–1112. https://doi.org/10.1021/acs.biochem.9b01038

Pu, Y., Ke, Y., & Bai, F. (2017). Active efflux in dormant bacterial cells – New insights into antibiotic persistence. *Drug Resistance Updates*, *30*, 7–14. https://doi.org/10.1016/j.drup.2016.11.002

Reznikoff, W. S. (1992). The lactose operon-controlling elements: A complex paradigm. *Molecular Microbiology*, *6*(17), 2419–2422. https://doi.org/10.1111/j.1365-2958.1992.tb01416.x

Roberfroid, S., Vanderleyden, J., & Steenackers, H. (2016). Gene expression variability in clonal populations: Causes and consequences. *Critical Reviews in Microbiology*, *42*(6), 969–984. https://doi.org/10.3109/1040841X.2015.1122571

Romano, A. C., Espana, E. M., Yoo, S. H., Budak, M. T., Wolosin, J. M., & Tseng, S. C. G. (2003). Different Cell Sizes in Human Limbal and Central Corneal Basal Epithelia Measured by Confocal Microscopy and Flow Cytometry. *Investigative Opthalmology & Visual Science*, *44*(12), 5125. https://doi.org/10.1167/iovs.03-0628

Ronneau, S., & Helaine, S. (2019). Clarifying the Link between Toxin–Antitoxin Modules and Bacterial Persistence. *Journal of Molecular Biology*, *431*(18), 3462–3471.

https://doi.org/10.1016/j.jmb.2019.03.019

Rothfield, L., Justice, S., & García-Lara, J. (1999). Bacterial Cell Division. *Annual Review of Genetics*, *33*(1), 423–448. https://doi.org/10.1146/annurev.genet.33.1.423

Rotman, E., & Seifert, H. S. (2014). The Genetics of *Neisseria* Species. *Annual Review of Genetics*, *48*(1), 405–431. https://doi.org/10.1146/annurev-genet-120213-092007

Rotman, E., Webber, D. M., & Seifert, H. S. (2016). Analyzing Neisseria gonorrhoeae Pilin Antigenic Variation Using 454 Sequencing Technology. *Journal of Bacteriology*, *198*(18), 2470–2482. https://doi.org/10.1128/JB.00330-16

RStudio Team. (2015). *RStudio: Integrated Development Environment for R*. Boston, MA. Retrieved from http://www.rstudio.com/

Sadarangani, M., Hoe, J. C., Makepeace, K., van der Ley, P., & Pollard, A. J. (2016). Phase variation of Opa proteins of Neisseria meningitidis and the effects of bacterial transformation. *Journal of Biosciences*, *41*(1), 13–19. https://doi.org/10.1007/s12038-016-9588-y

Sadarangani, M., Pollard, A. J., & Gray-Owen, S. D. (2011). Opa proteins and CEACAMs: Pathways of immune engagement for pathogenic *Neisseria*. *FEMS Microbiology Reviews*, *35*(3), 498–514. https://doi.org/10.1111/j.1574-6976.2010.00260.x

Safi, H., Gopal, P., Lingaraju, S., Ma, S., Levine, C., Dartois, V., Yee, M., Li, L., Blanc, L., Liang, H.-P. H., Husain, S., Hoque, M., Soteropoulos, P., Rustad, T., Sherman, D. R., Dick, T., & Alland, D. (2019). Phase variation in Mycobacterium tuberculosis glpK produces transiently heritable drug tolerance. *Proceedings of the National Academy of Sciences*, *116*(39), 19665–19674. https://doi.org/10.1073/pnas.1907631116

Sánchez-Contreras, M., Martín, M., Villacieros, M., O'Gara, F., Bonilla, I., & Rivilla, R. (2002). Phenotypic Selection and Phase Variation Occur during Alfalfa Root Colonization by *Pseudomonas fluorescens* F113. *Journal of Bacteriology*, *184*(6), 1587–1596. https://doi.org/10.1128/JB.184.6.1587-1596.2002

Sánchez-Romero, M. A., & Casadesús, J. (2018). Contribution of SPI-1 bistability to Salmonella enterica cooperative virulence: Insights from single cell analysis. *Scientific Reports*, *8*(1), 14875. https://doi.org/10.1038/s41598-018-33137-z

Sánchez-Romero, M. A., & Casadesús, J. (2020). The bacterial epigenome. *Nature Reviews Microbiology*, *18*(1), 7–20. https://doi.org/10.1038/s41579-019-0286-2

Sathiyamoorthy, K., Mills, E., Franzmann, T. M., Rosenshine, I., & Saper, M. A. (2011). The crystal structure of Escherichia coli group 4 capsule protein GfcC reveals a domain organization resembling that of Wza. *Biochemistry*, *50*(24), 5465–5476. https://doi.org/10.1021/bi101869h

Savidge, T. C. (2016). Epigenetic Regulation of Enteric Neurotransmission by Gut Bacteria. *Frontiers in Cellular Neuroscience*, *9*. https://doi.org/10.3389/fncel.2015.00503

Schreiber, F., Littmann, S., Lavik, G., Escrig, S., Meibom, A., Kuypers, M. M. M., & Ackermann, M. (2016). Phenotypic heterogeneity driven by nutrient limitation promotes growth in fluctuating environments. *Nature Microbiology*, *1*(6), 16055. https://doi.org/10.1038/nmicrobiol.2016.55

Schwan, W. R. (2011). Regulation of *fim* genes in uropathogenic *Escherichia coli*. *World Journal of Clinical Infectious Diseases*, *1*(1), 17. https://doi.org/10.5495/wjcid.v1.i1.17

Schwan, W. R., Beck, M. T., Hung, C. S., & Hultgren, S. J. (2018). Differential Regulation of *Escherichia coli fim* Genes following Binding to Mannose Receptors. *Journal of Pathogens*, *2018*, 1–8. https://doi.org/10.1155/2018/2897581

Sehn, J. K. (2015). Chapter 9—Insertions and Deletions (Indels). In S. Kulkarni & J. Pfeifer (Eds.), *Clinical Genomics* (pp. 129–150). Academic Press. https://doi.org/10.1016/B978-0-12-404748-8.00009-5

Seib, K. L., & Jennings, M. P. (2021). Chapter 19—Epigenetics of infectious diseases. In T. O. Tollefsbol (Ed.), *Medical Epigenetics (Second Edition)* (Vol. 29, pp. 407–424). Academic Press. https://doi.org/10.1016/B978-0-12-823928-5.00016-5

Silverman, M., & Simon, M. (1980). Phase variation: Genetic analysis of switching mutants. *Cell*, *19*(4), 845–854. https://doi.org/10.1016/0092-8674(80)90075-6

Smits, W. K., Kuipers, O. P., & Veening, J.-W. (2006). Phenotypic variation in bacteria: The role of feedback regulation. *Nature Reviews Microbiology*, *4*(4), 259–271. https://doi.org/10.1038/nrmicro1381

Spiers, A. J., Kahn, S. G., Bohannon, J., Travisano, M., & Rainey, P. B. (2002). Adaptive Divergence in Experimental Populations of *Pseudomonas fluorescens*. I. Genetic and Phenotypic Bases of Wrinkly Spreader Fitness. *Genetics*, *161*(1), 33–46. https://doi.org/10.1093/genetics/161.1.33

Stylianidou, S., Brennan, C., Nissen, S. B., Kuwada, N. J., & Wiggins, P. A. (2016).

SuperSegger: Robust image segmentation, analysis and lineage tracking of bacterial cells: Robust segmentation and analysis of bacteria. *Molecular Microbiology*, *102*(4), 690–700. https://doi.org/10.1111/mmi.13486

Sundar, S., & Singh, B. (2014). Identifying vaccine targets for anti-leishmanial vaccine development. *Expert Review of Vaccines*, *13*(4), 489–505. https://doi.org/10.1586/14760584.2014.894467

Sundqvist, M. (2014). Reversibility of antibiotic resistance. *Upsala Journal of Medical Sciences*, *119*(2), 142–148. https://doi.org/10.3109/03009734.2014.903323

Tauseef, I., Ali, Y. M., & Bayliss, C. D. (2013). Phase Variation of PorA, a Major Outer Membrane Protein, Mediates Escape of Bactericidal Antibodies by Neisseria meningitidis. *Infection and Immunity*, *81*(4), 1374–1380. https://doi.org/10.1128/IAI.01358-12

Thomassin, J.-L., Lee, M. J., Brannon, J. R., Sheppard, D. C., Gruenheid, S., & Moual, H. L. (2013). Both Group 4 Capsule and Lipopolysaccharide O-Antigen Contribute to Enteropathogenic Escherichia coli Resistance to Human α-Defensin 5. *PLOS ONE*, *8*(12), e82475. https://doi.org/10.1371/journal.pone.0082475

Torrey, H. L., Keren, I., Via, L. E., Lee, J. S., & Lewis, K. (2016). High Persister Mutants in Mycobacterium tuberculosis. *PLOS ONE*, *11*(5), e0155127. https://doi.org/10.1371/journal.pone.0155127

Trzilova, D., & Tamayo, R. (2021). Site-Specific Recombination – How Simple DNA Inversions Produce Complex Phenotypic Heterogeneity in Bacterial Populations. *Trends in Genetics*, *37*(1), 59–72. https://doi.org/10.1016/j.tig.2020.09.004

Tsai, C. N., & Coombes, B. K. (2019). The Role of the Host in Driving Phenotypic Heterogeneity in Salmonella. *Trends in Microbiology*, *27*(6), 508–523. https://doi.org/10.1016/j.tim.2019.01.004

Van den Bergh, B., Michiels, J. E., Wenseleers, T., Windels, E. M., Boer, P. V., Kestemont, D., De Meester, L., Verstrepen, K. J., Verstraeten, N., Fauvart, M., & Michiels, J. (2016). Frequency of antibiotic application drives rapid evolutionary adaptation of Escherichia coli persistence. *Nature Microbiology*, *1*, 16020. https://doi.org/10.1038/nmicrobiol.2016.20

van den Broek, D., Bloemberg, G. V., & Lugtenberg, B. (2005). The role of phenotypic

variation in rhizosphere Pseudomonas bacteria. *Environmental Microbiology*, *7*(11), 1686–1697. https://doi.org/10.1111/j.1462-2920.2005.00912.x

van der Woude, M. W. (2011). Phase variation: How to create and coordinate population diversity. *Current Opinion in Microbiology*, *14*(2), 205–211. https://doi.org/10.1016/j.mib.2011.01.002

van der Woude, M. W. (2017). Epigenetic Phase Variation in Bacterial Pathogens. In W. Doerfler & J. Casadesús (Eds.), *Epigenetics of Infectious Diseases* (pp. 159–173). Springer International Publishing. https://doi.org/10.1007/978-3-319-55021-3_7

van der Woude, M. W., & Bäumler, A. J. (2004). Phase and Antigenic Variation in Bacteria. *Clinical Microbiology Reviews*, *17*(3), 581–611. https://doi.org/10.1128/CMR.17.3.581-611.2004

Vaser, R., Sović, I., Nagarajan, N., & Šikić, M. (2017). Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Research*, *27*(5), 737–746. https://doi.org/10.1101/gr.214270.116

Vos, M. (2009). Why do bacteria engage in homologous recombination? *Trends in Microbiology*, *17*(6), 226–232. https://doi.org/10.1016/j.tim.2009.03.001

Voter, A. F., Callaghan, M. M., Tippana, R., Myong, S., Dillard, J. P., & Keck, J. L. (2020). Antigenic Variation in *Neisseria gonorrhoeae* Occurs Independently of RecQ-Mediated Unwinding of the *pilE* G Quadruplex. *Journal of Bacteriology*, *202*(3). https://doi.org/10.1128/JB.00607-19

Vujovic, F., Hunter, N., & Farahani, R. M. (2019). Notch pathway: A bistable inducer of biological noise? *Cell Communication and Signaling*, *17*(1), 133. https://doi.org/10.1186/s12964-019-0453-0

Wachter, J., & Hill, S. (2016). Positive Selection Pressure Drives Variation on the Surface-Exposed Variable Proteins of the Pathogenic Neisseria. *PLOS ONE*, *11*(8), e0161348. https://doi.org/10.1371/journal.pone.0161348

Wallecha, A., Correnti, J., Munster, V., & van der Woude, M. (2003). Phase Variation of Ag43 Is Independent of the Oxidation State of OxyR. *Journal of Bacteriology*, *185*(7), 2203–2209. https://doi.org/10.1128/JB.185.7.2203-2209.2003

Weigel, W. A., & Dersch, P. (2018). Phenotypic heterogeneity: A bacterial virulence strategy. *Microbes and Infection*, *20*(9–10), 570–577.

https://doi.org/10.1016/j.micinf.2018.01.008

Wellington, E. M., Boxall, A. B., Cross, P., Feil, E. J., Gaze, W. H., Hawkey, P. M., Johnson-Rollings, A. S., Jones, D. L., Lee, N. M., Otten, W., Thomas, C. M., & Williams, A. P. (2013). The role of the natural environment in the emergence of antibiotic resistance in Gram-negative bacteria. *The Lancet Infectious Diseases*, *13*(2), 155–165. https://doi.org/10.1016/S1473-3099(12)70317-1

Wells, T. J., Tree, J. J., Ulett, G. C., & Schembri, M. A. (2007). Autotransporter proteins: Novel targets at the bacterial cell surface. *FEMS Microbiology Letters*, *274*(2), 163–172. https://doi.org/10.1111/j.1574-6968.2007.00833.x

West, S. A., & Cooper, G. A. (2016). Division of labour in microorganisms: An evolutionary perspective. *Nature Reviews Microbiology*, *14*(11), 716–723. https://doi.org/10.1038/nrmicro.2016.111

Whitfield, C., & Roberts, I. S. (1999). Structure, assembly and regulation of expression of capsules in Escherichia coli. *Molecular Microbiology*, *31*(5), 1307–1319. https://doi.org/10.1046/j.1365-2958.1999.01276.x

Willenborg, J., Willms, D., Bertram, R., Goethe, R., & Valentin-Weigand, P. (2014). Characterization of multi-drug tolerant persister cells in Streptococcus suis. *BMC Microbiology*, *14*(1), 120. https://doi.org/10.1186/1471-2180-14-120

Wilmaerts, D., Windels, E. M., Verstraeten, N., & Michiels, J. (2019). General Mechanisms Leading to Persister Formation and Awakening. *Trends in Genetics*, *35*(6), 401–411. https://doi.org/10.1016/j.tig.2019.03.007

Wisniewski-Dyé, F., & Vial, L. (2008). Phase and antigenic variation mediated by genome modifications. *Antonie van Leeuwenhoek*, *94*(4), 493–515. https://doi.org/10.1007/s10482-008-9267-6

Wöhrmann, T., & Weising, K. (2011). In silico mining for simple sequence repeat loci in a pineapple expressed sequence tag database and cross-species amplification of EST-SSR markers across Bromeliaceae. *Theoretical and Applied Genetics*, *123*(4), 635–647. https://doi.org/10.1007/s00122-011-1613-9

Wood, T. K., Knabel, S. J., & Kwan, B. W. (2013). Bacterial Persister Cell Formation and Dormancy. *Applied and Environmental Microbiology*, *79*(23), 7116–7121. https://doi.org/10.1128/AEM.02636-13

Zambrano, M. M., & Kolter, R. (1996). GASPing for Life in Stationary Phase. *Cell*, *86*(2), 181–184. https://doi.org/10.1016/S0092-8674(00)80089-6

Zaslaver, A., Bren, A., Ronen, M., Itzkovitz, S., Kikoin, I., Shavit, S., Liebermeister, W., Surette, M. G., & Alon, U. (2006). A comprehensive library of fluorescent transcriptional reporters for Escherichia coli. *Nature Methods*, *3*(8), 623–628. https://doi.org/10.1038/nmeth895

Zelewska, M. A., Pulijala, M., Spencer-Smith, R., Mahmood, H.-T.-N. A., Norman, B., Churchward, C. P., Calder, A., & Snyder, L. A. S. (2016). Phase variable DNA repeats in Neisseria gonorrhoeae influence transcription, translation, and protein sequence variation. *Microbial Genomics*, *2*(8). https://doi.org/10.1099/mgen.0.000078

Žgur-Bertok, D. (2007). PHENOTYPIC HETEROGENEITY IN BACTERIAL POPULATIONS. *Acta Agriculturae Slovenica*, *1*, 8.

