Dissertation

in fulfilment of the requirements for the degree

Doctor rerum naturalium

of the Faculty of Mathematics and Natural Sciences

at the Christian Albrechts University of Kiel

Investigating the functional importance of biased codon usage within the bacterial species *Pseudomonas fluorescens*

Submitted by

Anuradha Mukherjee

from the Group of Microbial Evolutionary Dynamics

at the Max Planck Institute for Evolutionary Biology

Plön, January 2020

First examiner: Dr. Jenna Gallie (MPI for Evolutionary Biology)

Second examiner: Prof. Dr. Tal Dagan (University of Kiel)

Additional examiner: Prof. Dr. Hinrich Schulenburg (University of Kiel)

Chairperson: Prof. Dr. Regina Scherließ (University of Kiel)

Date of oral examination: 13.03.2020

Zusammenfassung

Der genetische Code ist degeneriert – das heisst die 20 proteinogenen Aminosäuren werden von 61 funktionellen Codons codiert. Man könnte daher erwarten, dass alle Codons für eine Aminosäure, auch synonyme Codons genannt, gleichermaßen verwendet werden. Es wird jedoch in allen biologischen Domänen die bevorzugte Verwendung von bestimmten synonymen Codons beobachtet. Dieses Phänomen nennt man verzerrte Codonverwendung oder codon bias. Solche synonymen Substitutionen, also Veränderungen des Nukleotidspiegels, die die Aminosäuresequenz nicht verändern, wurden weithin als selektiv neutral angesehen. Eine wachsende Zahl an Studien deuten jedoch auf etwas anderes hin - auch Substitutionen können nicht-neutrale und messbare Fitnesseffekte hervorrufen.

Experimentelle Studien mit randomisierter Codonverwendung in Genen von Prokaryonten und Eukaryonten zeigen große Fitnesseffekte, die von verändertem Wachstum über globale Veränderungen der Transkription und Translation bis hin zu unterschiedlicher Proteinproduktion und -funktion reichen. Die Auswirkungen synonymer Substitutionen treten besonders bei Bakterien auf, bei denen die Wachstumsrate häufig durch die Geschwindigkeit der Translation begrenzt wird. Diese Beobachtungen ergeben sich jedoch hauptsächlich aus der selektiven Anreicherung bestimmter Codons, ein Muster, das in natürlich entwickelten Genomen selten vorkommt. Um die funktionellen Implikationen von natürlich divergierenden Mustern der Codonverwendung zu untersuchen, haben wir ein essentielles und hoch exprimiertes Gen identifiziert, das ausschließlich synonyme Unterschiede zwischen drei Genomen des Bakteriums *Pseudomonas fluorescens* aufweist. Diese Varianten wurden zwischen diesen Genomen ausgetauscht, um die Auswirkungen auf die Genexpression und Fitness zu testen. Um die synonymen Allele zwischen den Genomen auszutauschen, habe ich eine narbenfreie Gentechnik optimiert, die routinemäßig für *Pseudomonas fluorescens* SBW25, A506 und Pf0-1 angewendet wird. Die resultierenden Mutanten unterschieden sich von den entsprechenden Wildtyp-Stämmen nur an einem Ort - dem Ort des Gens, das ausgetauscht wurde. Ich habe dann die Auswirkungen von Synonym-Substitutionen auf die Genexpression und Fitness getestet und konnte zeigen, dass der Austausch der etablierten Codon-Verwendungsmuster eines Gens von einem Stamm zu einem anderen erhebliche Auswirkungen auf die Genexpression (Transkription) hat.

Nachdem ich eine Technik zur Manipulation des Genoms von *P. fluorescens*-Stämmen (zusätzlich zu SBW25) optimiert hatte, untersuchte ich die genetische Basis eines der am besten untersuchten und am ausführlichsten charakterisierten Phänotypen von *P.*

fluorescens - die Fähigkeit, Matten an der Luft-Flüssigkeits-Grenzfläche zu bilden . Zunächst verglich ich den Phänotyp der Mattenbildung in A506, Pf0-1 und PICF7 mit dem Modellstamm SBW25; während SBW25, A506 und PICF7 Matten bildeten, tat Pf0-1 dies nicht. Um die genetischen Wege zur Mattenbildung in PICF7 zu identifizieren, entwickelte ich Werkzeuge für die Transposon-Suppressor-Analyse. Diese zeigten, dass PICF7 im Gegensatz zu SBW25, bei dem die von den wss-Genen kodierten Zellulosebiosynthesemaschinen zur Bildung von Matten verwendet werden, das Pel-Operon zur Synthese des Pel-Exopolysaccharids verwendet. Pel ist ein Exopolymer, von dem bekannt ist, dass es Biofilme (Matten) in *P. aeruginosa* bildet. Unter Verwendung genetischer Techniken stellte ich fest, dass A506 auch keine wss-Gene verwendet um Matten herzustellen, sondern das pga-Operon, welches in SBW25 selten dazu genutzt wird. Das Fehlen der Fähigkeit zur Mattenbildung in flüssigem Medium durch Pf0-1 ist faszinierend, da es einige der von *P. aeruginosa* bekannten Strukturgene besitzt und Berichten zufolge auf festen Oberflächen Matten bildet. Während die drei mattenbildenden Stämme (SBW25, A506 und PICF7) unterschiedliche Strukturgene ausnutzen, sind die zentralen Regulationswege, welche die Expression der jeweiligen Operons optimieren, gleich geblieben. Dies lässt auf einen modularen Mechanismus schließen, bei dem mehrere Strukturgene sich über die Genome hinweg gegenseitig ersetzen können.

Bisher beschränkte sich unser Verständnis der Genomentwicklung von *P. fluorescens* auf den Stamm SBW25. Mit der Entwicklung genetischer Werkzeuge im Rahmen dieser Arbeit haben wir nun die Möglichkeit, die vergleichenden Studien von *P. fluorescens* zu erweitern. Die Wirksamkeit dieser Werkzeuge wurde durch die Untersuchung natürlicher Veränderungen synonymer Codons sowie die Variation der molekularen Wege, die zur Besiedlung der Luft-Flüssigkeits-Grenzfläche genutzt werden, ausführlich demonstriert.

Abstract

The genetic code is degenerate – 20 proteinogenic amino acids are coded for by 61 functional codons. One might then expect all the codons for an amino acid, called synonymous codons, to be used equally. However, across all domains of life, preferential use of synonymous codons has been observed. This phenomenon is called codon bias. Synonymous substitutions, nucleotide level changes that do not modify the amino acid sequence, were widely viewed to be selectively neutral. Mounting evidence, however, indicates otherwise – synonymous substitutions can induce non-neutral and measurable fitness effects.

Experimental studies with randomised codon usage in genes from both prokaryotes and eukaryotes reveal large-scale fitness effects, ranging from altered growth, global changes in both transcription and translation and protein output and function. The effects of synonymous substitutions become particularly prominent in bacteria, where growth rate is often limited by the speed of translation. However, these observations mainly arise from selectively enriching certain codons, a pattern that is rare in naturally evolved genomes. To investigate the functional implications of naturally diverged patterns of codon usage we have identified an essential and highly expressed gene that exclusively exhibits synonymous differences across three genomes of *Pseudomonas fluorescens* bacteria. These variants were swapped between genomes to test for effects on gene expression and fitness. To swap the synonymous alleles between genomes, I have optimised a scar-free genetic engineering technique that is routinely used for SBW25, to A506 and Pf0-1. The resulting mutants varied from the corresponding wild type strains at *only* one locus – the locus of the gene that was swapped. I proceeded to test for the effects synonymous substitutions on gene expression and fitness and demonstrated that changing established codon usage patterns of a gene in one strain to that of another strain has considerable effects on gene expression (transcription).

Having optimised a technique for manipulating genomes of *P. fluorescens* strains (besides SBW25), I proceeded to examine the genetic basis of one of the most studied and extensively characterised phenotype of *P. fluorescens* – the ability to form mats at the air-liquid interface. First, I compared mat formation phenotype in A506, Pf0-1 and PICF7 to the model strain SBW25; while, SBW25, A506 and PICF7 formed mats, Pf0-1 did not. To identify the genetic routes to mat formation in PICF7, I developed tools for transposon suppressor analysis, which revealed that unlike SBW25, which utilises the cellulose biosynthetic machinery encoded by the *wss* genes to form mats, PICF7 makes use of the *pel*

operon to synthesise the Pel exopolysaccharide. Pel is an exopolymer known to form biofilms (mats) in *P. aeruginosa*. Using genetic engineering techniques I identified that A506 does not use *wss* genes either, but the *pga* operon to produce mats, a pathway seldom utilised in SBW25. The lack of mat forming ability in liquid medium by Pf0-1 is intriguing as it possesses some of the structural genes known from *P. aeruginosa* and is reported to form mats on solid surfaces. While different structural genes are exploited by the three mat forming strains (SBW25, A506 and PICF7), the central regulatory pathways that fine-tune the expression of the operons have remained the same. This is suggestive of a modular mechanism, wherein multiple structural genes can substitute for one another across genomes.

So far, our understanding of *P. fluorescens* genome evolution has been limited to SBW25. With the development of genetic tools within the scope of this thesis, we now have the opportunity to expand on the comparative study of *P. fluorescens* and the efficacy of these tools for the same has been amply demonstrated through the examination of the effects of changing naturally evolved synonymous codons as well as the variation in molecular routes exploited to colonise the air-liquid interface.

Table of Contents

List of Figures

List of Tables

Chapter 1 : Introduction

1.1 Genetic code is degenerate

The four bases – adenine (A), thymine (T), guanine (G) and cytosine (C) – are the building blocks of the universal genetic code. A consecutive set of three of these bases makes a codon, making 64 possible combinations of codons. Since the number of standard amino acids is 20, there is clearly more than one codon available for a given amino acid, thereby making the genetic code degenerate. As can be seen in Figure 1.1, amino acids are coded for by between one and six codons. Interestingly, no amino acids are encoded by five codons, and only one (isoleucine, I) is coded for by three codons.

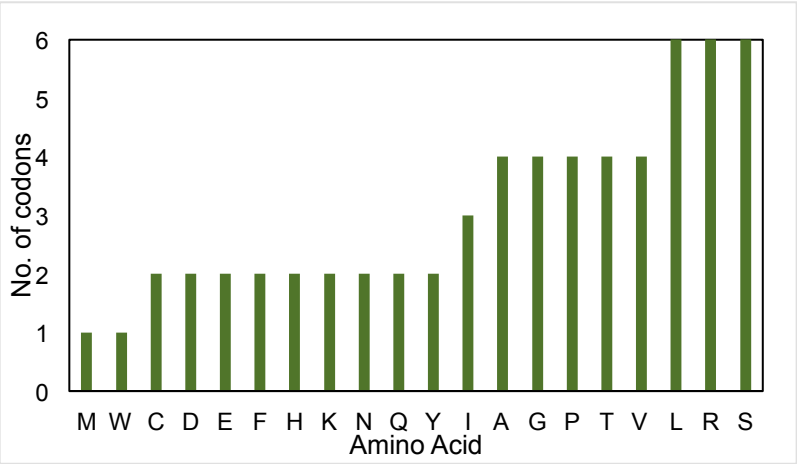


Figure 1.1: Graph showing the number of codons coding for each of the 20 standard amino acids. Almost all the amino acids are coded for by more than one codon; the exceptions are methionine (M) and tryptophan (W).

Multiple codons for an amino acid are called synonymous codons. Synonymous codons are not used equally; some codons are favoured over others, a phenomenon known as codon bias. Although this biased usage is observed in all domains of life (Sharp & Li, 1986; Chamary & Hurst, 2005; Qiu *et al.,* 2011; Goodman, Church, & Kosuri, 2013; Tello, Saavedra, & Spencer, 2013; Trotta, 2013; Chithambaram, Prabhakaran, & Xia, 2014a; Camiolo, Melito, & Porceddu, 2015; Boël *et al*., 2016; Agashe *et al.*, 2016; Narula *et al.*, 2019) there are still some gaps in our understanding of the factors that drive this bias. To

demonstrate the extent of biased usage, consider the example of amino acid leucine. Leucine has a set of six codons – TTG, TTA, CTG, CTA, CTC and CTT. From a naive point of view, one might expect all the codons to be used equally, *i.e.* each codon would be used 1/6th (16.67%) of the time (Figure 1.2A). But from a bioinformatics screen of the genome of *P. fluorescens* SBW25, a skewed usage is observed. In *Pseudomonas fluorescens* SBW25, the CTG codon is used preferentially (57.92%), while the other codons are used much less (Figure 2B).

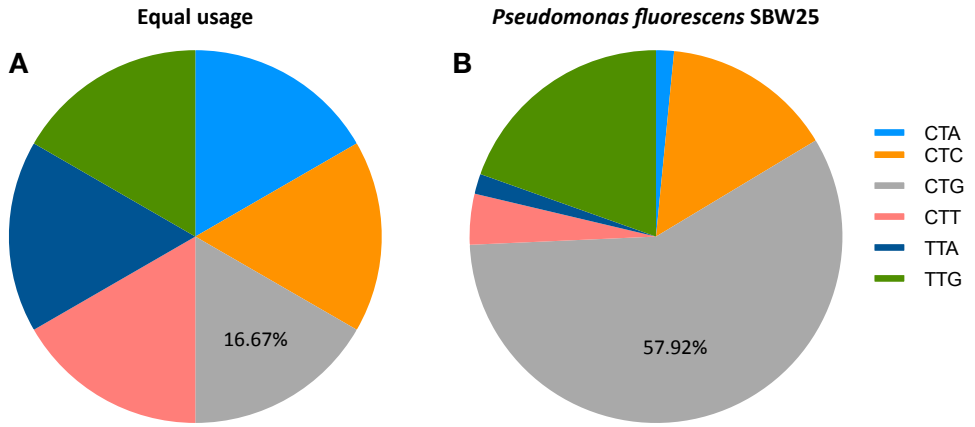


Figure 1.2: Genome-wide leucine codon usage. Leucine has a codon set of six (TTA, TTG, CTT, CTC, CTA, CTG). **A.** If these codons were used equally across a genome, one would expect each codon to be used approximately 1/6 of the time. However, real genomes show markedly biased leucine codon usage. **B.** This chart shows leucine codon usage in the bacterial genome *Pseudomonas fluroescens* SBW25, encoding a total of 231,672 leucine residues.

Codon bias exists not only at the organismal level but also at the level of genes. Understanding the factors that shape codon usage patterns across and within organisms may shed light on the evolutionary trajectory of genomes. There may be a number of factors that drive differences in codon usage patterns and determining the extent of influence each driver may help us understand better the evolution and maintenance of genomes. Not only the architecture of genomes but also the selection pressure that population of genomes respond to, can be defined better with a clearer assessment of drivers of codon biases.

1.1.1 Factors shaping codon bias

Improvements in and decreasing cost of sequencing has facilitated a thorough examination of genomes. Comparatively smaller size of bacterial genomes has made it

possible to sequence these in greater numbers as compared to eukaryotic genomes. The progress in the ability to study genomes is revealing patterns that were hitherto unidentified. One such characteristic genomic pattern is codon usage bias. However, as mentioned earlier, our understanding of the factors that have led to codon usage bias and help sustain it, are still lacking. A growing body of studies carried out in the last two decades, provides a glimpse into an array of probable factors that might have shaped codon biases in individual organisms.

1.1.1.1 Transfer RNA (tRNA) and its cytoplasmic pool

The intracellular transfer RNA (tRNA) pool has been implicated as a major contributor causing biased codon usage. It has been demonstrated that for highly expressed genes in *Escherichia coli*, codon use is strongly correlated by tRNA availability (Ikemura, 1981). The correlation between codon use and tRNA genes, however, becomes weaker for genes that have more regulatory functions, as these genes, the author suggests have low levels of protein production. Similar dependency pattern of codon use and tRNA genes have also been reported for the highly expressed genes in yeast (Ikemura, 1982). Nuclear genes of *Saccharomyces cerevisiae* were isolated by two-dimensional polyacrylamide gel electrophoresis and relative abundances of the tRNA genes were determined. These tRNA genes, like those in *E. coli* were found to be strongly correlated with codon preference. Moreover, codon usage is organism specific as has been mentioned in the previous section and it is believed to be so, as the use of a codon might depend on the available tRNAs (Ikemura, 1982). Binding with tRNAs depends on certain criteria – firstly, it could be dictated by the thiolation of uridine at the wobble position, which in turn would select for codons ending in A as compared to G-ending codons. Secondly, codons with A/U at the first and second positions would undergo optimal binding with anticodons when the third codon letter is C. Thirdly, presence of Inosine (I) at the wobble position of an anticodon leads to preference for C/U ending codons, thereby avoiding A. Usually at the first anticodon position U is favoured, as this base or a modification thereof can bind to all synonymous codons (Rocha, Eduardo, P.C., 2004). This factor becomes highly significant for fast growing bacteria with smaller genomes as these can then have a tRNA pool that is enriched with U as the first base of the anticodon. Consequently, the codon usage in such a bacterial system would be shaped accordingly.

Codon translation takes place in the A-site of the ribosome by tRNAs, therefore optimality or rarity of a codon can be defined in terms of the available tRNAs in the

cytoplasmic pool (Hanson & Coller, 2017). This cytoplasmic pool of tRNA can be referred to as the tRNA pool. These findings can be corroborated through bioinformatics studies as well, which have demonstrated that the tRNA pool can shape codon bias (Sabi & Tuller, 2014).

Codon usage bias correlates with tRNA usage but the extent to which this correlation holds may vary between organisms; prokaryotic genomes showed a comparatively weaker correlation in comparison to eukaryotic genomes. Moreover, the ratio of cognate to near-cognate tRNAs in the pool could also favour certain codons of some amino acids (Hanson & Coller, 2017). Some theoretical studies have tried to outline the factors that could largely contribute to codon usage bias. One such study demonstrated that in the 102 bacterial species that were studied, the tRNA genes in fast-growing bacteria were abundant but less diverse (Rocha, Eduardo, P.C., 2004). The smaller sets of tRNAs optimises the translational machinery to higher growth rates; limited diversity of tRNAs would allow ribosomes to be almost saturated even with the small pool of tRNAs. The smaller pool would in turn drive the co-evolution of the more preferred codons. The most preferred codons that evolve would in turn drive the selection of the cognate tRNAs, thereby leading to coevolution of codons and tRNAs. Additionally, set of tRNA genes have also been known to change in response to dramatic changes in codon composition (McDonald *et al.*, 2015). Bacteria often exchange genetic material through lateral gene transfer. When the amount of such genetic material is large and codon usage of the foreign DNA differs from the host, then the host may increase the number of tRNA genes to cope with influx of DNA with "altered" usage. This study reports, on the basis of correlation between tRNA genes and altered state codon usage in *E. coli* and *Salmonella* that tRNA genes undergo changes as well, in a manner that is proportionate to the changes in codon usage pattern.

Experimental studies have demonstrated that the choice of codons in an organism depends upon the availability of corresponding tRNAs in the tRNA pool (Dittmar, Goodenbour, & Pan, 2006). Expression of tRNA genes varies between human tissues and due to the central role that tRNAs have in protein translation, this study reports that the differences in tRNA abundances indicate translational control. Such variation may also reflect the changes in translation that occur with changes in developmental stages of an organism. Another study showed that in *S. cerevisiae* there was a higher propensity of codons that share the same tRNA to appear in closer proximity (Cannarozzi *et al*., 2010). Therefore the translation rate and the choice and order of codons in a gene were determined by the ability of tRNAs to translate multiple codons. This allows for faster and more efficient translation and might be a major determinant for choice of codon and the order of codons. Theoretical work also lends support to the role of tRNA availability and rate of codon translation (Rudorf &

Lipowsky, 2015). Availability of mature tRNAs, ternary complexes (aminoacyl-tRNA+GTP+ribosome), and charging of discharged tRNA, all contribute to codon elongation rate. Together, all three aforementioned factors contribute to the local concentration of mature tRNA that is able to decode the codon of interest. The elongation rate therefore, becomes specific to the identity of not only the codon of interest, but also the neighbouring codons. Accurate and efficient decoding of codons would depend on the local concentration of cognate tRNAs. Otherwise, the translational output may be affected, as incorporation of near-cognate or non-cognate aminoacyl-tRNAs would lead to misreading errors. This is turn would also increase the rate of codon elongation and thereby protein synthesis. Increased translation times would disadvantage organisms with fairly short generation times, such as bacteria.

Not only the tRNA pool but also chemical modifications of a tRNA can define which codons may be used in an organism. Cognate tRNAs can recognise and bind to the corresponding codons by Watson-Crick base pairing. However, certain chemical modification on these tRNAs allow wobble pairing (Blanchet *et al.*, 2018). Such post-transcriptional modifications are mainly observed in the anticodon loop of the tRNA – either in the anticodon itself or in the bases in the immediate vicinity of the anticodon. In the case of lysine, two codons are available, AAA and AAG. The same tRNA, tRNA-Lys-UUU, translates both codons. In most bacterial species, tRNA-Lys-UUU is the only tRNA available in the pool and it can decode both the cognate AAA and the near-cognate AAG due to the modifications in the anticodon of tRNA-Lys-UUU. A separate study has also demonstrated that in addition to chemical modifications in the tRNA, the codon-anticodon pairing rate and specificity can be a governing factor in selecting for a particular synonymous codon (Ran & Higgs, 2010).

1.1.1.2 Messenger RNA (mRNA) stability and toxicity

Among other factors presumed to shape codon biases, structure and stability of mRNA is considered to an important contributor (Goodman *et al.*, 2013). Positioning of optimal and rare codons in an mRNA transcript can affect its folding, which in turn will define its stability. This study focusing on the effect of N-terminal bases of an mRNA showed that replacing the favoured codons with rare substitutes, increased translation efficiency. This increase has been attributed to secondary structure formation (disruption) at the N-terminal ends. A study with yeasts describes how rare codons at the 5' end of the mRNA transcript increase translational efficiency by providing a slow ramp (Tuller *et al.*, 2010a). The ramp is

hypothesised to provide a measure of translational control by regulating the initiation rate. This way there is less probability of ribosome crowding when the elongation phase of translation takes place. Ribosome profiling of yeast genes revealed that about 50 bases in the 5' end of mRNAs have rare codons (Hanson & Coller, 2017). These codons slow down elongation, thereby allowing ribosomes downstream to space out evenly and not cause jams along the mRNA during elongation. Additionally, it has been demonstrated in budding yeast that stretches of non-optimal codons can recruit the DEAD-box helicase that activates Dhh1 (Radhakrishnan *et al.*, 2016). Activated Dhh1 initiates mRNA decay. Therefore, it becomes important to control the number as well as the placement of optimal and rare codons in the mRNA transcript. These factors help in maintaining the stability of the mRNA as well as the secondary structures.

Synonymous substitutions have also been described to have toxic effects (Mittal *et al.*, 2018). The authors show that some synonymous variants of GFP when over expressed in *E. coli*, were detrimental to cell growth. A stretch of 132 nucleotides in the one of the "toxic" variants appeared to be responsible for the toxic effects. Although the mechanism of toxicity was not well understood but it seemed that the mRNA itself may be responsible for the toxicity. The toxic effects of the mRNA were attributed to the combination of synonymous codons that made up the mRNA fragment. Therefore, not only the structure of mRNA but potential toxic effects of certain interactions between codons can lead to some codons being preferred over others. What is essential to understand is that although individual stable and unstable mRNA transcripts may vary in codon identities, the transcripts that are stable, usually have an overwhelming number of optimal codons while the unstable transcripts are enriched in non-optimal codons (Presnyak *et al.*, 2015). This work in yeast describes the correlation between enrichment of optimal codons and stability of mRNA transcripts. It was observed that some codons were preferentially enriched in the stable mRNAs, which corresponded to one third of all codons. The unstable mRNA transcripts on the other hand had a majority of non-optimal codons that roughly equalled two-thirds of the all protein coding codons.

1.1.1.3 Protein folding

Synonymous substitutions do not affect the primary sequence of proteins but are speculated to have significant effects on protein folding. Protein folding has been demonstrated to be a significant contributor of codon bias, especially in bacterial species

(Cortazzo *et al.*, 2002). Using *E. coli* as a model system, it has been demonstrated that replacing the rare codons with the optimal codons not only increased the rate of protein production, but also the rate of protein misfolding. Co-translational protein folding is especially common in bacteria, which makes it imperative that the placement and choice of codon be such that protein production is maximised not only in terms of the enhanced rate but only also accuracy with respect to the ultimate structure. Protein function, in addition to protein folding can be affected by changing synonymous codons. Changed codon use of the *dper* protein in *Drosophila*, affected not only the folding of the circadian clock regulation protein, but also the function of the protein (Fu *et al.*, 2016). This protein is enriched in non-optimal codons at N-terminal region as well as in the unstructured regions of the protein. Substitution of these non-optimal codons with optimal codons resulted in altered structure of dPER; changes in structure were determined by trypsin digestion – the wild type protein was more prone to digestion while the optimised dPER was more resistant. Moreover, temperature mediated precipitation demonstrated that optimised dPER precipitation rate was higher than the wild type dPER. Additionally, circadian rhythms were impaired in the flies that carried the optimised dPER. The flies with wild type dPER displayed regular circadian phenotype (bimodal morning and evening spikes in activity), while the flies transgenic for single copy of optimised dPER exhibited lower amplitudes in spikes and transgenic flies for two copies of optimised dPER exhibited severely reduced phenotypes. Therefore *in vivo*, codon optimisation may lead to changes not only in structure but also impaired function and hence codon use might evolve to optimise functionality of proteins, but not necessarily to incorporate optimal codons. Non-optimal codons can have important implications to protein folding and hence function of the protein.

1.1.1.4 Environmental factors

The above factors are intrinsic drivers of codon bias. In other words, the conditions that are present within an organism, are shaping its codon preference. However, all of these intrinsic factors evolve in response to environmental changes, which in turn might drive codon preference. Evolution of these intrinsic factors in response to changing environment is greatly perceived in the case of host pathogen interaction. The set of tRNA genes of an organism can change dramatically in response to influx of large amounts of foreign DNA (McDonald *et al.*, 2015). Horizontal gene transfer is a phenomenon that widely occurs in prokaryotes. In the event that horizontally transferred DNA is of a considerable amount, then positive correlation between codon usage of incoming DNA and proportional enrichment of

corresponding tRNA genes is observed. Codon usage biases of viral genomes were observed to have been shaped by their interactions with the respective hosts. For instance, in the study by (Wong *et al.*, 2010), the codon usage bias of the human and avian influenza viruses observed to have diverged since these were specialised to the human and avian hosts respectively. The synonymous codon preference in both virus and its host were found to be similar overall. This is insightful since influenza viruses rely heavily on host translational machinery and hence host adaptation would shape the codon usage pattern of these viruses. Similar observations have been made for viral codon usage adaptation to host with respect to honey bees and honey bee associated viruses as well (Chantawannakul & Cutler, 2008). From the pattern of convergence of honey bee virus codon use to that of honey bees, it is apparent that such optimization allows for successful establishment and maintenance within the host.

Other environmental conditions, such as nutrient availability, may also influence the codon preferences of genes within an organism. Regulatory or stress response genes that are required only under certain conditions of growth and are not necessarily designed for maximal protein output may have differential codon use patterns as compared to highly expressed genes. For instance, amino acid starved conditions can lead to altered charging of tRNAs which may cause the otherwise rare codons to become optimal (Dittmar *et al.*, 2005). Usually, under amino acid restrictive conditions, tRNA isoacceptors corresponding to rare codons get preferentially charged; these classes of tRNAs are called starvation-insensitive tRNAs. Amino acid limitation is a condition of stress and hence to keep in tune with the availability of tRNAs, a number of stress response genes or genes with regulatory roles under conditions of amino acid starvation, are preferentially enriched in the non-optimal codons. This would ensure that these genes are translated during reduced amino acid availability, commensurate with the presence of starvation insensitive tRNAs. Among other similar studies on the effects of nutrient availability on genome evolution, one particular study describes through modelling, that nitrogen limitation could influence the choice for synonymous codons in both bacterial and eukaryotic parasites (Seward & Kelly, 2016).

1.1.2 Measuring codon usage bias

Quantifying the extent of codon usage bias has been the focus of much work (Sharp & Li, 1987; Berg, 1997; Carbone, Zinovyev, & Kepes, 2003; Reis, Savva, & Wernisch, 2004; (Sabi & Tuller, 2014; Sabi, Daniel & Tuller, 2016). The Codon Adaptation Index (CAI) and tRNA Adaptation Index (tAI) are two of the methods that are used extensively to measure the bias in

codon usage. Both these indices provide a correlation between codon content and the translational output of genes to determine which genes are most optimised to the codon usage profile of a genome.

1.1.2.1 Codon Adaptation Index (CAI)

There has been increasing evidence that the highly expressed genes show the most codon bias (Sharp & Li, 1987). Thus, genes that are composed of the most frequent codons are expected to provide an extreme example of an organisms' particular codon bias. In order to study the effects of biased codon usage on translational speed and hence protein expression, a parameter – the codon adaptation index, or CAI – is used (Carbone *et al*., 2003). CAI assigns a numerical value to each gene of a genome based on the translational bias on gene expression. The CAI values – and, by correlation, the expected expression level – of each gene can then be compared with those of a representative pool of genes from the organism. This method limits the representative pool to the highly expressed genes; organisms for which knowledge about gene functions is scarce are not currently used for computation of CAI. In 2003 Carbone *et al*. proposed an improvement of the CAI whereby genes are screened for the highest values in the CAI scale; the authors used CAI as a measure to detect the dominant genes and codon biases (Carbone *et al*., 2003).

1.1.2.2 tRNA Adaptation Index (tAI)

To understand the correlations between codon usage bias of genes and their expression levels, tAI is being increasingly used (Tuller *et al.*, 2010b; Tuller *et al.*, 2010a; Wong *et al.*, 2010; Tello *et al*., 2013; Presnyak *et al.*, 2015; Quax *et al.*, 2015; Tello, Avalos, & Orellana, 2018). This index measures tRNA usage by coding sequences (Reis *et al.*, 2004). There remains a strong positive correlation between tRNA gene copy numbers and tRNA abundances, as well as codon choices across some genomes (Ikemura, 1999; Duret, 2000). This measure, therefore, uses tRNA usage of a gene as a proxy for determining translational selection. Translational selection is pursuant to how well a particular gene is adapted to the cellular tRNA pool. An absolute adaptiveness value is determined for each codon as a function of the number of tRNA isoacceptors that can identify the codon in question, the tRNA gene copy number and the selective constraint of codon-anticodon pairing efficiency. This is followed by computation of the relative adpativeness value of each codon; this is obtained by taking a ratio of the absolute adaptiveness value of a codon to the maximum

possible value of absolute adaptiveness. Finally, the tAI of a gene is computed as a geometric mean of all relative adaptiveness values of the codons in the gene.

1.1.3 Non neutral effects of synonymous codons

1.1.3.1 Synonymous mutations exhibit equivalent degree of fitness effects as non-synonymous mutations

The previous section clearly indicates that codon usage patterns observed in different genomes have a functional importance. It follows then, that altering these intrinsic patterns may have an effect on the functional output of genomes. Since synonymous codon substitutions do not alter the protein sequence, it was thought that these substitutions or mutations are largely neutral or undergo weak selection. However, a mounting body of evidence suggests otherwise. In particular, a study of fitness effects of synonymous mutations in ribosomal structural proteins of *Salmonella* shows a distribution similar to that of non-synonymous mutations (Lind, Berg, & Andersson, 2010). The hypothesis suggested in this work is that changes in mRNA structure and/or stability may be causal to the large fitness effects of synonymous substitutions. In another study, that investigates the emergence of mutations in response to selective pressure, the authors have found that out of three mutations that arose in an isogenic population of *P. fluorescens* SBW25, two were synonymous mutations (Bailey *et al.*, 2014). The mutations arose in a glucose transporter gene (*gtsB*) in the population under nutrient limiting conditions (minimal medium), wherein glucose was the sole carbon source. Although, the non-synonymous mutation arose first, it was replaced within the next 200 generations by the first synonymous mutation, which was further replaced in another 200 generations by the second synonymous substitution. Through competitive fitness it was confirmed that both synonymous substitutions provided a fitness advantage with respect to the ancestor and the distribution of fitness effects of these synonymous substitutions were within the range of that provided by non-synonymous substitutions. Therefore, synonymous mutations can provide a fitness advantage that rivals that of non-synonymous mutations and can spontaneously arise under selective pressure. The compelling aspect of the study however, is the observation that the first synonymous substitution (A15A) that spontaneously emerged in the laboratory population was the most common mutation detected in *gtsB* gene of closely related strains of Pseudomonads, many of which are environmental isolates. This indicates that even in natural environments non-neutral effects of synonymous mutations might be perceived.