Zhang, Z., Claessen, D., & Rozen, D. E. (2016). Understanding Microbial Divisions of Labor. *Frontiers in Microbiology*, *7*. https://doi.org/10.3389/fmicb.2016.02070

Zhang, Z., Du, C., de Barsy, F., Liem, M., Liakopoulos, A., van Wezel, G. P., Choi, Y. H., Claessen, D., & Rozen, D. E. (2020). Antibiotic production in *Streptomyces* is organized by a division of labor through terminal genomic differentiation. *Science Advances*, *6*(3), eaay5781. https://doi.org/10.1126/sciadv.aay5781

Zhao, L., Chen, X., Xu, X., Song, G., & Liu, X. (2009). Analysis of the AIDA-I gene sequence and prevalence in Escherichia coli isolates from pigs with post-weaning diarrhoea and oedema disease. *The Veterinary Journal*, *180*(1), 124–129. https://doi.org/10.1016/j.tvjl.2007.10.021

Zhou, K., Aertsen, A., & Michiels, C. W. (2014). The role of variable DNA tandem repeats in bacterial adaptation. *FEMS Microbiology Reviews*, *38*(1), 119–141. https://doi.org/10.1111/1574-6976.12036

Chapter 5: Appendix

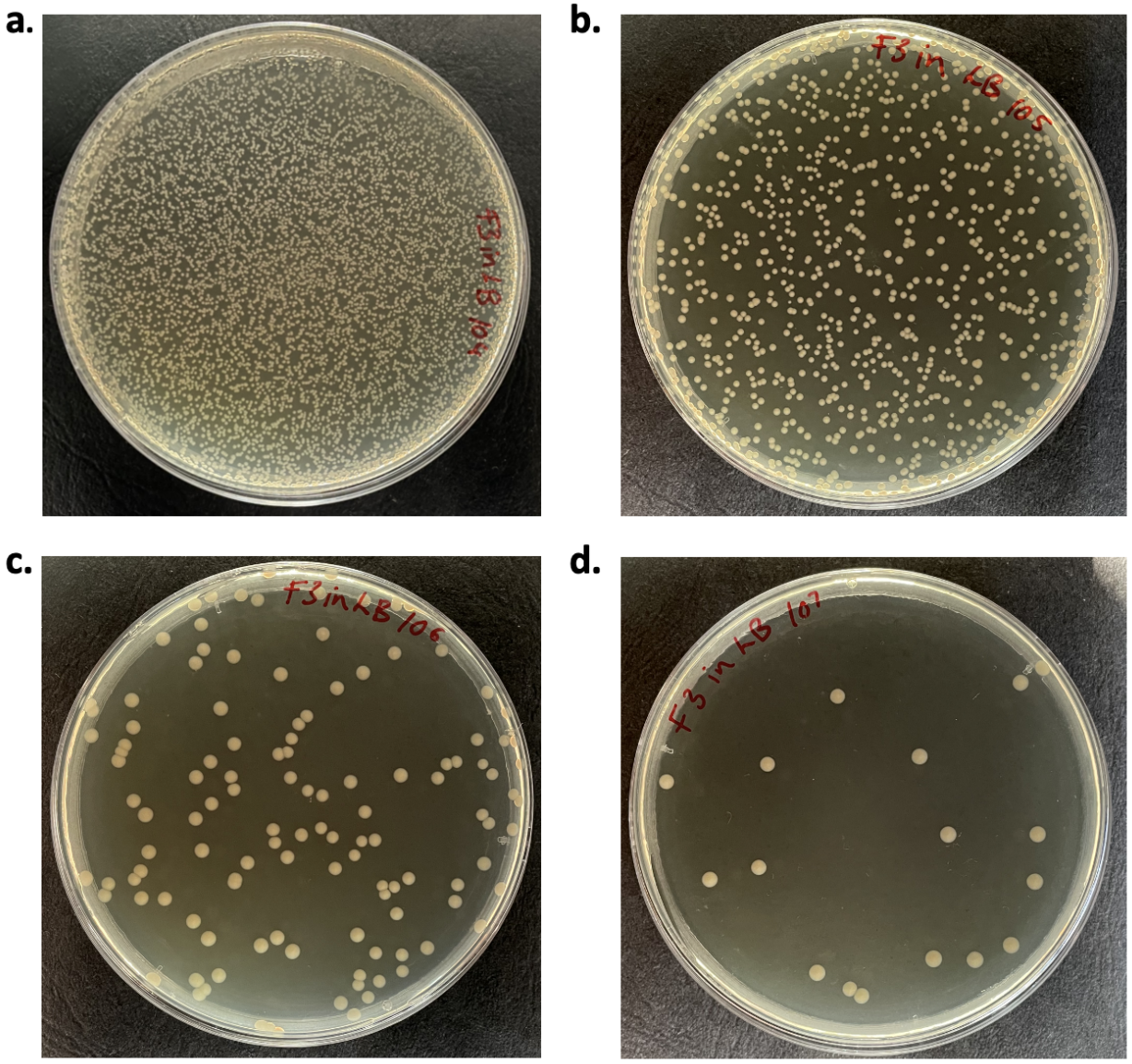


Figure 5.1 SC375 cells grown on LB agar plates form only compact colonies. Overnight cultures of SC375 plated at dilutions **(a)** 1.0e4-fold **(b)** 1.0e5-fold **(c)** 1.0e6-fold and **(d)** 1.0e7-fold and grown overnight at 37°C for 48 hours.

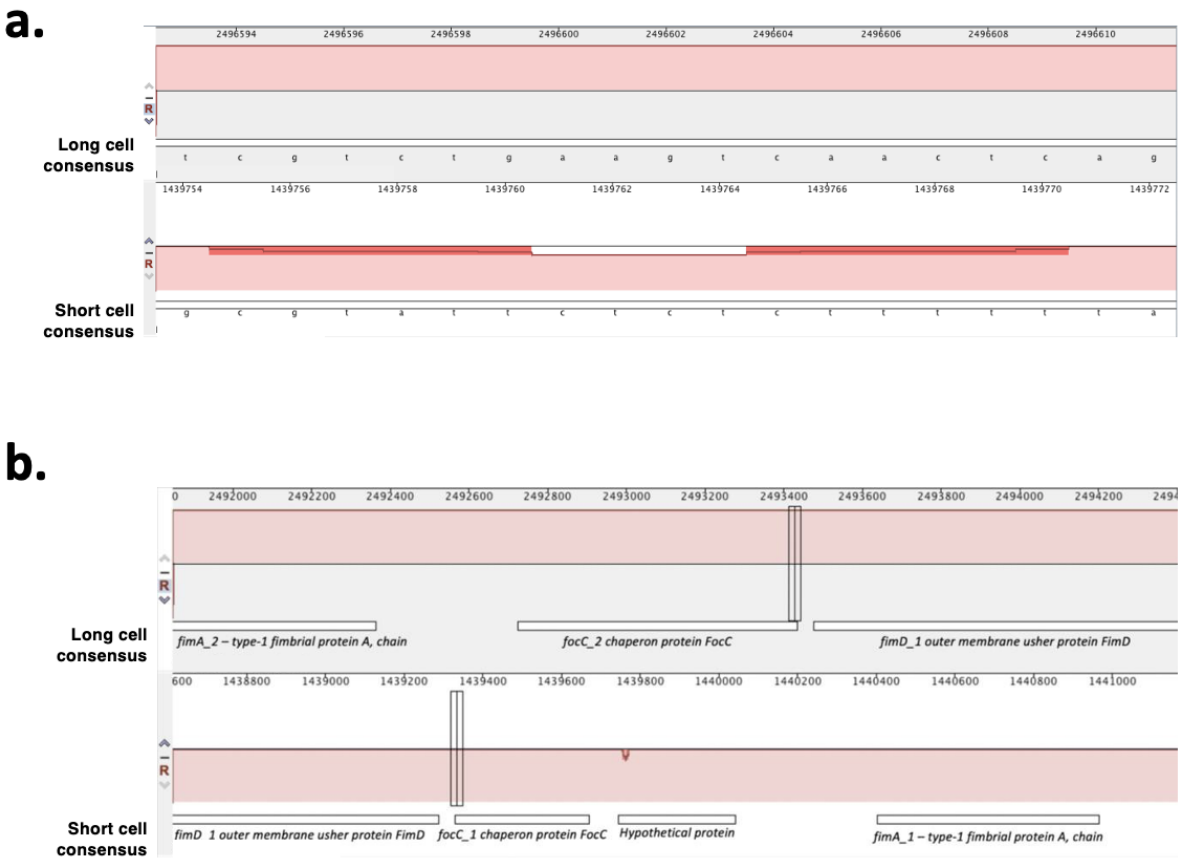


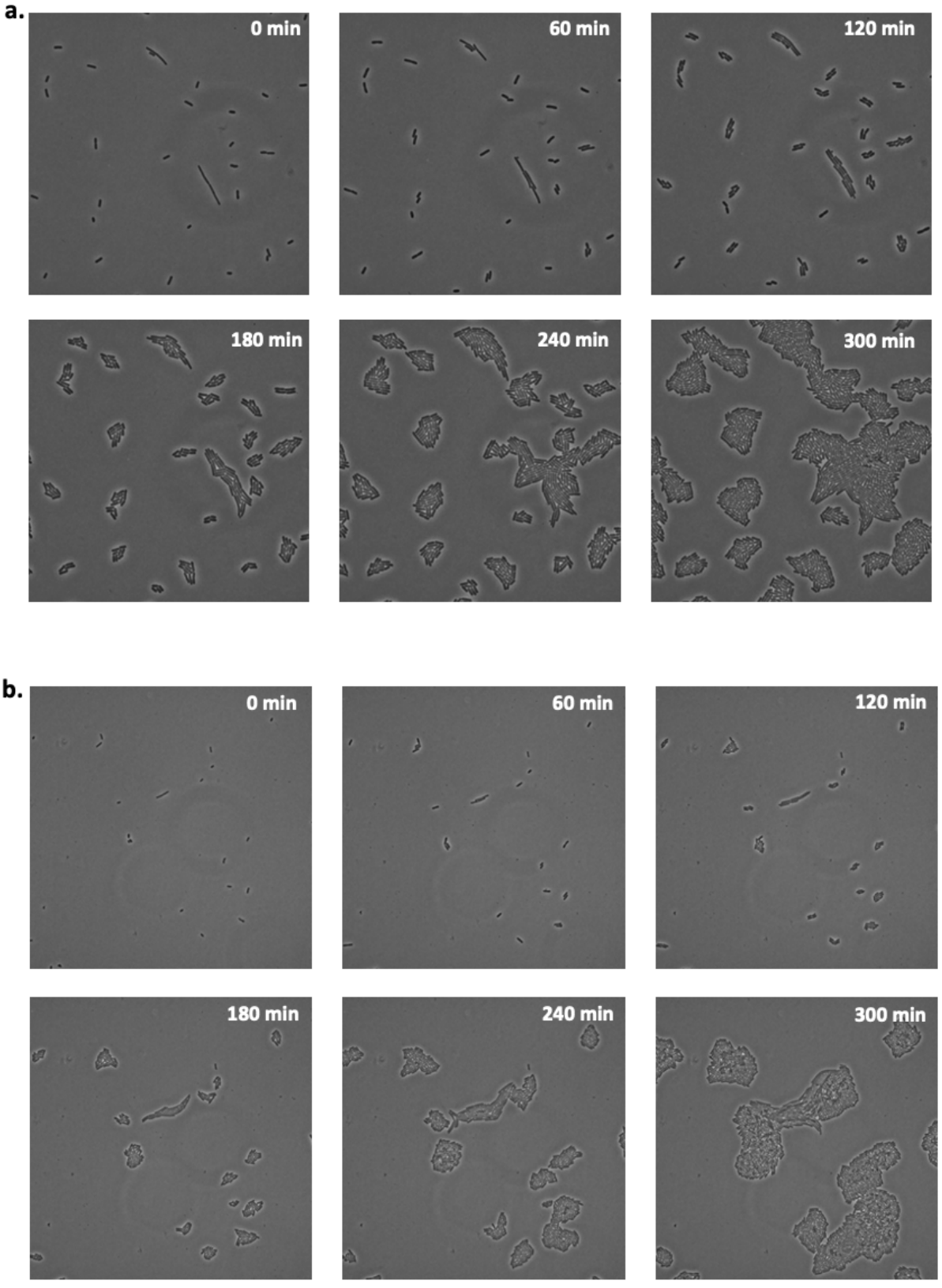
Figure 5.2 Short cells showed a nucleotide deletion within the chaperon protein FocC in the fim operon. This deletion was not present in all individual sequences, so we attributed it to alignment error. **(a)** the 4 nucleotide CTCT deletion within the chaperon protein **(b)** the truncated FocC protein and its corresponding hypothetical protein.

Table 5.1 RNA sequencing results showing genes that are positively regulated in the long cells compared to the short cells. RNA-sequencing results were analysed using RESeq2. baseMean refers to the average of the normalised count values – the average amount of RNA present in both sample types for that gene. Log2foldchange, is the fold change observed for that gene in the sample, lfcSE is standard error, stat is the test statistic.

locus	Gene	Description	baseMean	log2foldchange	lfcSE	stat	pvalue
AKNKLGHJ_05082	aidA-I	AIDA-I autotransporter	42996.78043	2.54068334	0.237655084	10.69063321	1.13E-26
AKNKLGHJ_05083	hypothetical protein		3337.365989	2.776120802	0.227417971	12.20713028	2.85E-34
AKNKLGHJ_05084	hypothetical protein		6430.254298	2.881326503	0.205770809	14.00260084	1.50E-44
AKNKLGHJ_05085	hypothetical protein		1116.760144	2.332827004	0.175978283	13.25633464	4.15E-40
AKNKLGHJ_05089	hypothetical protein		8904.950944	2.822099433	0.172126532	16.3954935	2.06E-60
AKNKLGHJ_05090	hypothetical protein		55963.62867	3.005115904	0.2347294	12.80246913	1.59E-37
AKNKLGHJ_05091	hypothetical protein		2228.624936	2.996687156	0.239323989	12.52146588	5.70E-36
AKNKLGHJ_05092	hypothetical protein		50.63306262	2.478125045	0.2670847	9.278423846	1.72E-20
AKNKLGHJ_05093	hypothetical protein		2748.956261	2.786971623	0.179062847	15.56420927	1.27E-54
AKNKLGHJ_05094	pdeL_3	Cyclic di-GMP phosphodiesterase	4107.945216	2.882358713	0.190119427	15.16077947	6.43E-52

Table 5.2 RNA sequencing results showing genes that are negatively regulated in the long cells compared to the short cells. RNA-sequencing results were analysed using RESeq2. baseMean refers to the average of the normalised count values – the average amount of RNA present in both sample types for that gene. Log2foldchange, is the fold change observed for that gene in the sample, lfcSE is standarderror, stat is the test statistic.

locus	Gene	Description	baseMean	log2foldchange	lfcSE	stat	pvalue
AKNKLGHJ_00140	gfcA	Threonine-rich inner membrane protein	15722.42581	-2.272071194	0.165102343	-13.76159271	4.34E-43
AKNKLGHJ_00141	gfcB_1	putative lipoprotein	2812.87995	-2.825327129	0.146256257	-19.31764967	3.82E-83
AKNKLGHJ_00142		hypothetical protein	1875.416138	-2.803578091	0.118541869	-23.65053047	1.17E-123
AKNKLGHJ_00143		hypothetical protein	4954.361906	-2.763053144	0.12181467	-22.68243343	6.68E-114
AKNKLGHJ_00144	kpsD	Polysialic acid transport protein	3506.152513	-2.768886392	0.140451968	-19.714116	1.63E-86
AKNKLGHJ_00145	etp	Low molecular weight protein-tyrosine-phosphatase	1194.816994	-2.705903457	0.136737119	-19.78909226	3.70E-87
AKNKLGHJ_00146	etk	Tyrosine-protein kinase	8683.119639	-2.385800657	0.131315229	-18.16849944	9.17E-74
AKNKLGHJ_00675	glnK	Nitrogen regulatory protein P-II 2	295.2743532	-2.24006634	0.519854599	-4.309024759	1.64E-05
AKNKLGHJ_01291	fimH_1	Type 1 fimbrin D-mannose specific adhesin	1855.07455	-2.048344734	0.313648942	-6.530692316	6.55E-11
AKNKLGHJ_01292	fimG_1		950.7805839	-2.496581888	0.511055117	-4.885151929	1.03E-06
AKNKLGHJ_01293	fimF_1		603.9540329	-2.580661707	0.50976852	-5.062418739	4.14E-07
AKNKLGHJ_01294	fimD_1	Outer membrane usher protein	5639.679514	-2.535624899	0.453474156	-5.591553267	2.25E-08
AKNKLGHJ_01295	fimD_2	Outer membrane usher protein	935.1715364	-2.739933226	0.31547988	-8.684969795	3.79E-18
AKNKLGHJ_01296	fimC	Chaperone protein	4858.831434	-2.545159211	0.307437691	-8.278618025	1.25E-16
AKNKLGHJ_01297	lpfA_1	putative major fimbrial subunit LpfA	7684.975359	-2.439807823	0.336069142	-7.259838881	3.88E-13
AKNKLGHJ_01298		Fimbrial subunit type 1	57771.07352	-2.428718	0.429923534	-5.649185983	1.61E-08


a.
0 min
60 min
120 min
180 min
240 min
300 min
b.
0 min
60 min
120 min
180 min
240 min
300 min

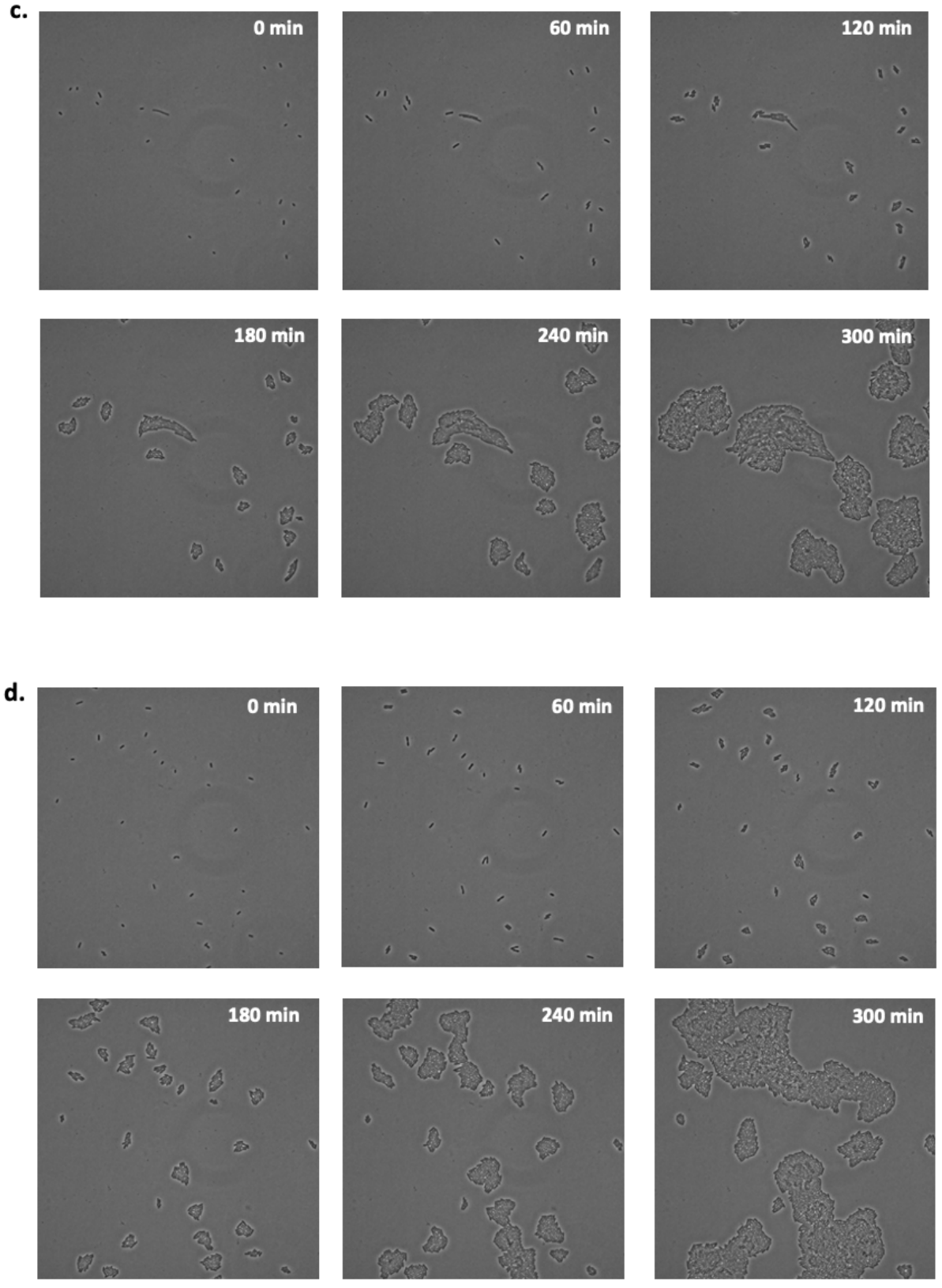


Figure 5.3 Still images from time lapse video series of SC375 growth over a period of 300 minutes. Each set of pictures (**a,b,c** and **d**) is a different video started from a different diluted culture, all cultures were grown in the same conditions.

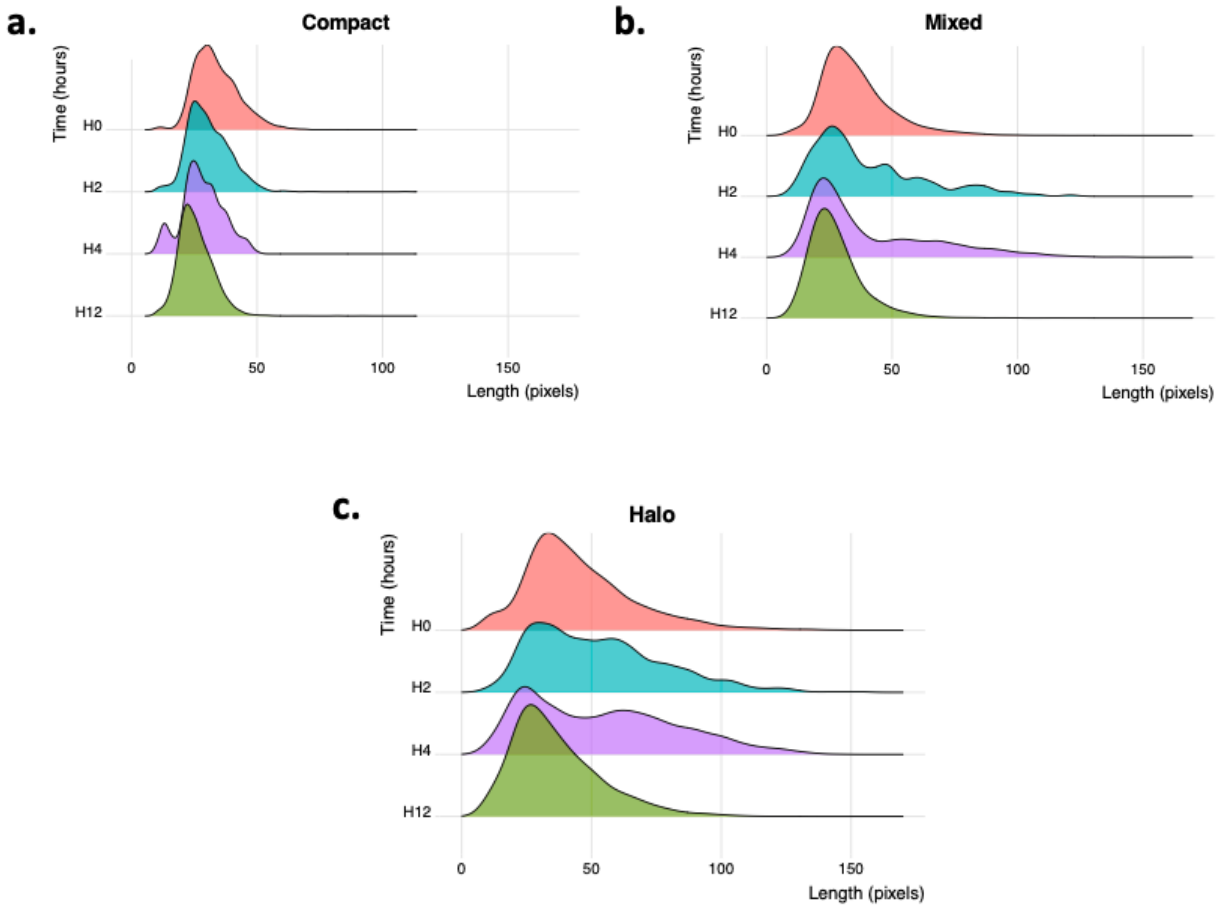
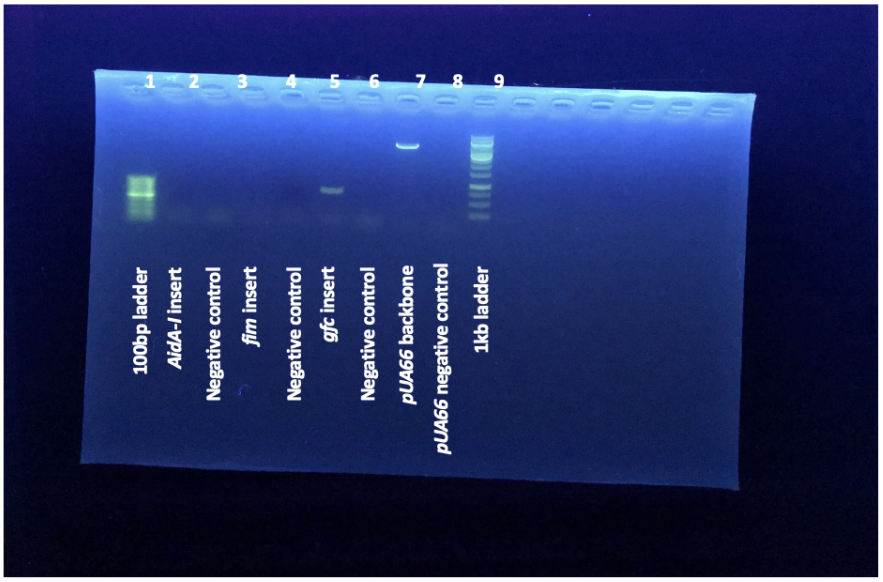


Figure 5.4 Cell length by colony type over time. Distribution of cell lengths in each colony type after sampling of liquid cultures grown for 0 hours, 2 hours, 4 hours and 12 hours, in **(a)** compact colonies, **(b)** mixed colonies and **(c)** halo colonies.

a.



b.