1.1.3.2 Synonymous mutations cause ribosomal stalling

Fitness effects of changing synonymous codons have also been demonstrated in organisms other than *E.coli, Salmonella* and yeast, thereby showing expansive fitness consequences of synonymous substitutions (Agashe *et al.*, 2013). Six synonymous variants of an integral enzyme producing gene, *fae* of *Methylobacterium etroquens* exhibited a range of fitness effects. The six synonymous variants had different degrees of synonymous substitutions – all frequent (AF), all rare (AR), conserved sites rare (CO), variable sites rare (VA), randomly picked (RN, 50% rare) and active sites rare (AC). FAE is crucial for growth on methanol and *fae* synonymous variant strains, including the wild type (WT) strain were grown and competed on methanol containing medium. WT had the highest growth rate, while the ones that had the maximum proportion of frequent codons – AF and AC – had the lowest growth rate. The strains that had the same proportion of rare codons – CO, VA and RN – all with 50% rare codons, did not exhibit similar growth rates; CO had higher growth rate among all three. The combination of rare codons at conserved residues (CO) was less detrimental than rare codons at variable residues (VA). In addition to growth, mRNA, protein quantities as well enzyme activity were determined. Surprisingly, the constructs with larger proportion of rarer codons did not necessarily exhibit the least amount of mRNA or protein or enzyme activity. The range of differences in mRNA and protein quantities as well as enzyme activity were found to be in fact correlated to internal Shine-Dalgarno sequences. The synonymous codon substitutions generated multiple internal Shine-Dalgarno sequences, which caused increased ribosomal pausing and consequently less protein production.

1.1.3.3 Synonymous substitutions alter proteome wide translational efficiency

To study the effect of manipulating the prevalence of a particular codon, frequent codons of arginine were replaced with the rare CGG codon in a subgroup of highly expressed genes of *E. coli* (Frumkin *et al.*, 2018). This codon was selected since it is the only arginine codon in *E. coli* that has single-copy tRNA that cannot translate other arginine codons. Translation efficiencies of each of the recoded genes were determined by normalisation of protein output to its corresponding mRNA output. Only one of the recoded genes was found to have reduced efficiencies, which indicates that possibly the recoding caused an increased translational demand for the CGG codon. To test this possibility, translation efficiencies of non-recoded genes were determined, of which 455 genes exhibited elevated translational efficiencies while 566 genes displayed reduced translational efficiencies, as compared to the recoded strain.

Remarkably, the genes that demonstrated lower translational efficiencies were enriched in CGG codons (five or more occurrences), while those that had elevated translational efficiencies were enriched in the CGU, CGC and CGA codons. To identify the cause of the reduction in translational efficiency, translation initiation rates were determined with the use of Ribosome Binding Site calculator; the initiation rates were found to be higher for the recoded genes. The proteome wide alterations, however, could be ameliorated by augmentation of the tRNA (anticodon) supply for CGG by mutation of one of the four copies of tRNAACG to CCG. As is evident, synonymous substitutions might perturb not only local changes in mRNA structure and/or stability and translation rates, but also output across the proteome. However, the perturbations can be corrected by proportionate changes in the tRNA pool, further lending support to the role of tRNA abundances in influencing codon usage.

Altogether, all of the above mentioned studies indicate that synonymous mutations have non-neutral effects. Not only do synonymous substitutions occur in laboratory conditions, but also there is evidence that such substitutions may occur in nature (Bailey *et al.*, 2014). Synonymous substitutions are capable of causing fitness effects that are equivalent to those of non-synonymous substitutions and can induce fitness effects by modulating protein expression and translational efficiency (Frumkin *et al.,* 2018), altered mRNA structure and stability (Bailey *et al*., 2014) and ribosome stalling (Agashe *et al.*, 2013), among others.

1.2 Bacterial experimental systems

Bacteria are relatively simple organisms, most of which have short mean doubling times and many of which have single chromosomes. Bacteria with single chromosomes permit comparative ease of genome manipulation, thereby providing a platform to observe the effects of the targeted manipulations. Bacteria usually undergo asexual reproduction; this provides an opportunity to replicate experiments with isogenic populations. Additionally, bacterial cultures can be stored, almost indefinitely at -80ºC and revived from a state of suspended animation. These factors make bacteria ideal experimental systems for genomic manipulation.

1.2.1 The *Pseudomonas fluorescens* experimental system

Pseudomonas fluorescens is a divergent bacterial species that currently encompasses about 213 genome-sequenced strains isolated from plant and soil environments around the world. Our laboratory currently uses four *P. fluorescens* strains: SBW25 (isolated from a sugarbeet leaf in the UK; (Rainey & Bailey, 1996)), PICF7 (from the roots of nursery-bred olive plants; (Mercado-Blanco *et al.*, 2004)), A506 (from a pear leaf East Coast USA; (Wilson & Lindow, 1993)) and Pf0-1 (from loam soil West Coast USA; (Compeau *et al.*, 1988)). These four strains show different degrees of evolutionary relatedness, SBW25 and PICF7 being more closely related than the others (Figure 1.3). The codon use will therefore have diverged between the four strains, with SBW25 and PICF7 showing most similar codon use patterns. In other words, differences in codon use would reflect how closely related the four strains are.

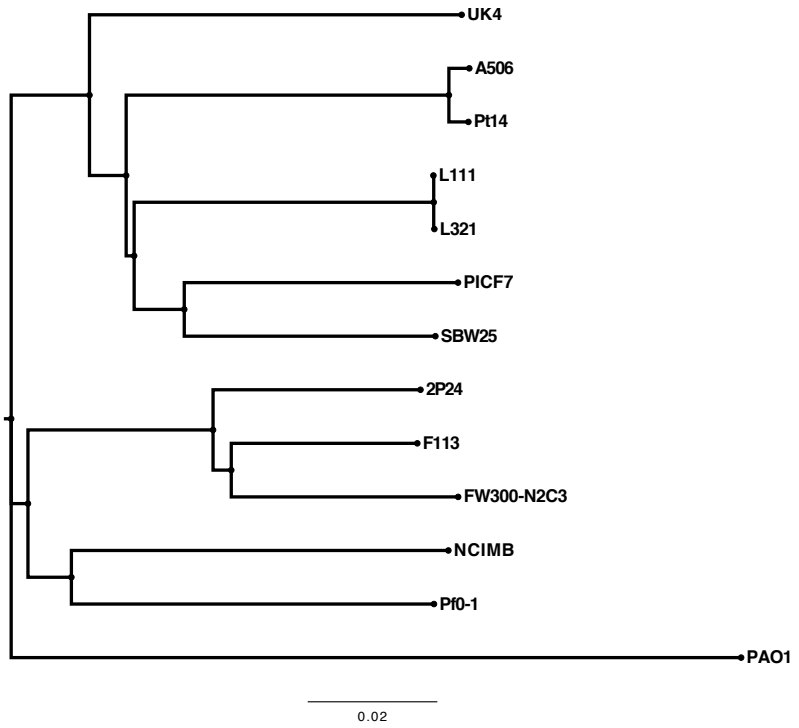


Figure 1.3: Phylogenetic tree for strains of *Pseudomonas fluorescens*. Phylogenetic tree for strains of *Pseudomonas fluorescens* constructed using the Neighbor, a PHYLIP distance joining program that uses Saitou and Nei's "Neighbor Joining Method," and UPGMA. As can be seen, SBW25 and PICF7 are most closely related. The entire genome was used for constructing the tree.

Pseudomonas as a genus is well known for secretion of exopolysaccharides that form the structural basis of mats (biofilms) (Rainey & Travisano, 1998, Spiers *et al.*, 2002; Friedman & Kolter, 2004b; Chang *et al*., 2007; Monds *et al.*, 2007; Ueda & Wood, 2009; Ueda & Saneoka, 2014; Cárcamo-Oyarce, *et al.*, 2015; Ghods *et al.*, 2015; Díaz-Salazar *et al*., 2017; Farias, Olmedilla, & Gallegos, 2019). This is a complex group of organisms that is frequently associated with both plants and animals and range between commensals and pathogens (Silby *et al.*, 2011).

1.2.2 Mat formation in the genus *Pseudomonas*

Mat formation or biofilm formation is a phenotypic attribute that has been reported in a number of *Pseudomonas* species (Spiers *et al.*, 2003; Friedman & Kolter, 2004b; Chang *et al*.,

2007; Ueda & Wood, 2009; Ueda & Saneoka, 2014; Cárcamo-Oyarce *et al*., 2015; Ghods *et al*., 2015; Díaz-Salazar *et al.*, 2017; Farias *et al*., 2019). While cellulose is the structural component of mats in SBW25 (Spiers *et al*., 2002), a number of other exopolymeric substances (EPS) are also described to be contributors to mats in other species of *Pseudomonas*. For instance, Pel and Psl, two EPSs that are products of the *pel* and *psl* loci, respectively, are known to be important contributors to mat formation in *P. aeruginosa* (Friedman & Kolter, 2004b; Colvin *et al*., 2012; Jones & Wozniak, 2017; Marmont *et al.*, 2017a). In the SBW25 experimental system, it has been observed that in the absence of the *wss* locus (cellulose biosynthesis genes), other EPS synthesising genes, such as PGA (product of the *pga* operon), form mats (Lind, Farr, & Rainey, 2017). In the same study, another instance of mat formation was reported when loss-of-function mutations in *nlpD* gene were found (cell division gene). In *P. fluorescens* Pf0-1, on the other hand, the *lap* locus was implicated in mat formation (Monds *et al*., 2007; Newell, Monds, & O'Toole, 2009; Newell *et al.*, 2011a; Boyd *et al.*, 2012). It is therefore evident, that different organisms may have different approaches to forming mats. Similar phenotypic solutions may be obtained by use of dissimilar adhesion materials. But what is to be noted is that many of these EPSs are generated in response to the activity of the two-component signal transduction pathways, that act to regulate levels of bis-(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP), a cyclic nucleoside, which targets one or more of the EPS biosynthetic genes. This suggests modularity in realisation of the mat phenotype – various structural genes respond to the same regulation (c-di-GMP levels) and might potentially be substituted with one another.

C-di-GMP was first discovered to have a role in activating the cellulose biosynthetic genes of *Gluconacetobacter xylinum* (Ross *et al*., 1987) and has since been acknowledged as an extensively used bacterial secondary messenger that activates cellulose biosynthetic enzymes in various bacterial species (Goymer *et al*., 2006; Malone *et al*., 2007; Römling, Galperin, & Gomelsky, 2013; Zorraquino *et al.*, 2013; Morgan, McNamara, & Zimmer, 2014; Petersen, Mills, & Miller, 2019). C-di-GMP is synthesised from two molecules of GTP through the activity of diguanylate cyclases (DGCs), and is degraded to pGpG by phosphodiesterases (PDEs). The DGC and PDE activities are associated with the GGDEF (Chan *et al.*, 2004; Paul *et al*., 2004) and EAL or HD-GYP domains (Schmidt, Ryjenkov, & Gomelsky, 2005; Hengge, 2009), respectively. The GGDEF and EAL/HD-GYP domains derive their names from the amino acid motifs that are characteristic of these domains. These amino acid motifs are crucial to the enzymatic activity of the corresponding proteins.

DGC activity and hence c-di-GMP levels are modulated by a two-component signal transduction system (Nixon, Ronson, & Ausubel, 1986). A two-component signal

transduction system consists of a histidine kinase (HK) and a response regulator (RR). The HK senses an environmental signal and transfers an inorganic phosphate to the RR (such as a DGC). This results in activation of RR and production of elevated levels of c-di-GMP. Figure 1.4 provides a summary of the activities of all the enzymes – HK, RR (DGC) and PDE – that regulate c-di-GMP levels in the context of *wss* genes.

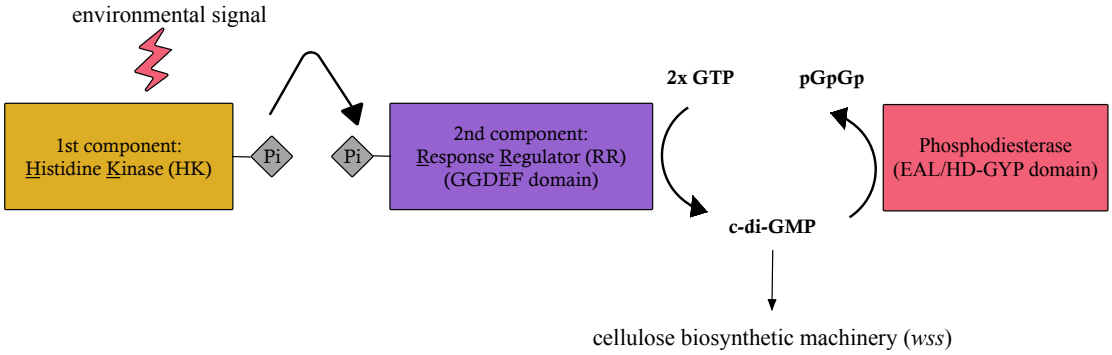


Figure 1.4: Summary of the pathways that regulate c-di-GMP levels in a bacterial cell. C-di-GMP levels are increased by the concerted activity of histidine kinases (HK) and response regulators (RR), when an environmental signal is relayed as an activation signal from the former to the latter. C-di-GMP levels are regulated by the antagonistic activity of diguanylate cyclases (DGCs, also the RR) and phosphodiesterases (PDEs), which produce and degrade c-di-GMP, respectively.

1.2.3 Mat formation in *Pseudomonas fluorescens* SBW25

In the context of *P. fluorescens,* mat formation has been extensively studied in SBW25. The structural component of the mats and the pathways that lead to formation of mat in SBW25 are known. SBW25 as an experimental system has been used largely to understand the mechanistic details of mat formation at the air-liquid interface. The molecular pathways involved in mat formation in SBW25 have been explored and a mounting body of information exists for the same. Similar phenotypes can be expected in the other related strains of *P. fluorescens* – PICF7, A506 and Pf0-1 – as these are, to a greater (or lesser) degree evolutionarily related to SBW25. An understanding of the pathways implicated in mat formation in SBW25, will aid in identifying the routes to mat formation in the other strains of *P. fluorescens* as well. The following sections describe the observation (mat formation) and outline the loci that participate in activating the genes that finally generate mats at the air-liquid interface.

1.2.3.1 SBW25 forms mats and diversifies in static microcosms

SBW25 produces smooth colonies when grown on agar plates of King's medium B (KB). When seeded into liquid culture of KB in a glass vial (microcosm) and grown under static conditions for 72 hours, the smooth colonies evolve to form variant genotypes. Static conditions generate a heterogeneous environment within the microcosm and the variant genotypes appear to prefer specific niches of this heterogeneous environment. These variant genotypes produce distinct colony morphologies when grown on KB agar plates – smooth (SM) that colonise the broth, wrinkly spreader (WS) that colonies the air-liquid interface and fuzzy spreader (FS) that form weak rafts at the air-liquid interface which later collapse to the bottom of the microcosm, respectively (Figure 1.5; (Rainey & Travisano, 1998; Ferguson, Bertels, & Rainey, 2013).

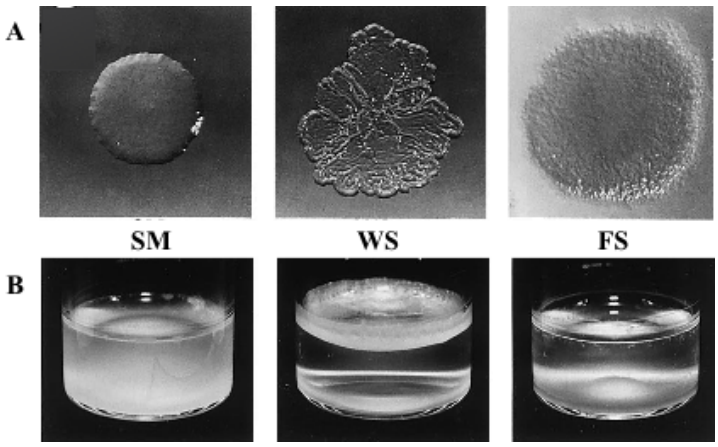


Figure 1.5: Differences in niche adaptation and colony morphology in *P. fluorescens* SBW25. Following propagation for 72 hours in a spatially heterogeneous environment, three distinct colony morphotypes are visible – smooth (SM), wrinkly spreader (WS) and fuzzy spreader (FS). Each of the colony morphs specialises in a particular niche when cultured in KB microcosms and form characteristic colonies when grown on KB plates. Adapted from Rainey and Travisano (1998)

1.2.3.2 Wrinkly spreader phenotype maps to the wss locus

The phenotype of each of the niche specialists isolated from an SBW25 static microcosm can be distinguished at the level of colony morphology, which provides the opportunity to use assays to identify the genetic basis of these colony morphotypes. The colonies implicated in the formation of mats at the air-liquid interface in SBW25 are the wrinkly spreaders (WS; Figure 1.4). Transposon suppressor analysis of the WS colonies revealed insertions in two key

loci – *wss* and *wsp* (Spiers *et al*., 2002). The transposon mutants were found to have lost the ability to colonise the air-liquid interface as well as form WS.

As mentioned above, disruption of the *wss* genes leads to loss of the wrinkly spreader phenotype and the ability to colonise the air-liquid interface. The *wss* (wrinkly spreader structural) genes form a ten gene operon – *wssA-J* (Figure 1.6). *In silico* examination of these genes revealed that *wss* genes share a great degree of homology with the cellulose synthesis genes of *G. xylinum* (Spiers *et al*., 2002). WssA was found to be similar to both YhjQ (from *E. coli*) as well as MinD, cell division and cell polarity determining proteins, respectively. This suggests that WssA might play a role in localisation of the cellulose synthesis machinery. Likewise, WssJ also has a possible localisation role as it was found to be similar to WssA. WssB, WssC and WssE share homology with the cellulose synthase complex, while WssD shares homology with the cellulase gene of *G. xylinum*, which suggests that these four genes might collectively form the cellulose synthesis machinery in SBW25. Lastly, WssF-WssI were found to be similar to the acetylation genes of *P. aeruginosa* PAO1 and hence is predicted to participate in cellulose acetylation. (Spiers et al., 2003). Together, these genes produce an acetylated cellulosic polymer (ACP), which has been demonstrated to be the structural basis of SBW25 mats by promoting adhesion between WS cells at the air-liquid interface (Spiers et al., 2002; Spiers et al., 2003).

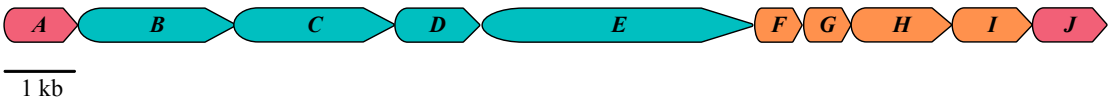


Figure 1.6: Representation of the *wss* locus. The genes in pink – *wssA* and *wssJ* – predicted to have localisation roles. WssBCDE (cyan) form the cellulose synthesis complex, while WssFGHI participate in cellulose modification.

1.2.3.3 Different molecular pathways can activate the wss genes

Transposon suppressor analysis of the WS colonies was shown to result in insertions in both the *wss* and *wsp* genes (Spiers *et al*., 2002). The *wsp* (wrinkly spreader phenotype) locus is highly conserved in all pseudomonads, and spans nearly 8.5 kb and consists of seven genes – *wspA-F* and *wspR* (Bantinaki *et al*., 2007). Wsp proteins share homology with the Che proteins of *E. coli*; the Che proteins play a role in chemotaxis (Parkinson, 2003). Given the shared homology between the Wsp and Che proteins, the following model for Wsp has been suggested (Bantinaki *et al*., 2007). The Wsp proteins act in concert to produce c-di-GMP,

when a methylated WspA relays a signal to WspE, which in turn is activated by phosphorylation. WspE can phosphorylate either WspF or WspR; phosphorylation of WspF – a methylesterase – leads to deactivation of WspA (removal of methanol), while phosphorylation of WspR – response regulator with GGDEF domain – leads to c-di-GMP production. WspC (methyltransferase) methylates WspA, while WspB and WspD are scaffold proteins.

Having identified the *wsp* genes and their role in inducing colonisation of the air-liquid interface, a strategy was devised to identify if additional routes might be involved in mat formation (McDonald *et al.*, 2009). The first step was to determine if mats are formed by an SBW25 strain derived from the deletion of the *wsp* locus - SBW25Δ*wsp*. The WS isolated from the mat forming SBW25Δ*wsp* strain was subjected to transposon suppressor analysis that revealed insertions in another locus – *aws* (alternate wrinkly spreader; (McDonald *et al.*, 2009)). The *aws* locus consists of three genes, *awsX, awsR* and *awsO*. The suggested model for Aws pathway leading to c-di-GMP synthesis involves signal reception and transduction by AwsO, which is predicted to be associated to the outer membrane due to presence of the OmpA (outer membrane porin) domain. The signal is relayed to AwsX, suggested periplasmic protein on account of the presence of a proteolytic cleavage site. Deletion of *awsX* revealed that this gene is also the negative regulator of *awsR*. Finally, AwsR that contains both the histidine kinase and response regulator (DGC) associated domains (HAMP and GGDEF, respectively), synthesises c-di-GMP. The GGDEF domain additionally harbours an R**D motif, suggestive of allosteric binding between AwsR and c-di-GMP.

Another independent mutational route to wrinkly spreader phenotype (and mats) was identified when SBW25 strain with deletions of *wsp* and *aws* was subjected to transposon suppressor analysis (McDonald *et al.*, 2009). The locus identified was *mws* (mike's wrinkly spreader) and it comprises of a single gene, *mwsR*. It produces a single large protein with multiple domains – a transmembrane domain for signal reception, two signal transduction domains for signal relay, one GGDEF domain that synthesises c-di-GMP and an EAL domain that degrades c-di-GMP.

Given the each of these mutational routes have an integrated negative regulation system, the obvious course to keep the respective paths constitutively functional is to cause loss-of-function mutations in the negative regulators – *wspF, awsX* and in the EAL domain of *mwsR*. Indeed such mutations have been observed and reported (McDonald *et al*., 2009).

1.3 Insights into codon bias evolution

1.3.1 Are synonymous substitutions neutral?

Codon usage bias, a pervasive phenomenon that has been extensively studied in the last four odd decades, has been observed both within and between genomes and across the tree of life (Ikemura, 1985; Bulmer, 1991; Qin *et al.*, 2004; Kimchi-Sarfaty *et al*., 2007; LaBella *et al.*, 2019). What remains unclear however, is why a preferential use of codons exists. Synonymous substitutions that change codons but not amino acids are commonly thought to be under weak selection, and hence have often been used as a neutral baseline measure for selection tests. A mounting body of evidence nevertheless indicates, that altering established codon usage patterns can have substantial effects that range from growth, mRNA structure and quantities, and protein expression and function (Chamary & Hurst, 2005; Chamary, Parmley, & Hurst, 2006; Peris *et al.*, 2010; Plotkin & Kudla, 2011; Agashe *et al*., 2013; Boël *et al*., 2016; Buhr *et al*., 2016; Kelsic *et al*., 2016; Frumkin *et al*., 2018). The extent of fitness effects of synonymous substitutions has in fact shown to be similar to those of non-synonymous substitutions (Lind *et al*., 2010; Bailey *et al*., 2014; Lebeuf-Taylor *et al.*, 2019). Given that these experimental observations and theoretical predictions are made in different organisms and for different genes, it is likely that synonymous mutations have a substantial role in genome evolution.

1.3.2 What shapes codon bias?

While clear indication remains elusive as to proportionate contribution of different factors to shaping codon bias, some contributing factors have been suggested based on empirical observations made in codon bias studies. Population genetic models posit that codon bias is a result of a balance between AT or GC mutational biases of genomes and selection (Bulmer, 1991). The influence of selection on codon composition is evidenced for highly expressed genes (Ikemura, 1985). Likewise, along a gene, the extent of selection for particular codons varies as well (Agashe *et al.*, 2013; Mittal *et al*., 2018). Theoretical correlations and experimental evidences indicate that tRNA availability (Ikemura, 1981; McDonald et al., 2015), tRNA modifications (Novoa *et al.*, 2012), mRNA structure and stability (Gu, Zhou, & Wilke, 2010; Hia *et al*., 2019), protein folding and function (M. Zhou *et al*., 2013; Buhr *et al*.,

2016) and translational efficiency (Cannarozzi *et al*., 2010; Frumkin *et al*., 2018) may be some of the factors that influence codon bias evolution.

1.3.3 Evolutionary origins of codon usage bias: a new approach

Most studies in the field of codon bias evolution so far, attempt to understand the functional implication of changing codon usage patterns by introduction of specific and targeted synonymous substitutions (Lind *et al*., 2010; Agashe *et al.*, 2013; Brandis & Hughes, 2016; Kelsic *et al*., 2016; Frumkin *et al*., 2017). Some other studies have employed an experimental evolution approach to infer the degree of impact synonymous mutations may have on the fitness of an organism (Bailey *et al*., 2014).

In this study, designed by Dr. Jenna Gallie, the attempt was to understand the functional consequences of codon bias evolution in closely related strains of *P. fluorescens*. Dr. Frederic Bertels screened genomes of three evolutionary diverged strains of *P. fluorescens* – SBW25, A506 and Pf0-1 – in order to identify genes that exclusively exhibit synonymous differences. Three genes were identified from the screen – *glyQ* (glycine tRNA synthetase gene), *acpP* (lipid metabolism gene) and *rpsJ* (30S ribosomal protein gene). The *glyQ* gene exhibits the maximum number of synonymous differences between the strains and hence was selected for further work (Table 1.1). This gene is present in single copy in the genome and is integral to producing mature glycine tRNAs. Therefore large number of changes in the gene is expected to cause some fitness effects.

Gene	Function	Length (bp)	No. of synonymous mutations		
			SBW25 vs A506	SBW25 vs Pf0-1	A506 vs Pf0-1
glyQ	Glycine tRNA ligase	954	39	54	67

Table 1.1: Synonymous differences in *glyQ* between SBW25, A506 and Pf0-1. Glycine tRNA ligase, *glyQ,* has a large number of synonymous differences between strains SBW25, A506 and Pf0-1 and hence was selected for use in the experimental approach of the project.

A novel approach was adopted for this study; *glyQ* alleles would be swapped in non-native backgrounds, thereby simulating a set of strains of the same background that differed

in codon usage for a single gene (Figure 1.7). A crucial difference in this study with respect to the ones mentioned above is that here, we are attempting to understand the factors that may have lead to codon bias evolution in closely related strains of *P. fluorescens*. The naturally evolved synonymous differences are employed to arrive at an understanding of the functional role these synonymous differences might have. Both experimental and bioinformatics approaches were applied to study the pattern of codon use and role of synonymous codons in the physiology of the bacterium, respectively.

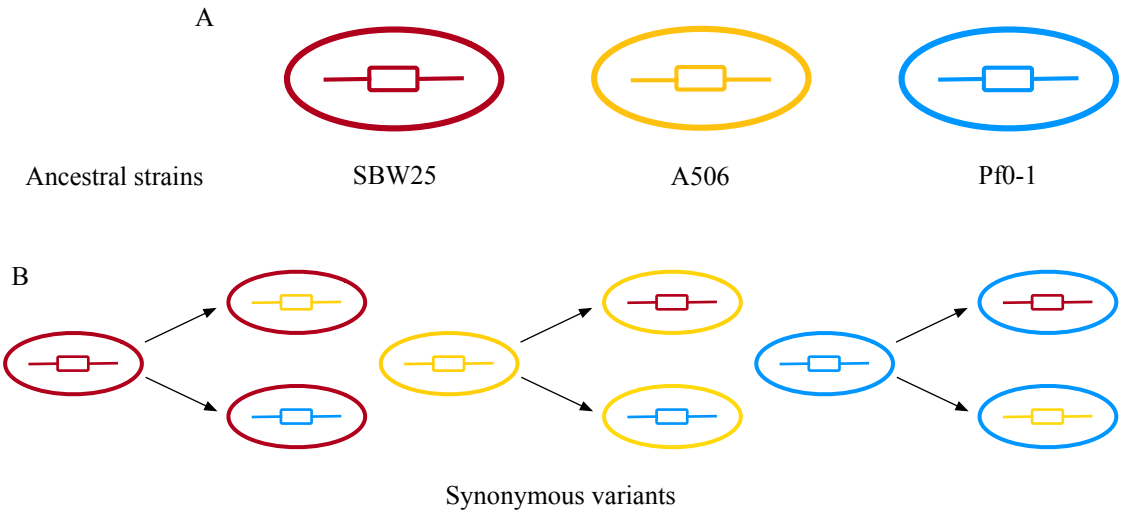


Figure 1.7: Experimental approach for construction of synonymous mutants of *glyQ*. A. This panel is representative of the ancestors that were used for construction of the mutants. The circle represents the cell, while the rectangle with tails represents the gene of interest, *glyQ*. **B.** This panel is representative of the final expected mutants – each of the *glyQ* alleles are swapped into non-native backgrounds. For each background we have three synonymous variants – one wild type variant and two mutant variants.

1.3.4 Synonymous substitutions result in non-neutral effects

The synonymous variants of *glyQ* show significant differences in mRNA quantities and in competitive fitness, although no discernible differences in growth rates were observed. One likely explanation is that the synonymous differences affect either mRNA structure or stability and hence the change in quantity of mRNA in the cell. However, the change in mRNA quantities may not be affecting the ultimate translational output dramatically and hence no appreciable growth differences are observed. But when the strains are competed, the

differences in fitness due to *glyQ* change become more apparent for some of the mutants. Given that codon bias is thought to be a balancing act between mutational biases and selection and/or drift, genetic drift may be the driving force for codon preference observed in *glyQ* of *P. fluorescens.*

Given that this work was done in *P. fluorescens,* of which SBW25 has been studied extensively with respect to mat formation, the ability of the wild type as well *glyQ* mutants to form mats was also investigated. An additional wild type strain, PICF7, was also used for these experiments.

1.4 Summary and aims of the current study

Not all codons (called synonymous codons) for an amino acid are used equally. A preferential use, called codon bias has been observed from viruses to humans. It has been traditionally thought that synonymous codon changes are neutral. However, a growing body of work over the last few decades clearly indicates that synonymous codons are not as neutral as were once thought to be. Although there is no clear understanding of degree of influence one or more contributors may have in shaping codon use, a number of factors have been implicated in forging codon usage bias. Cellular pool of tRNA, stability of mRNA, protein folding and function are among the factors that have been identified to shape codon bias. These processes appear to be affected to a greater or lesser degree, in turn, when established codon use patterns have been altered. In the scope of this thesis, an attempt was made to understand the functional consequences of changing synonymous codons. To this end, synonymous codons that have evolved over time were engineered into non-native backgrounds. In other words, a gene that has evolved to exclusively exhibit synonymous differences was swapped with its allele in non-native background. This would provide an insight into the functional role of synonymous substitutions. The aims of this thesis therefore were as follows:

1. Screening of the genome of *P. fluorescens* for codon usage patterns using bioinformatics technique. The entire genome and genes of interest – *glyQ, acpP, rpsJ* and the diguanylate cyclases and histidine kinases of the *wsp, aws* and *mws* loci – were screened for codon usage patterns (Chapter 3).

2. Intensive phenotypic investigation of engineered synonymous mutants and the cognate wild type strains. This was achieved using growth assays in rich as well as minimal medium, competitive fitness in King's medium B, mRNA quantification using qRT-PCR and mat formation assay (Chapter 4).

3. Exploring the genetic underpinnings of mat formation in closely related strains of *P. fluorescens*. This was achieved by evolving mat formation in SBW25, PICF7, A506 and Pf0-1, followed by identification of molecular routes using whole genome sequencing, transposon suppressor analysis and genetic engineering (Chapter 5).

Chapter 2 : Materials and Methods

2.1 Materials

2.1.1 Media and culture conditions

P. fluorescens was routinely cultured in King's medium B (KB – 10 ml glycerol, 20 g Proteose Peptone No. 3 (liquid cultures, Condalab) or tryptone (plates), 1.95 g $K_2HPO_4.3H_2O$ and 6.1 ml of 1M $MgSO_4.7H_2O$, (King, Ward, & Raney, 1954)) or Lysogeny Broth (LB – 10 g tryptone, 5 g yeast extract and 10 g NaCl, (Bertani, 1951)) and grown at 28ºC. For growth assays, *P. fluorescens* was cultured additionally, in Terrific Broth (TB – 12 g tryptone, 24 g yeast extract, 4 ml glycerol and 100 ml phosphate salts containing 23.1 g KH_2PO_4 and 125.4 g of K_2HPO_4, (Green and Sambrook, 2012)) or Super Broth (SB – 32 g tryptone, 20 g yeast extract, 5 g NaCl and 5 ml 1N NaOH, (Atlas, R.M., 2004)) or minimal M9 medium – 34 g Na_2HPO_4, 15 g KH_2PO_4, 2.5 g NaCl, 5 g NH_4Cl, 100 µl 1M $CaCl_2.2H_2O$, 2 ml 1M $MgSO_4.7H_2O$ and 20% (w/v) of glucose, (Sambrook, *et al.,* 1989)). For making plates 1.5% bacteriological agar was added to each of the media. Overnight cultures were grown for 16 hours (or 24 hours where indicated) while shaking at 200 rpm in 13 ml tubes containing 4 ml media (or 5 ml of media where indicated).

E. coli were routinely cultured in LB at 37ºC. All growth media were purchased from Sigma-Aldrich, unless otherwise mentioned. Incubation was done in incubators from New Brunswick.

2.1.2 Bacterial strains

Bacterial strains used are listed below in Table 2.1. All strains listed in the table were stored indefinitely at -80ºC in 31.1% (v/v) glycerol saline.