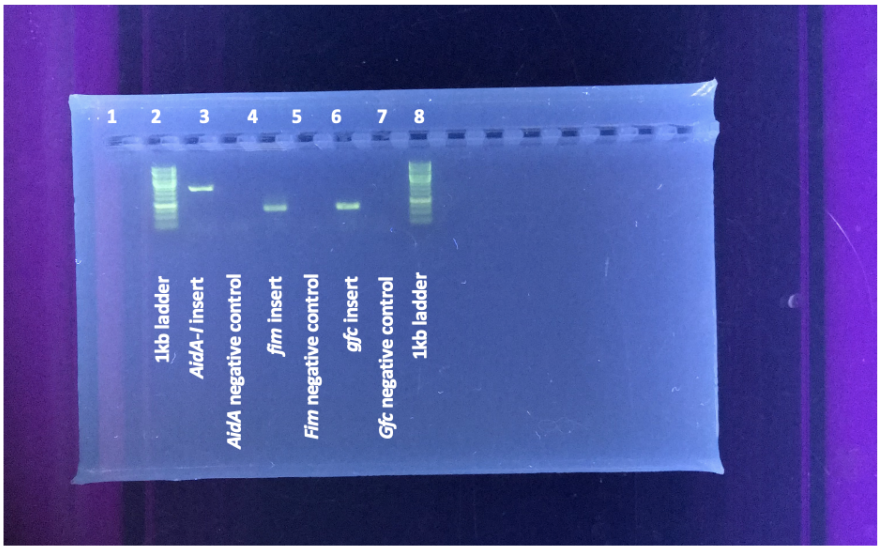


Figure 5.5 Confirmation of successful PCR amplification of promoter regions for SC375 genomic sequences. Results of PCR reactions run on a 1% agarose gel. **(a)** Successful amplification of the pUA66-backbone is shown in lane 7. Other PCR reactions for inserts proved unsuccessful. **(b)** Amplification of all 3 inserts was successful.

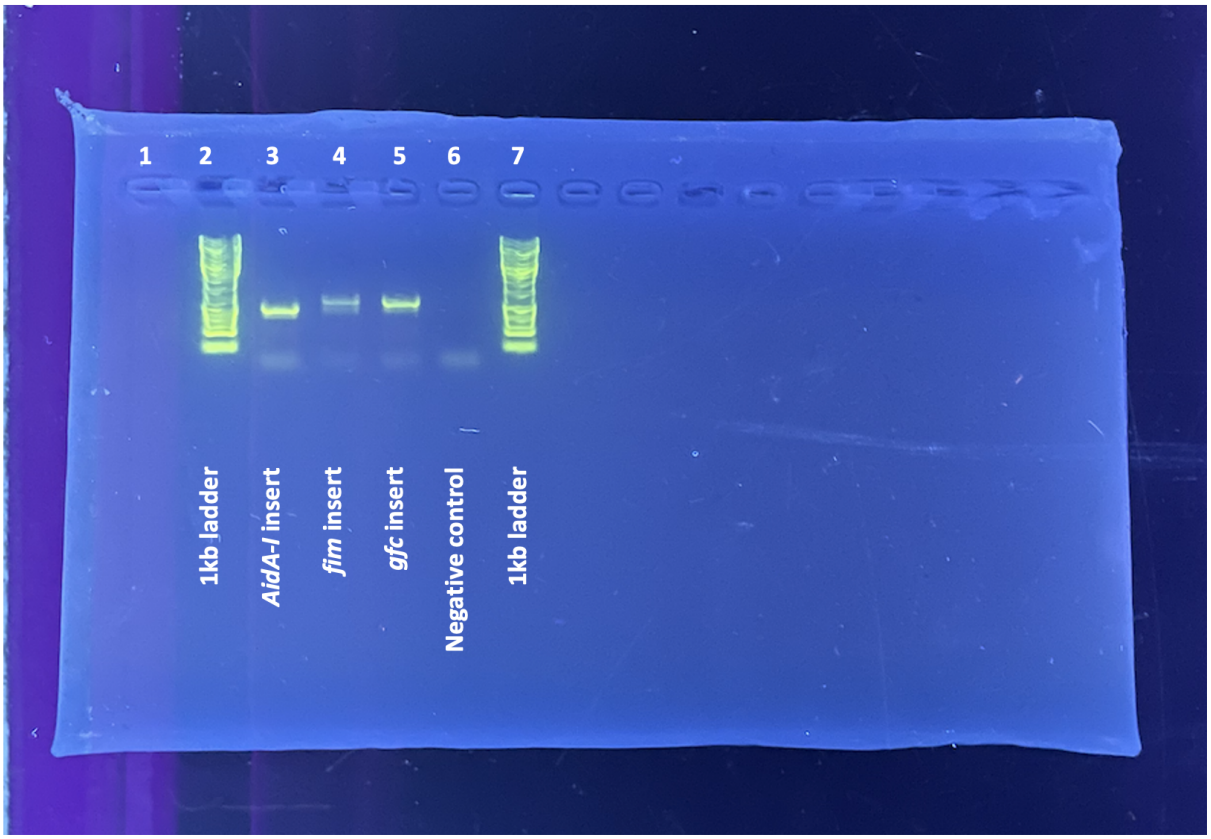


Figure 5.6 Results of PCR to check transformation of plasmid constructs into electrocompetent cells was successful. PCR was performed using overnight cultures of transformed colonies as PCR templates. Lanes from left to right: 1kb ladder, AidA-I insert, fim insert, gfc insert, negative control (ddh_20), 1 kb ladder. Bands for gfc and fim were of similar expected lengths. The AidA-i insert is expected to be 2000bp long, this band did not show a band of the expected length and therefore was not deemed to be a successful transformation.

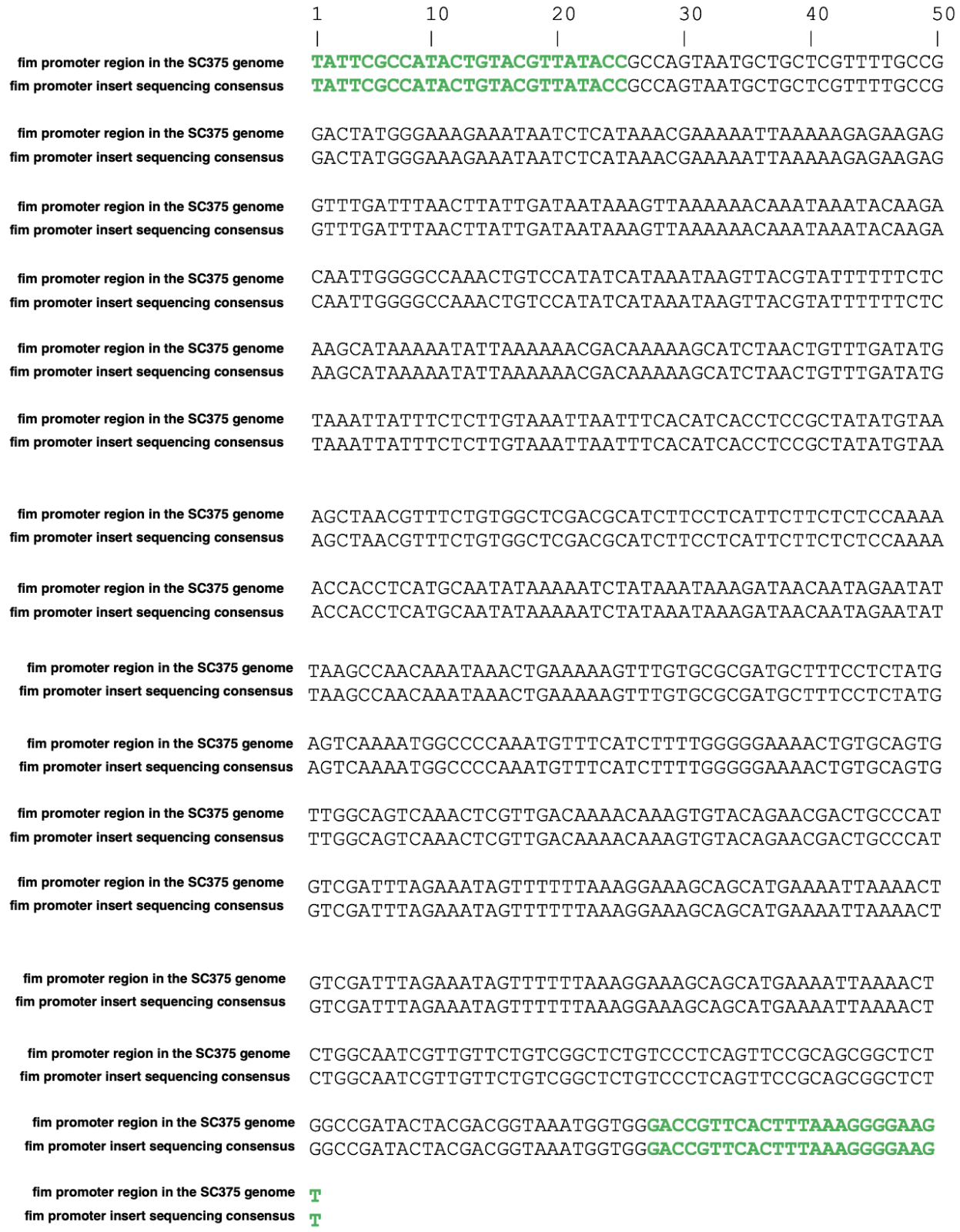


Figure 5.7 Alignment of the fim promoter region within the SC375 genome and sequencing results. Green nucleotides show the primer binding sites.

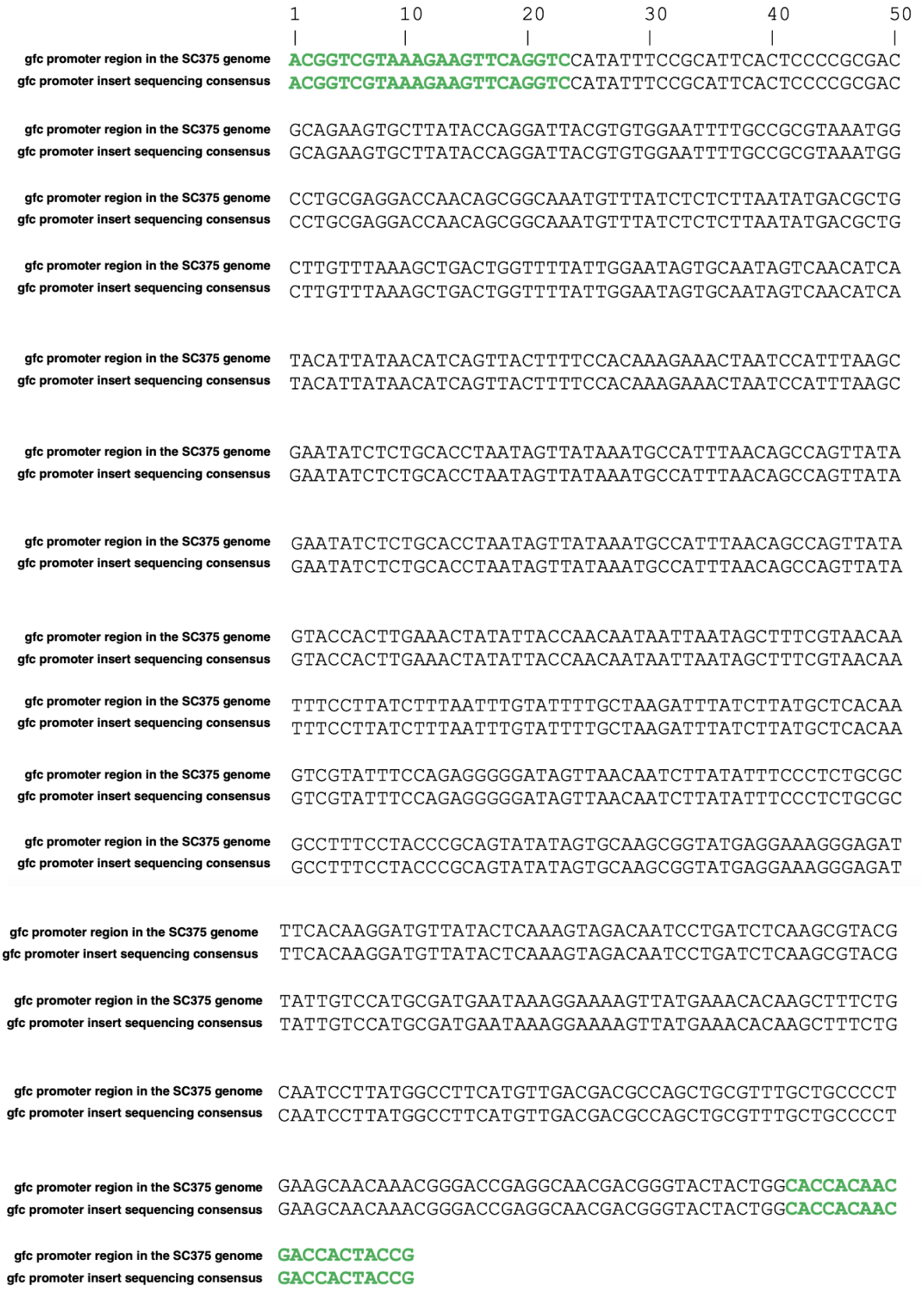


Figure 5.8 Alignment of the gfc promoter region within the SC375 genome and sequencing results. Green nucleotides show the primer binding sites.