Strains	Genotype and properties	Reference
Pseudomonas fluorescens		
SBW25	Ancestral (wild type) strain isolated from the leaf of a sugar beet plant grown at the University Farm, Wytham, Oxford in 1989.	(Rainey & Bailey, 1996)
A506	Ancestral (wild type) strain isolated from a pear tree leaf in California USA in 1993. It has a 57 kb plasmid and is resistant to rifampicin (mutation in *rpoB*) and a single nucleotide insertion in *rpoS*, resulting in frameshift.	(Loper *et al*., 2012)
Pf0-1	Ancestral (wild type) strain isolated from loam soil in Sherborn, Massachusetts, USA in 1987. Pf0-1has a mutation in *gacA*	(Silby *et al*., 2009; Loper *et al*., 2012)
PICF7	Ancestral (wild type) strain isolated from the roots of nursery bred olive plants from Córdoba province in Southern Spain in 2004.	(Mercado-Blanco *et al*., 2004; Martínez-García *et al*., 2015)
LSWS	Niche-specialist mutant evolved from SBW25. Wrinkly genotype resulting from a mutation in *wspF* (A901C)	(Spiers *et al*., 2002)
FS	Fuzzy spreader isolated from the smooth SBW25 propagated in a static microcosm for seven days	(Rainey & Travisano, 1998)
SBW25-*lacZ*	SBW25 containing a neutral, chromosomal *lacZ* marker. Used as the competitor genotype in fitness experiments.	(Zhang & Rainey, 2007)
SBW25_*glyQ*_A506	SBW25 with A506 *glyQ* gene (39 mutations)	This study
SBW25_*glyQ*_Pf0-1	SBW25 with Pf0-1 *glyQ* gene (54 mutations)	This study
A506_*glyQ*_SBW25	A506 with SBW25 *glyQ* gene (39 mutations)	This study
A506_*glyQ*_Pf0-1	Pf0-1 with SBW25 *glyQ* gene (67 mutations)	This study
Pf0-1_*glyQ*_SBW25	Pf0-1 with SBW25 *glyQ* gene (54 mutations)	This study
Pf0-1_*glyQ*_A506	Pf0-1 with A506 *glyQ* gene (67 mutations)	This study
Part SBW25_*glyQ*_A506	SBW25 with part of the A506 *glyQ* gene (bases 1-750), and part of the SBW25 *glyQ* gene (bases 751-954)	This study

Strains	Genotype and properties	Reference
Part Pf0-1_*glyQ*_SBW25	Pf0-1 with part of the Pf0-1 *glyQ* gene (bases 1-420), and part of the SBW25 *glyQ* gene (bases 421-954)	This study
Part A506_*glyQ*_Pf0-1	A506 with part of the A506 *glyQ* gene (bases 1-581 and 836-954), and part of the Pf0-1 *glyQ* gene (bases 580-835)	This study
SBW25_*glyQ*_middle_fast	Three out of four of the ten central codons (bases 502-534) of SBW25 *glyQ* gene changed to the most used codon in SBW25	This study
SBW25_*glyQ*_end_slow	Four out of eight of the ten terminal codons (bases 919-948) of SBW25 *glyQ* gene changed to the least used codon in SBW25	This study
SBW25 mat morph 1-10	10x wrinkly strains forming mat at the air liquid interface in KB under static conditions derived from SBW25	This study
A506 mat morph 1-10	10x webby strains forming mat at the air liquid interface in KB under static conditions derived from A506	This study
PICF7 mat morph 1-10	10x wrinkly strains forming mat at the air liquid interface in KB under static conditions derived from PICF7 followed by transposon mutagenesis with IS-Ω-Km/hah	This study
A506Δ*psl*	A506 with the *psl* operon deleted	This study
A506Δ*pga*	A506 with the *pga* operon deleted	This study
A506Δ*pgapsl*	A506Δ*pga* with the *psl* operon deleted	This study
A506_*wspF*_LoF	Loss of Function (LoF) mutation in *wspF* gene of A506 (CTGATGGACCTGATC, Δ151-165)	This study
Escherichia coli		
DH5α-*λpir*	*supE*44, Δ*lacU*169, *hsdR*17, *recA*1, *endA*1, *gyrA*96, *thi*-1, *relA*1, *λpir*	Invitrogen
S17-1 *λpir*	*recA*1, *endA*1, *thiE*1, *pro*-82, *hsdR*17	(de Lorenzo, Cases, Herrero, &

		Timmis, 1993)

Table 2.1: Name and properties of bacterial strains used in this study.

2.1.3 Plasmids and transposons

The plasmids and transposons used in this study are tabulated below (Table 2.2)

Name	Properties	Reference
Plasmids		
pCR™8/GW/TOPO®	Spe^R, pUC ori, 2.8 kb sequencing plasmid	Invitrogen
pUIC3	Tc^R, *mob, oriR6K, bla,* Δpromoter-*lacZY*	(Rainey, 1999)
Transposons		
IS-Ω-Km/hah	Km^R, ColE1 *ori, npt* promoter, *loxP*	(Giddens *et al*., 2007)

Table 2.2: Name and properties of plasmids and transposons used in this study.

2.1.4 Antibiotics, enzymes and reagents

Antibiotics were used at the following concentrations – tetracycline (Tc) 12.5, 50 (liquid cultures) or 100 μg/ml (plates), Tc was prepared by dissolving in 1:1 ethanol:water, kanamycin (Km) 100 μg/ml, spectinomycin 100 μg/ml and cycloserine 800 μg/ml (all dissolved in water). Nitrofurantoin ((E)-1-[(5-nitro-2-furyl)methylideneamino]imidazolidine-2,4-dione, NF) was dissolved in dimethyl sulfoxide (DMSO) to achieve a final concentration of 100 μg/ml and was used to inhibit growth of *E. coli*. 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-gal, also from Sigma-Aldrich) was dissolved in dimethylformamide and used at 60 μg/ml as a chromogenic indicator of β-galactosidase activity. Taq polymerase (GoTaq®) and Pfu polymerase (Promega) were used for PCRs at concentrations as mentioned in the respective section. Restriction enzymes were purchased from Promega and digestions were carried out at 37ºC as per the manufacturer's protocols. T4 DNA ligase (Promega) was used for ligation as per the manufacturer's protocols. A combinatorial enhancer solution (CES) was used to improve the efficiency of PCR reactions and added where indicated. It is composed of the following components – 2.7M betaine, 6.7 mM dithiothreitol (DTT), 6.7% (v/v) DMSO and 55 μg/ml of bovine serum albumin (Ralser *et al*., 2006). All antibiotics and chemicals were purchased from Sigma-Aldrich, unless otherwise stated.

2.1.5 DNA and RNA extraction materials

Extractions of genomic DNA and plasmid DNA were done using the DNeasy® Blood & Tissue kit and QIAprep® Spin Miniprep kit, respectively from Qiagen. RNA extraction was done using the RNeasy® Mini kit, also from Qiagen.

2.1.6 Primers and oligos

Primers were synthesised by Metabion. Synthesized primers were resuspended in deionized water to achieve a concentration of 100 pmol/µl and stored at -20ºC. Primers were used at a final working concentration of 5 pmol/µl (or 10 pmol/µl as specified). Oligos for constructing the *faster*, *slow* and *flip* alleles of *glyQ* were constructed using the gBlocks® Gene Fragments service from IDT. Oligos were reconstituted by centrifuging the vial for 5 seconds at 300xg, followed by addition of nuclease free water to achieve a concentration of 100 ng/µl. The vial was briefly (15 seconds) vortexed, incubated at 50ºC for 20 minutes, followed by vortexing for 15 seconds and spinning down the contents. Resuspended oligos were stored at -20ºC. The primers used in this study are tabulated below (Table 2.3)

Name	Sequence (5'➔3')	Target
Allelic replacements		
SBW25_glyQ_f	CTTGCCGATGTCCTTCAGGCTC	SBW25 *glyQ*
SBW25_glyQ_r	CACAGCCTTGTCGTAGACGCTG	SBW25 *glyQ*
A506_glyQ_f	CTTGCCGATGTCCATCAGGCTG	A506 *glyQ*
A506_glyQ_r	CAGGCGCAGGTTGAAGTCTTC	A506 *glyQ*
Pf01_glyQ_f	GGACTTGCCGATGTCTTTCAGG	Pf0-1 *glyQ*
Pf01_glyQ_r	GATCGTTGAAGGCTTCGAGCTTC	Pf0-1 *glyQ*
glyQ_mut_1af	GAAGATCTCGCAGCCCAAGTAACA GGTGAATTTC	A506 *glyQ*
glyQ_mut_1ar	CCAGGAAATCTTGAGCACTCATTG TG	A506/Pf0-1 *glyQ*
glyQ_mut_2af	CACAATGAGTGCTCAAGATTTCCT GG	SBW25/Pf0-1/A506 *glyQ* surrounding bases
glyQ_mut_2ar	GAAGATCTGTTACCGGCGATGATC TGCTG	SBW25/A506/Pf01 *glyQ* surrounding bases and pCR™8/GW/TOPO®

Name	Sequence (5'➔3')	Target
glyQ_mut_1bf	GAAGATCTCAGGTTGATCCGCACCGTGTC	SBW25 *glyQ* surrounding bases
glyQ_mut_1br	GAAATTCACCTGTTACTTGGGCTGCG	SBW25/A506 *glyQ* surrounding bases
glyQ_mut_2bf	CGCAGCCCAAGTAACAGGTGAATTTC	pCR™8/GW/TOPO®
glyQ_mut_1cf	GAAGATCTCGCAGCCCAAGAAACAGGTGAATTTC	Pf0-1 *glyQ*
glyQ_mut_1df	GAAGATCTGAGGTTGATCCGCACGGTATCG	A506 *glyQ* surrounding bases
glyQ_mut_1ef	GAAGATCTGGTGAATTTCGTGAGCCAGCCTAC	Pf0-1/SBW25 *glyQ*
glyQ_mut_1cr	GTAGGCTGGCTCACGAAATTCACC	A506/Pf0-1 *glyQ* surrounding bases
glyQ_mut_2cf	GGTGAATTTCGTGAGCCAGCCTAC	pCR™8/GW/TOPO®
glyQ_mut_1dr	CCAGAAAATCTTGAGCACTCATTGTGC	SBW25/A506 *glyQ*
glyQ_mut_2df	GCACAATGAGTGCTCAAGATTTTCTGG	Pf0-1 *glyQ* surrounding bases
glyQ_mut_1ff	GAAGATCTCAGGTTGATCCGCACCACATC	Pf0-1 *glyQ* surrounding bases
glyQ_seqmut_f	GAGAATTCGATCGTGTGTGTGGTGC	SBW25 *glyQ* slow fragment
glyQ_seqmut_r	GAGAATTCGTGCGCTTGCTGATCA	SBW25 *glyQ* slow fragment
glyQ_seqmut_f1	GAAGATCTGATCGTGTGTGTGGTGC	SBW25 *glyQ* faster/flipped fragment
glyQ_seqmut_r1	GAAGATCTGTGCGCTTGCTGATCA	SBW25 *glyQ* faster/flipped fragment
glyQ_mut_1hr	GTTCAACCAGGAAATCTTGAGCACTCA	SBW25 *glyQ* faster/slow/flipped fragment
glyQ_mut_2hf	CAAGATTTCCTGGTTGAACTGGGCAC	SBW25 *glyQ* faster/slow/flipped fragment
glyQ_mut_2hr	TGTTACTTGGGCTGCGATTTAAAGAGC	SBW25 *glyQ* faster/slow/flipped fragment

Name	Sequence (5'➔3')	Target
glyQ_start_f	GAAGATCTCTTCGACTTCGCGCACCAG	SBW25 *glyQ* surrounding bases
glyQ_middle_f	GAAGATCTCAGCCCAAGTAACAGGTGAATTTCGTG	SBW25 *glyQ*
glyQ_middle_r	GAAGATCTGACTTTCTTCGCTTCGAACTTCAGG	SBW25 *glyQ*
glyQ_end_r	GAAGATCTGAACACCACGTTCTGCAGGC	SBW25 *glyQ* surrounding bases
glyQ_start_fast_2	GAAGGTGCGCACGGCCGGGGTCGGCTG	SBW25 *glyQ*
glyQ_start_fast_3	CAGCCGACCCCGGCCGTGCGCACCTTC	SBW25 *glyQ*
glyQ_start_fast_4	CGTGATAAGCCATGTACAGCGGATC	SBW25 *glyQ*
glyQ_start_slow_2	GGTACGCACAGCTGGTGTAGGTTGAGACACGAAATTC	SBW25 *glyQ*
glyQ_start_slow_3	GAATTTCGTGTCTCAACCTACACCAGCTGTGCGTACC	SBW25 *glyQ*
glyQ_middle_fast_2	CATGGCCAGGCGTTCCAGGCCGTAGGTG	SBW25 *glyQ*
glyQ_middle_fast_3	CACCTACGGCCTGGAACGCCTGGCCATG	SBW25 *glyQ*
glyQ_middle_fast_4	GACAACTGGGAATCGCCGACC	SBW25 *glyQ*
glyQ_middle_slow_2	CAGGTACATAGCTAGTCTCTCTAGTCCGTAGGTG	SBW25 *glyQ*
glyQ_middle_slow_3	CACCTACGGACTAGAGAGACTAGCTATGTACCTG	SBW25 *glyQ*
glyQ_end_fast_2	GTGCAGCTTCCAGCTTGGCCAGCACTTCATC	SBW25 *glyQ*
glyQ_end_fast_3	GATGAAGTGCTGGCCAAGCTGGAAGCTGCAC	SBW25 *glyQ*
glyQ_end_slow1_2	GCCTCCAACTTAGCTAGTACCTCATCACGCAG	SBW25 *glyQ*
glyQ_end_slow1_3	CTGCGTGATGAGGTACTAGCTAAGTTGGAGGC	SBW25 *glyQ*
glyQ_end_slow_4	GCACGCCAGCAATACATCCTG	SBW25 *glyQ*
SBW25_glyQ_1	GAAGATCTGACAGGTTGATCCGCACCGTG	SBW25 *glyQ* surrounding bases

Name	Sequence (5'➔3')	Target
SBW25_glyQ_2	AAGTACCACACAACCTTGCTCGG	SBW25 *glyQ* surrounding bases
SBW25_flip_f2	CCGAGCAAGGTTGTGTGGTACTTCAA CCATATGACATGGAGGTCGGAGC	SBW25 *glyQ* flipped fragment
SBW25_flip_r2	GAGTTTGGCGAGAACCTCGTCTC	SBW25 *glyQ* flipped fragment
SBW25_slow_r2	GAAGATCTGCGACGCTCGTTGGCATC	SBW25 *glyQ* flipped/slow fragment
pga_del_f1	GAAGATCTGATGCTGGTGCTGACCAA GCTC	A506 *pga* operon surrounding bases
pga_del_r1	GAGCTTCCTTACACACGGAAAGCC	A506 *pga* operon surrounding bases
pga_del_f2	GGCTTTCCGTGTGTAAGGAAGCTCGC GAACGCAATGCCCAAAGC	A506 *pga* operon surrounding bases
pga_del_r2	GAAGATCTCGAGTTTGCCTTCCACCA CCAG	A506 *pga* operon surrounding bases
psl_del_f1	GAAGATCTGAACAGCTCCCTGGCATT GTTTCG	A506 *psl* operon surrounding bases
psl_del_r1	CAGATCCTCAATACACCACCAGAACG	A506 *psl* operon surrounding bases
psl_del_f2b	CGTTCTGGTGGTGTATTGAGGATCTG GCCGGTTCCCACCAGAACG	A506 *psl* operon surrounding bases
psl_del_r2	GAAGATCTGGCTTCCTGACCTACGCT TCTCA	A506 *psl* operon surrounding bases
wspF_del_f1	GAAGATCTGGACAAGCTGCTCAATAC CGGC	A506 *wspF*
wspF_del_r1	CCGTCCATCACCGGCATGATCAGGTC CGGGGTCAGCTC	A506 *wspF*
wspF_del_f2	ATGCCGGTGATGGACGG	A506 *wspF*
wspF_del_r2	GAAGATCTGCGCAATTCTATCCAGTG GGCG	A506 *wspF*
GW1	GTTGCAACAAATTGATGAGCAATGC	pCR™8/GW/TOPO®
GW2	GTTGCAACAAATTGATGAGCAATTA	pCR™8/GW/TOPO®
Quantitative PCR (qPCR)		
rpoD_qRTPCR_1	CAACGAAGTAGACGAAAGCTG	SBW25 *rpoD*

Name	Sequence (5'➔3')	Target
rpoD_qRTPCR_2	GACGGTTGATGTCCTTGATCTC	SBW25 *rpoD*
rpoD_qRTPCR_5	GAGATCAAGGACATCAACCGT	Pf0-1 *rpoD*
rpoD_qRTPCR_6	TTCTTGGCGATGGAGATCAC	Pf0-1 *rpoD*
glyQ_RT_F1	CGTACCTTCCAAGACTTGATCC	*glyQ*
glyQ_RT_r1	CTACTTCCATATCGTAGGGCTG	*glyQ*
glyQ_RT_F2	GTACCTTCCAAGACTTGATCCT	*glyQ*
glyQ_RT_r2	CTACTTCCATATCGTAGGGCTG	*glyQ*
recA_RT_F2	CGTAATGCGTATGGGCGAT	*recA*
recA_RT_r2	TAGATTTCAACGATACGGCCTTT	*recA*
Identification of mutations in the mat forming strains		
mwsR_f	TTGTCCAAAGTCACGCCGCC	*mwsR* in SBW25 evolved mat forming mutants
mwsR_r	TCAGTCGAACATGAACAGCGCATCA	*mwsR* in SBW25 evolved mat forming mutants
mwsR_mut_f1	TGGTGCTGGTGTGGGACG	*mwsR* in SBW25 lines 2 & 3
mwsR_mut_r2	GGCCATTTCCTTGAGCATGCG	*mwsR* in SBW25 lines 2 & 3
mwsR_mut_f2	CGCATGCTCAAGGAAATGGCC	*mwsR* in SBW25 lines 1 & 5
mwsR_mut_r1	GAACTGTCGGGCGGAGATGTTC	*mwsR* in SBW25 lines 1 & 5
wspF_f	CGCACGGGCATGAGGATTG	*wspF* in SBW25 lines 4 & 8
wspF_r	CGCGAGAGGTCATGTTCATTTGGC	*wspF* in SBW25 lines 4 & 8
awsR_f	ATGAGCGGCGCCAAGGT	*awsR* in SBW25 line 6
awsR_f1	TAGCCTGTTGCGCTTCCTGC	*awsR* in SBW25 line 7
awsR_r	CGTTGACGACGGGGTTCTCC	*awsR* in SBW25 lines 6 & 7
wspA_f	GTGAAGAACTGGACCTTGCGCC	*wspA* in SBW25 evolved mat forming mutants
wspA_f1	ATGCACTCGGTGATGGGCG	*wspA* in SBW25 line 9
wspA_r	TCAGACTTTGAAACGCGACACGC	*wspA* in SBW25 line 9
wspE_f	ATGACCCCCGACCAGATGC	*wspE* in SBW25 evolved mat forming mutants
wspE_f1	CGTTCGGTGGACAAACTGCTGAA	*wspE* in SBW25 line 10
wspE_r	TCATGCCCGTGCGCCT	*wspE* in SBW25 line 10
wspA_PICF7_f	GTGAAGAACTGGACCTTGCGCC	*wspA* in PICF7 line 1
wspA_PICF7_r	TCAGACTTTGAAACGCGACACGC	*wspA* in PICF7 line 1
wspF_PICF7_f	GGGCATGAGGATTGCGATCGT	*wspF* in PICF7 lines 2,3,8,9

		& 10
wspF_PICF7_r	TCGTCGGTCTTGATGTCGTCGATC	*wspF* in PICF7 lines 2,3,8,9 & 10
awsX_PICF7_f	ATGAACGTGGCTGTCGCGT	*awsX* in PICF7 line 4
awsX_PICF7_r	TCATGACGCAGGCGCTCG	*awsX* in PICF7 line 4
awsO_PICF7_f	TTGTTCTCGACCGCGCGTT	*awsO* in PICF7 line 6
awsO_PICF7_r	TCAGTCCGCGATCACCACAATC	*awsO* in PICF7 line 6
mwsR_PICF7_f	TTGTCCAAAGTCACGCCGCC	*mwsR* in PICF7 evolved mat forming mutants
mwsR_PICF7_r	TCAGTCGAACATGAACAGCGCATC	*mwsR* in PICF7 evolved mat forming mutants
mwsR_PICF7_f1	CGGCGACCGTATGCTCAAGG	*mwsR* in PICF7 line 7
mwsR_PICF7_r1	GCTTGCCGTCGCCACTGAA	*mwsR* in PICF7 line 7
Transposon mutagenesis		
Tn*phoA*-II	GTGCAGTAATATCGCCCTGAGCA	IS-Ω-Km/hah
CEKG 2A	GGCCACGCGTCGACTAGTACNNNNNNNNNNAGAG	Non-specific
CEKG 2B	GGCCACGCGTCGACTAGTACNNNNNNNNNNACGCC	Non-specific
CEKG 2C	GGCCACGCGTCGACTAGTACNNNNNNNNNNGATAT	Non-specific
Hah-1	ATCCCCCTGGATGGAAAACGG	IS-Ω-Km/hah
CEKG 4	GGCCACGCGTCGACTAGTAC	5' end of CEKG 2A, B, & C

Table 2.3: Names, sequences and target regions of the primers used. Incorporated *Bgl*II and *Eco*RI restriction sites are underlined.

2.1.7 Materials for production of chemically competent cells

Filter sterilized TBFI and TBFII buffers were required to produce chemically competent *E. coli* cells. TBFI buffer is composed of 30 mM KOAc, 100 mM RbCl and 10 mM $CaCl_2$ at a pH of 5.8, while TBFII buffer is composed of 10 mM MOPS, 10 mM RbCl, 75 mM $CaCl_2$ and 15% (v/v) glycerol, at a pH of 7.0.

2.1.8 Materials for microscopy

Colony visualisation were done using Leica MS-5 dissection microscope. Microscopic images of colonies were captured using a VisiCam® 1.3 (1.3 Megapixel) USB camera with a 0.45X eyepiece adapter (VWR). Where needed, the images were adjusted and measured using Fiji, Version 2.0.0-rc-69/1.52p (Schindelin *et al.*, 2012). Mr. Michael Schwarz imaged microcosms using Canon EOS 550D with a Canon EF-S lens.

2.1.9 Growth assay materials

Growth assays were conducted in 96-well plates (Sarstedt) and the optical density of the cultures were read in an Epoch^{TM}2 Microplate Spectrophotometer from BioTek® with Gen5 analysis software. To prevent condensation on the lid of the plates, the Breathe-Easy® (Sigma-Aldrich) sealing membrane was used instead.

2.2 Methods

2.2.1 Genomic DNA extraction

Genomic DNA was extracted using the DNeasy® Blood & Tissue kit. The method has been adapted from the manufacturer provided protocol. 500 µl of overnight culture was pelleted by centrifugation for 10 minutes at 13,000 rpm and resuspended in 180 µl buffer ATL. 20 µl of proteinase K was added to the resuspension, mixed by inverting the tube four times and then incubated at 56ºC, 700 rpm on a thermomixer (Eppendorf) for 1 hour. Subsequently, the tube was vortexed for 15 seconds, 200 µl of buffer AL was added and mixed by inverting the tube four times. 200 µl of ethanol was added and mixed by inverting the tube four times. The resulting mixture was added to the DNeasy® Mini spin column and centrifuged at 13,000 rpm for 1 minute. The column was placed in a fresh 2 ml collection tube, 500 µl of buffer AW1 was added and centrifuged at 13,000 rpm for 1 minute. Following this, the column was placed in another fresh 2 ml collection tube, 500 µl of buffer AW2 was added and centrifuged at 13,000 rpm for 3 minutes. The flow through was discarded and the column was centrifuged at 13,000 rpm for 1 minute. Subsequently, the column was placed in a 1.5 ml Eppendorf Tube®, 100 µl of nuclease-free water heated to 60ºC was added to the column and incubated at room temperature for 1 minute. The tube was then centrifuged at 13,000 rpm for 1 minute. In cases where the sample was used for genome sequencing, quality of the sample was checked using agarose gel electrophoresis and spectrophotometry (260/280 ratio: 1.8 – 2.0)

2.2.2 Polymerase chain reaction (PCR)

2.2.2.1 Standard PCR

The components in a standard 25 µl reaction were: 5 µl of 5x PCR buffer, 5 µl of 5x CES, 3 µl of 25 mM $MgCl_2$, 1 µl of 10 mM dNTP mix (Quanta), 2 µl of each of 5 pmol/µl of forward and reverse primers, 0.25 µl of Taq polymerase and 30-100 ng of DNA, volume made up to 25 µl with deionized water. Reaction conditions were: initial template denaturation at 94ºC for 10 minutes, followed by 30 cycles of denaturation for 30 seconds at 94ºC, annealing for 30 seconds at 56-60ºC and strand elongation for 1 minute per kilobase (kb) of target DNA at 72ºC. For cloning into pCR™8/GW/TOPO®, a final extension step of

5 minutes at 72ºC was performed prior to cooling the sample indefinitely to 4ºC. Polymerase chain reactions were carried out using a Biometra TRIO 48 thermocycler (Analytik Jena).

2.2.2.2 Strand overlap extension PCR (SOE-PCR)

Constructs for integration into genome – genomic allele swaps (Chapter 4) and genomic deletions (Chapter 5) – were carried out using SOE-PCR (Ho *et al.*, 1989) technique. Briefly, the process involves amplification of two overlapping PCR products that are subsequently annealed in a second PCR reaction. This generates a single PCR product, which is further amplified. For genomic deletions, DNA on either side of the targeted locus was amplified; amplification was performed using the standard PCR techniques (see section 2.2.2.1). To increase fidelity of amplification, Pfu polymerase was used. This requires 2 minutes per kb of target DNA for strand elongation. Primers that complemented for approximately 25 bases were used. For genomic allele swaps, DNA of the target locus, in addition to DNA on either side of the target locus was amplified using the standard PCR techniques with Pfu polymerase. For both deletions and allele swaps, equal amounts of each of the standard PCR products were used as templates in a SOE-PCR reaction that contained – 5 µl of 10x Pfu Polymerase Buffer, 1 µl of 10 mM dNTP mix, 4 µl of each of the 5 pmol/µl of forward and reverse primers, 2 µl of Pfu polymerase and 1 µl of template, volume made up to 50 µl with deionized water. PCR cycling conditions were the same as employed for a standard PCR.

2.2.2.3 Arbitrary primed-PCR (AP-PCR)

Transposon-chromosome junctions were amplified (Chapter 5) using the AP-PCR technique. This is a two-step semidegenerate PCR method developed by (Manoil, 2000) and adapted by (Jacobs *et al*., 2003). The first round of PCR was a degenerate PCR reaction that contained 4 µl of 5x PCR buffer, 1 µl 10 mM dNTP mix, 2.4 µl of 25 mM $MgCl_2$, 2 µl each of 10 pmol/µl primer, 0.5 µl Taq polymerase and 3 µl template, volume made up to 20 µl with deionized water. Reaction conditions were: initial template denaturation at 94ºC for 10 minutes, followed by 6 cycles of denaturation of 30 seconds at 94ºC, annealing for 30 seconds at 42ºC and strand elongation for 3 minute at 72ºC, wherein the annealing temperature was decreased by 1ºC with every cycle. This was followed by 25 cycles of 30 seconds at 94ºC, 30 seconds at 60ºC and 3 minutes at 72ºC. The PCR products from this round were diluted with 80 µl of deionized water and used as template for the second round of PCR. The second round

of PCR contained 4 µl of 5x PCR buffer, 1 µl 10 mM dNTP mix, 2.4 µl of 25 mM $MgCl_2$, 2 µl each of 10 pmol/µl primer, 0.5 µl Taq polymerase and 2 µl template, volume made up to 20 µl with deionized water. Reaction conditions were: initial template denaturation at 94ºC for 10 minutes, followed by 30 cycles of denaturation of 30 seconds at 94ºC, annealing for 30 seconds at 65ºC and strand elongation for 3 minute at 72ºC. The sample was cooled to 4ºC and sent for sequencing to Macrogen Europe (Amsterdam, The Netherlands).

2.2.3 Cloning and transformation

2.2.3.1 Plasmid purification, digestion and ligation

Overnight cultures of *E. coli* were used for plasmid DNA extraction using the QIAprep® Spin Miniprep kit (Qiagen). Plasmid extraction was done using the manufacturer provided protocol. Plasmid extraction was followed by digestion of both plasmid and the desired insert at 37ºC with appropriate restriction enzyme(s). Additionally, the plasmid was treated with thermosensitive alkaline phosphatase to prevent vector re-ligation. Ligation reaction was carried out by incubation at room temperature for 1 hour, followed by overnight incubation at 4ºC with T4 DNA ligase in a final volume of 10 µl with 7.5 µl of insert, 1 µl of digested plasmid, 1 µl of ligation buffer and 0.5 µl of T4 DNA ligase. Cloning into pCR™8/GW/TOPO® was performed using the TOPO® TA cloning protocol which was adapted from the manufacturer's protocol. Before cloning into pCR™8/GW/TOPO®, the PCR product was used for poly-A tailing containing 25 µl of template, 6 µl of 25 mM $MgCl_2$, 3 µl of 5x PCR buffer, 1 µl of 10 mM dNTP mix and 0.1 µl of Taq polymerase, incubated at 72ºC for 20 minutes followed by cooling to 4ºC, indefinitely. TOPO® TA cloning was performed with 4 µl of poly-A tailed PCR product, 1 µl of Salt solution and 1 µl of vector gently mixed and incubated at room temperature for 10-15 minutes. The cloned product was used for transformation and remainder was stored at -20ºC.

2.2.3.2 Production and transformation of chemically competent cells

Chemically competent cells of *E. coli* DH5α-*λpir* and *E. coli* S17-1 *λpir* were produced from overnight cultures that were grown to mid-log phase in 200 ml LB medium. 200 ml culture was divided into four 50 ml aliquots in Falcon tubes and pelleted by centrifugation at 3,000 rpm, 4ºC for 10 minutes. Pellets were resuspended in 1 ml ice-cold TBFI buffer and

combined. Volume was made up to 15 ml with TBFI buffer and incubated on ice for 60 minutes. This was followed by centrifugation at 3,000 rpm, 4ºC for 10 minutes, and resuspension of pellet in 4 ml of ice-cold TBFII. The resulting suspension was stored at -80ºC in 50 µl aliquots until further use.

Transformation was performed by thawing the appropriate competent cells on ice for 5 minutes, followed by addition of 4 µl of either intact plasmid or ligation mix. The reactions were incubated on ice for 10 minutes, heat-shocked at 42ºC for 30 seconds, following which the heat-shocked cells were incubated on ice for another 2 minutes and 250 µl LB medium was added. The mixture was incubated at 37ºC for 60 minutes to recover cells. After incubation, the cells were plated on appropriate selective medium.

2.2.4 Gel electrophoresis

Agarose gels were prepared with 1% w/v of agarose and 1x TAE buffer. TAE is composed of 40 mM Tris Base, 20 mM acetic acid and 1 mM of EDTA Sodium salt dehydrate. SYBR™ Safe DNA Gel Stain (Invitrogen) was added to the liquefied agarose for visualisation of DNA under UV light. Samples were loaded with 6X Blue/Orange Loading Dye (Promega) and gels were run at 90 V until sufficient separation of the gel loading indicator dyes were obtained. Gels were viewed using a UV transilluminator in a gel documentation system. Where needed, DNA bands were excised under blue light using sterile droppers and DNA was extracted from the gel using the QIAquick® Gel Extraction kit (Qiagen).

2.2.5 DNA sequencing

2.2.5.1 Sanger sequencing

Plasmid DNA and PCR products were purified using the QIAprep® Spin Miniprep kit and QIAquick® PCR Purification kit (Qiagen), respectively. Purified DNA samples were used for setting up sequencing reaction as follows – 2 µl of 5x BigDye® buffer, 0.5 µl BigDye® Terminator, 1 µl of 5 pmol/µl primer and 1 µl of template, volume made up to 10 µl with HPLC water. Reactions were set up in MicroAmp™ Fast Optical 96-well Reaction Plate, 0.1 ml (Applied Biosystems™, ABI) and run in Applied Biosystems™ Veriti Thermal Cycler.

The reaction conditions were – initial sample denaturation at 96ºC for 1 minute, followed by amplification by 30 cycles of denaturation at 96ºC for 10 seconds, annealing at 56ºC for 15 seconds, strand elongation at 60ºC for 4 minutes, followed by cooling to 4ºC, indefinitely. On completion of the PCR reaction, samples were cleaned using BigDye® XTerminator™ Purification kit as per the manufacturer's instructions: 47 µl of SAM™ Solution and 10 µl of XTerminator™ Solution for 10 µl of PCR product per well. The plates were sealed with MicroAmp™ Clear Adhesive Film (Applied Biosystems™, ABI) and vortexed for 30 minutes at the maximum speed of the Taitec MicroMixer E-36 (Taitec). The plate was then centrifuged at 1,000x g for 2 minutes and submitted to the Sequencing Unit at the Max Planck Institute for Evolutionary Biology, Plön, Germany. Sequence traces were analysed using 4Peaks version 1.7.1 (Mekentosj) and mapped to reference sequence using Geneious® version 10.1.3.

2.2.5.2 Whole genome sequencing

Whole genome sequencing was achieved using high-throughput, amplification based MiSeq System from Illumina® by Dr. Sven Künzel at the Max Planck Institute for Evolutionary Biology, Plön, Germany. The samples were sequenced as paired end with a run of 2x300 bp and an average of 1,000,000 reads per sample. The mean coverage is 50X. Additionally, another run of sequencing was done for A506 wild type strain using MiSeq as paired end with a run of 2x150 bp with an average of nearly 3,000,000 reads per sample and mean coverage of nearly 75X. The sequence traces were analysed using *breseq* (Deatherage & Barrick, 2014) and Geneious® version 10.1.3.