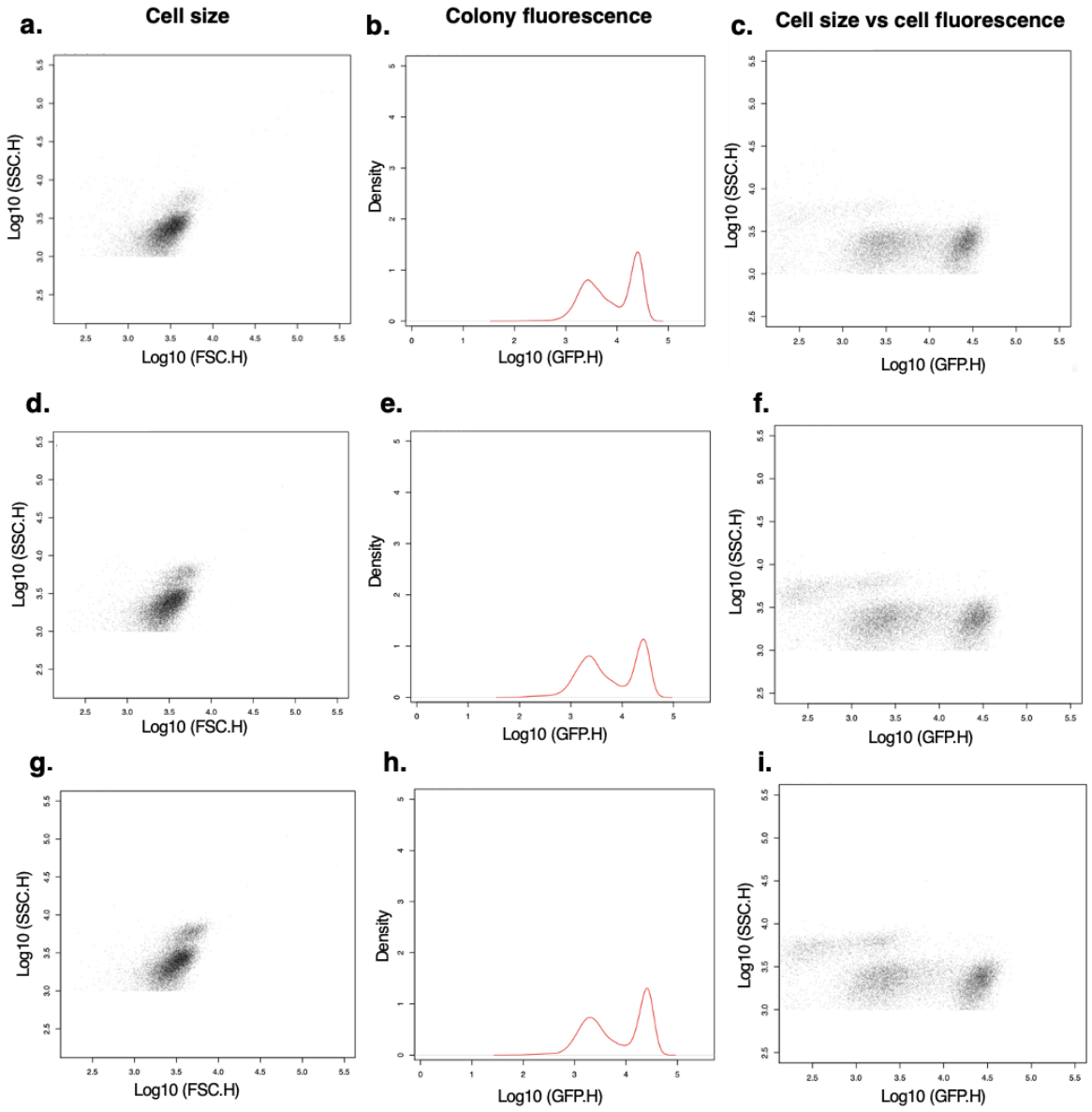


Figure 5.9 Triplicate fluorescent resuspended colonies of SC375 cells transformed with the pgfc::GFP plasmid. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within three fluorescent colonies. **(a) (b) (c)** Cell size **(b) (e) (h)** colony fluorescence and **(c) (f) (i)** cell fluorescence of cells within a fluorescent colony. Coupled with the results of our negative control, we determined that this lower fluorescence value was due to cell autofluorescence from exposure to the 470 nm light used to measure GFP.

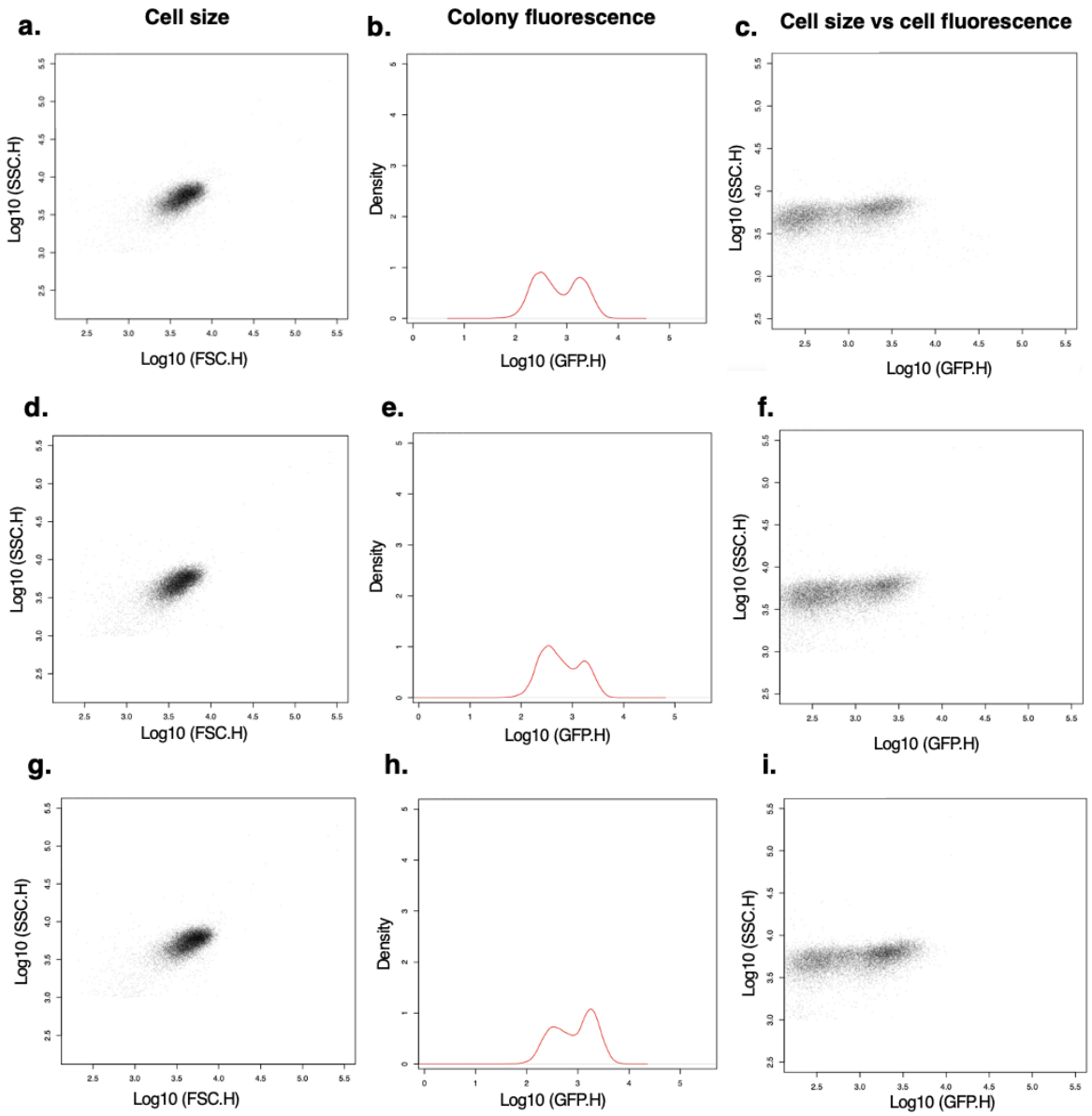


Figure 5.10 Triplicate non-fluorescent resuspended colonies of SC375 cells transformed with the pgfc::GFP plasmid. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within three fluorescent colonies. **(a) (b) (c)** Cell size **(b) (e) (h)** colony fluorescence and **(c) (f) (i)** cell fluorescence of cells within a fluorescent colony. Coupled with the results of our negative control, we determined that this lower fluorescence value was due to cell autofluorescence from exposure to the 470 nm light used to measure GFP.

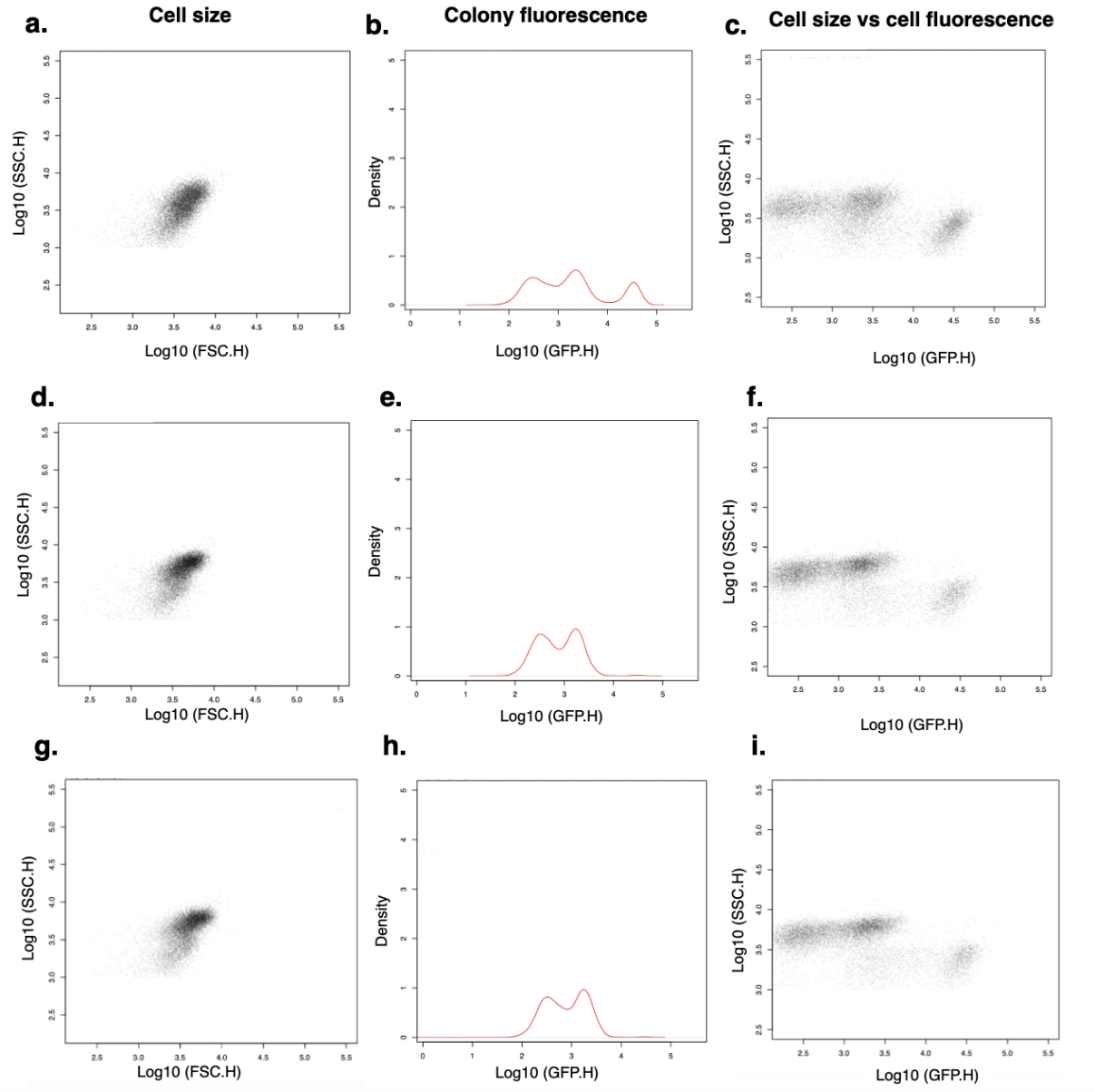


Figure 5.11 Triplicate mixedt resuspended colonies of SC375 cells transformed with the pgfc::GFP plasmid. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within three fluorescent colonies. **(a) (b) (c)** Cell size **(b) (e) (h)** colony fluorescence and **(c) (f) (i)** cell fluorescence of cells within a fluorescent colony. Coupled with the results of our negative control, we determined that this lower fluorescence value was due to cell autofluorescence from exposure to the 470 nm light used to measure GFP.

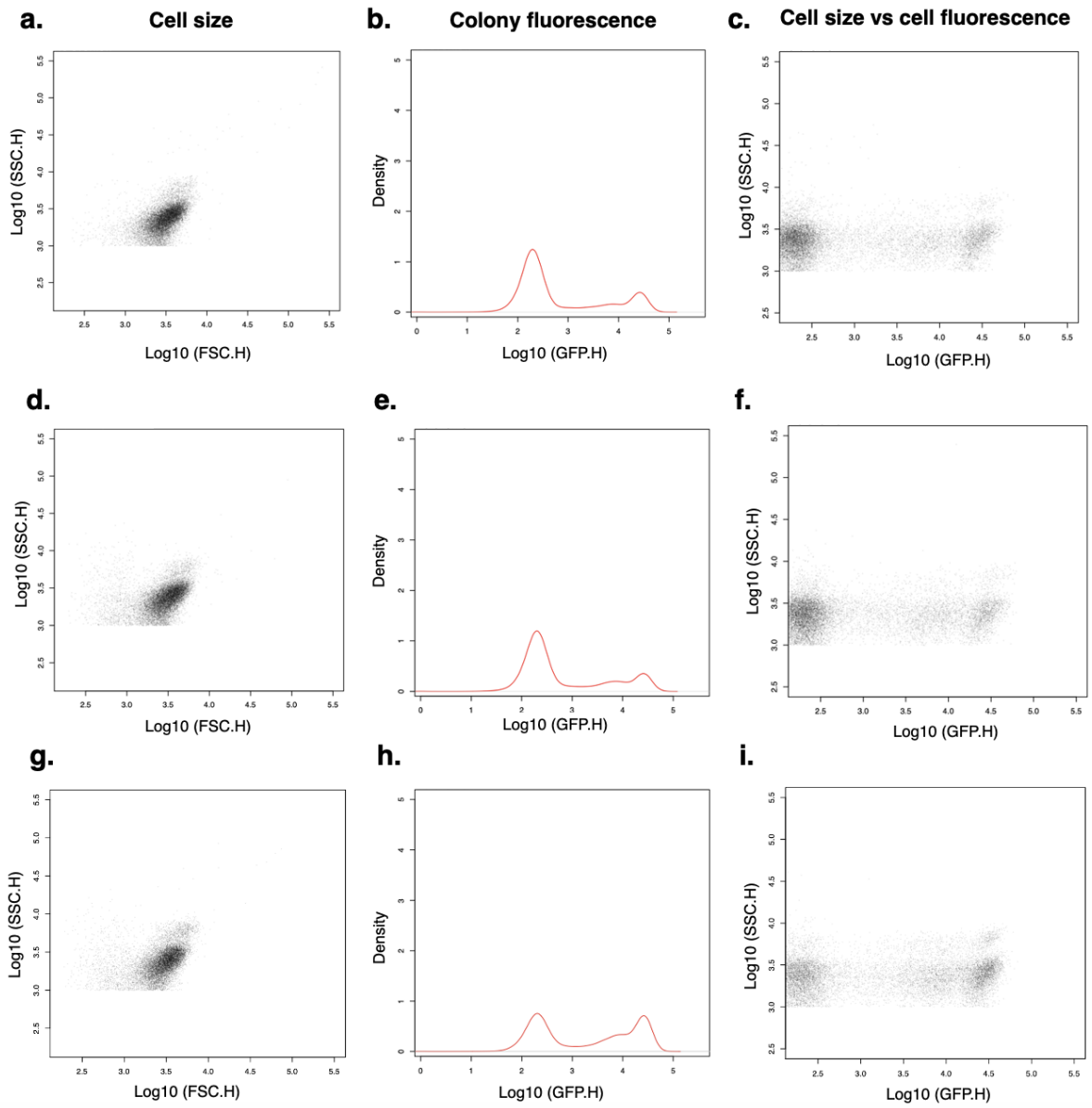


Figure 5.12 Triplicate fluorescent resuspended colonies of SC375 cells transformed with the pfim::GFP plasmid. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within three fluorescent colonies. **(a) (b) (c)** Cell size **(b) (e) (h)** colony fluorescence and **(c) (f) (i)** cell fluorescence of cells within a fluorescent colony. Coupled with the results of our negative control, we determined that this lower fluorescence value was due to cell autofluorescence from exposure to the 470 nm light used to measure GFP.

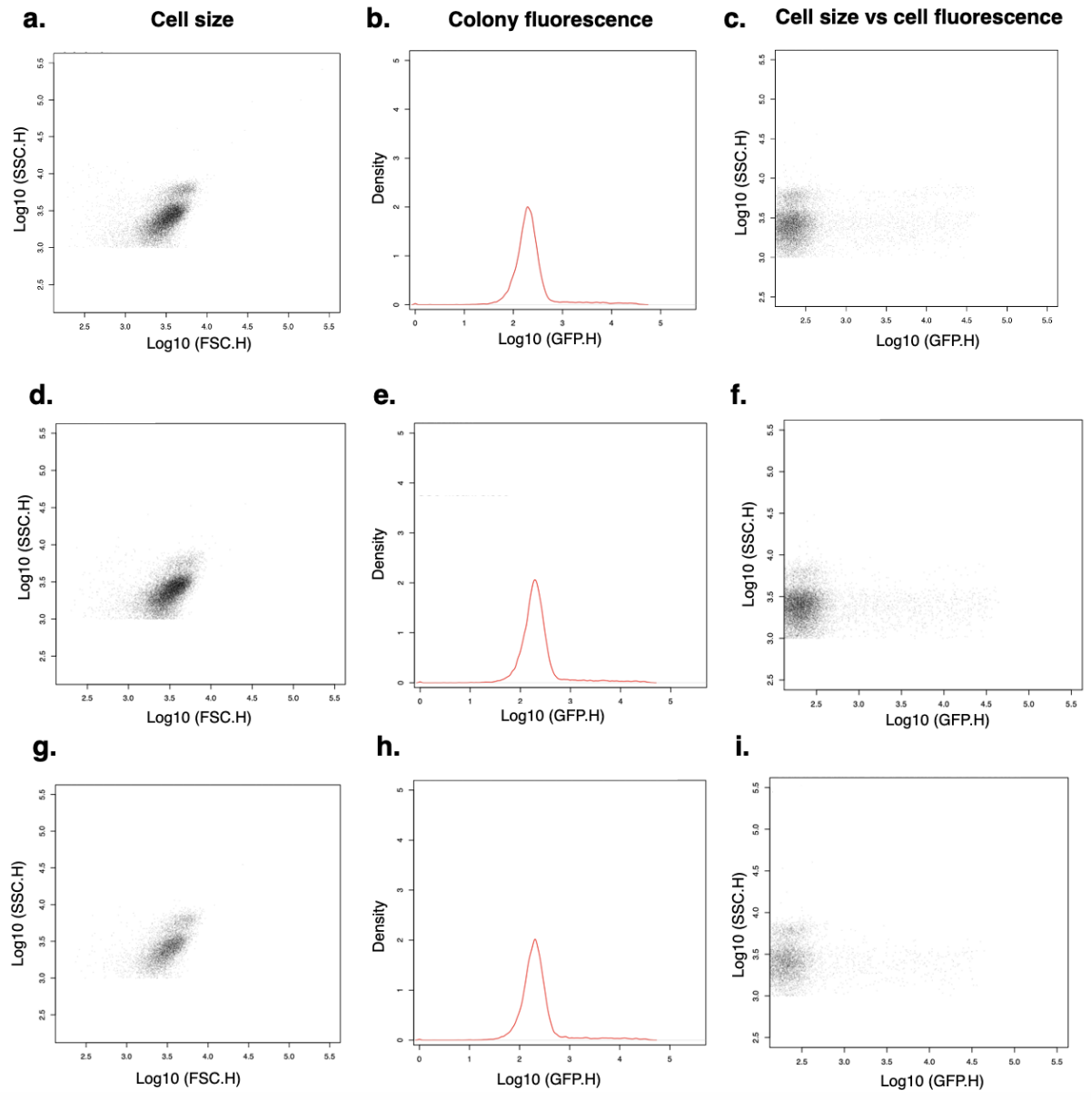


Figure 5.13 Triplicate fluorescent resuspended colonies of SC375 cells transformed with the pfim::GFP plasmid. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within three fluorescent colonies. **(a) (b) (c)** Cell size **(b) (e) (h)** colony fluorescence and **(c) (f) (i)** cell fluorescence of cells within a fluorescent colony. Coupled with the results of our negative control, we determined that this lower fluorescence value was due to cell autofluorescence from exposure to the 470 nm light used to measure GFP.

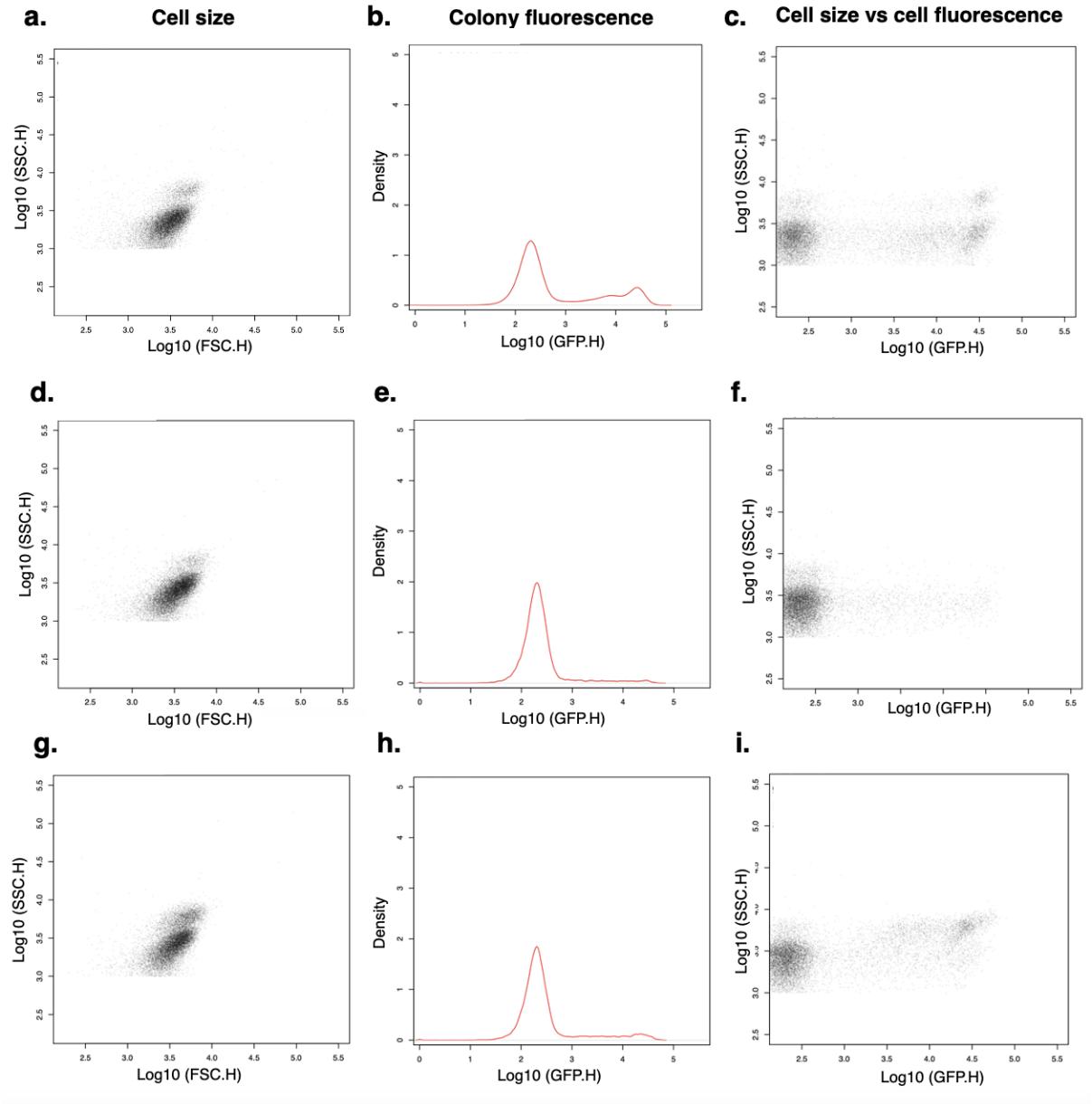


Figure 5.14 Triplicate fluorescent resuspended colonies of SC375 cells transformed with the pfim::GFP plasmid. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within three fluorescent colonies. **(a) (b) (c)** Cell size **(b) (e) (h)** colony fluorescence and **(c) (f) (i)** cell fluorescence of cells within a fluorescent colony. Coupled with the results of our negative control, we determined that this lower fluorescence value was due to cell autofluorescence from exposure to the 470 nm light used to measure GFP.

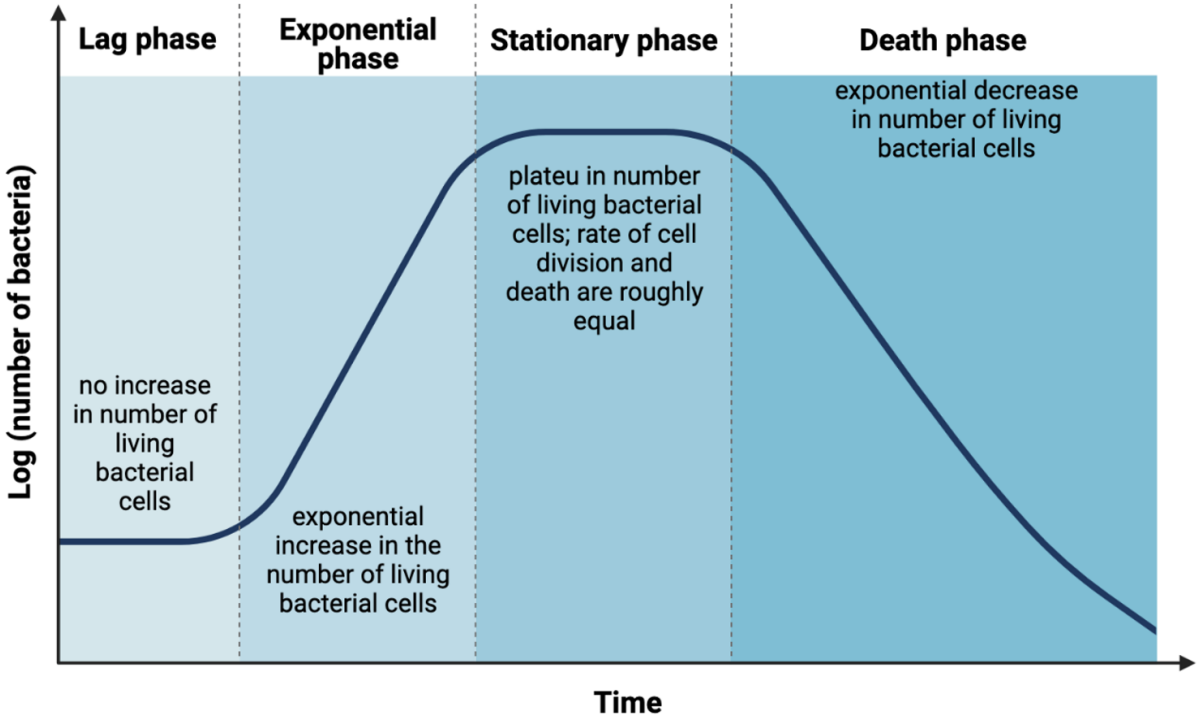


Figure 5.15 Bacterial growth phases. Bacteria have 4 main phases of growth: lag phase, exponential phase, stationary phase and death phase. Adapted from (BioRender, n.d.).