2.2.6 Bi-parental conjugation

E. coli donor and *P. fluorescens* recipient strains were grown in LB medium containing appropriate antibiotics, overnight. 1 ml of the recipient culture was heat shocked at 45ºC for 20 minutes (for SBW25) or not (for A506, Pf0-1, PICF7) and pelleted by centrifugation at 13,000 rpm for 2 minutes. Simultaneously, 300 µl of donor culture was pelleted by centrifugation at 13,000 rpm for 2 minutes. Both pellets were resuspended in 500 µl of LB (washing) and pelleted by centrifugation at 13,000 rpm for 2 minutes. The resulting pellets were resuspended in 200 µl of LB and mixed thoroughly. Cells were pelleted a final time (13,000 rpm, 2 minutes) before resuspension in 30 µl of LB. This concentrated bacterial

resuspension was spotted and spread onto the centre of a pre-warmed LB agar plate and allowed to dry for 10 minutes at room temperature. The plate was incubated at 28ºC for 24 hours. After incubation, the inoculum was removed and resuspended in 500 ul of LB and serially diluted in Ringer's solution before plating onto LB agar plate containing Tc (to select against the donor plasmid, pUIC3), NF (to select against *E. coli*) and X-gal. After incubation at 28ºC for 48 hours, up to ten individual colonies were selected, grown in LB medium with Tc (to grow transconjugants), overnight and stored at -80ºC.

2.2.7 Two-step allelic exchange

A two-step allelic exchange protocol (Kitten *et al.*, 1998) was used for targeted genomic engineering (Chapters 4 and 5). This method was used without enrichment for the construction of the mutants with the background of A506, Pf0-1 and PICF7, as wells as SBW25 (natural *glyQ* allele swaps). However, all other mutant constructions with the SBW25 background were done with an enriched two-step allelic exchange method; this is because growing A506, Pf0-1 and PICF7 cells have not been successfully isolated after subjection to enrichment with the antibiotic cycloserine (personal communication, Dr. Jenna Gallie). With SBW25 cells the enrichment step increases the probability of obtaining the desired mutants. As the name suggests, this method involves two steps – in the first step, upon entering the *P. fluorescens* cells, pUIC3 integrates into the host chromosome *via* recombination at a locus homologous to that of its cloned insert. This is realized through mobilisation of the desired pUIC3 construct from an *E. coli* donor into the desired *P. fluorescens* recipient (see section 2.2.6). Next, 10 µl of a combination of ten transconjugant *P. fluorescens* overnight cultures was used to inoculate 200 ml LB in a 1 L flask, without antibiotic to allow excision of the chromosomally integrated pUIC3. After 16 hours of incubation at 28ºC (200 rpm), the overnight culture was pelleted (13,000 rpm, 2 minutes) and resuspended in 500 µl of Ringer's solution. For the two-step allelic replacement with enrichment, 10 µl of a combination of ten transconjugant *P. fluorescens* overnight cultures was used to inoculate 200 ml M9 minimal medium in a 1 L flask. Following 16 hours of incubation at 28ºC (200 rpm), cells from the overnight culture were harvested (13,000 rpm, 2 minutes) and washed in 400 µl of LB. 400 µl of the resulting concentrated cell suspension was transferred to 20 ml fresh LB in a 250 ml flask and incubated for 45 minutes at 28ºC (200 rpm). 12.5 µg/ml Tc was added to the culture and incubated for another 2 hours at 28ºC (200 rpm). This selected for growth of cells that had incorporated the pUIC3 vector (pUIC3 has a tet^R gene) into the host chromosome. This was followed by addition of cycloserine, an antibiotic that kills growing *Pseudomonas* cells,

to the culture and incubated for another 5 hours at 28ºC (200 rpm). This step kills the Tc resistant cells and selects cells that had undergone a second round of homologous recombination. The second round of recombination is a rare event that leads to loss of the chromosomally integrated vector. Cells with successful excision of the complete vector also lose Tc resistance (*tet*^{R} gene) and are white on the selective medium (loss of *lacZ* gene present on pUIC3). For both with and without enrichment methods, cells were harvested by centrifugation (13,000 rpm, 2 minutes), washed with Ringer's solution and serially diluted. The dilutions were spread on LB agar plates containing 60 µg/ml X-gal. The plates were incubated at 28ºC for 48 hours after which white colonies were picked and purified by re-streaking on fresh LB agar plates with X-gal. Incorporation of the desired DNA fragment was checked by PCR and confirmed by Sanger sequencing (see section 2.2.5.1).

2.2.8 RNA extraction

Cellular RNA was extracted using the RNeasy® Mini kit (Qiagen) and performed as per the manufacturer's instructions. Firstly, 400 µl of overnight culture was transferred to 20 ml KB in 250 ml flask. Bacteria were grown up to the mid-log phase (OD_{600} ~ 0.5-0.6). To 500 µl of the mid-log phase culture, 1 ml of RNAprotect Bacteria Reagent (Qiagen) was added and mixed by vortexing for 5 seconds. The mixture was incubated at room temperature for 5 minutes prior to centrifugation at 5,000x g for 10 minutes. The resulting pellet was resuspended in a mixture of 10 µl of Proteinase K and 200 µl TE buffer, vortexed for 10 seconds and incubated at 20ºC in a thermomixer for 2 minutes. After incubation, a mixture of 700 µl of RLT buffer and β-mercaptoethanol (10 µl β-mercaptoethanol per ml of RLT uffer) was added to the tube and vigorously vortexed. To this 500 µl of 99% ethanol was added and mixed by pipetting. 700 µl of the mix was transferred to an RNeasy® Mini Spim column and centrifuged at 13,000 rpm for 15 seconds. This was followed by addition of 350 µl of RW1 buffer to the RNeasy® column, centrifugation at 13,000 rpm for 15 seconds, addition of 10 µl of DNase I and 70 µl of RDD buffer and incubation at room temperature for 15 minutes. 350 µl of RW1 buffer was added to the column, incubated at room temperature for 5 minutes prior to centrifugation at 13,000 rpm for 15 seconds. The column was placed in a new collection tube, 500 µl of RPE buffer was added and centrifuged at 13,000 rpm for 15 seconds. RPE buffer addition was repeated but the tube was centrifuged for 2 minutes. The column was placed in a 1.5 ml collection tube, 50 µl of RNase-free water was added to the column, incubated at room temperature for 1 minute and finally centrifuged at 13,000 rpm for 1

minute. RNA was stored at -80ºC. RNA concentration was checked using spectrophotometry (260/280: 2.0-2.2).

2.2.9 Complementary DNA (cDNA) synthesis

Each of the RNA samples was diluted to obtain the same starting concentration of RNA (150 ng/µl) for cDNA synthesis. Reverse transcription of RNA to obtain cDNA was done using the High-Capacity cDNA Reverse Transcription Kit. For each sample a no reverse transcriptase (-RT) control was set up; this control allows to check if the reverse transcription and therefore cDNA synthesis from RNA has worked or genomic DNA has been amplified. The reaction contained a 2x RT master mix composed of 2 µl of 10x RT buffer, 0.8 µl of 25x dNTP mix (100 mM), 2 µl of 10x RT Random Primers, 1 µl of MultiScribe™ Reverse Transcriptase and1 µl of RNase inhibitor, volume made up to 10 µl with nuclease-free water. To 10 µl of 2x RT master mix 10 µl of RNA sample was added, mixed by pipetting and tubes were sealed. The reaction was run in a thermal cycler at 25ºC for 10 minutes, 37ºC for 120 minutes, 85ºC for 5 minutes and cooling to 4ºC, indefinitely. The resulting cDNA was diluted in nuclease free water 10^2 fold and stored at -20ºC.

2.2.10 Transposon mutagenesis

Transposon mutagenesis was carried out based on the method as described in (Giddens *et al*., 2007), using a bi-parental conjugation protocol that was set up between an *E. coli* donor carrying the IS-Ω-Km/hah transposon on the pSCR001 plasmid and the wrinkly spreader *P. fluorescens* PICF7 recipient (see section 2.2.6). This method was used for mutagenesis of the wrinkly spreaders of PICF7 (Chapter 5). On a selective medium of LB agar containing NF (select against *E. coli* donor) and Km (select against plasmid), transconjugants were spread; while the wrinkly spreader produces wrinkly colonies on this medium, the transconjugants deficient in the wrinkly spreader phenotype were picked, purified, grown in liquid cultures at 28ºC (200 rpm) and stored at -80ºC. The transposon insertion site was determined by AP-PCR (see section 2.2.2.3) and purified using 0.1 µl ExoI and 0.2 µl TSAP, volume made up to 5 µl using deionized water. The mixtures were incubated at 37ºC for 90 minutes followed by heat inactivation at 85ºC for 15 minutes. The purified DNA samples were sent to Macrogen Europe (Amsterdam, The Netherlands). The sequence data was mapped to *P. fluorescens* PICF7 genome using Artemis software (Rutherford *et al.*, 2000) and Geneious® version

10.1.3. For unannotated transposon insertion sites, possible gene identities were determined by inferring similarities from BLASTP alignments of the Basic Local Alignment Search Tool (BLAST) bioinformatics program from Sanger Institute.

2.2.11 Phenotypic assays

2.2.11.1 Measurement of growth rates

P. fluorescens strains were purified from glycerol stocks by streaking on the relevant medium agar plates that was being tested for growth dynamics. The plates were incubated at 28ºC for 48 hours (or 24 hours for nutritionally enriched medium such as TB or SB) and overnight cultures were set up in 200 µl of relevant liquid medium in a 96- well microtitre plate, each well inoculated with a single colony. A well randomization Java script written by Dr. Frederic Bertels was used for used for randomization of samples on a plate. The 96-well microtitre plate with the appropriate, randomized inoculum was incubated at 28ºC (200 rpm) for 16 hours. For growth measurements at different temperatures, the appropriate temperature (24ºC, 28ºC, 32ºC or 34ºC) was used for both agar plates as well liquid cultures for overnight incubation. 2 µl of the overnight cultures were transferred to the relevant 198 µl fresh medium in a 96-well microtitre plate. The microtitre plates were sealed with the Breathe-Easy® sealing membrane to prevent condensation on the lid and thereby erroneous readings. The optical density of the cultures at 600 nm (OD_{600}) over a period of 48 hours at 5 minute intervals (with 5 seconds prior to each reading) were determined using Epoch^{TM}2 Microplate Spectrophotometer from BioTek® with Gen5 analysis software. Replicates of each assay were obtained and the mean and standard error values determined. The growth rate of each strain was determined using the *growthcurver* package in R (version 3.4.0) (Sprouffske & Wagner, 2016).

2.2.11.2 Niche preference in static microcosms

The ability of *P. fluorescens* strains to form mats in KB microcosms under static conditions was determined using a method adapted from (Rainey & Travisano, 1998). *P. fluorescens* strains were purified from glycerol stocks by streaking on KB agar plates and incubated at 28ºC for 48 hours. Individual colonies were inoculated into 6 ml KB in 25 ml vials (microcosms) and vortexed for 15 seconds. Ten replicates for each strain were set up.

The caps were loosened and the microcosms were incubated at 28ºC for 72 hours under static conditions to allow development of a heterogeneous environment within the microcosm. Following incubation for 72 hours, the microcosms were photographed (by Mr. Michael Schwarz), vigorously vortexed for 1 minute and serially diluted up to 10^6 fold in Ringer's solution. 50 µl and 100 µl of dilutions of each replicate were spread on KB agar plates and incubated at 28ºC for 48 hours. Colonies were visualized microscopically and photographs of the colonies were acquired.

2.2.11.3 Dissection microscopy

Colonies were visualized under the Leica MS5 dissection microscope at a magnification of 1.6X and images were photographed using VisiCam® camera from VWR and the VisiCam Image Analyser software.

2.2.11.4 Competitive fitness

P. fluorescens strains were purified from glycerol stocks by streaking on KB agar plates and incubated at 28ºC for 48 hours. Eight individual colonies were selected and used to inoculate 4 ml KB in 15 ml Falcon tubes. Tubes were incubated for 16 hours (200 rpm) at 28ºC and vortexed for 1 minute. Fresh 4 ml KB medium in Falcon tubes were inoculated with 4 ul of the vortexed, overnight culture and incubated at 28ºC (200 rpm) for 24 hours. Another round of transfer was done after 24 hours. The cultures were diluted to 10^5 fold and 50 µl and 100 µl were spread on LB agar plates with 60 µg/ml of X-gal. Genotypes were distinguished on the basis of a neutral *lacZ* marker in SBW25 (Zhang & Rainey, 2007). For A506 and Pf0-1, the competitive fitness determination was indirect as the corresponding neutral *lacZ* marker in either of these strains was not available. Each of the A506 and Pf0-1 background strains – both mutants and wild types – were competed against SBW25 strain containing the neutral *lacZ* marker. The relative fitness of each of the A506/Pf0-1 mutant strains (relative fitness computed with respect to SBW25-*lacZ*) was compared to the relative fitness of the A506/Pf0-1 wild type strains (relative fitness computed with respect to SBW25-*lacZ*)

Relative fitness differences during competition experiments were calculated using statistical methods devised by (Lenski *et al.*, 1991). The density of each competitor was calculated at the beginning and end of each competition and these were used to compute the

Malthusian parameter for each strain (natural logarithm of the final density of each strain divided by the corresponding initial density). Relative fitness was quantified as a ratio of Malthusian parameters of competing strains.

2.2.11.5 Quantitative PCR (qPCR)

Quantification of *glyQ* transcripts was performed relative to the quantities of *rpoD* and *recA* in each of the *P. fluorescens* with *glyQ* allele swap as well as the wild type. First, a standard curve was set up by mixing 5 µl of all wild type (SBW25, A506 and Pf0-1) cDNA preparations for each of the background. The mixes were serially diluted up to 10^5 fold. The remaining cDNA as well as the –RT preparations (see section 2.2.9) of each of the mutants and wild type were serially diluted up to 10^3 fold; these would serve as the template for qPCR. A no template control (NTC) was also set up, to check formation of primer dimer; the template was substituted with nuclease-free water in the NTC. Primer pairs were tested at 300 mM and 900 nM for optimal amplification on serially diluted cDNA. The reaction was set up in MicroAmp™ Fast Optical 96-well Reaction Plate, 0.1 ml (Applied Biosystems™, ABI) using 10 µl of Fast SYBR™ Green Master Mix, 1 µl each of the forward and reverse primers and 5 µl of diluted cDNA, volume made up to 20 µl with nuclease free water. Plates were sealed with MicroAmp™ Optical Adhesive Film and run by denaturation at 95ºC for 15 seconds followed by amplification by 40 cycles of denaturation at 95ºC for 1 second and annealing and extension at 60ºC for 20 seconds. This was followed by a run at 95ºC for 15 seconds, 60ºC for 15 seconds and 95ºC for 15 seconds that dissociates the amplified DNA on the basis of size. This is indicative of (i) if amplification has occurred in the –RT controls (indicative of genomic DNA amplification and not cDNA) and (ii) if primer dimers are forming (DNA dissociation in the NTC). The run was performed in the 7900HT Fast Real-Time PCR System with 384-Well Block Module (Thermo Fisher Scientific). Relative quantities were determined using the standard curve. Fold change was determined by taking a ratio of the relative *glyQ* quantities of the mutant strains to that of the corresponding wild type strains.

2.2.12 Statistical analyses

Mean, median, standard deviation and standard error were calculated using Microsoft Office Excel 2011. ANOVA (Analysis of Variance) was used for growth rate measurements in Chapter 4 to determine if the population means were unequal. In cases where the assumptions of normality and homoscedasticity (equal variances) were not met, Kruskal-Wallis test by ranks (non-parametric test) was used. Parametric one-sample *t*-tests and two-samples *t*-tests were used in Chapter 4 to detect deviation of the population mean from a specified value and to determine if the population means were unequal, respectively. All the above-mentioned statistical tests were performed using R 2017 Version 3.4.0.

2.2.13 Calculation of adaptation indices

2.2.13.1 Calculation of CAI

Codon Adaptation Index (CAI) was computed using the *cusp* and *cai* function of The European Molecular Biology Open Software Suite (EMBOSS), a free Open Source software analysis package. EMBOSS version 6.6.0 was used for computing genome-wide CAI of the four strains of *P. fluorescens* – SBW25, A506, Pf0-1 and PICF7, as well as the CAI of *glyQ*, *rpsJ* and *acpP*. CAI of the histidine kinase and diguanylate cyclase genes of the *wsp, aws* and *mws* pathways in the four strains was also determined using this package.

2.2.13.2 Calculation of tAI

A species specific tRNA Adaptation Index (tAI) was determined using the program stAIcalc (Sabi & Tuller, 2014; Sabi *et al*., 2016). This is an improvement over the tAI method described by (Reis et al., 2004) as stAIcalc takes into account the species specific wobble weights. A weight for each tRNA-codon interaction is determined by a function that takes into account the number of tRNA isoacceptors for a particular codon, the gene copy number of tRNAs and the selective constraint which determines the effectiveness of a tRNA-codon interaction. Selective constraint ranges between 0 and 1, which correspond to exact interaction to no interaction, respectively. Geometric mean of the relative weight of each codon-anticodon interaction is then computed by the program (relative to the maximum

weight of interaction); this is the tAI for a particular gene. A method called the hill-climbing approach has been applied to determine the selective constraint; this method seeks to boost the correlation between CUB (codon usage bias) and tAI. The program allows selecting the desired input value for hill-climbing stringency. This stringency value is meant to be an estimate of the number of selective constraint values (s-values) required for optimal tAI determination. I have used the maximum possible value for hill-climbing stringency as it permits the program to inspect the most number of s-values and hence a better scope for estimation of tAI.

Chapter 3 : Bioinformatics characterisation of codon use in *P. fluorescens*

3.1 Introduction

Bioinformatics is a powerful tool that helps make predictions about the possible outcomes of an experimental approach. In my PhD project, I have attempted to use the bioinformatics approach in order to make predictions about the experimental system. I have investigated codon bias in a naturally evolved model system: the bacterial species *Pseudomonas fluorescens*. *P. fluorescens* is a divergent bacterial species that currently encompasses about 100 genome-sequenced strains isolated from plant and soil environments around the world. Our laboratory currently uses four *P. fluorescens* strains: SBW25 (isolated from a sugarbeet leaf in the UK; (Rainey & Bailey, 1996)), PICF7 (from the roots of nursery-bred olive plants (Mercado-Blanco *et al*., 2004)), A506 (from a pear leaf East Coast USA; (Wilson & Lindow, 1993)) and Pf0-1 (from loam soil West Coast USA; (Compeau *et al*., 1988)). These four strains show different degrees of evolutionary relatedness, SBW25 and PICF7 being more closely related than the others (see Figure 1.3). The codon use will therefore have diverged between the four strains, with SBW25 and PICF7 showing more similar codon use patterns than with either A506 or Pf0-1. In other words, differences in codon use would reflect how closely related the four strains are.

One phenotypic feature that can be investigated in the four strains of *P. fluorescens* is their ability to form mats through the two-component signal transduction pathway. Mat formation is a characteristic feature of many species of *Pseudomonas* and is a phenotype that has often been reported for both wild as well as laboratory strains. Within the scope of my PhD project, I am studying mat formation phenotype in the four different strains of *P. fluorescens,* as an attempt to obtain an understanding of the genetic underpinnings of a distinctive phenotype such as mats. As the name suggests, the two-component signal transduction pathway involves two components – histidine kinases (HK) and response regulators (RR) such as diguanylate cyclases (DGCs) (Nixon *et al.*, 1986; Mascher, Helmann, & Unden, 2006; Pérez-Mendoza & Sanjuán, 2016) (see Figure 1.4). Therefore, the similarity (or differences) in the mat formation phenotype of the strains can be investigated in part through the bioinformatics characterization of the genes that make up the HKs and DGCs in these strains.

Additionally, bioinformatic screening of three strains of *P. fluorescens* – SBW25,

A506 and Pf0-1 – revealed exclusive synonymous differences in only three genes. Dr. Frederic Bertels performed the screen. It would be insightful to see how much the codon usage of these three genes – *glyQ, acpP* and *rpsJ* – differs from that of the genome. I carried out experimental work with *glyQ* only, as this is the gene that shows the maximum number of synonymous differences between the strains and hence is predicted to possibly translate into the largest fitness effects (Table 3.1)

Gene	Function	Length (bp)	No. of synonymous mutations		
			SBW25 vs A506	SBW25 vs Pf0-1	A506 vs Pf0-1
glyQ	Glycine tRNA ligase (translation)	954	39	54	67
acpP	Acyl carrier protein (lipid metabolism)	237	2	9	9
rpsJ	30S Ribosomal protein S10 (translation)	312	0	8	8

Table 3.1: Synonymous differences in *glyQ, acpP* and *rpsJ* between SBW25, A506 and Pf0-1. Glycine tRNA ligase, *glyQ,* has a large number of synonymous differences between strains SBW25, A506 and Pf0-1 and hence was selected for use in the experimental approach of the project.

Two other methods that are better suited to understanding codon use across a gene with respect to its genome are the Codon Adaptation Index (CAI) and tRNA Adaptation Index (tAI). CAI first described in the paper by Sharp and Li (Sharp & Li, 1987), has been used extensively as a tool to not only determine how well the genes of an organism are adapted to the genome (Carbone *et al*., 2003; Xia, 2007), but also how well genes of a pathogen might be adapted to the host (Chithambaram, Prabhakaran, & Xia, 2014a; Chithambaram, Prabhakaran, & Xia, 2014b; Khandia *et al*., 2019). It has been used as a quantitative measure to ascertain if certain codons might have a selective advantage as compared to others. For instance, in the case of a host-pathogen interaction, it has been observed that at times viral genomes may reflect very closely the codon usage patterns of the host (Khandia *et al*., 2019). This has been particularly observed in the case of influenza virus strains, wherein the codon usage patterns of mammalian viruses and avian viruses have evolved to match the codon usage patterns of the corresponding hosts (Wong *et al*., 2010). This suggests that codon usage may evolve with respect to selective pressures and the degree of such adaptation can be determined with CAI.

tAI is a more recent method (Reis *et al*., 2004) but already has been implemented as a tool to calculate the translation efficiency of a protein coding sequence with respect to the available tRNAs – both cognate and wobble tRNAs (Sabi & Tuller, 2014), (Sabi *et al*., 2016). This approach measures the how well the codon usage of genes is adapted to the tRNA genes in the genome. In other words, tAI determines the adaptation of genes to the translational demands (tRNA availability) in the genome.

3.2 Aims

1. To determine genome wide codon usage in *P. fluorescens* SBW25, PICF7, A506 and Pf0-1
2. To determine codon usage of the histidine kinases (HKs) and diguanylate cyclases (DGCs) in the four strains of *P. fluorescens*.
3. To determine codon usage of genes of interest – *glyQ, acpP* and *rpsJ*
4. To measure Codon Adaptation Index of the four strains and the three genes of interest.
5. To measure tRNA Adaptation Index of *glyQ, acpP* and *rpsJ.*

3.3 Results

3.3.1 Genome wide codon usage across all four strains of *P. fluorescens* is highly conserved

This study was motivated with the aim to investigate how codon usage differs between closely related species. A codon usage table for genomes of *P. fluorescens* SBW25, *P. fluorescens* PICF7, *P. fluorescens* A506, *P. fluorescens* Pf0-1, *P. aeruginosa* PAO1 and *E. coli* B REL606 was computed using the *cusp* function of The European Molecular Biology Open Software Suite (EMBOSS), a free Open Source software analysis package. EMBOSS version 6.6.0 was used. This was useful for identifying the most preferred codon for amino acids in each species, and also for comparing the relative proportions of different codons used across each genome. *E. coli* tends to show the greatest divergence in codon usage (highlighted

in red in Table 7.1, in Appendix 1.1), while *P. aeruginosa* particularly differs from *P. fluorescens* with respect to the codon for glutamic acid (E). Differences in codon usage between all of the tested *P. fluorescens* strains and the two outgroups are represented graphically below in Figure 3.1.

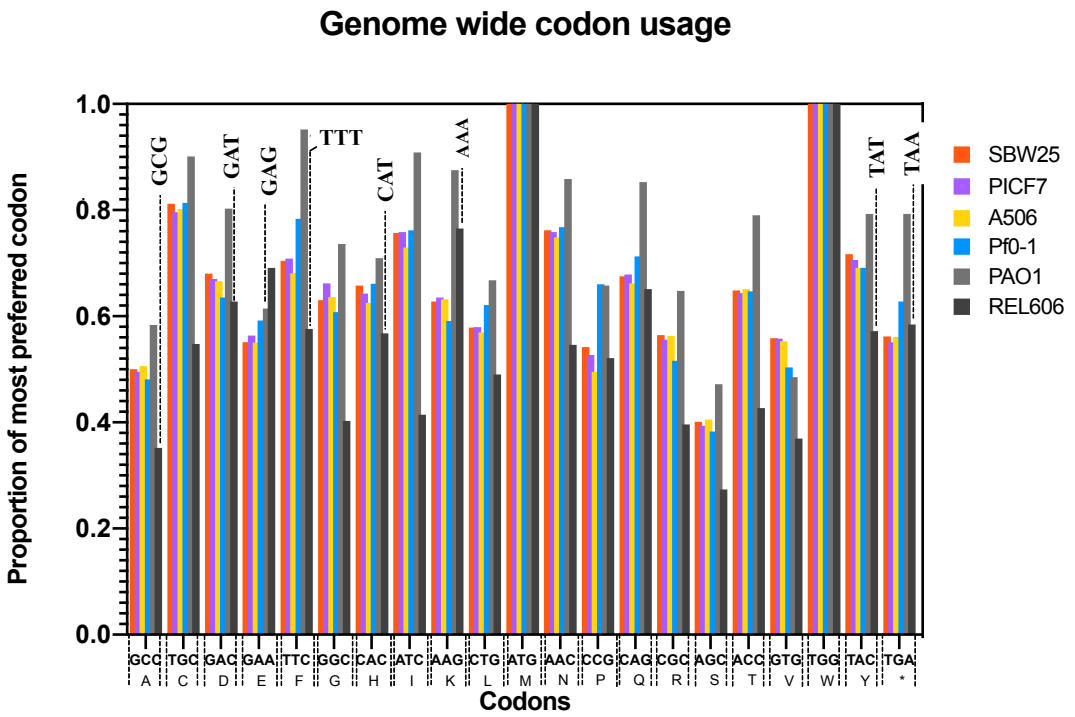


Figure 3.1: Genome-wide usage of the preferred codon for each amino acid in *P. fluorescens* SBW25, PICF7, A506 and Pf0-1 as compared to outgroups, *P. aeruginosa* PAO1 and *E. coli* B REL606. For M and W, which are encoded by only one codon, the frequency is exactly the same across all strains; for all other amino acids, some degree of difference in codon usage proportion is observed across all *P. fluorescens* strains. The differences in codon preference in the outgroups have been indicated in the figure. *=STOP codon.

3.3.2 Codon usage of histidine kinases and diguanylate cyclases in the four strains of *P. fluorescens* exhibit similar patterns

Since correlation between codon usage and transitional output patterns have been reported (Hanson & Coller, 2017), it therefore stands to reason that codon usage has between genes of the same organism might be different. In fact, in the study by (Plotkin, Robins, & Levine, 2004) variation in codon usage patterns of specific genes was observed in different mammalian tissues. The main explanation for the observed variation was attributed to differences in translational demands by different tissues and therefore modulation of protein

output in accordance with the developmental stage of humans. In the scope of my project therefore, genes across each of the four bacterial species were studied. Some of the genes that were screened for differences in codon usage pattern were the HKs and DGCs of the four strains, as mat formation phenotype is a trait common to many species of *Pseudomonas*. Additionally, I also investigated the codon use of the three genes that exclusively exhibit variation in synonymous codon use between SBW25, A506 and Pf0-1. Codon usage table for all of the genes was computed using the *cusp* function of The European Molecular Biology Open Software Suite (EMBOSS), a free Open Source software analysis package. EMBOSS version 6.6.0 was used (see section 2.2.13.1). Figures 3.2 and 3.3 shows the genome wide codon usage as well as codon usage for each of HKs and DGCs of the *wsp* operon in *P. fluorescens* SBW25, PICF7, A506, and Pf0-1, respectively. The codon usage these genes vary in terms of both the preferred codon and its usage proportion across the genes. The amino acids for which the codons differ with respect to the genome wide codon usage have been labelled in the figures.

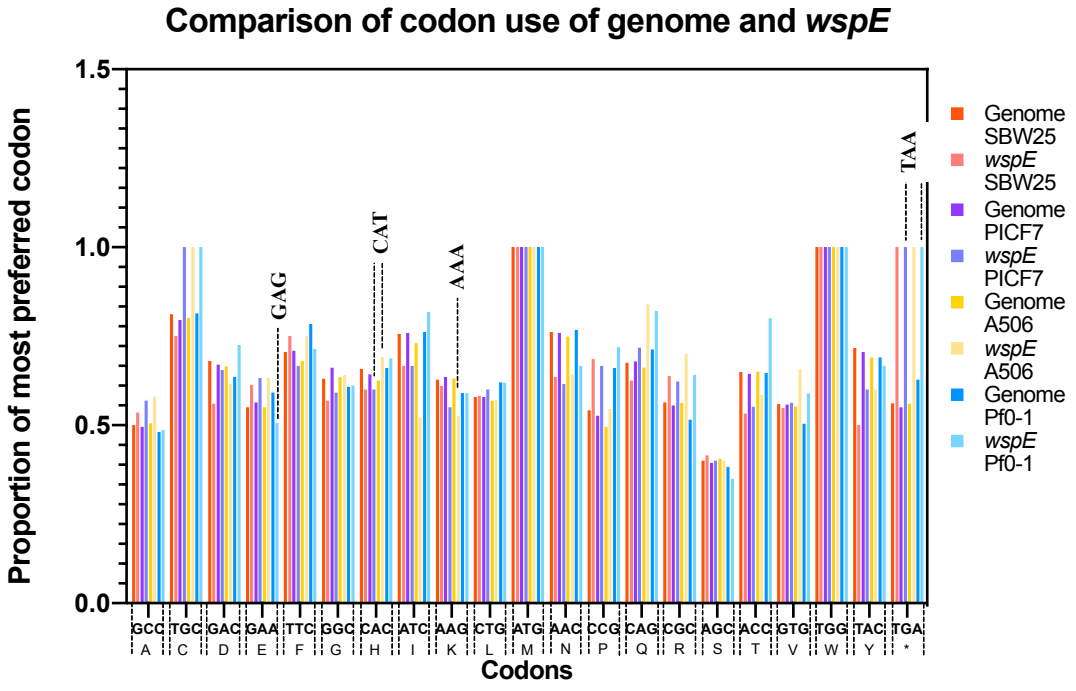


Figure 3.2: Comparison of the codon use across the genome and *wspE* genes of *P. fluorescens* SBW25, PICF7, A506 and Pf0-1. For most amino acids the most preferred codon remains the same across the genomes as well as for *wspE*. The amino acids for which the most preferred codon differs with respect to the genome are labeled – GAG instead of GAA for Asp (E) in Pf0-1 *wspE*; CAT instead of CAC for His (H) in both PICF7 and A506 *wspE*; AAA instead of AAG for Lys (K) in A506 *wspE* and TAA instead of TGA for STOP (*) in both PICF7 and Pf0-1 *wspE*. For M and W, which are

encoded by only one codon, the frequency is exact in all three; for all other amino acids, some degree of difference is observed across all three strains. *=STOP codon

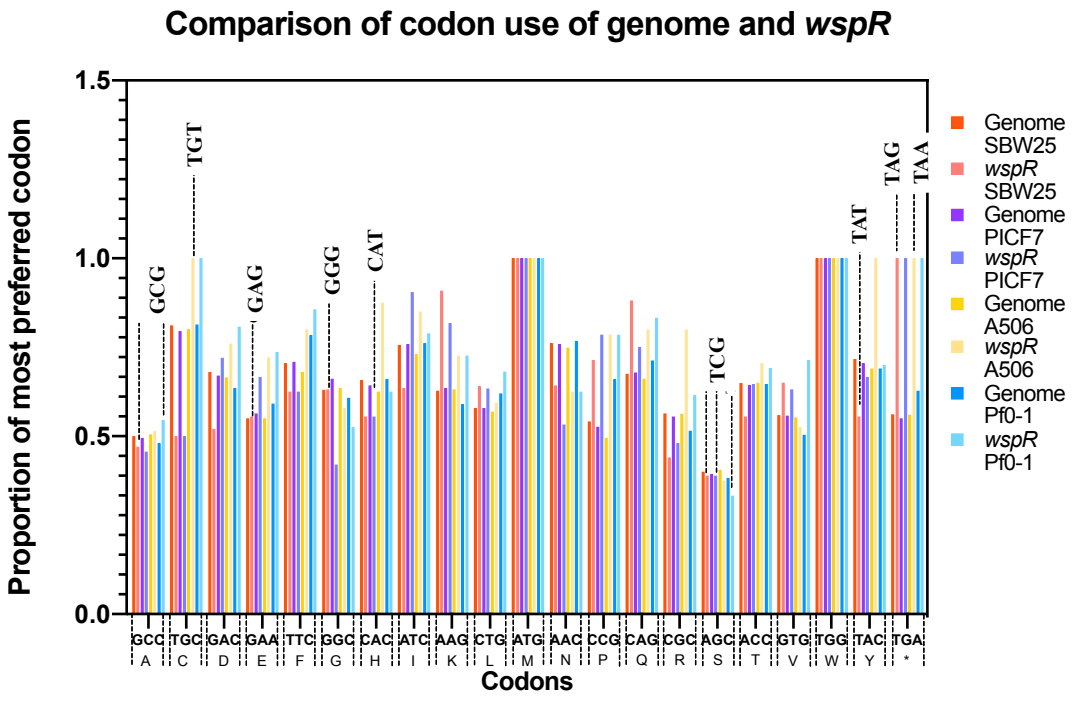


Figure 3.3: Comparison of the codon use across the genome and *wspR* genes of *P. fluorescens* SBW25, PICF7, A506 and Pf0-1. For most amino acids the most preferred codon remains the same across the genomes as well as for *wspR*. The amino acids for which the most preferred codon differs with respect to the genome are labeled – GCG instead of GCC for Ala (A) in both SBW25 and Pf0-1 *wspR*; TGT instead of TGC for Cys (C) in A506 *wspR*; GAG instead of GAA for Asp (E) in SBW25 *wspR*; GGG instead of GGC for Gly (G) in SBW25 *wspR*; CAT instead of CAC for His (H) in both PICF7 *wspR*; TCG instead of AGC for Ser (S) in SBW25, PICF7 and A506 *wspR*; TAT instead of TAC for Tyr (Y) in SBW25 *wspR* and TAG and TAA instead of TGA for STOP (*) in SBW25 and A506 *wspR*, respectively. For M and W, which are encoded by only one codon, the frequency is exact in all three; for all other amino acids, some degree of difference is observed across all three strains. *=STOP codon.