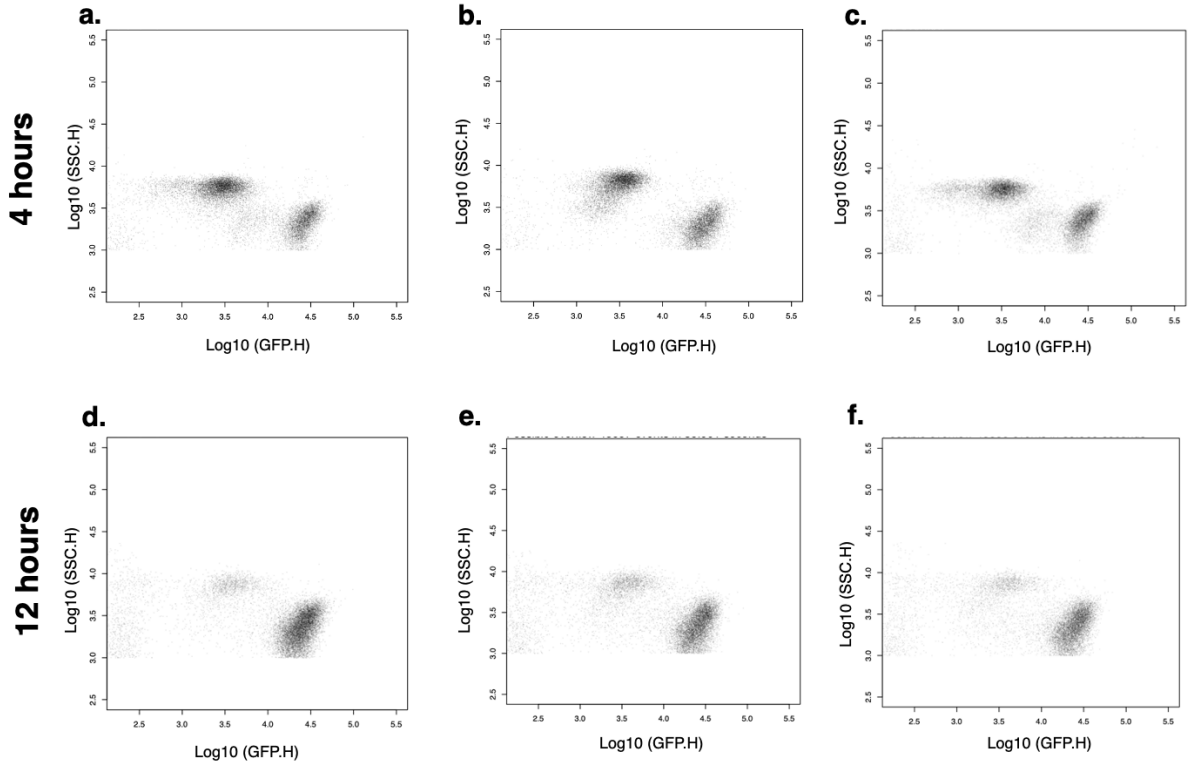


Figure 5.16 Fluorescence and cell size of triplicate pgfc::GFP transformed strains over time. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within three liquid cultures. Cell size and fluorescence was measured after 4 hours **(a) - (c)** and 12 hours **(d) - (f).**

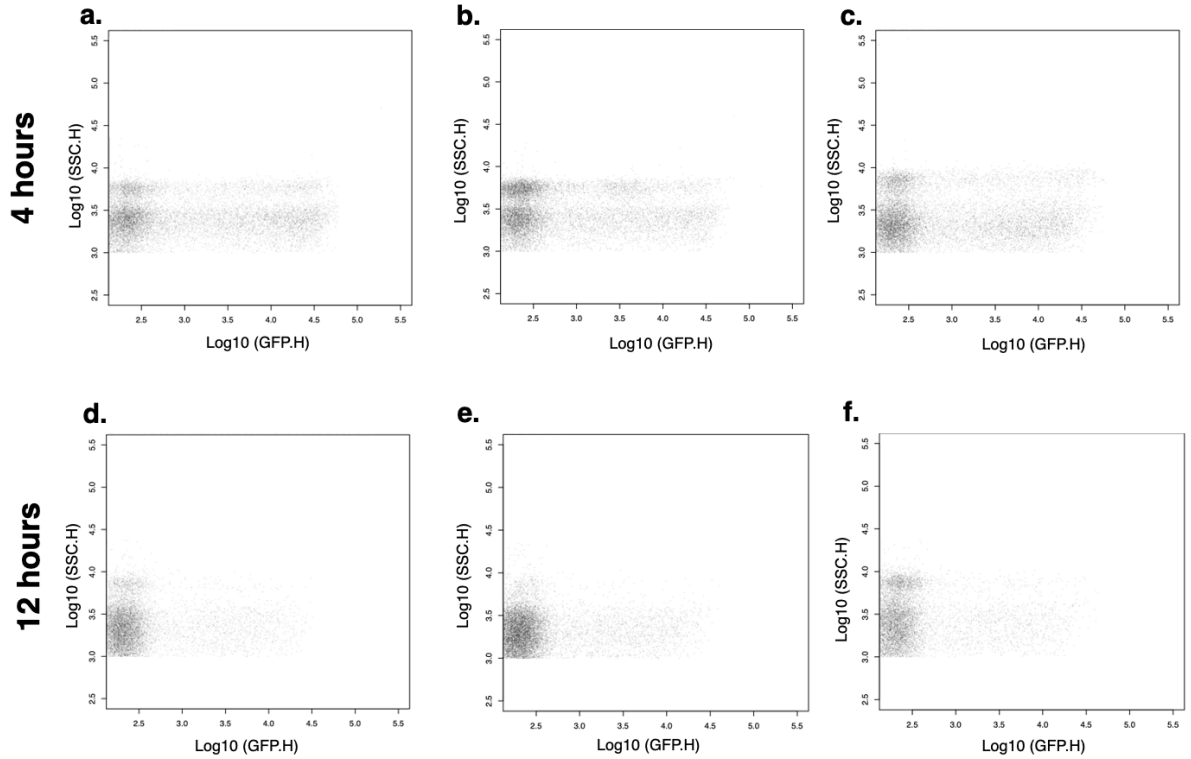
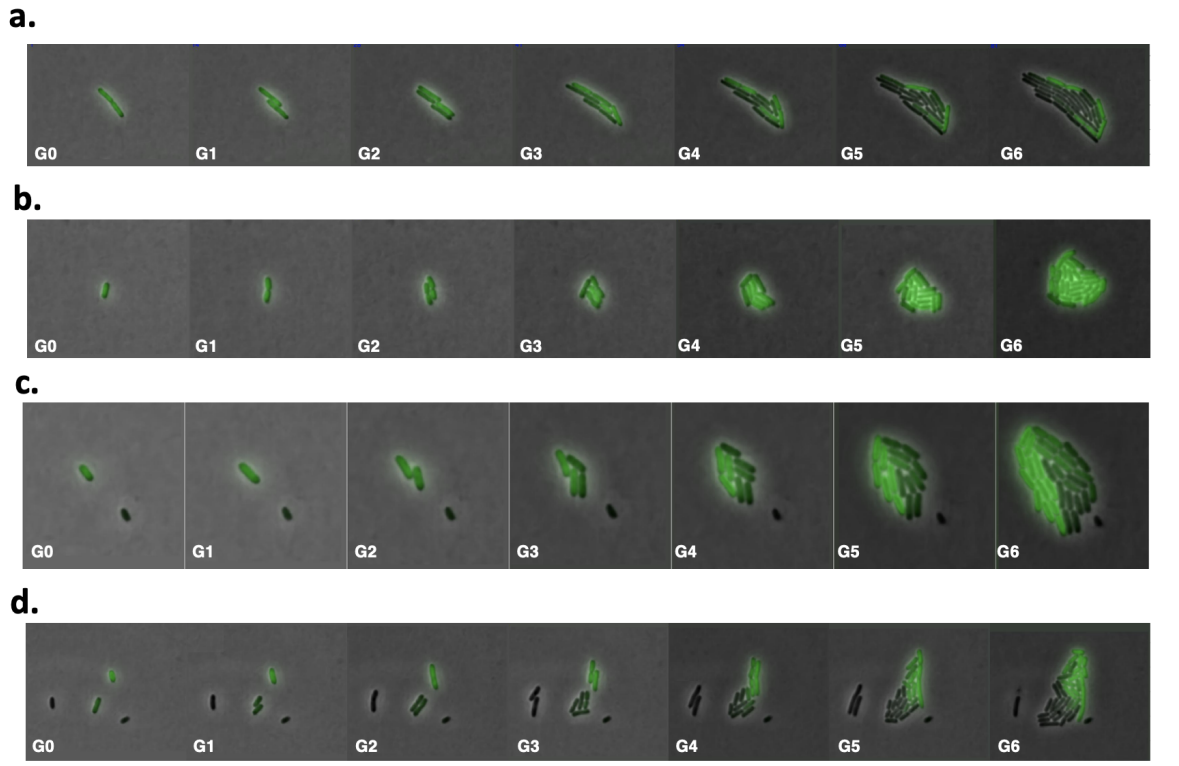
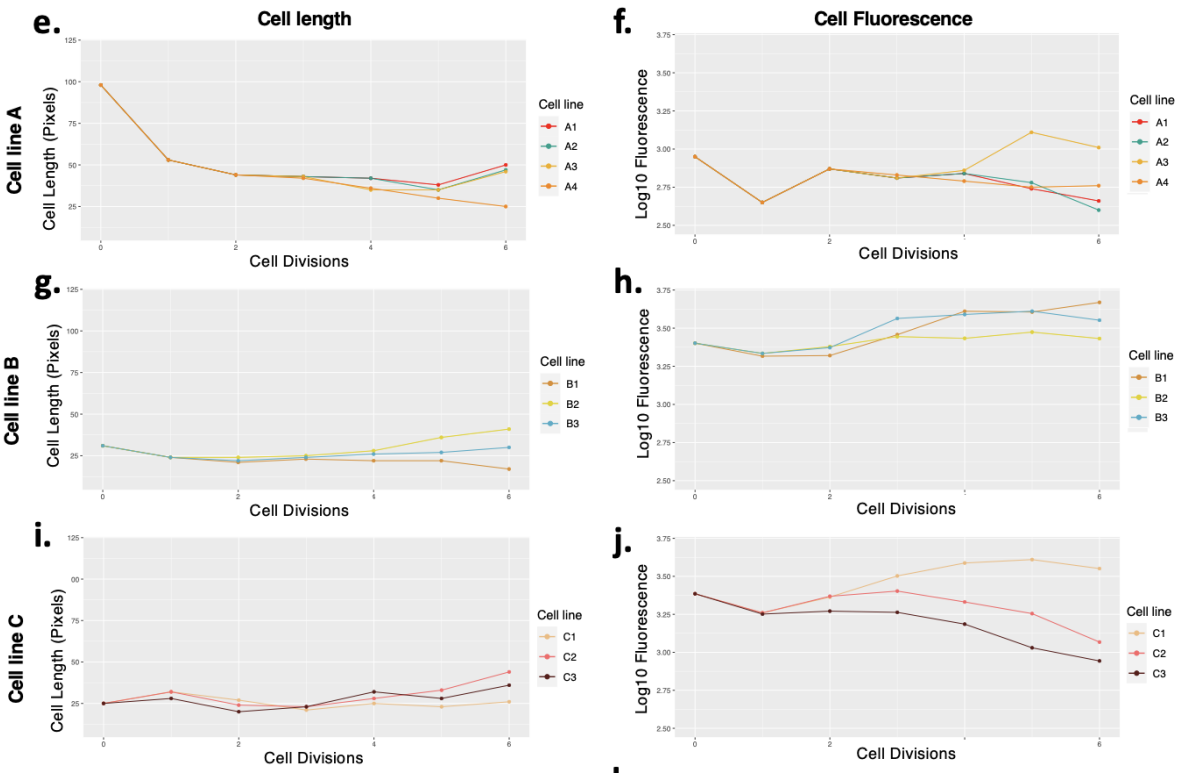


Figure 5.17 Fluorescence and cell size of triplicate pfim::GFP transformed strains over time. Flow cytometry was used to measure cell size (FSC), complexity (SSC) and fluorescence (GFP) within three liquid cultures. Cell size and fluorescence was measured after 4 hours **(a) - (c)** and 12 hours **(d) - (f).**


a.
G0
G1
G2
G3
G4
G5
G6
b.
G0
G1
G2
G3
G4
G5
G6
c.
G0
G1
G2
G3
G4
G5
G6
d.
G0
G1
G2
G3
G4
G5
G6


e.
Cell length
Cell line A
Cell Length (Pixels)
Cell Divisions
Cell line
A1
A2
A3
A4
f.
Cell Fluorescence
Log10 Fluorescence
Cell Divisions
Cell line
A1
A2
A3
A4
g.
Cell line B
Cell Length (Pixels)
Cell Divisions
Cell line
B1
B2
B3
h.
Log10 Fluorescence
Cell Divisions
Cell line
B1
B2
B3
i.
Cell line C
Cell Length (Pixels)
Cell Divisions
Cell line
C1
C2
C3
j.
Log10 Fluorescence
Cell Divisions
Cell line
C1
C2
C3

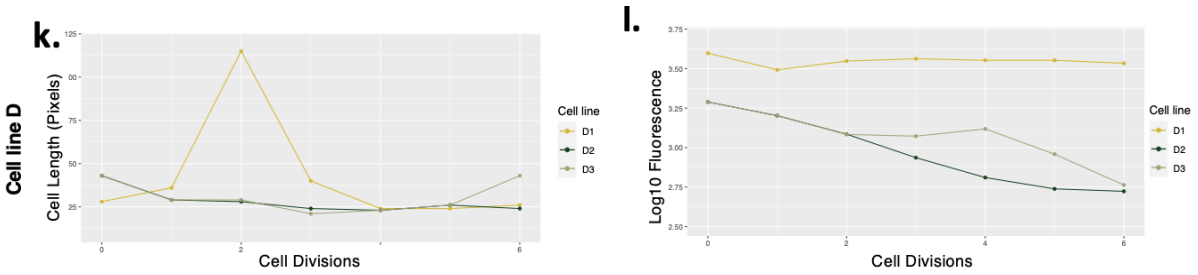
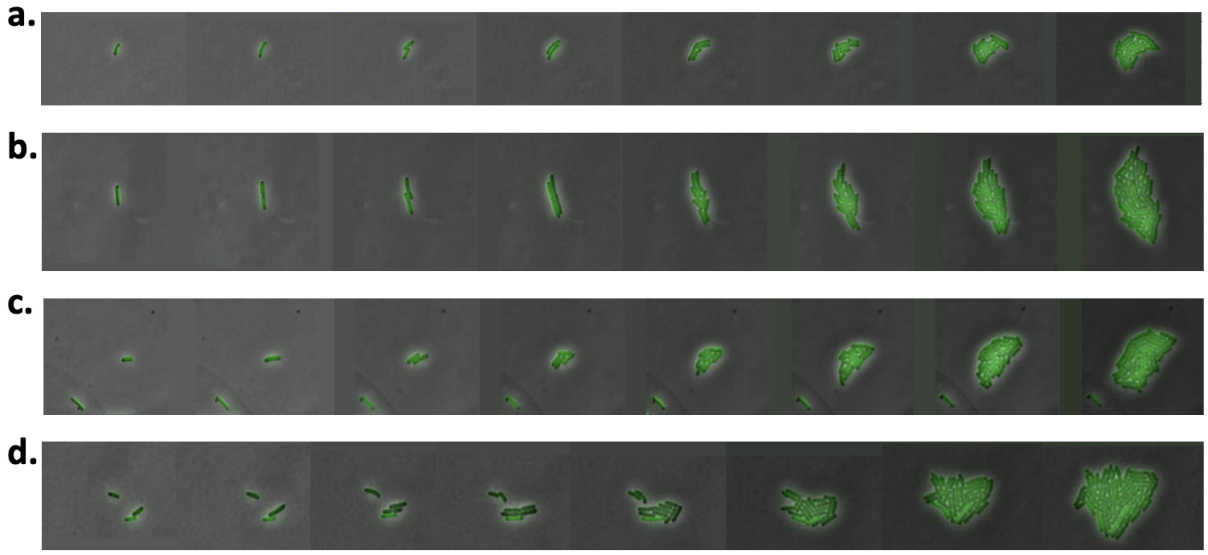


Figure 5.18 Cell length and mean fluorescence changes of pfim::GFP transformed colonies over time. Sample time lapse videos of cell growth from diluted overnight cultures, each collection of images (**a, b, c** and **d**) is from a different overnight culture. Cell segmentation data plotted as line graphs of cell length (**e, g, i** and **k**) and Log10 of mean cell fluorescence (**f, h, j** and **l**). For clarity the division times for all cells has been aligned, so in some cases there are an unequal number of time steps between each cell division.


a.
b.
c.
d.