Not only are these codon use differences observed in the HKs and DGCs of the *wsp* operon, but also in the corresponding genes of the *aws* operon (Figures 3.4 and 3.5). Again, the amino acids for which the codons differ with respect to the genome wide codon usage have been labelled in the figures. Pf0-1 does not harbour the *aws* operon and hence this is not shown on the graphs.

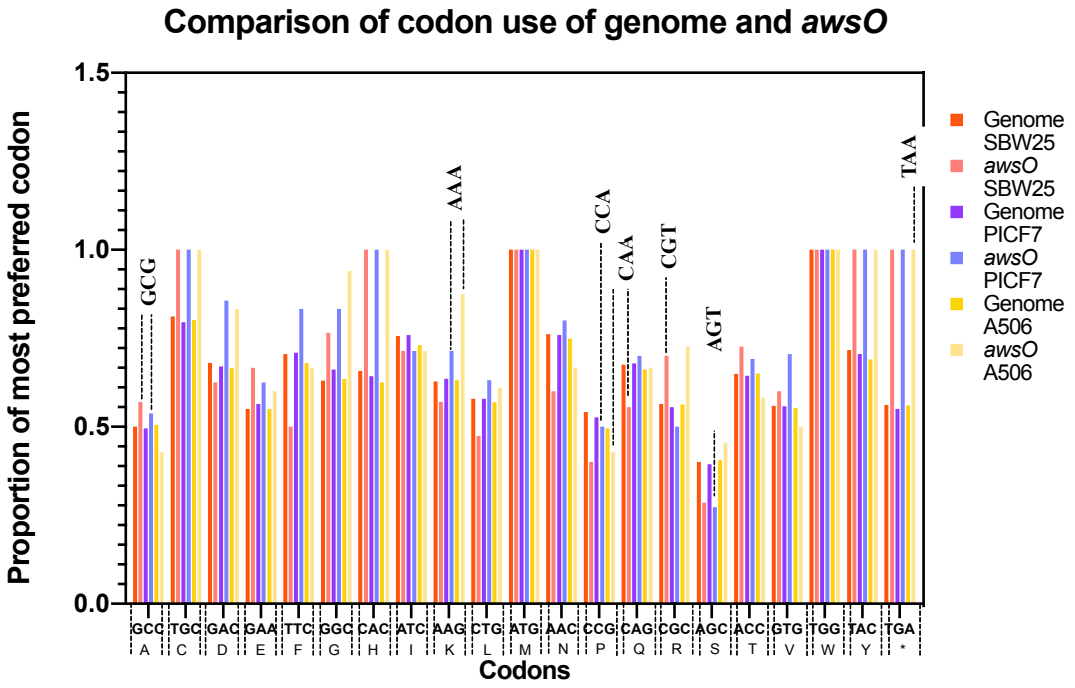


Figure 3.4: Comparison of the codon use across the genome and *awsO* genes of *P. fluorescens* SBW25, PICF7 and A506. For most amino acids the most preferred codon remains the same across the genomes as well as for *awsO*. The amino acids for which the most preferred codon differs with respect to the genome are labeled – GCG instead of GCC for Ala (A) in both SBW25 and PICF7 a*wsO*; AAA instead of AAG for Lys (K) in both PICF7 and A506 *awsO*; CCA instead of CCG for Pro (P) in both PICF7 and A506 *awsO*; CAA instead of CAG for GlN (Q) in SBW25 *awsO*; CGT instead of CGC for Arg (R) in SBW25 *awsO;* AGT instead of AGC for Ser (R) in PICF7 *awsO* and TAA instead of TGA for STOP (*) in A506 *awsO*, respectively. For M and W, which are encoded by only one codon, the frequency is exact in all three*;* for all other amino acids, some degree of difference is observed across all three strains. *=STOP codon.

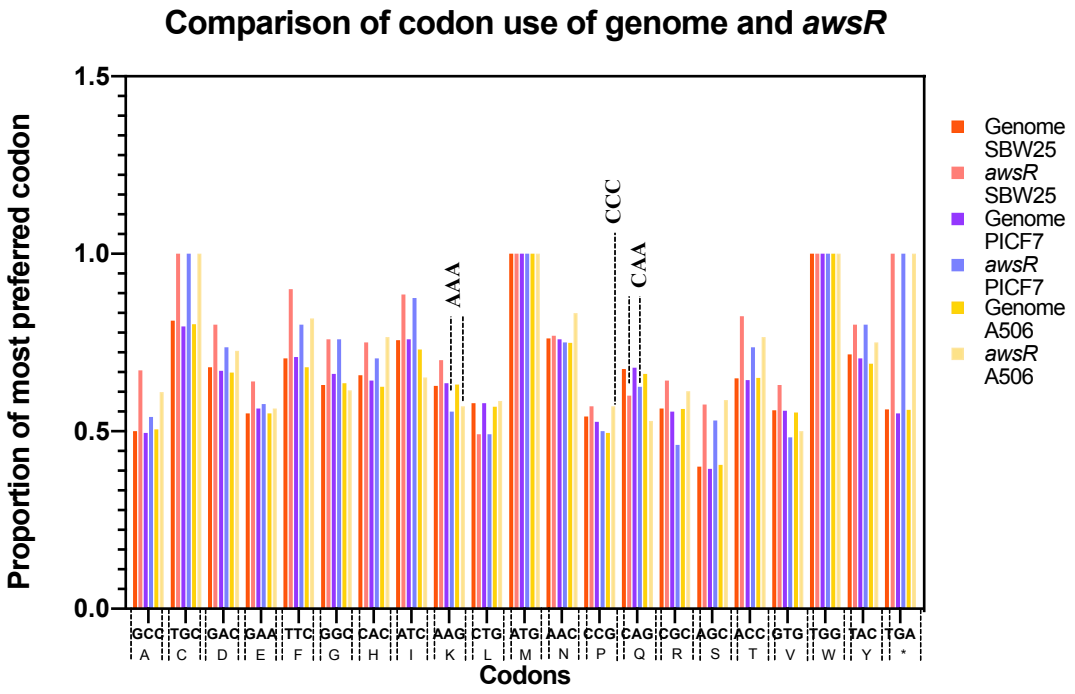


Figure 3.5: Comparison of the codon use across the genome and *awsR* genes of *P. fluorescens* SBW25, PICF7 and A506. For most amino acids the most preferred codon remains the same across the genomes as well as for *awsO*. The amino acids for which the most preferred codon differs with respect to the genome are labeled – AAA instead of AAG for Lys (K) in both PICF7 and A506 *awsR*; CCC instead of CCG for Pro (P) in A506 *awsO*; CAA instead of CAG for GlN (Q) in both SBW25 and PICF7 *awsR*. For M and W, which are encoded by only one codon, the frequency is exact in all three; for all other amino acids, some degree of difference is observed across all three strains. *=STOP codon.

The negative regulator of the *wsp* operon, *wspF,* also shows differences in codon use across the four strains of *P. fluorescens* (Figure 3.6). Although across all of the strains most of the codons used are the same for all amino acids, the proportions of each of these codons across the strains for *wspF* are different.

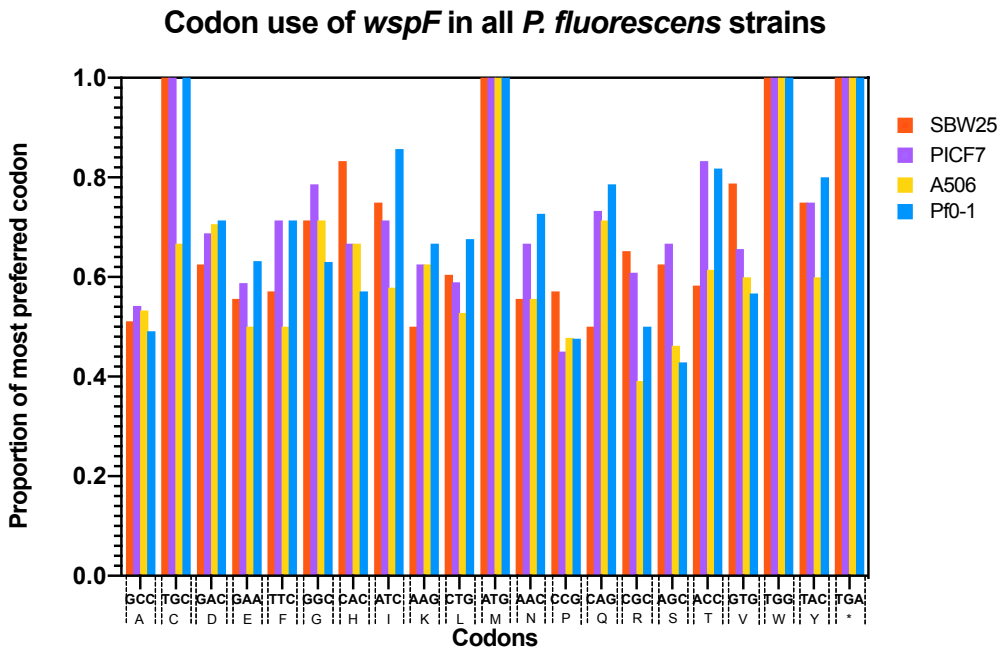


Figure 3.6: Codon usage of *wspF* gene for each amino acid in *P. fluorescens* SBW25, PICF7, A506 and Pf0-1. For M and W, which are encoded by only one codon, the frequency is exact in all three; for all other amino acids, some degree of difference is observed across all three strains. *=STOP codon.

There are certain advantages associated with production of mats and these become more apparent under various environmental stressors (Davey & O'Toole, 2000; Spiers *et al.*, 2013). The genes involved in mat formation may not be the most highly expressed but considering that the strains of *P. fluorescens* used in this study had been isolated from the wild, a plethora of environmental stresses must have been met by these strains and the genome (and individual genes) would have been shaped by these encounters. However, each of the strains was isolated from a different environmental niche and the expectation is that although the codon adaptation of the genes from each strain would match the host (genome), there would be intrinsic differences in the codon adaptation between genomes (and the genes therein). To illustrate this point, codon adaptation index (CAI) of *wspF* was computed and has been compared between the strains (Figure 3.7). CAI computed using the *cusp* and *cai* functions of The European Molecular Biology Open Software Suite (EMBOSS), a free Open Source software analysis package. EMBOSS version 6.6.0 was used (see section 2.2.13.1).

CAI of *wspF* in all of the strains is different, which is indicative of the different evolutionary forces that might have shaped these organisms. It also reflects that similar environmental conditions may results in similar optimal codon usage. From Figure 3.7 it appears that among the four strains, *wspF* in A506 is most adapted to its genome wide codon usage, with a CAI value of 0.694 while that in PICF7 *wspF* is least adapted to the codon usage of PICF7 genome, with a CAI value of 0.668. SBW25 and Pf0-1 had very similar CAI values of 0.687 and 0.686, respectively. Considering that all of the strains were isolated from different environmental niches, it is not surprising to see the CAI of *wspF* is different in the different strains.

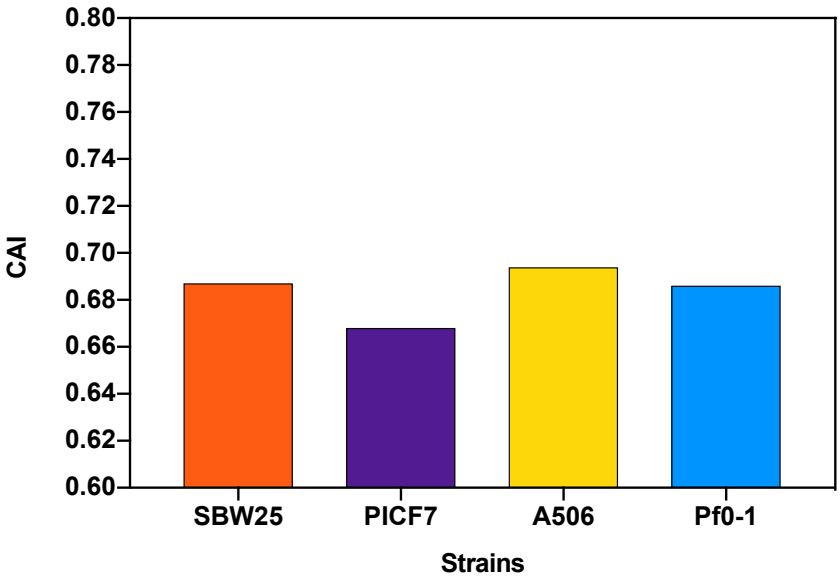


Figure 3.7: Codon Adaptation Index (CAI) of *wspF* of four strains of *P. fluorescens.* The CAI of *wspF* in all of the four strains is different. CAI of Pf0-1 and SBW25 are most similar, while that of A506 is the highest and PICF7 is the least.

3.3.3 Codon usage of the genes of interest – *glyQ, acpP* and *rpsJ* – exhibit some degree of differences in codon preference

As mentioned in section 3.1, the genomes of the three strains of *P. fluorescens* – SBW25, A506 and Pf0-1 – were screened for genes that exclusively have synonymous differences. The screen yielded three candidate genes – *glyQ, acpP* and *rpsJ*. The aim of the project is to understand the functional implications of biased codon usage on the physiology of a bacterial species. This will be achieved through experimentally swapping the naturally evolved alleles of *glyQ*, in non-native backgrounds.

The codons and their respective proportions for each of the genes and the backgrounds would indicate how well the codon use of the genes conforms to that of the genome. Codon usage table for the three genes was computed using the *cusp* function of The European Molecular Biology Open Software Suite (EMBOSS), a free Open Source software analysis package. EMBOSS version 6.6.0 was used (see section 2.2.13.1). Figure 3.9 shows the codon usage and proportion of each preferred codon for each of the genes – *glyQ, acpP* and *rpsJ* – as compared to the codon usage across the genome of SBW25. As can bee seen in Figure 3.8, the codon use in each of the genes is different from the genome wide codon usage in SBW25.

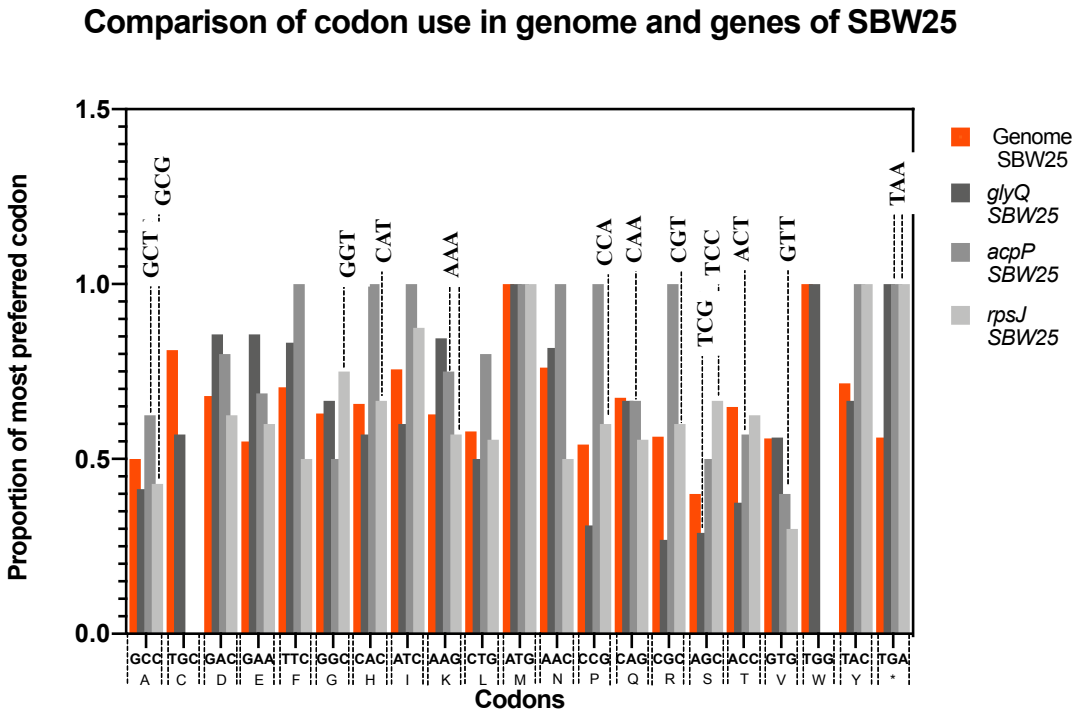


Figure 3.8: Codon usage of *glyQ*, *acpP* and *rpsJ* of SBW25 as compared to the genome wide codon usage in SBW25. For M and W, which are encoded by only one codon, the frequency is exact in all three; for all other amino acids, some degree of difference is observed across all three genes. *=STOP codon. For some of the amino acids, both the proportion as well the preference of codon is different. These have been marked on the plot with the preferred codon.

Again, not only are these inter gene differences observed in SBW25, but are seen in A506 and Pf0-1 as well, as seen in Figures 3.9 and 3.10, respectively.

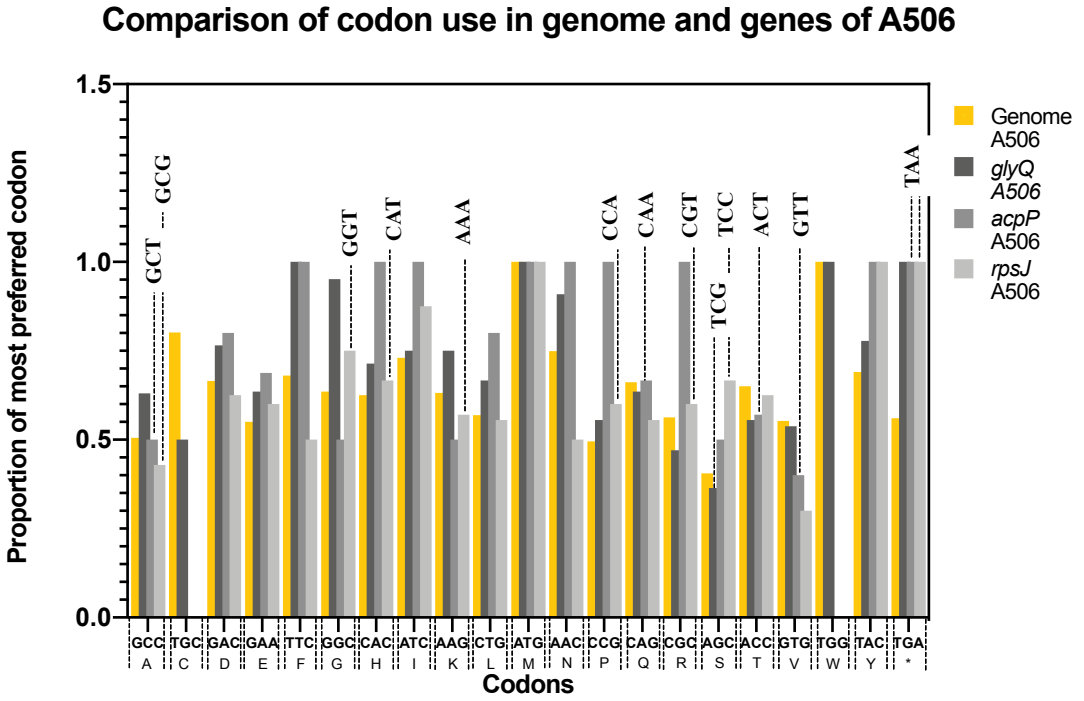


Figure 3.9: Codon usage of *glyQ*, *acpP* and *rpsJ* of A506 as compared to the genome wide codon usage in A506. For M and W, which are encoded by only one codon, the frequency is exact in all three; for all other amino acids, some degree of difference is observed across all three genes. *=STOP codon. For some of the amino acids, both the proportion as well the preference of codon is different. These have been marked on the plot with the preferred codon.

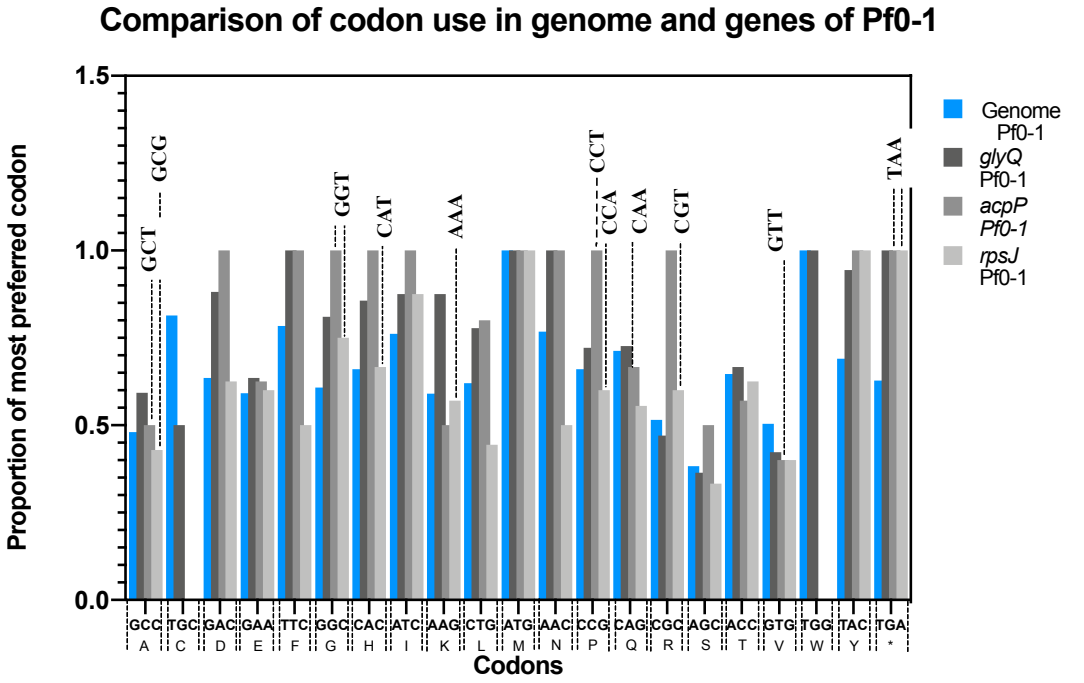


Figure 3.10: Codon usage of *glyQ*, *acpP* and *rpsJ* of Pf0-1 as compared to the genome wide codon usage in Pf0-1. For M and W, which are encoded by only one codon, the frequency is exact in all three; for all other amino acids, some degree of difference is observed across all three genes. *=STOP

codon. For some of the amino acids, both the proportion as well the preference of codon is different. These have been marked on the plot with the preferred codon.

Of the three genes, *glyQ* was used for the experimental work and hence a comparison of the codon use of *glyQ* between SBW25, A506 and Pf0-1 might shed light on the possible outcomes of swapping *glyQ* alleles into non-native backgrounds. Figure 3.11 shows a comparison of *glyQ* codon use between SBW25, A506 and Pf0-1. Not surprisingly, the codon usage pattern of *glyQ* is different in the three strains – although most of the preferred codons for *glyQ* in the three strains are the same in all, the proportion of those codons are quite different. Some of the starker differences are in terms of proportion of codon use of alanine (A), glutamic acid (E), glycine (G), histidine (H), lysine (L), proline (P), arginine (R), threonine (T) and tyrosine (Y). For most of the amino acids, the largest differences are between the proportions of the preferred codon between SBW25 and Pf0-1, hinting towards a higher probability of observing the largest fitness effects upon swapping *glyQ* alleles between these two strains.

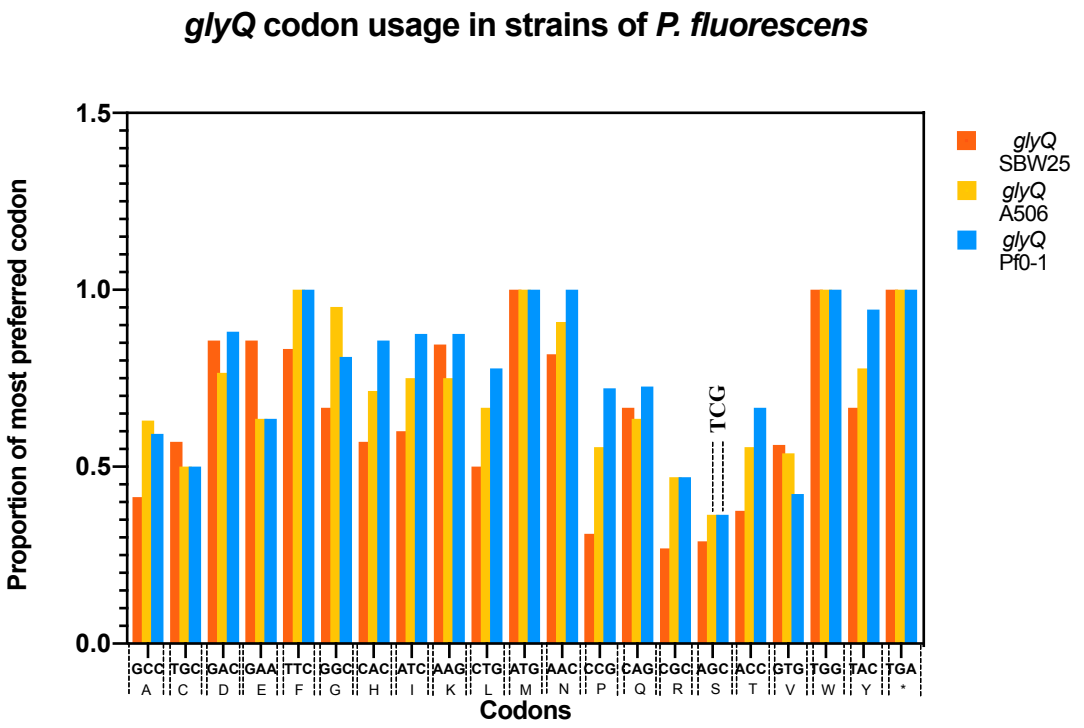


Figure 3.11: Comparison of *glyQ* codon use in the three strains of *P. fluorescens*. Again, for M and W, which are encoded by only one codon, the frequency is exact in all three; for all other amino acids, some degree of difference is observed across all three genes. *=STOP codon. For most amino acids, the

most preferred codons were the same but the proportions varied, with much stronger variation in the case of alanine, glutamic acid, glycine, histidine, to name a few.

3.3.4 Codon Adaptation Index (CAI) varies across genes and genomes

Codon Adaptation Index (CAI) has often been used to measure how well the codons of a particular gene are adapted to the genome wide codon usage in an organism. It is the most widely used method for measuring codon usage bias. It is a quantitative method for predicting the level of expression of a protein coding sequence based on its codons.

I have applied this method for quantifying the codon usage bias of the genomes of the four strains to predict the allele swaps that might have the largest effects on gene function. The allele swapping however was restricted to three strains only – SBW25, A506 and Pf0-1. For instance, if *glyQ* is swapped from A506 into SBW25 and SBW25 codon adaptation index is higher than that of A506, then in a particular environment the native SBW5-*glyQ* combination might have higher *glyQ* levels than the non-native SBW25-*glyQ* combination. If the cellular levels of *glyQ* are extrapolated for growth, since *glyQ* generates the protein glycine tRNA ligase and has the potential to effect global expression of glycine rich genes as a consequence of affecting the glycine-tRNA pool, it might be said that the native SBW5-*glyQ* combination might grow faster than the non-native SBW25-*glyQ*.

For computing both the genome-wide CAI and the CAI of the genes, I have used the *cusp* and *cai* functions of The European Molecular Biology Open Software Suite (EMBOSS) (see section 2.2.13.1)

3.3.4.1 Codon Adaptation Index of the genomes

The CAIs obtained for each of the genomes were plotted on one graph as four cumulative frequency distributions (Figure 3.12). This figure thus compares the CAI values of all four strains and gives an indication of the degree of relatedness of the strains. For example the lines representing SBW25 and A506 are almost overlapping throughout the distribution. This could indicate that for the majority of protein coding genes, the two strains have similarly optimized codon usage (and possibly also similar gene expression levels under various conditions).

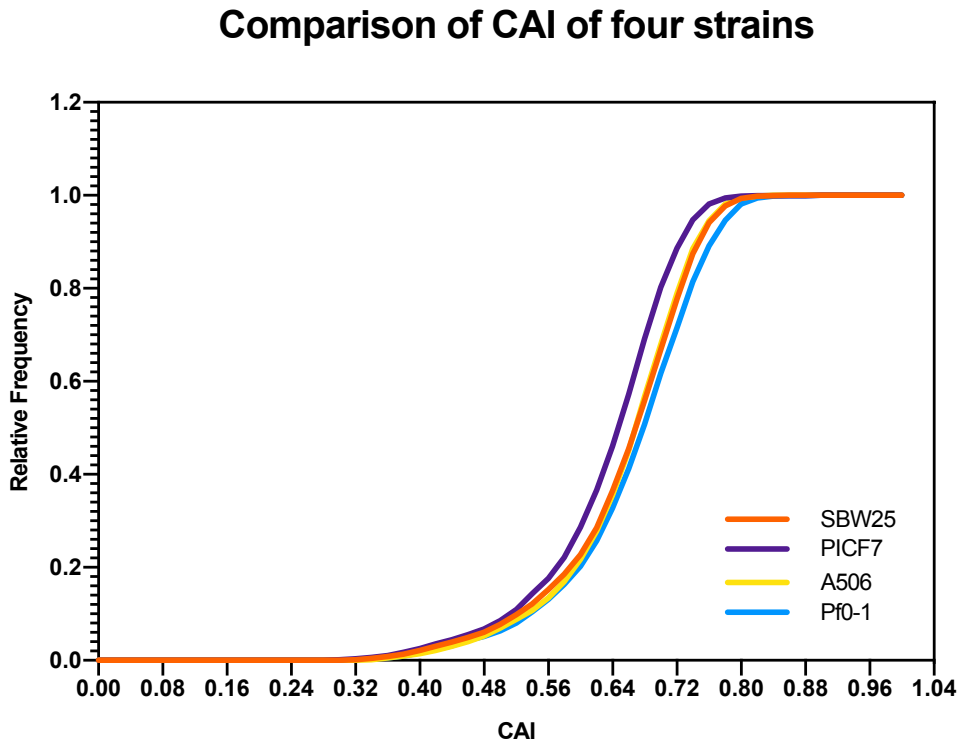


Figure 3.12: Cumulative frequency of CAI across the protein-coding genomes of *P. fluorescens* SBW25, PICF7 A506 and Pf0-1. The figure shows the relatedness of the four strains with respect to optimal codon usage. SBW25 and A506 appear to be most closely related, with PICF7 and Pf0-1 differing to a greater extent.

3.3.4.2 Codon adaptation index of the three genes of interest

The CAIs of the individual genes, *glyQ, acpP* and *rpsJ*, were also calculated. Then, each strain's set of three CAIs was plotted against that of the other two strains (Figure 3.13). In this process, a higher CAI value corresponds to more "optimal" codon usage (and hence may indicate higher expression of a gene). First, the analysis showed a high degree of correlation between the three genes in the four strains. More specifically, the CAI value for *glyQ* is around 0.72 in SBW25, PICF7 as well as A506 and 0.78 in Pf0-1, hinting that *glyQ* may be more highly expressed in Pf0-1 as compared to the three other strains. The *acpP* gene has a lower value in SBW25, PICF7 and Pf0-1 than A506, indicating the gene may be more highly expressed in A506 than the others. As evident from the plots 3.13C, 3.13E and 3.13F, *rpsJ* gene has the lowest CAI in Pf0-1. This may indicate that *rpsJ* has the least expression in Pf0-1 as compared to the other strains and from the CAI plots it may appear that among the three genes of interest, *rpsJ* may have the lowest expression in all of the four strains.

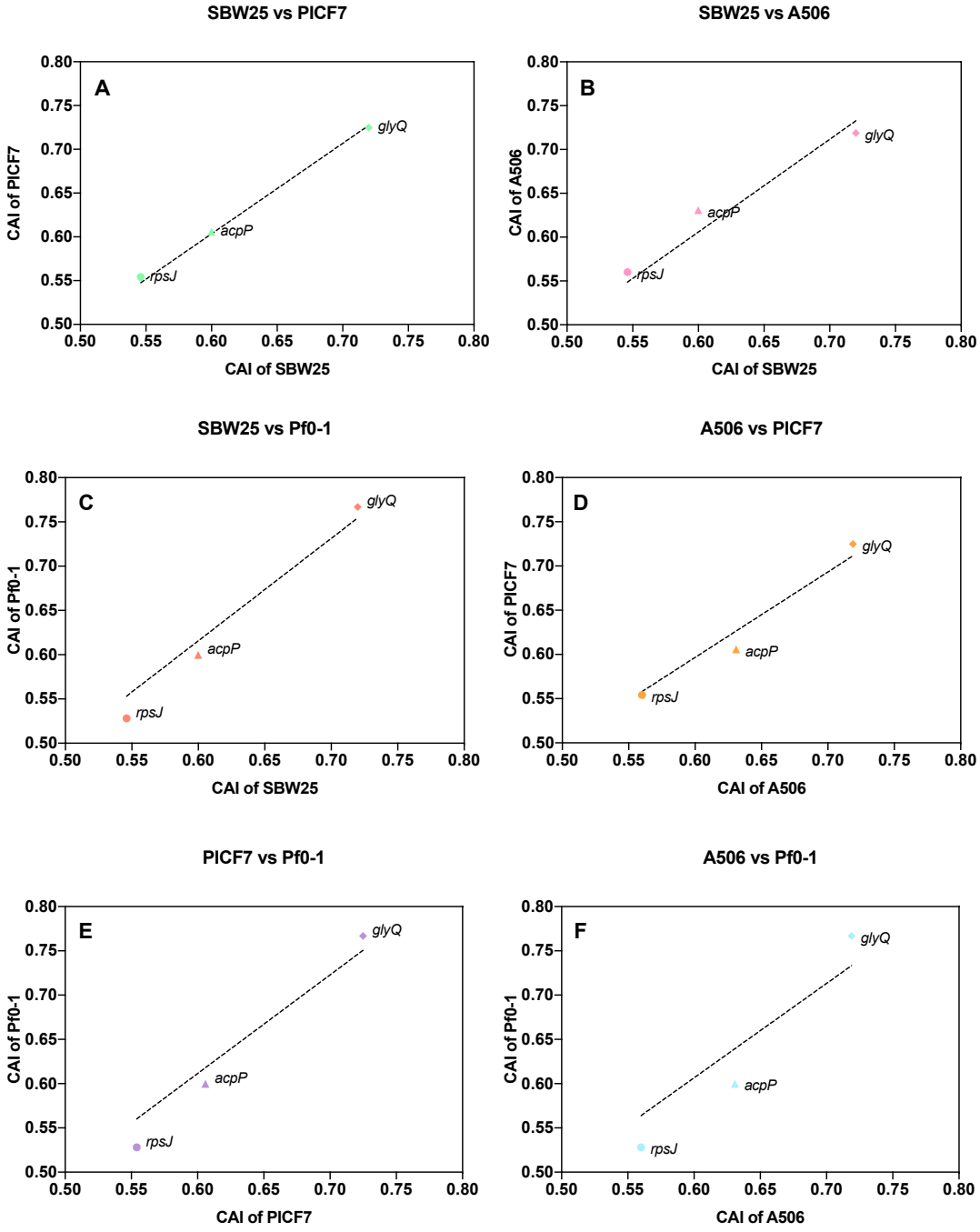


Figure 3.13: A comparison of the CAI of *glyQ, acpP* and *rpsJ* in each pair of *P. fluorescens* SBW25, PICF7, A506 and Pf0-1. The CAI of the genes was compared for two strains at a time. Values for *glyQ, acpP* and *rpsJ* are around 0.75, 0.6 and 0.54 respectively. A. CAI of SBW25 vs PICF7. B. CAI of SBW25 vs A506. C. CAI of SBW25 vs Pf0-1. D. CAI of A506 vs PICF7. E. CAI of PICF7 va Pf0-1. F. CAI of A506 vs Pf0-1.

3.3.5 tRNA Adaptation Index (tAI) varies across genomes

tRNA Adaptation Index (tAI) has been recently used extensively across various studies to determine if the codon usage of a particular gene is well correlated with the tRNA availability such that the translational output is optimum (Tuller *et al*., 2010a; Tello *et al*., 2013; Sabi & Tuller, 2014; Presnyak *et al*., 2015). I have used the stAI calc program by (Sabi & Tuller, 2014; Sabi et al., 2016) and used the maximum stringency possible to determine the tAI (see section 2.2.13.2). This program is an improvement over the previous method as it computes a species-specific weight for the tRNA and codon interactions; both Watson-Crick as well as wobble.

A low tAI value suggests that the gene is possibly less adapted to the tRNA pool of the organism, while a higher value is suggestive of better adaptability. Figure 3.14 shows a distribution of the tAI values across the SBW25 genome. The tAI values across the SBW25 genome range from 0.237 to 0.509. As can be seen in Figure 3.14, the tAI values for *glyQ* and *awsO* are in the range of 0.34 and as per the distribution, this places the genes towards a higher tAI value. This is suggestive that *glyQ* and *awsO* might be better adapted to the tRNA content in SBW25. On the other hand, *rpsJ* has the least tAI value (0.296) among the genes that are indicated on the figure. This is suggestive that *rpsJ*, unlike the two aforementioned genes might be least adapted to the tRNA content of SBW25. The other genes – *acpP, wspE, wspF, wspR* and *awsR* – lie within the lower and higher range set by *rpsJ* and *glyQ/awsO*, respectively. Table for tAI values are in Appendix 1.6.

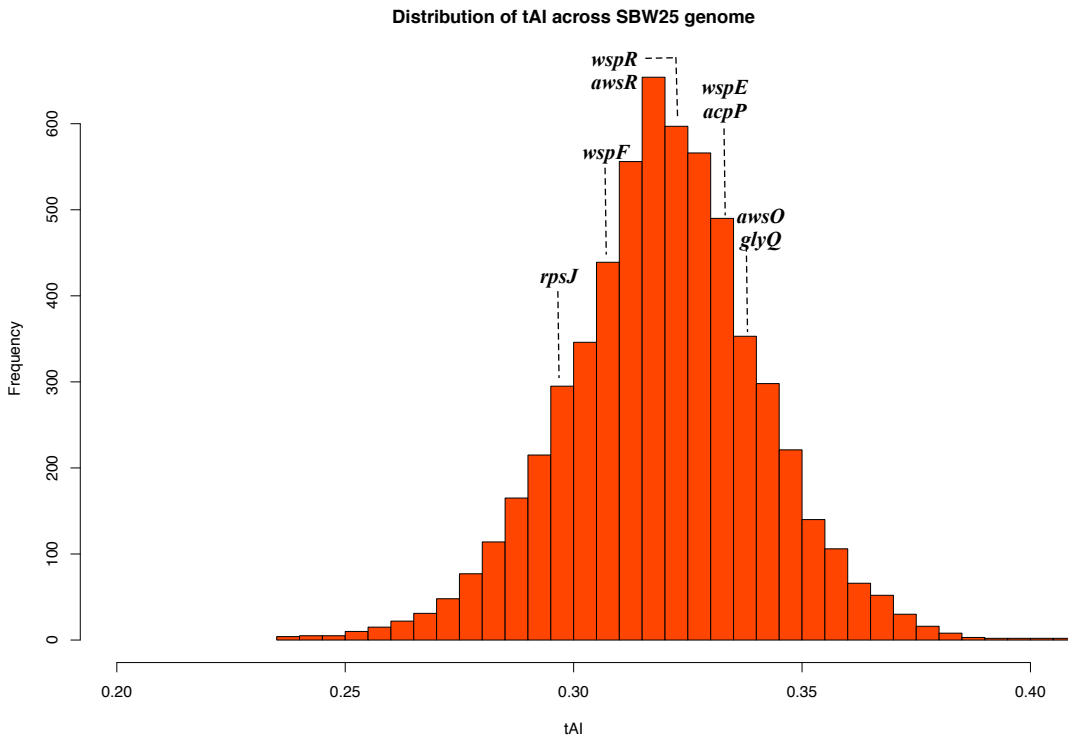


Figure 3.14: Distribution of tAI across the genome of SBW25. While *rpsJ* has a lower tAI value with respect to the other genes, *glyQ* and *awsO* have the higher tAI values. The other genes fall in between these two limits.

In the case of A506 (Figure 3.15), we see a different scenario. While, *glyQ* still has the higher tAI value (0.278) among the genes that were examined in the scope of this study, the lower tAI value is now exhibited by *wspR* (0.211). This is suggestive that *glyQ* might be most adapted to the tRNA content of A506, while *wspR* might be the less adapted to the tRNA content. The variation in the tAI distribution pattern, however, might be explained in part by the fact that the program used for computing the tRNA adaptation indices computes this parameter based on species-specific codon-anticodon interaction. Depending on the number and role of both Watson-Crick and wobble interactions in A506, the tAI of the genes may differ from that of another organism. Of course, the degree of translational requirement of the gene within an organism may also depend on the factors that have shaped its genome evolution. This is turn might be reflected in the tRNA adaptation index – less translation need/output might correspond to a lower tAI value. The tAI values for the genes in A506 are also lower in general than the corresponding values for SBW25.

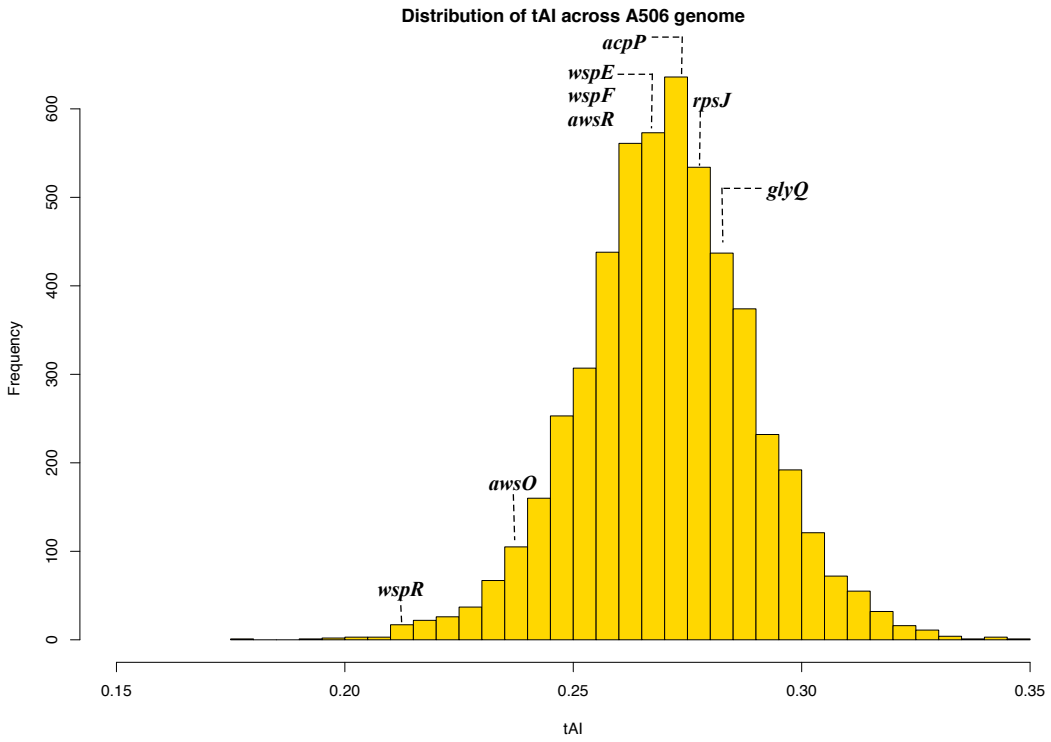


Figure 3.15: Distribution of tAI across the genome of A506. While *wspR* has a lower tAI value with respect to the other genes, *glyQ* has the higher tAI value. The other genes fall in between these two limits. The tAI values range from 0.179 to 0.348

In Pf0-1, the tAI values range from 0.14 to 0.426. Again, the pattern in Pf0-1 differs from that of both SBW25 and A506 (Figure 3.16), which can be in part explained by the method used and also possibly by the evolutionary forces that shape the genome of Pf0-1. In other words, niche-specific factors may shape the tRNA availability, gene codon content and the codon-anticodon interactions therein. In Pf0-1, the gene exhibiting the higher tAI values among others is *wspF*, while the lower values are exhibited by *wspE* and *wspR*. Again, suggestive that while *wspF* might be the more adapted, *wspE* and *wspR* might be less adapted to the tRNA content of Pf0-1. The other genes have intermediate tAI values between these limits. Pf0-1 does not contain the *aws* locus and hence genes of this operon are not mentioned in Figure 3.16.

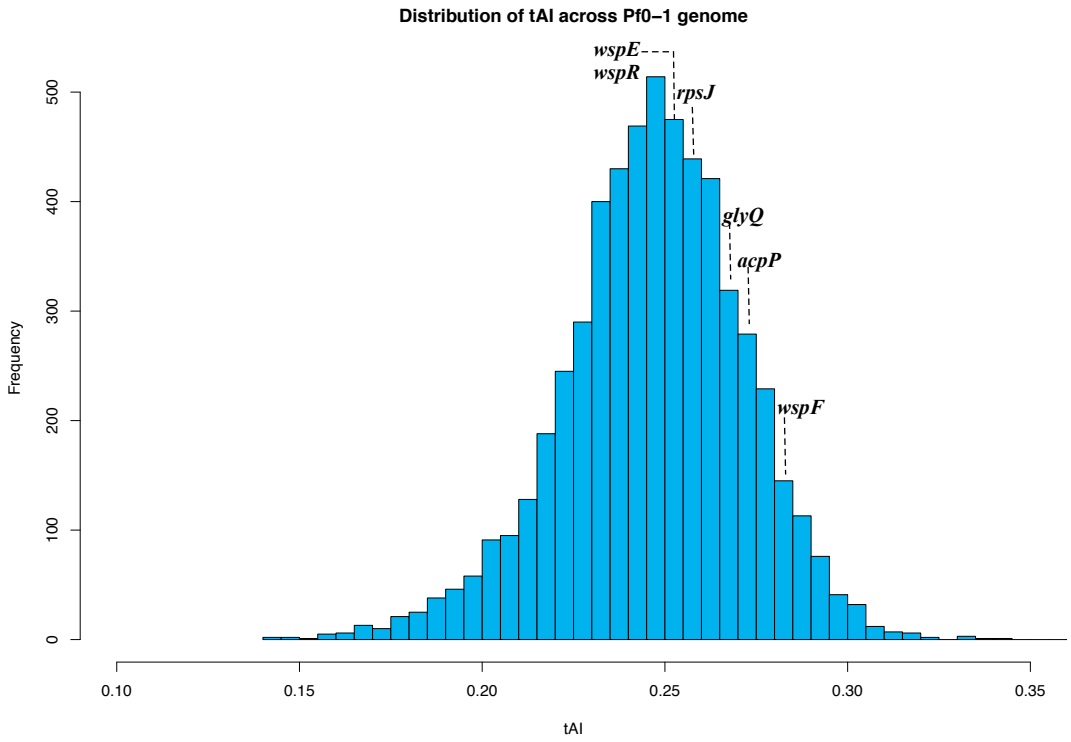


Figure 3.16: Distribution of tAI across the genome of Pf0-1. While *wspE/wspR* has a lower tAI value with respect to the other genes, *wspF* has the higher tAI value.

The tAI values across PICF7 genome ranges from 0.141 and 0.357 and the higher tAI value is exhibited by *wspF* (0.29) while the lower tAI value is exhibited by *awsR* (0.202); as iterated before, this could be indicative of *wspF* and *awsR,* to be more and less adapted to the tRNA content in PICF7. The variation in tAI distribution across PICF7 genome differs from that of three other strains (Figure 3.17). The reason to that is probably similar as has been mentioned for both A506 and Pf0-1 – number and role of strain-specific codon-anticodon interactions might lead to differences in tAI.

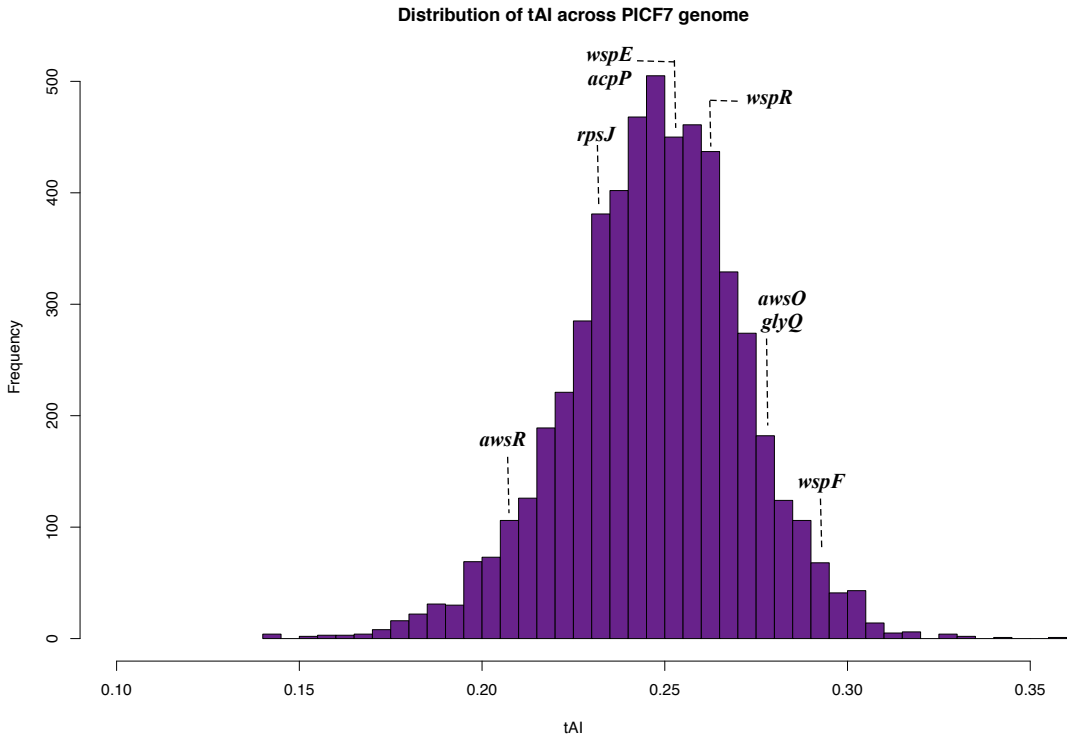


Figure 3.17: Distribution of tAI across the genome of PICF7. While *awsR* has a lower tAI value with respect to the other genes, *wspF* has the higher tAI value.

3.4 Discussion

3.4.1 Codon usage in *P. fluorescens* SBW25, PICF7, A506 and Pf0-1

As has been reported for several organisms before, we see varied codon usage patterns in the four strains of *P. fluorescens*. There are genome wide differences in codon usage patterns. The codon usage patterns in SBW25 and PICF7 are more similar, while the variation in codon usage increased with decreasing relatedness of the strains. The two outgroups – *P. aeruginosa* PAO1 and *E. coli* B REL606, showed the most variation with respect to SBW25. *P. aeruginosa* had more similarity in codon usage pattern with the strains of *P. fluorescens* as opposed to *E. coli.* These results conform to the evolutionary distance of each of the strains and of outgroups to the strains. The preferences for the codons for every amino acid across the four genomes was similar but the degree of preference varied across them. The

environment from which these strains were isolated might have shaped the intrinsic differences in codon use across the four strains. The variation in codon use reflects not only the evolution of the strains in distinct environments, but also the degree of relatedness of the strains.

3.4.2 Codon usage of histidine kinases and diguanylate cyclases in *P. fluorescens*

As mentioned in section 3.3.2, mat formation phenotype in a trait that is distinctive of many species of *Pseudomonas*. The codon usage patterns of some of the genes that are central to this pathway in *P. fluorescens* were studied. The HKs and DGCs of each of the strains, not surprisingly show differences in codon usage patterns as compared to the genomes. Overall the preferential codons for most amino acids are the same. For some amino acids, such as alanine, glutamine and arginine in the *awsO* gene of SBW25, the codon that is preferred most is different from that of the genome wide preference. Exact reason for why this exists is unknown, but it may be speculated that in the natural environment of SBW25, this codon preference in *awsO* contributes to mRNA stability, translational speed and/or accuracy or protein folding or a certain combination of all the factors.

3.4.3 Codon usage of the genes of interest

The genes of interest include *glyQ, acpP* and *rpsJ*. As seen for HKs and DGCs, each of these also shows variations in codon usage as compared to the genome. Among the three genes, *glyQ* appears to have the most similar codon usage with respect to the genome. While the variation in *acpP* and *rpsJ* are greater. One explanation could be that *glyQ* is an essential gene that is required in the exponential growth phase of the bacteria. It codes for glycine tRNA ligase and hence has the capacity to modulate the cellular tRNA pool for glycine. Consequently, a codon use pattern that matches the genome is advantageous, as that would enable fast translation of the protein under exponential growth.

The other two genes, *acpP* and *rpsJ,* code for acyl-carrier protein and 30S ribosomal protein S10, respectively. The argument for need in exponential growth stands for *rpsJ* as well, if not for *acpP*. The 30S ribosomal protein S10 is coded for by *rpsJ* and this protein has an essential role in translational elongation as it can cross link with tRNAs in the P-site of the

ribosome (Riehl *et al.*, 1982). S10 has also been shown to be an important component of the antitermination complex; S10 and NusB together form the antitermination complex for antitermination of the *rrn* operon (Luo *et al*., 2008; Burmann *et al*., 2010). S10 either forms the antitermination complex or is part of the ribosomal complex, but cannot do both simultaneously. S10 is also an unstructured protein, which endows it with the flexibility to adapt to varying functional needs – translation or transcription. Since protein folding can be affected by the codon composition, it might be that the codon use pattern of *rpsJ* is a reflection of the (un)structural demands of the protein.

3.4.4 Codon Adaptation Index

The codon adaptation index would reflect how well the genes are adapted to the codon use patterns of the genome.

3.4.4.1 Codon Adaptation Index of the genomes

From the phylogenetic tree it appears that SBW25 and PICF7 are more closely related, but the CAI indicates differently. As mentioned before, CAI is a measure that reflects how well adapted a gene might be to the genome wide codon usage. Similar CAI values across genomes might indicate that such genomes have evolved to have very similar codon usage patterns and might have been shaped by similar evolutionary forces, leading to them being more closely related. The CAI of SBW25 and A506 are very close and almost overlap. This is not surprising considering that both SBW25 and A506 were isolated from leaf surfaces of plants. The environmental niche of the two strains is similar and the metabolic requirements of these two strains would have much more overlap than with either PICF7 or Pf0-1. The CAI values reflect exactly that – genes of A506 and SBW25 may have comparable translational outputs.

3.4.4.2 Codon Adaptation Index of the three genes of interest

CAI of each gene of every genome was computed with respect to the corresponding genome. Across all of the strains, *rpsJ* has the lowest CAI, while *glyQ* has the highest value. The low CAI value of *rpsJ* is indicative of the codon usage pattern of the gene that differs

extensively from the genome wide codon usage. The same argument stands for *glyQ* – high CAI value is demonstrative of the codon usage pattern of *glyQ* that closely mirrors that of the genome. CAI of *acpP* is intermediate of the other two genes, corresponding to the intermediate level of variation in codon use of *acpP* as compared to the genome. In other words, *glyQ* is best adapted to the genome wide codon use, *rpsJ* the least, while *acpP* can be predicted to have some level of translational output that is intermediate to the other two.

The effects of *glyQ* allele swap can be predicted by comparing the CAI of the gene in the three (SBW25, A506 and Pf0-1) backgrounds. Pf0-1 has the highest CAI value for *glyQ,* it might be predicted that, *glyQ* from A506 or SBW25 in Pf0-1 background would be growing slowly as compared to the native *glyQ*; if *glyQ* translation is a proxy for growth. The effect may be exacerbated under conditions of faster growth. However, Pf0-1 *glyQ* in SBW25 (or A506) background would be growing faster than *glyQ* in the native background.

3.4.5 tRNA Adaptation Index

The tRNA adaptation index is a measure that provides information of how well the codons of a particular gene are adapted to the tRNA genes of an organism. tRNA gene copy number is used as proxy for the tRNA availability in the cell.

The tAI values for each of the genes that were studied in the scope of this chapter exhibited variation across the four genomes. This is not surprising considering that he tRNA gene availability in the four genomes is different. Moreover, given that the stAI calc method employs species-specific weights for codon-anticodon interactions, the variation in tAI across genomes is bound to be due to intrinsic differences in tRNA and codon contents. *glyQ* consistently exhibits tAI towards a higher range among all the genes studied. This might mean that *glyQ* being a single-copy highly expressed gene has to be well adapted to the tRNA availability, so that it may be translated with both speed and accuracy. On the other hand, *acpP*, a lipid metabolism gene exhibits tAI values ranging from intermediate to a higher range. This is suggestive that this gene may not be the best adapted to the tRNA availability as it might be required when lipid sources have to be metabolised. A number of other lipid metabolism genes also exist with *P. fluorescens* and hence *acpP* alone might not have to be the best optimized to the tRNA content.

For the other genes, *rpsJ, wspE, wspF, wspR, awsO* and *awsR,* there is much more variation in the tAI values across the genome. The observed variation might be a reflection of how the genome – both codon usage and tRNA gene availability – might have evolved in the four strains and shaped by the interaction encountered in the respective environments.

The effects of *glyQ* allele swap can be predicted by comparing the tAI of the gene in the three (SBW25, A506 and Pf0-1) backgrounds. The tAI value of *glyQ* is the highest in SBW25 (0.336) as compared to the A506 (0.278) and Pf0-1 (0.266). This might mean that if *glyQ* were to be swapped from a native to a non-native background, then a change in *glyQ* expression might be expected due to an engineered change in codon availability with respect to tRNA availability.

Chapter 4 : Altering codon usage in the *glyQ* gene of *P. fluorescens*

4.1 Introduction

Glycyl-tRNA synthetase, also known as glycine tRNA ligase (GlyRS) is one of the amino acyl-tRNA synthetases that catalyse the attachment of an amino acid to their cognate tRNAs. GlyRS therefore charges tRNAs with glycine. In eubacteria, such as *E. coli* and *P. fluorescens*, this enzyme is encoded has two subunits – an α-subunit and a β-subunit (Ostrem, Biochemistry, 1974; Keng *et al.*, 1982; Webster *et al.*, 1983). The active enzyme is an α2/β2 tetramer and in *E. coli* the α and β subunits mRNAs are transcribed as a single unit (Webster *et al*., 1983). The two subunits are coded for by the their respective genes – *glyQ* for the α-subunit, while *glyS* for the β-subunit. GlyRS belongs to the Class II family of tRNA synthetases (Valencia-Sánchez *et al*., 2016). Class II aminoacyl-tRNA synthetases are mainly dimeric or multimeric enzymes and contain three conserved motifs. The active site of Class II enzymes has a core of antiparallel β strand.

The reaction catalyzed by the glycyl-tRNA synthetase is as follows:

$$\text{aa-tRNA}^{\text{Gly}} + \text{glycine} + \text{ATP} \xrightarrow{\text{GlyRS}} \text{aa-glycyl-[tRNA}^{\text{Gly}}\text{]} + \text{AMP} + \text{PP}_i$$

The α-subunit of GlyRS in the bacterial species *Aquefix aeolicus* contains the aminoacylation site but needs the β-subunit for activity (Valencia-Sánchez *et al.*, 2016). The β-subunit on the other hand, has the tRNA recognition elements. The α-subunit alone is able to activate glycine, which is the first step for aminoacylation, but complete activity of the enzyme requires both subunits. However, in studies with *E. coli* glycyl-tRNA synthetase complexes, it has been shown that the β-subunit contributes largely to the aminoacylation activity of the holoenzyme, while the contribution of the α-subunit is small (Nagel *et al.*, 1984; Freist, Logan, & Gauss, 1996). The α-subunit has a major role in binding glycine and ATP (Freist *et al*., 1996). Both subunits are therefore required for activity of the enzyme and large mutations in either can be expected to have an effect on the activity of the holoenzyme.

Moreover, while investigating the essential genes of *E. coli*, it has been demonstrated that *glyQ* is essential while *glyS* is not (Gerdes *et al*., 2003; Baba *et al.*, 2006; Goodall *et al*.,

2018). The single-gene knockout mutants of *glyQ* from the Keio collection showed the knockout to be lethal (no growth in LB), while that of *glyS*, although a central gene, was not lethal (Baba *et al*., 2006). The argument for *glyS* was that the genome probably harbours a second copy of *glyS* that could have arisen due to a gene duplication event. Another independent study using Transposon-directed insertion site sequencing (TraDIS), showed that both *glyQ* and *glyS* are essential (Goodall *et al*., 2018). An insertion index was determined by normalizing the number of transposon insertions to the size of the gene. Genes associated with lower transposon insertions were defined as essential genes. Crosschecking of these genes with the Keio collection revealed that *glyQ* was found in both the TraDIS and Keio studies to be essential.

There is ample experimental evidence that the glycine tRNA synthetase genes are required for the growth and replication of bacteria. Moreover, functional analysis of these genes has revealed that inhibition of the activities of the protein product of the genes leads to a reduction in activity of the holoenzyme. Further, a bioinformatic screen for genes with exclusively synonymous differences between *P. fluorescens* SBW25, A506 and Pf0-1 revealed only three genes: *acpP, rpsJ* and *glyQ* (done in conjunction with Dr Frederic Bertels). This result indicates that there is large selective pressure acting against amino acid changes in *glyQ*.

In the light of its biological role, and the experimental evidence that it is required for survival, highly expressed, and highly conserved in *P. fluorescens*, *glyQ* provides an opportunity to understand the functional importance of the evolution of differing codon usage. In this chapter, alteration of codon usage will be achieved by swapping the native *glyQ* gene from each of the *P. fluorescens* SBW25, A506 and Pf0-1 backgrounds into the two non-native backgrounds. The resulting set of mutants will be phenotypically characterized with respect the *glyQ* expression, growth and fitness. These experiments are expected to provide insight into the codon use in *glyQ* in each strain, and thereby contribute to wider knowledge about the evolution of codon use.

4.2 Aims

1. To construct six *P. fluorescens* SBW25, A506 and Pf0-1 mutants, each containing a non-native copy of *glyQ*. The end result will be a set of nine strains (six mutants and three wild types).

2. To construct strains with "designed" *glyQ* alleles.
3. To measure and compare the growth rates and fitness of each genotype in aim 1 and 2.
4. To quantify the expression of each *glyQ* allele in its native and two non-native backgrounds using qRT-PCR.
5. To characterise mat forming abilities of each of the wild type strains and the *glyQ* mutants.

4.3 Results

4.3.1 Construction of natural alleles of *glyQ* – naturally evolved synonymous variants of *glyQ* were swapped into non-native backgrounds

The scar-free mutants were constructed using strand overlap extension PCR (SOE-PCR, (Ho *et al*., 1989) and a two-step allelic exchange protocol to replace the gene with its allele from a non-native background (see section 2.2.2.2 and 2.2.7). Six genotypes were constructed using this technique (Table 4.1); these genotypes contain *glyQ* from one of its two non-native backgrounds. The final product of the engineering process is an engineered strain containing a single, mutant copy of the gene in question in the place of the native gene, with no other genetic changes ("scars"; see Figure 4.1). Further, each individual engineering process should result in a wild type strain that has been through the engineering process but retained the original copy of the gene. These "reconstructed wild type" strains will make excellent controls for phenotypic effects of the engineering process.

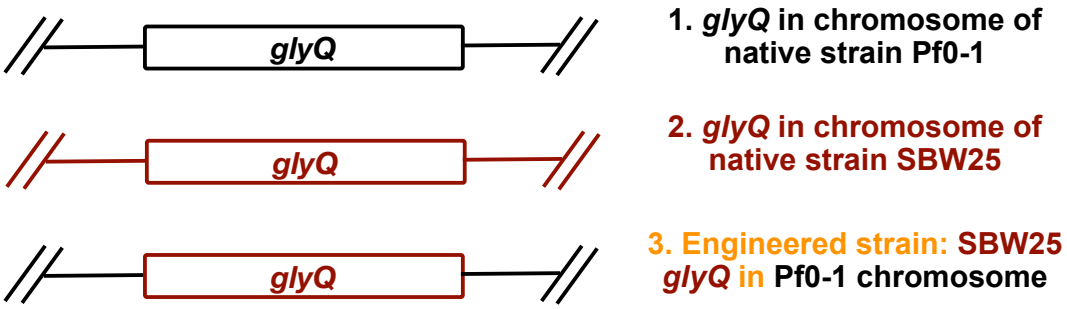


Figure 4.1: Replacement of the SBW25 *glyQ* allele into the Pf0-1 genome. This engineering process results in a "clean" replacement of the native allele with the non-native allele. No other genetic changes occur, meaning downstream phenotypic differences can be attributed to the mutations of interest.

In addition to the six mutant genotypes, three partial mutants were also isolated that are the by-products of the engineering process. These are described as partial mutants since these contain part of the gene from one strain and the remaining part from another strain.