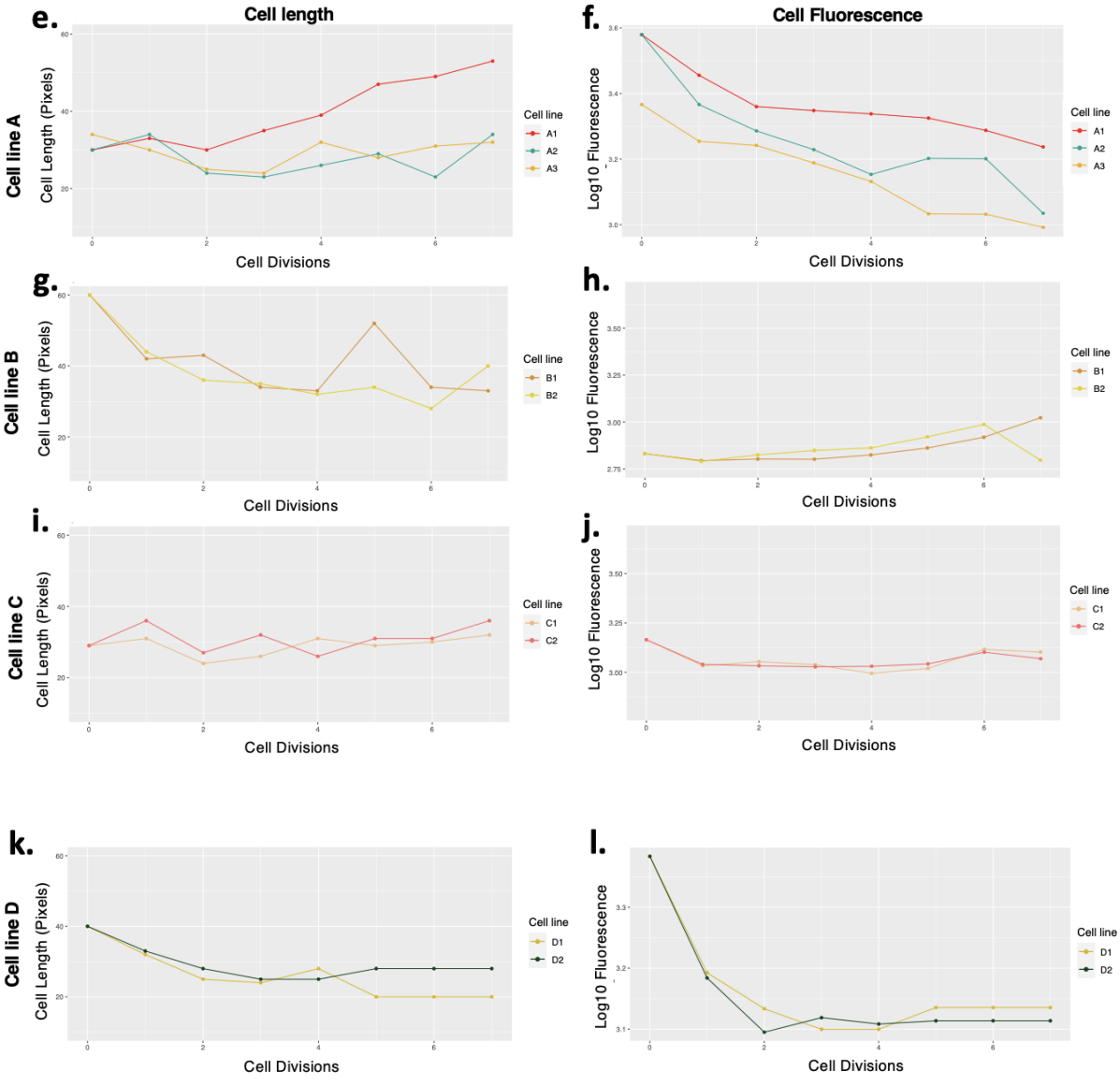

e.
Cell length
Cell line A
Cell Length (Pixels)
Cell Divisions
Cell line
A1
A2
A3
f.
Cell Fluorescence
Log10 Fluorescence
Cell Divisions
Cell line
A1
A2
A3
g.
Cell line B
Cell Length (Pixels)
Cell Divisions
Cell line
B1
B2
h.
Log10 Fluorescence
Cell Divisions
Cell line
B1
B2
i.
Cell line C
Cell Length (Pixels)
Cell Divisions
Cell line
C1
C2
j.
Log10 Fluorescence
Cell Divisions
Cell line
C1
C2
k.
Cell line D
Cell Length (Pixels)
Cell Divisions
Cell line
D1
D2
l.
Log10 Fluorescence
Cell Divisions
Cell line
D1
D2

Figure 5.19 Cell length and mean fluorescence changes of pgfc::GFP transformed colonies over time. Sample time lapse videos of cell growth from diluted overnight cultures, each collection of images (**a, b, c** and **d**) is from a different overnight culture. Cell segmentation data plotted as line graphs of cell length (**e, g, i** and **k**) and Log10 of mean cell fluorescence (**f, h, j** and **l**). For clarity the division times for all cells has been aligned, so in some cases there are an unequal number of time steps between each cell division.

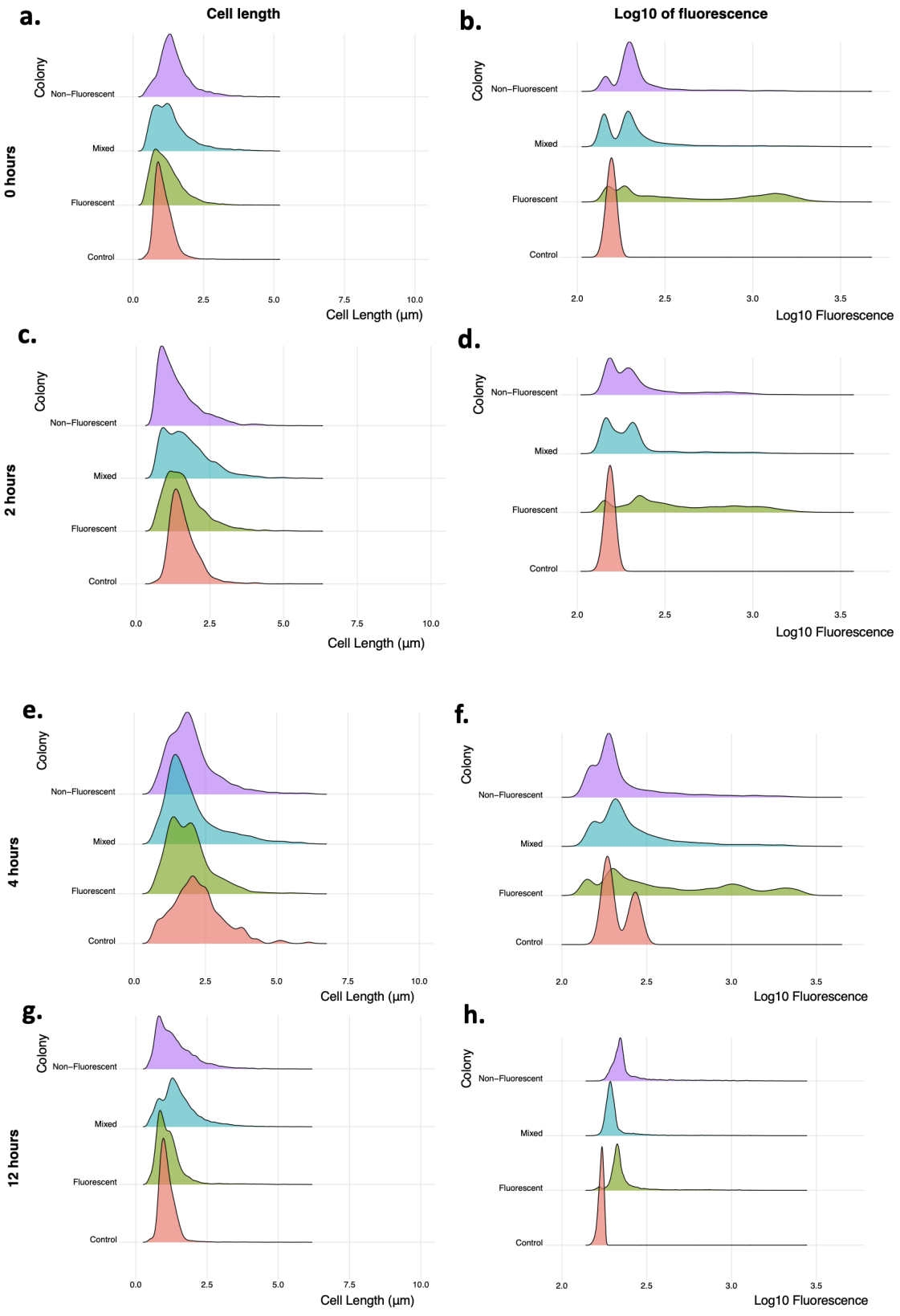

a.
Cell length
b.
Log10 of fluorescence
c.
d.
e.
f.
g.
h.
0 hours
2 hours
4 hours
12 hours
Colony
Non−Fluorescent
Mixed
Fluorescent
Control
0.0
2.5
5.0
7.5
10.0
Cell Length (µm)
2.0
2.5
3.0
3.5
Log10 Fluorescence

Figure 5.20 Cell length and fluorescence changes of pfim::GFP transformed cells over time, plotted by colony type. We grew colonies for a period of 12 hours, with samples taken at 0, 2, 4 and 12 hours to be photographed using microscopy. Data was plotted as GGridges of cell length (**a, c, e** and **g**), and Log10 of mean fluorescence (**b, d, f** and **h**).

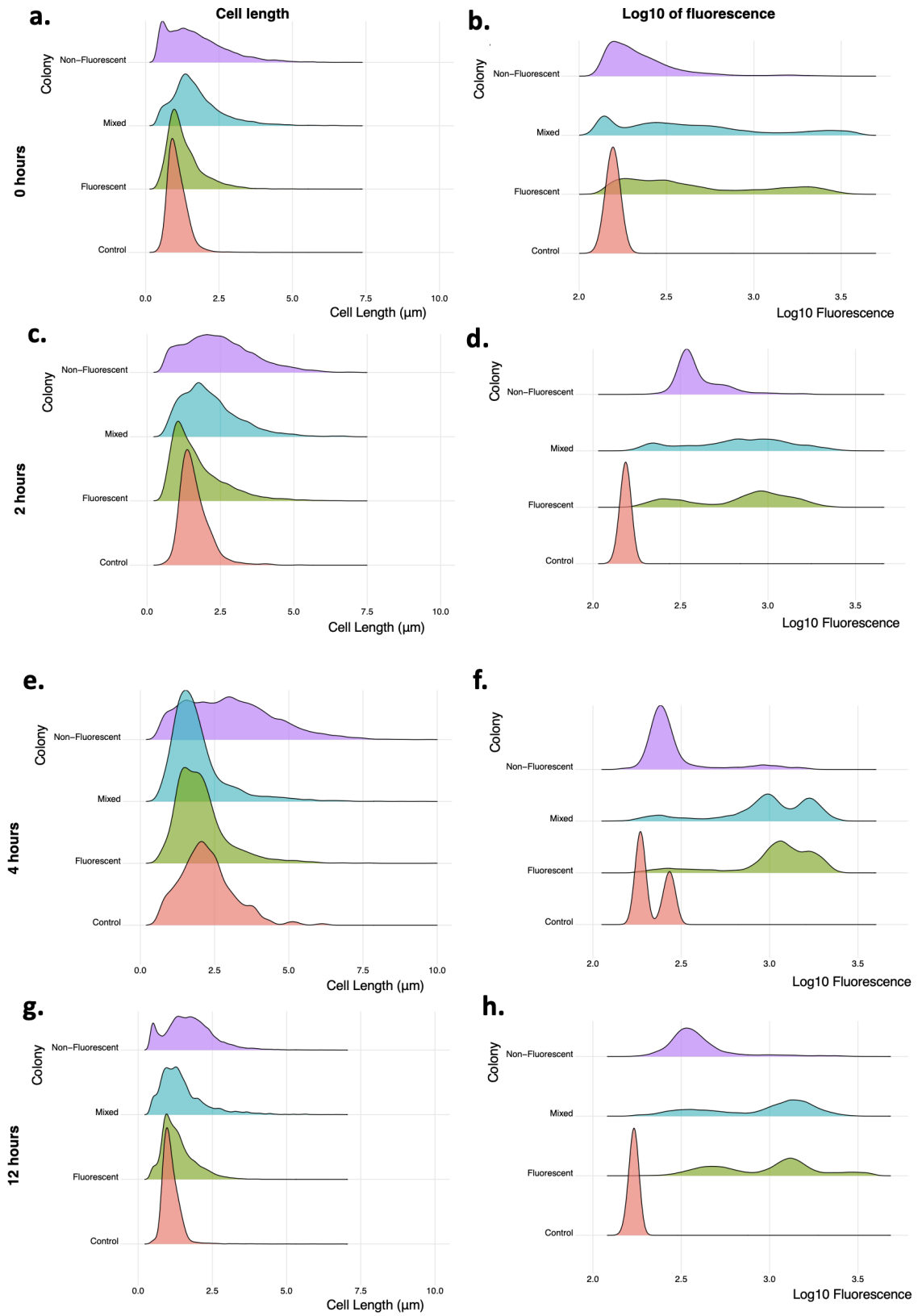

a.
Cell length
b.
Log10 of fluorescence
0 hours
c.
d.
2 hours
e.
f.
4 hours
g.
h.
12 hours
Colony
Non−Fluorescent
Mixed
Fluorescent
Control
0.0
2.5
5.0
7.5
10.0
Cell Length (µm)
2.0
2.5
3.0
3.5
Log10 Fluorescence

Figure 5.21 Cell length and fluorescence changes of pgfc::GFP transformed cells over time, plotted by colony type. We grew colonies for a period of 12 hours, with samples taken at 0, 2, 4 and 12 hours to be photographed using microscopy. Data was plotted as GGridges of cell length (**a, c, e** and **g**), and Log10 of mean fluorescence (**b, d, f** and **h**).