Strain	Name	Description
1	SBW25_glyQ_A506	SBW25 with A506 *glyQ* (39 mutations)
2	SBW25_glyQ_Pf0-1	SBW25 with Pf0-1 *glyQ* (54 mutations)
3	A506_glyQ_SBW25	A506 with SBW25 *glyQ* (39 mutations)
4	A506_glyQ_Pf0-1	A506 with Pf0-1 glyQ (67 mutations)
5	Pf0-1_glyQ_SBW25	Pf0-1 with SBW25 *glyQ* (54 mutations)
6	Pf0-1_glyQ_A506	Pf0-1 with A506 *glyQ* (67 mutations)
16	SBW25_glyQ_A506_part	SBW25 with part of the A506 *glyQ* gene (bases 1-750), and part of the SBW25 *glyQ* gene (bases 751-end)
17	Pf01_glyQ_SBW25	Pf0-1 with part of the SBW25 *glyQ* gene (Pf01 *glyQ* bases 1-420, SBW25 *glyQ* bases 421-end)
18	A506_glyQ_Pf01_part	A506 with part of the Pf01 *glyQ* gene (bases 582-835), and part of the A506 *glyQ* gene (bases 1-581, 836-end)

Table 4.1: A list of genotypes to be constructed by genetic engineering. All of the above strains are complete. The ones highlighted in yellow are the complete mutant strains that were constructed; the ones in purple are the three partial mutants that were obtained at the end of the engineering process. The number of mutation in parenthesis indicates the number of synonymous differences between the non-native and native alleles.

4.3.2 Construction of strains with designed alleles of *glyQ*

Various combinations of synonymous mutations were designed for *glyQ*. In this project, the aim was to identify the functional role, if any, of synonymous codons and determine the effect on the physiology of the bacteria when synonymous mutations are introduced. To begin with, three variants of *glyQ* were designed – *faster glyQ* (first 17 codons changed to the most used codon), *slow glyQ* (all of the codons changed to the least used codon) and *flip glyQ* (all codons changed to codon of opposite orientation; the most used codons altered to the rare codons and vice versa). All of these variants were manufactured using the gBlocks service of Integrated DNA Technologies (IDT). None of these variants could be successfully integrated into the chromosome; despite several attempts, all the colonies obtained at the end of the engineering process were the wild type variants. However, it was demonstrated using mutant-specific *glyQ* PCRs that the desired mutants were present in the culture. The fact that the desired genotypes were present, but could not rise to a sufficient level for successful isolation,

suggests that these *glyQ* alleles are highly deleterious (or possibly lethal). This could be due to the large number of synonymous changes introduced simultaneously. Alternatively, it may be because the beginning of genes are often most intolerant to synonymous changes (due to RNA secondary structure influencing the start of translation). Each of the three alleles made changes in this part of the *glyQ* gene (Kelsic *et al.*, 2016; Frumkin *et al*., 2017).

The next step was to construct alleles of *glyQ* wherein the alterations in codon use were not as drastic as the *faster, slow* or *flip*. The idea behind these constructions was to identify which process of protein synthesis is most affected by changes in synonymous codons. The beginning of the gene is most crucial for translation initiation, while the middle of the gene plays a significant role in rate of elongation. The terminal part of the gene usually harbours signals for translation termination. Different codons might have different roles to play in each of these gene segments. The "faster" codons may be more essential in the middle of the gene since this is the region that has the more important role to play in speed of translation. With the presence of "fast" codons, the gene would be translated faster and may be beneficial for the bacteria that need fast translation of the gene. Therefore, two variants each for each of the regions of *glyQ* were designed – start-fast, start-slow, middle-fast, middle-slow, end-fast and end-slow – wherein, the fast and slow were the variants with exclusively "fast" or "slow" codons for the specific region. Of these variants, only the middle-fast and end-slow could be successfully constructed.

4.3.3 Growth rates and competitive fitness of each of the *glyQ* synonymous mutants

4.3.3.1 Growth in Lysogeny broth

The constructed mutants were tested for growth in different growth media. Firstly, growth in Lysogeny broth (LB) was carried out. This is a regular laboratory growth medium that is used for growing bacteria such as *E. coli* and *P. fluorescens.* Glycerol stock of each strain was streaked on agar plates and plates were incubated at 28ºC for 48 hours, in order to obtain isolated colonies. Each isolated colony was used to set up overnight cultures that were incubated at 28ºC, 200rpm for 16 hours. The overnight culture was then used to set up the growth assay in the EpochTM2 Microplate Spectrophotometer from BioTek® Instruments. The assay was conducted over 48 hours. The growth rate was computed using *growthcurver*

(Sprouffske & Wagner, 2016), an R package designed to compute growth parameters for growth assays set up in 96-well plates (see section 2.2.11.1).

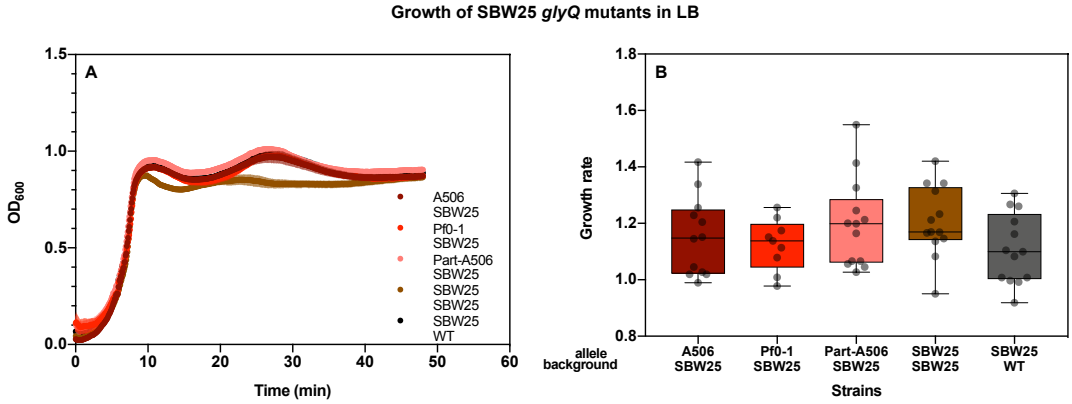


Figure 4.2: Swapping *glyQ* alleles in the SBW25 background does not affect growth rate of mutants in LB. A. OD vs Time plot of growth of SBW25 *glyQ* mutants in LB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 13 technical replicates. No significant difference was observed in the growth rate of the mutants when compared to the cognate WT (ANOVA, p= 0.10661)

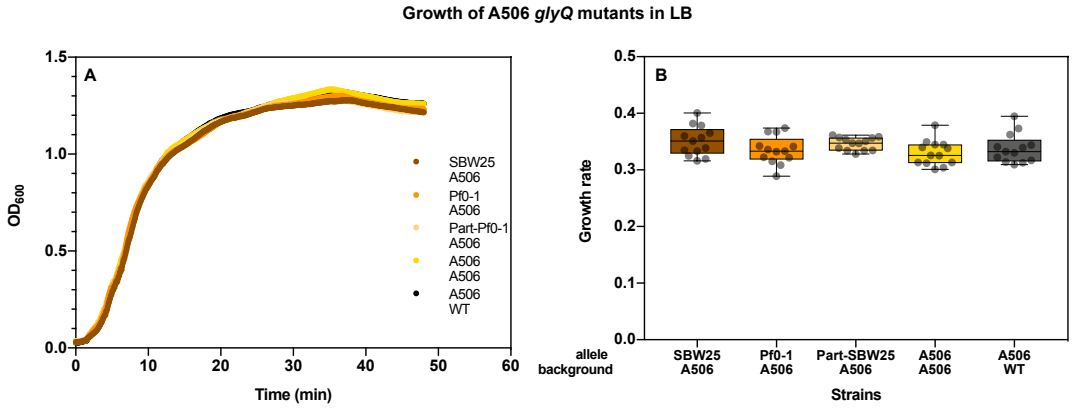


Figure 4.3: Swapping *glyQ* alleles in the A506 background does not affect growth rate of mutants in LB. A. OD vs Time plot of growth of A506 *glyQ* mutants in LB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 13 technical replicates. No significant difference was observed in the growth rate of the mutants when compared to the cognate WT (ANOVA, p= 0.14388)

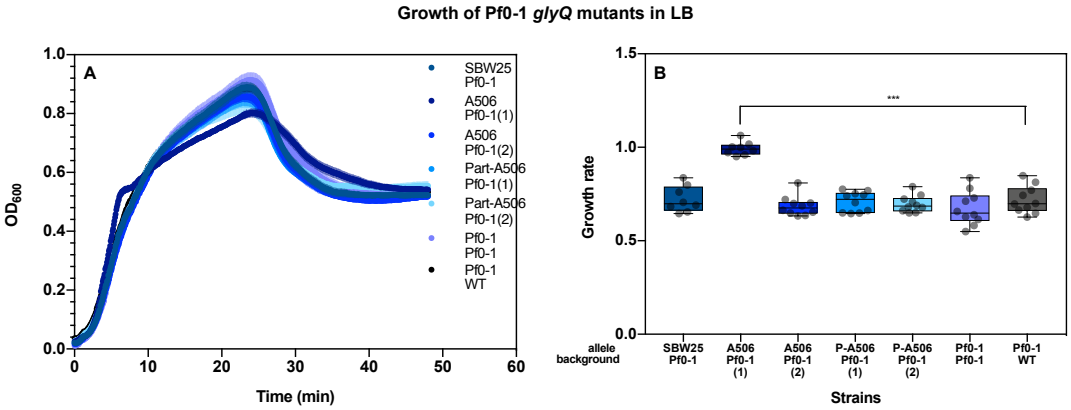


Figure 4.4: Swapping *glyQ* alleles in the Pf0-1 background may affect growth rate of mutants in LB. A. OD vs Time plot of growth of Pf0-1 *glyQ* mutants in LB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 10 technical replicates. Significant difference in the growth rate was when A506 Pf0-1(1) was compared to the WTs and each of the other mutants and partial mutants. (ANOVA, p= 0)

In Figures 4.2 and 4.3, it can be seen that none of the mutants of either SBW25 or A506 backgrounds showed any significant differences in growth rate when compared to the cognate WTs. However, among the Pf0-1 background *glyQ* mutants, one of the mutants, A506 *glyQ* in Pf0-1 background (A506 Pf0-1(1)) showed a significant difference in growth rate when compared to all other Pf0-1 background genotypes (Figure 4.4). It also showed a significant difference in growth rate when compared to its biological replicate (A506 Pf0-1(2)). This could imply that A506 Pf0-1(1) has acquired some background mutations, other than ones engineered via *glyQ* allele swap and hence behaves differently in Lysogeny broth (LB) when compared to its biological replicate. It is therefore possible that the difference observed for A506 Pf0-1(1) is due to other suspected alterations elsewhere in the genome.

4.3.3.2 Growth in King's medium B

The next medium tested for growth was King's medium B (KB). Like LB, KB is also a regular laboratory growth medium that is routinely used to grow *P. fluorescens*. The assay was set up as described in section 4.3.3.1. Growth rates were computed using *growthcurver*.

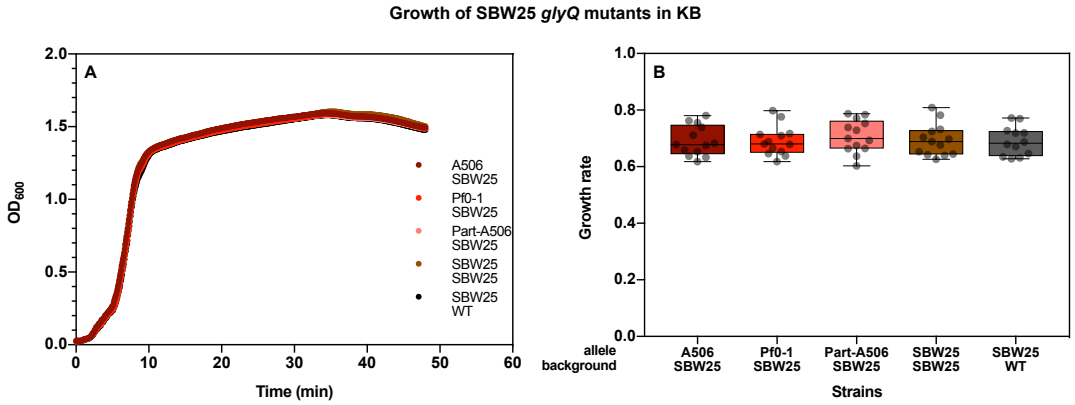


Figure 4.5: Swapping *glyQ* alleles in the SBW25 background does not affect growth rate of mutants in KB. A. OD vs Time plot of growth of SBW25 *glyQ* mutants in KB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 13 technical replicates. No significant difference was observed in the growth rate of the mutants when compared to the cognate WT (Kruskal-Wallis, p= 0.6568)

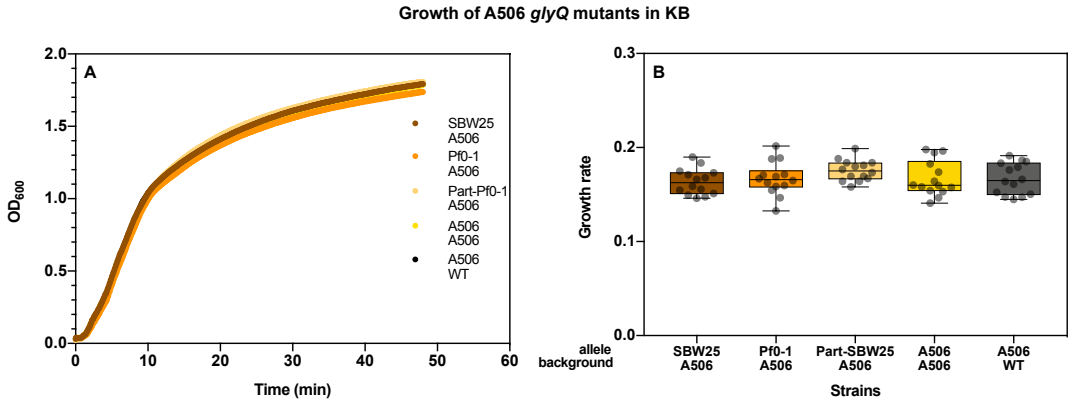


Figure 4.6: Swapping *glyQ* alleles in the A506 background does not affect growth rate of mutants in KB. A. OD vs Time plot of growth of A506 *glyQ* mutants in KB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 14 technical replicates. No significant difference was observed in the growth rate of the mutants when compared to the cognate WT (ANOVA, p= 0.35381)

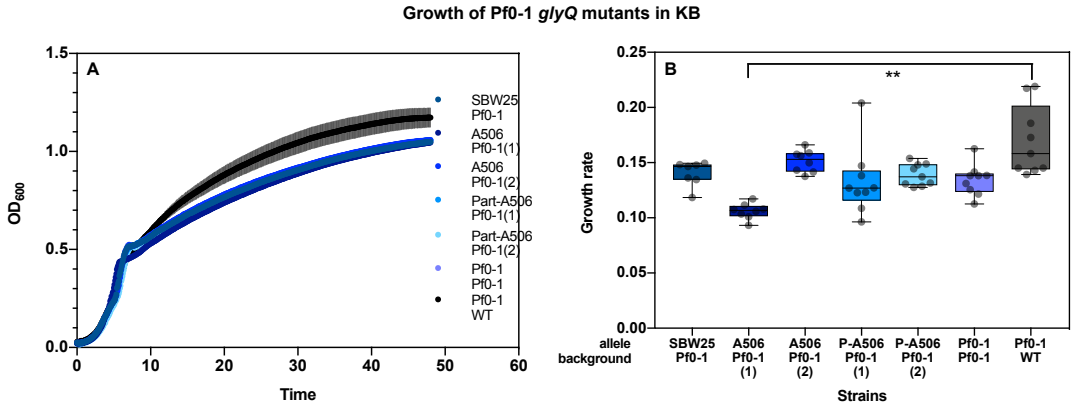


Figure 4.7: Swapping *glyQ* alleles in the Pf0-1 background may affect growth rate of mutants in KB. A. OD vs Time plot of growth of Pf0-1 *glyQ* mutants in KB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 9 technical replicates. At least one of the strains shows a significant difference from the cognate WT (Kruskal-Wallis, p= 0.0004326)

As can be seen from Figures 4.5 and 4.6, none of the mutants of either SBW25 or A506 backgrounds showed any significant differences in growth rate when compared to the cognate WTs. However, among the Pf0-1 background *glyQ* mutants, at least one of the mutants showed a significant difference in growth rate when compared to the cognate WT (Figure 4.7). From the plot 4.7B, it appears that strain A506 Pf0-1(1) is in all probability the strain that shows a difference in growth rate with respect to the WT Pf0-1. As mentioned in section 4.3.3.1, this could be attributed to mutations that could have arisen during the mutant construction, in the genomic background other than the *glyQ* region.

4.3.3.3 Growth in minimal medium M9

Since none of the mutants showed any difference in growth rates with respect to the cognate wildtypes in either LB or KB, the next step was to grow the constructed genotypes in a more stressful environment in order to resolve small differences, if any, between the mutants and wildtypes. Both LB and KB are media that are rich in carbon and nitrogen sources and there is no need for the organism to synthesise any sugars or amino acids. Contrariwise, M9 is composed of ammonium and magnesium salts with one carbon source (usually glucose) and therefore provides enough stress for the organism in the context of having to synthesise all essential building blocks, such as amino acids.

The growth assay was set up by purifying colonies from glycerol stocks on M9 medium supplemented with 20% glucose (agar plates) and incubating the plates at 28ºC for 72 hours. Next, single colonies were inoculated into 200 µl of M9 medium supplemented with 20% glucose in a 96-well plate and incubated at 28ºC at 200 rpm for 16 hours (overnight). The overnight culture was then used to set up the growth assay in the EpochTM2 Microplate Spectrophotometer from BioTek® Instruments. The assay was conducted over 48 hours. The growth rate was computed using *growthcurver*((Sprouffske & Wagner, 2016), an R package designed to compute growth parameters for growth assays set up in 96-well plates (see section 2.2.11.1).

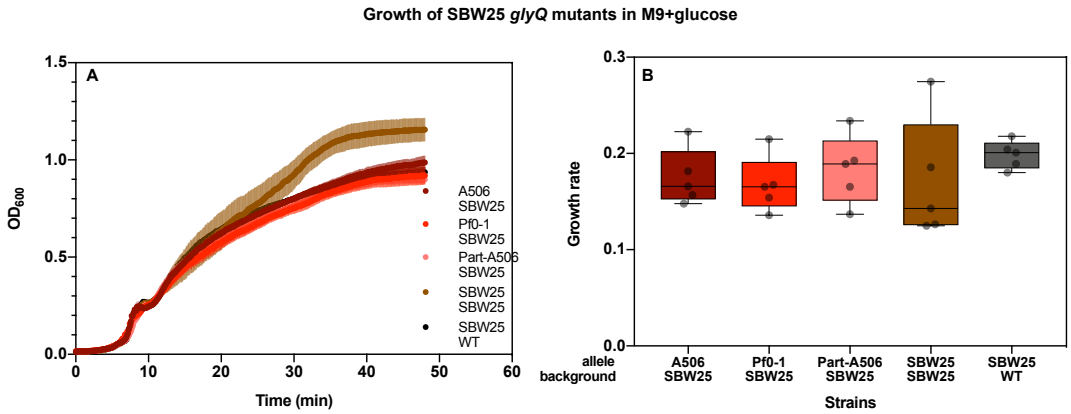


Figure 4.8: Swapping *glyQ* alleles in the SBW25 background does not affect growth rate of mutants in M9+glucose. A. OD vs Time plot of growth of SBW25 *glyQ* mutants in M9+glucose over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 5 technical replicates. No significant difference was observed in the growth rate of the mutants when compared to the cognate WT (Kruskal-Wallis, p= 0.3804)

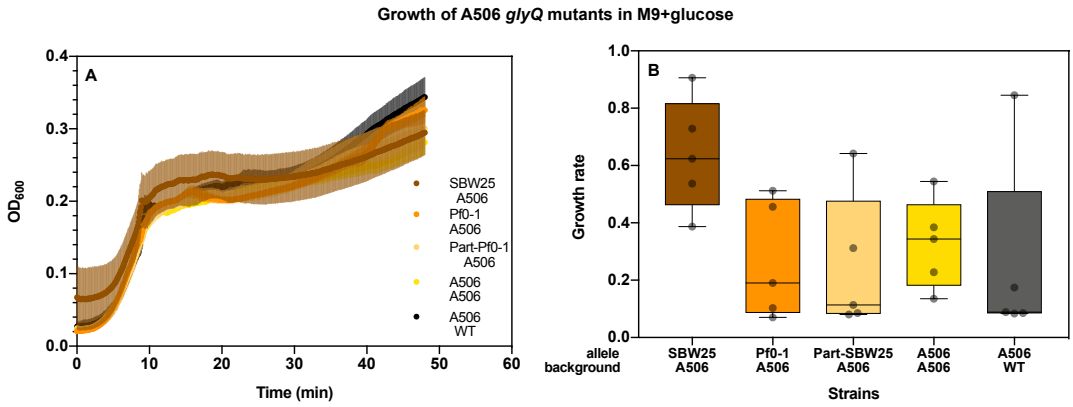


Figure 4.9: Swapping *glyQ* alleles in the A506 background does not affect growth rate of mutants in M9+glucose. A. OD vs Time plot of growth of A506 *glyQ* mutants in M9+glucose over a period of

48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 5 technical replicates. No significant difference was observed in the growth rate of the mutants when compared to the cognate WT (Kruskal-Wallis, p= 0.08513)

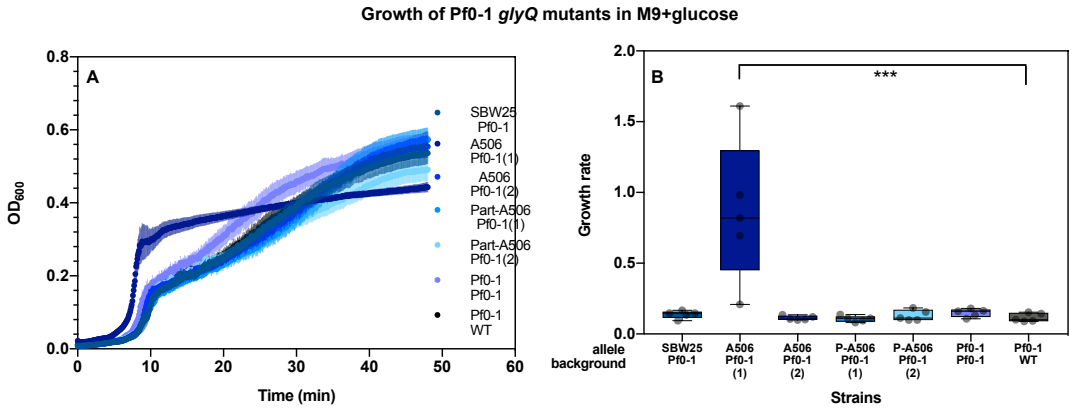


Figure 4.10: Swapping *glyQ* alleles in the Pf0-1 background may affect growth rate of mutants in M9+glucose. A. OD vs Time plot of growth of Pf0-1 *glyQ* mutants in M9+glucose over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 5 technical replicates. At least one of the strains shows a significant difference from the cognate WT (Kruskal-Wallis, p= 0.007358)

None of the mutants of either SBW25 or A506 backgrounds showed any significant differences in growth rate when compared to the cognate WTs as seen in Figure 4.8 and Figure 4.9, respectively. However, among the Pf0-1 background *glyQ* mutants, yet again, at least one of the mutants showed a significant difference in growth rate when compared to the cognate WT. From the plot 4.10B, it appears that strain A506 Pf0-1(1) is in all probability the strain that shows a difference in growth rate with respect to the WT Pf0-1. As mentioned in section 4.3.3.1, this could be attributed to mutations that could have arisen during the mutant construction, in the genomic background other than the *glyQ* region.

4.3.3.4 Growth in Terrific broth

Terrific Broth (TB) is a nutritionally rich medium that is used for faster growth of bacteria. It has higher amounts of tryptone and yeast extract that pushes the bacteria to grow much faster than regular lab media such as LB or KB. This could create a stress of high

growth consequently, pushing translation to the extreme limits thereby forcing small (dis)advantages in translation (if any) to show up as a fitness effect.

As none of the *glyQ* mutants in any of the non-native backgrounds showed any differences in fitness in either LB or KB, TB was used to resolve smaller differences in growth. The growth assay was set up by purifying colonies from glycerol stocks on TB agar plates and incubating the plates at 28ºC for 24 hours. Next, single colonies were inoculated into 200 µl of TB in a 96-well plate and incubated at 28ºC at 200 rpm for 16 hours (overnight). The overnight culture was then used to set up the growth assay in the EpochTM2 Microplate Spectrophotometer from BioTek® Instruments. The assay was conducted over 48 hours. The growth rate was computed using *growthcurver* (Sprouffske & Wagner, 2016), an R package designed to compute growth parameters for growth assays set up in 96-well plates (see section 2.2.11.1).

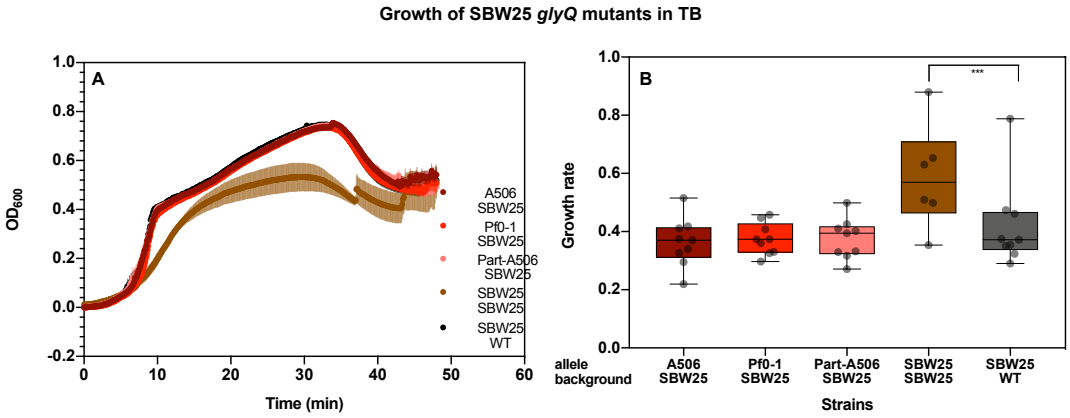


Figure 4.11: Swapping *glyQ* alleles in the SBW25 background may affect growth rate of mutants in TB. A. OD vs Time plot of growth of SBW25 *glyQ* mutants in TB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 9 technical replicates. Significant difference was observed in the growth rate of the mutants when compared to the reconstructed WT (SBW25 SBW25) (ANOVA, p= 0.00028)

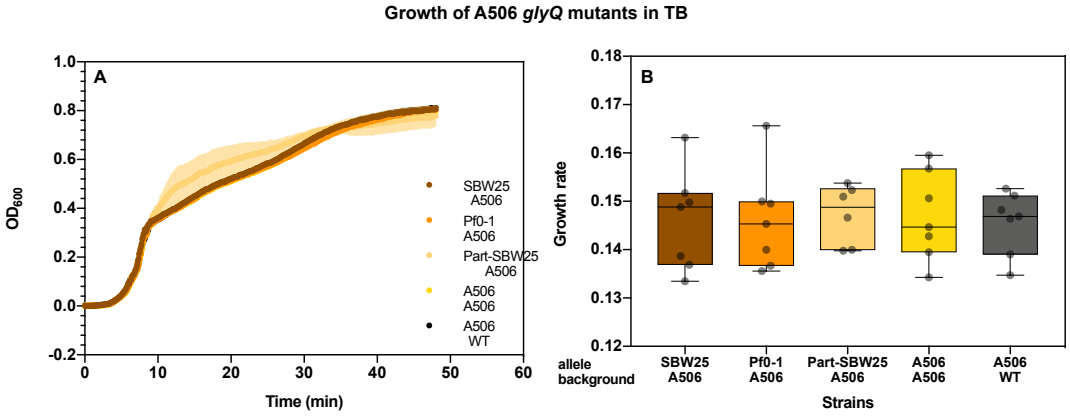


Figure 4.12: Swapping *glyQ* alleles in the A506 background does not affect growth rate of mutants in TB. A. OD vs Time plot of growth of A506 *glyQ* mutants in TB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 7 technical replicates. No significant difference was observed in the growth rates of the mutants as compared to the cognate WTs (ANOVA, p= 0.99716)

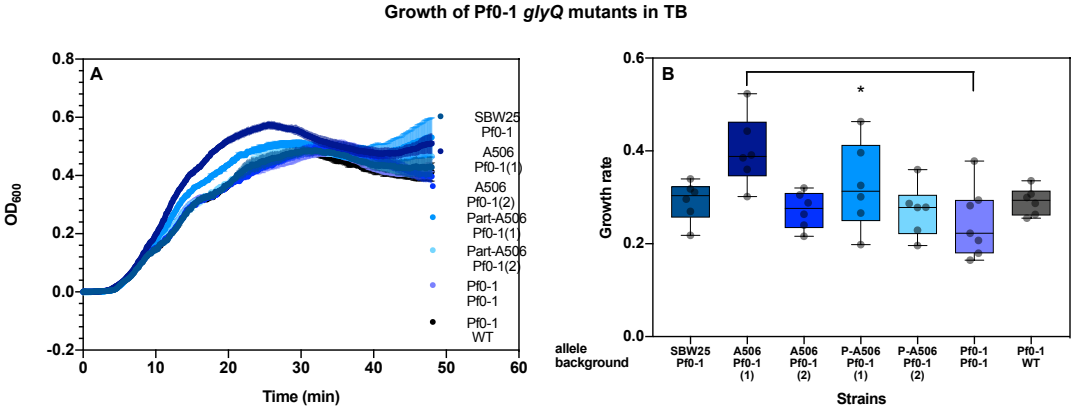


Figure 4.13: Swapping *glyQ* alleles in the Pf0-1 background may affect growth rate of mutants in TB. A. OD vs Time plot of growth of Pf0-1 *glyQ* mutants in TB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 7 technical replicates. At least one strain shows a significant difference in growth rate. (Kruskal-Wallis, p= 0.01871)

Among the SBW25 background strains, all of the mutants (both complete and partial) and SBW25 wild type have a significant difference from the engineered wild type (SBW25 SBW25) (see Figure 4.11). Since this strain should be exactly the same as the SBW25 wild type strain with regards to the *glyQ* sequence, any differences between the reconstructed wild type and SBW25 wild type strain could be due to other mutations that may have arisen in the genomic background during the strain construction process.

In the A506 background strains mutants and cognate wild types do not show any significant difference in growth rates (Figure 4.12). As can be seen in Figure 4.13, among Pf0-1 background strains, the first biological replicate of A506 *glyQ* in Pf0-1 background (A506 Pf0-1(1)) seems to be growing with a higher rate as compared to the other strains of the same background. This could be the one contributing to the significant differences in growth rate.

4.3.3.5 Growth in Super broth

Super Broth (SB) is a nutritionally rich medium that is used for faster growth of bacteria. Like Terrific Broth, SB also has higher amounts of tryptone and yeast extract that pushes the bacteria to grow much faster than regular lab media such as LB or KB. Therefore, this could create a stress of high growth as well, consequently causing higher translation rates thereby forcing small (dis)advantages in translation (if any) to show up as a fitness effect.

This medium was also used for the same reason as TB – creating an environment of faster growth that might exacerbate small differences in growth. The growth assay was set up by purifying colonies from glycerol stocks on SB agar plates and incubating the plates at 28ºC for 24 hours. Next, single colonies were inoculated into 200 µl of SB in a 96-well plate and incubated at 28ºC at 200 rpm for 16 hours (overnight). The overnight culture was then used to set up the growth assay in the EpochTM2 Microplate Spectrophotometer from BioTek® Instruments. The assay was conducted over 48 hours. The growth rate was computed using *growthcurver* (Sprouffske & Wagner, 2016), an R package designed to compute growth parameters for growth assays set up in 96-well plates (see section 2.2.11.1).

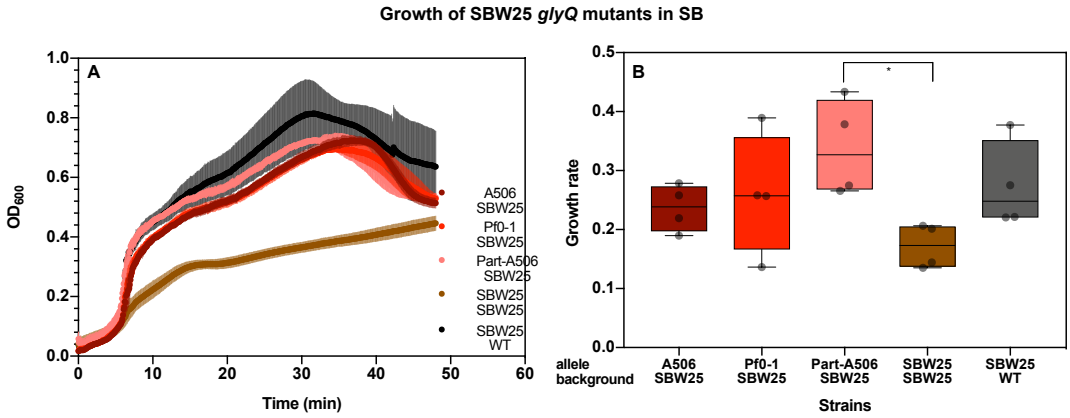


Figure 4.14: Swapping *glyQ* alleles in the SBW25 background does not affect growth rate of mutants in SB. A. OD vs Time plot of growth of SBW25 *glyQ* mutants in SB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 4 technical replicates. No significant difference in growth rate was observed between the mutants and the cognate wild types (ANOVA, p= 0.06261), although a TUKEY Honest Significant Difference comparison shows that the partial mutant (Part-A506 SBW25) and engineered wild type (SBW25 SBW25) show a significant difference in growth (p= 0.0350), even though such a significant difference is not seen between partial mutant and SBW25 wild type.

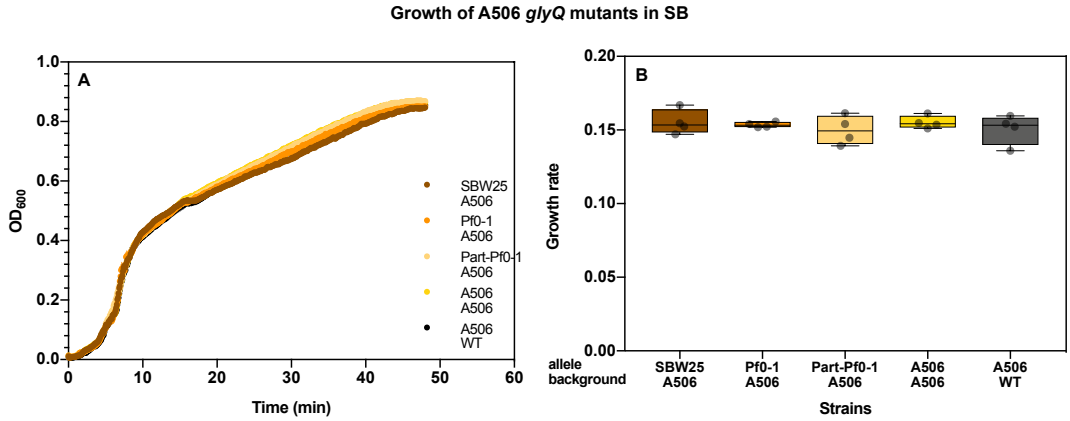


Figure 4.15: Swapping *glyQ* alleles in the A506 background does not affect growth rate of mutants in SB. A. OD vs Time plot of growth of A506 *glyQ* mutants in SB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 4 technical replicates. No significant difference in growth rate was observed between the mutants and the cognate wild types (ANOVA, p= 0.77288).

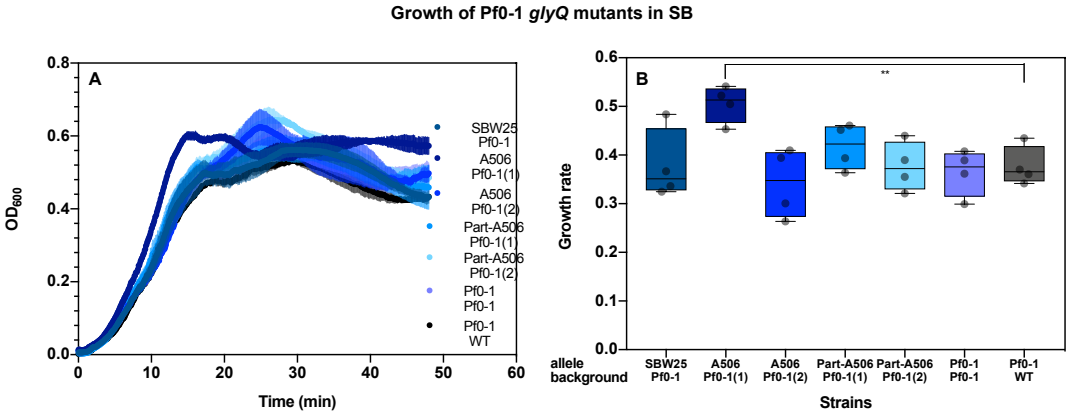


Figure 4.16: Swapping *glyQ* alleles in the Pf0-1 background affects growth rate of one of the mutants in SB. A. OD vs Time plot of growth of Pf0-1 *glyQ* mutants in SB over a period of 48 hour. Each point is the reading taken at every 5 minutes. **B.** Maximal average growth rate (milliOD/minute; y-axis) was measured using the R package, *growthcurver*. Each bar is a mean 4 technical replicates. Significant difference in growth rate was observed between the A506 *glyQ* mutant in Pf0-1 background (A506 Pf0-1(1) and the cognate wild types (ANOVA, p= 0.00823).

As can be seen in Figure 4.14A, the engineered SBW25 wild type (SBW25 SBW25) behaves much differently in terms of growth as compared to the other strains, including SBW25 wild type. Since the reconstructed wild type and the SBW25 wild type are both meant to be the same strain, including the *glyQ* gene, therefore the differences in growth may be attributed to background mutations in the genome that may have arisen during strain construction. Although, no significant differences in growth can be observed between the reconstructed wild type and SBW25 wild type, the partial mutant (Part-A506 SBW25) and the reconstructed wild type show a weakly significant difference in growth rate (p= 0.0350, Figure 4.14B).

None of the A506 background mutants show a significant difference in growth rate when compared to the cognate wild type strains (Figure 4.15B)

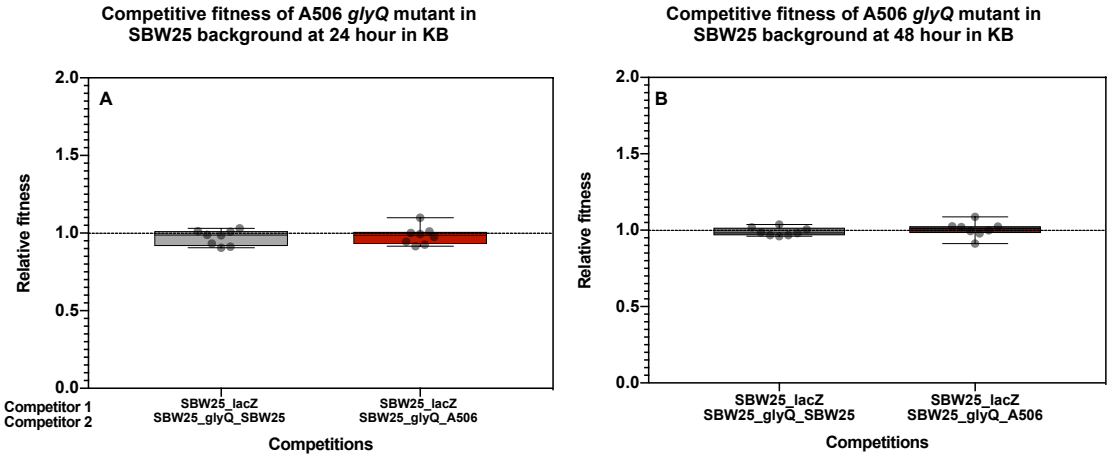
While, among the Pf0-1 background strains, only the first biological replicate of the A506 *glyQ* mutant in Pf0-1 background (A506 Pf0-1(1)) behaves significantly differently from others. It also differs from the cognate second biological replicate (A506 Pf0-1(2)) and shows a significant difference in growth rate. This could also be due to some mutation that may have arisen during the construction of the strain as both the replicates are meant to have the same genotype (Figure 4.16).

4.3.3.6 Competitive fitness in King's medium B

In this assay the SBW25 background *glyQ* mutants and their cognate engineered wild type strains were competed against an ancestral SBW25 that has a neutral *lacZ* marker (Zhang & Rainey, 2007). Eight independent replicates were set up for each competition in KB medium by mixing equal proportions of each competitor and the final plating of 10^5 fold dilution of culture was done by spreading 50 µl and 100 µl on LB agar plates with 60 µg/ml of X-gal. The competitors were distinguished on the basis of their ability to utilise or not utilise X-gal (blue and white colonies, respectively; see section 2.2.11.4)

Statistical methods devised by (Lenski *et al.*, 1991) were used to calculate the relative fitness values during competition experiments. The density of the competitors was calculated at 0 hours, 24 hours and 48 hours and these values were used to calculate Malthusian parameters for each strain (see section 2.2.11.4)

Figure 4.17 shows the relative fitness of the A506 *glyQ* mutant and Pf0-1 *glyQ* mutant in the SBW25 background. Neither of these mutants showed any significant difference in fitness as compared to the ancestral SBW25 with the neutral *lacZ* marker (SBW25_*lacZ*).



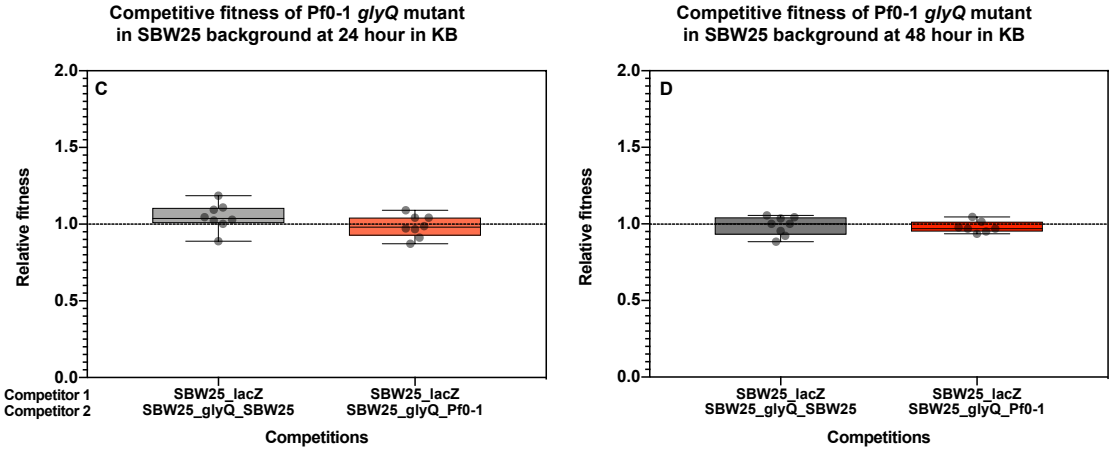


Figure 4.17: No difference in relative fitness of the non-native *glyQ* mutants in SBW25 background. Box plot of the fitness of the SBW25 wild type with the neutral *lacZ* marker relative to the reconstructed wild type (SBW25_*lacZ* vs SBW25_*glyQ*_SBW25) and the fitness of SBW25 wild type with the neutral *lacZ* marker relative to the engineered mutant (SBW25_*lacZ* vs SBW25-_*glyQ*_A506/Pf0-1). All fitness assays were performed in shaking conditions. Values greater than 1 (dashed line) indicate a higher relative fitness of the first competitor. **A**. Relative fitness of the competitors at the end of 24 hours (SBW25_*lacZ* vs SBW25_*glyQ*_SBW25, t-test, p= 0.1516 and SBW25_*lacZ* vs A506 *glyQ* in SBW25 background; t-test, p=0.4470). **B**. Relative fitness of the competitors at the end of 48 hours. (SBW25_*lacZ* vs SBW25_*glyQ*_SBW25, t-test, p= 0.3508 and SBW25_*lacZ* vs A506 *glyQ* in SBW25 background; t-test, p=0.7592). **C.** Relative fitness of the competitors at the end of 24 hours (SBW25_*lacZ* vs SBW25_*glyQ*_SBW25, t-test, p= 0.1681 and SBW25_*lacZ* vs Pf0-1 *glyQ* in SBW25 background; t-test, p=0.5976). **D**. Relative fitness of the competitors at the end of 48 hours. (SBW25_*lacZ* vs SBW25_*glyQ*_SBW25, t-test, p= 0.5784 and SBW25_*lacZ* vs Pf0-1 *glyQ* in SBW25 background; t-test, p=0.1901).

In assay with A506 wild type strains, A506 background *glyQ* mutants and their cognate engineered wild type strains were competed against an ancestral SBW25 that has a neutral *lacZ* marker (Zhang & Rainey, 2007). Additionally, the ancestral A506 wild type strain (ancestral strain that was used for construction of the mutants) was also competed against the ancestral SBW25 with the neutral *lacZ* marker. Eight independent replicates were set up for each competition in KB medium by mixing equal proportions of each competitor and the final plating of 10^5 fold dilution of culture was done by spreading 50 µl and 100 µl on LB agar plates with 60 µg/ml of X-gal. The competitors were distinguished on the basis of their ability to utilise or not utilise X-gal (blue and white colonies, respectively; see section 2.2.11.4)

Statistical methods devised by (Lenski *et al*., 1991) were used to calculate the relative fitness values during competition experiments. The density of the competitors was calculated at 0 hours, 24 hours and 48 hours and these values were used to calculate Malthusian parameters for each strain (see section 2.2.11.4). None of the mutants showed a significant difference in fitness when compared to the wild type A506.

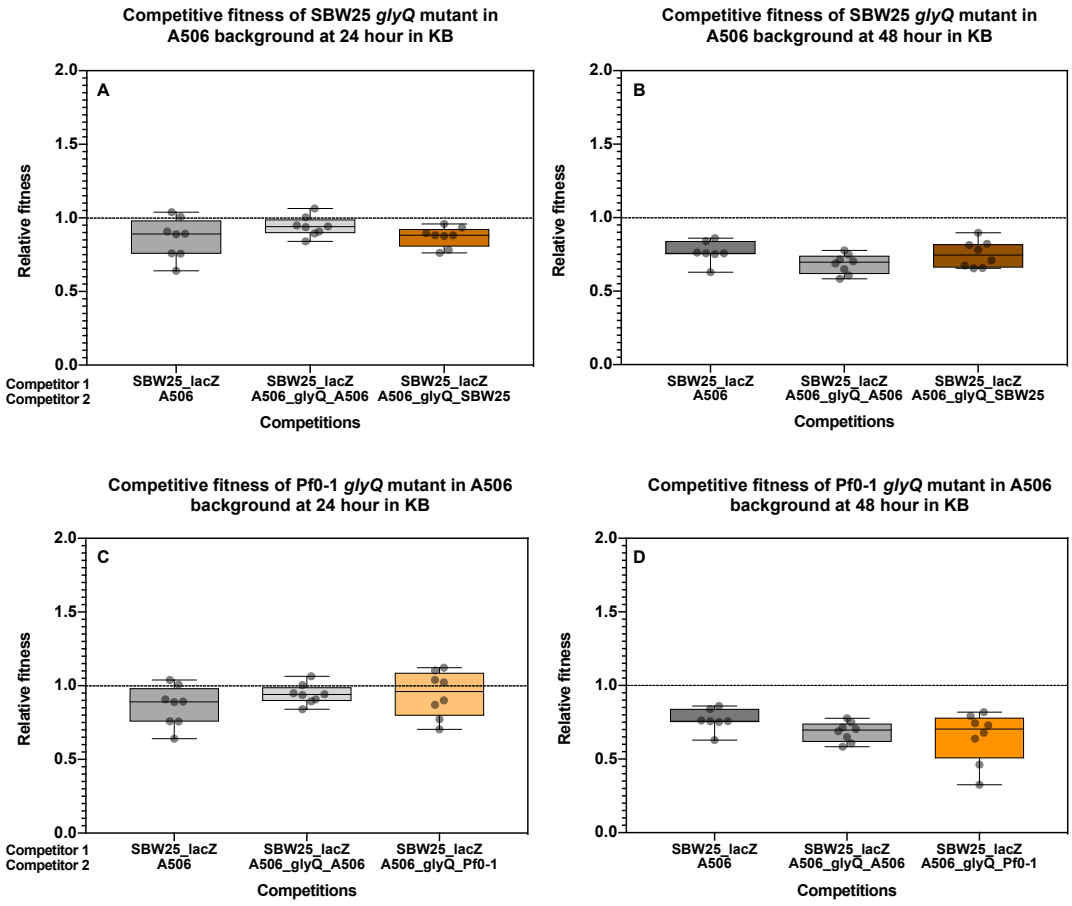


Figure 4.18: No difference in relative fitness of the non-native *glyQ* mutants in A506 background. Box plot of the fitness of the A506 wild type relative to the reconstructed wild type (A506 vs A506_*glyQ*_A506) and the fitness of A506 wild type relative to the engineered mutant (A506 vs A506_*glyQ*_SBW25/Pf0-1). All fitness assays were performed in shaking conditions. Values greater than 1 (dashed line) indicate a higher relative fitness of the first competitor. **A.** Relative fitness of the competitors at the end of 24 hours (SBW25_*lacZ*/A506 vs A506/A506_*glyQ*_A506, t-test, p= 0.1396 and SBW25_*lacZ*/A506 vs A506/SBW25 *glyQ* in A506 background; t-test, p=0.6228). **B.** Relative fitness of the competitors at the end of 48 hours. (SBW25_*lacZ*/A506 vs A506/A506_*glyQ*_A506, t-test, p= 0.3508 and SBW25_*lacZ*/A506 vs A506/SBW25 *glyQ* in A506 background; t-test, p=0.7592). **C**. Relative fitness of the competitors at the end of 24 hours (SBW25_*lacZ*/A506 vs A506/A506_*glyQ*_A506, t-test, p= 0.1681 and SBW25_*lacZ*/A506 vs A506/Pf0-1 *glyQ* in A506 background; t-test, p=0.5976). **D**. Relative fitness of the competitors at the end of 48 hours.

(SBW25_*lacZ*/A506 vs A506/A506_*glyQ*_A506, t-test, p= 0.5784 and SBW25_*lacZ*/A506 vs A506/Pf0-1 *glyQ* in A506 background; t-test, p=0.1901).

In assay with Pf0-1 wild type strains, Pf0-1 background *glyQ* mutants and their cognate engineered wild type strains were competed against an ancestral SBW25 that has a neutral *lacZ* marker (Zhang & Rainey, 2007). Additionally, the ancestral Pf0-1 wild type strain (ancestral strain that was used for construction of the mutants) was also competed against the ancestral SBW25 with the neutral *lacZ* marker. Eight independent replicates were set up for each competition in KB medium by mixing equal proportions of each competitor and the final plating of 10^5 fold dilution of culture was done by spreading 50 µl and 100 µl on LB agar plates with 60 µg/ml of X-gal. The competitors were distinguished on the basis of their ability to utilise or not utilise X-gal (blue and white colonies, respectively; see section 2.2.11.4)

Statistical methods devised by (Lenski *et al*., 1991) were used to calculate the relative fitness values during competition experiments. The density of the competitors was calculated at 0 hours, 24 hours and 48 hours and these values were used to calculate Malthusian parameters for each strain (see section 2.2.11.4). For the Pf0-1 backgrounds, no significant difference in competitive fitness of the SBW25 *glyQ* in Pf0-1 background is observed as compared to the wild type strain. But both the biological replicates A506 *glyQ* in Pf0-1 background shows a significant difference in growth rate as compared to the wild type strain. A506 and Pf0-1 *glyQ* show the largest number of synonymous differences – 67 (see Table 3.1) - as compared to the other combinations. It is possible that 67 synonymous differences are large enough to cause changes at the level of translational output (protein synthesis rate and efficiency), which in turn is reflected in the fitness of the mutant relative to the wild type. The mutants in fact are more fit as compared to the wild type strain; one possible explanation could be that the codon position and order in A506 *glyQ* allow faster translation and may be more correlated with the tRNA availability in Pf0-1. In the light of this result, the lack of significant differences in fitness of the Pf0-1 *glyQ* in A506 background mutant relative to its cognate wild type strain is unexpected. But, it is possible that the reverse combination (Pf0-1 *glyQ* in A506) and the corresponding wild type (A506) are both equivalently correlated to the tRNA availability in A506 and hence the translation in the two is equivalent. The remaining gene and background combinations (A506/Pf0-1 *glyQ* in SBW25, SBW25/Pf0-1 *glyQ* in A506 and SBW25 *glyQ* in Pf0-1) do not show significant difference in fitness as compared to the wild type strains.

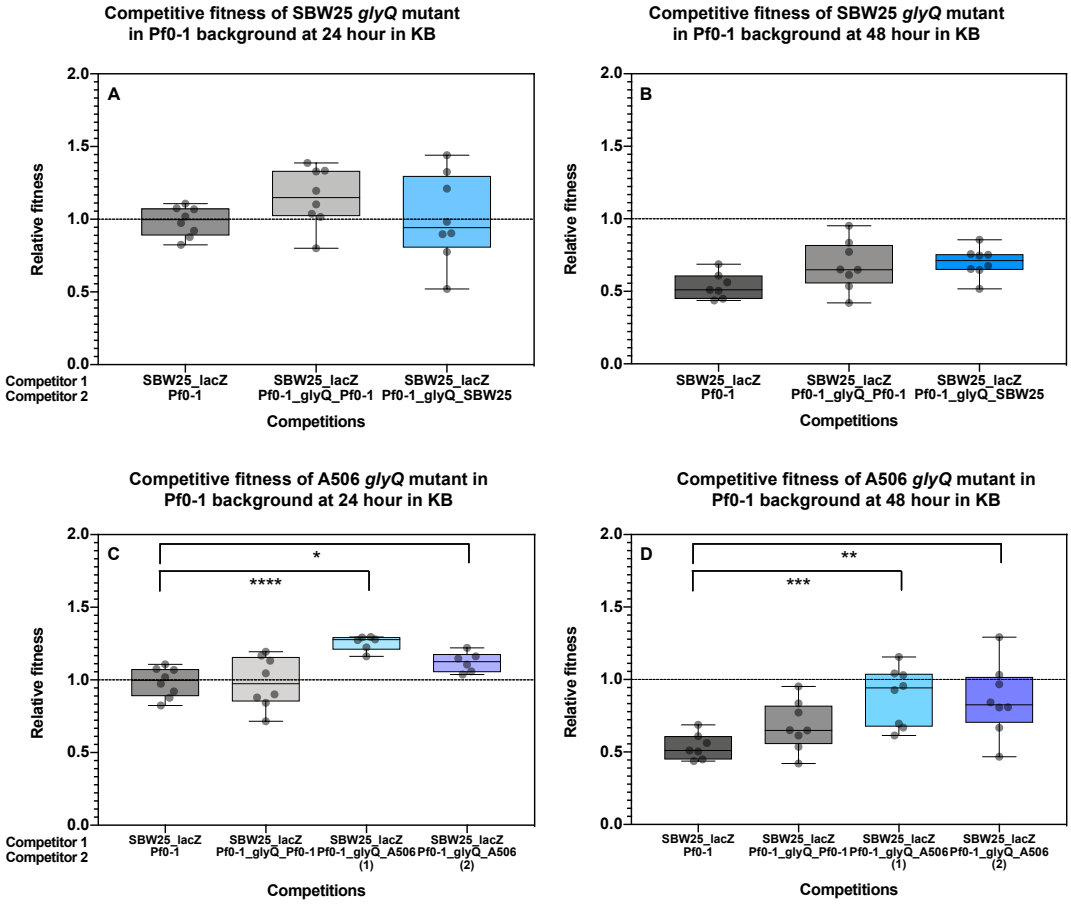


Figure 4.19: Mutants in the background of Pf0-1 show significant difference in growth rate with respect to wild type Pf0-1. Box plot of the fitness of the Pf0-1 wild type relative to the reconstructed wild type (Pf0-1 vs Pf0-1_*glyQ*_Pf0-1) and the fitness of Pf0-1 wild type relative to the engineered mutant (Pf0-1 vs Pf0-1_*glyQ*_SBW25/A506). All fitness assays were performed in shaking conditions. Values greater than 1 (dashed line) indicate a higher relative fitness of the first competitor. **A.** Relative fitness of the competitors at the end of 24 hours (SBW25_*lacZ*/Pf0-1 vs SBW25_*lacZ*/Pf0-1_*glyQ*_Pf0-1, t-test, p= 0.1574, SBW25_*lacZ*/Pf0-1 vs SBW25_*lacZ*/SBW25 *glyQ* in Pf0-1 background; t-test, p=0.9102). **B.** Relative fitness of the competitors at the end of 48 hours. (SBW25_*lacZ*/Pf0-1 vs SBW25_*lacZ*/Pf0-1_*glyQ*_Pf0-1, t-test, p=0.3401, SBW25_*lacZ*/Pf0-1 vs SBW25_*lacZ*/SBW25 *glyQ* in Pf0-1 background; t-test, p=0.7562). **C**. Relative fitness of the competitors at the end of 24 hours (SBW25_*lacZ*/Pf0-1 vs SBW25_*lacZ*/Pf0-1_*glyQ*_Pf0-1, t-test, p=0.1574, SBW25_*lacZ*/Pf0-1 vs SBW25_*lacZ*/A506(1)/(2) *glyQ* in Pf0-1 background; t-test, p<0.0001 and p=0.0144, both biological replicates, respectively). **D**. Relative fitness of the competitors at the end of 48 hours. (SBW25_*lacZ*/Pf0-1 vs SBW25_*lacZ*/Pf0-1_*glyQ*_Pf0-1, t-test, p=0.3401, SBW25_*lacZ*/Pf0-1 vs SBW25_*lacZ*/A506(1)/(2) *glyQ* in Pf0-1 background; t-test, p=0.0010 and p=0.0060, both biological replicates, respectively).

4.3.4 Relative quantification of *glyQ* expression

Although weakly significant differences in growth rate were observed between one of the mutants and cognate wild types of the Pf0-1 background strains under fast growth conditions, these differences have in all likelihood arisen due to mutations that may have emerged in the genomic background during strain construction. Such weakly significant differences under conditions of fast growth can also be seen between the reconstructed wild type of SBW25 and SBW25 wild type, which further implies that differences between these two strains that are meant to be identical genotypically could be due to mutations in the genome other than the *glyQ* region (since the *glyQ* region was confirmed to be the same in the reconstructed wild type and SBW25 wild type by Sanger sequencing). It is therefore possible that the synonymous alterations effected in the *glyQ* gene by replacing the gene with its alleles in non-native background do not have large enough impact to affect growth.

Although effects on growth are not seen, it may not indicate that synonymous changes have no effect on the physiology of the bacteria at all. In order to understand better the effect of the changes on the bacteria relative quantification of *glyQ* messenger RNA (mRNA) was performed. Synonymous mutations are known to affect mRNA stability – some of the mutations are known to make the mRNA more stable while others are known to make the mRNA more prone to degradation. Therefore, by quantifying the mRNA levels of *glyQ* against the mRNA levels of a housekeeping gene that has not been altered and is known to be present in stable quantities throughout the growth phase of *P. fluorescens,* we can understand how the synonymous alterations in *glyQ* affect its mRNA stability.

For quantification of *glyQ*, three biological replicates of each strain were used for RNA extraction. The strains were grown overnight in KB liquid culture and the culture was used for RNA extraction (see section 2.2.8). For each of the three biological replicates, three technical replicates were used for cDNA synthesis (see section 2.2.9) and quantitative PCR (qPCR). Two reference genes – *rpoD* and *recA* – were used. Quantification of *glyQ* was done relative to the quantities of *rpoD* and *recA* (see section 2.2.11.5). Each bar on the graphs is therefore an average of three technical replicates.

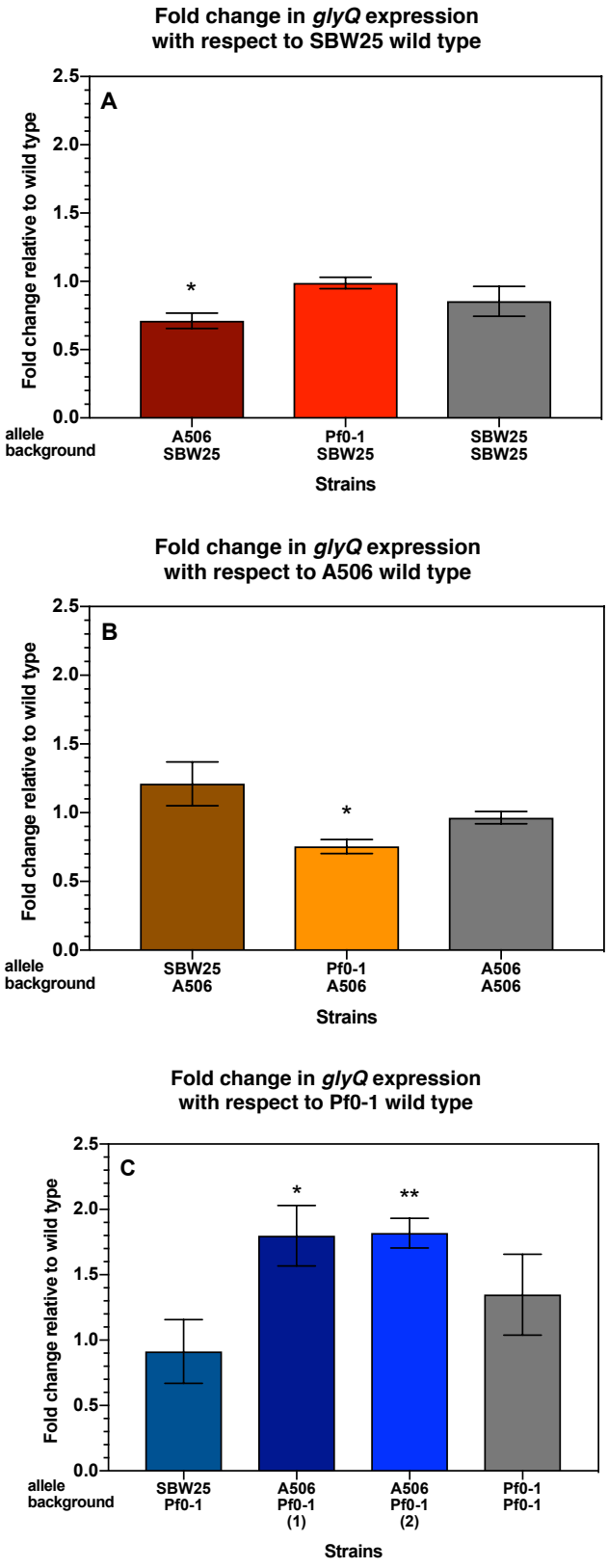


Figure 4.20: Weakly significant difference in mRNA quantities of *glyQ* mutants. Plot of the relative expression levels of *glyQ* as compared to the quantities of *rpoD* and *recA* in all of the *glyQ* mutants and corresponding wild type strains. **A**. Weakly significant difference in the relative quantity of *glyQ* in A506 *glyQ* mutant in SBW25 background as compared to SBW25 wild type (A506 SBW25

vs SBW25 WT; ANOVA, p= 0.0249) **B**. Weakly significant difference in the relative quantity of *glyQ* in Pf0-1 *glyQ* mutant in A506 background as compared to A506 wild type (Pf0-1 A506 vs A506 WT; *t*- test, p= 0.0161). **C. S**ignificant difference in the relative quantity of *glyQ* in A506 *glyQ* mutant in Pf0-1 background as compared to Pf0-1 wild type, as well as compared to SBW25 *glyQ* in Pf0-1 strain (A506 A506 Pf0-1(2) vs Pf0-1 WT and vs SBW25 Pf0-1; ANOVA, p= 0.0084)

As can be seen in Figure 4.20A, the A506 *glyQ* mutant in SBW25 background has lower relative amounts of *glyQ* mRNA as compared to the SBW25 wild type strain. While, for the A506 backgrounds, not only does the Pf0-1 *glyQ* mutant in A506 background have significantly lower relative *glyQ* amount as compared to the A506 wild type strain, but also compared to SBW25 *glyQ* mutant in A506 background (Figure 4.20B). In the case of the Pf0-1 background strains, A506 *glyQ* in Pf0-1 background showed significantly higher relative mRNA quantities of *glyQ* as compared to Pf0-1 wild type strains. The higher relative expression of *glyQ* in the A506 *glyQ* mutants in Pf0-1 background echoes the competitive fitness results. This mutant (both biological replicates) shows higher fitness relative to the wild types strain. It is possible that in A506 *glyQ* mutant in Pf0-1 background, higher amounts of *glyQ* is being expressed, which in turn translates to higher amounts of protein. This strain also showed higher amounts of *glyQ* mRNA as compared to SBW25 *glyQ* mutant in Pf0-1 background (Figure 4.20C).

4.3.5 Mat formation of the strains of *P. fluorescens*

Mat formation ability is an important and characteristic phenotype exhibited by various species of *Pseudomonas* and has been extensively studied in *P. fluorescens* SBW25 (Spiers *et al*., 2003; Friedman & Kolter, 2004b; Chang *et al*., 2007; Ueda & Wood, 2009; Ueda & Saneoka, 2014; Cárcamo-Oyarce *et al.*, 2015; Ghods *et al*., 2015; Díaz-Salazar *et al.*, 2017; Farias *et al*., 2019). Therefore, this phenotype was investigated for the wild type strains as well as the phenotypes with the aim to characterise as well as observe the effects of synonymous mutations on mat formation behaviour. SBW25 mat formation ability was reproduced while mat formation in A506, Pf0-1 and an additional strain, PICF7 were investigated. The mutants did not show any differences in mat formation behaviour with respect to the cognate wild type strains, but characteristic variation in mat phenotype was observed between the different wild type strains. Knowledge of mat formation ability and strategies by the A506, Pf0-1 and PICF7, in a spatially heterogeneous environment is still

unreported and hence in the scope of this thesis, mat formation phenotype and routes leading to it were studied. This part of the work was extensively performed and warrants a separate chapter. Therefore the results of this work are made available in the following chapter.

4.4 Discussion

4.4.1 Construction of natural alleles of *glyQ* – naturally evolved synonymous variants of *glyQ* were swapped into non-native backgrounds

In order to understand the role that synonymous mutations may have on the function of a gene, naturally evolved synonymous alleles of a tRNA synthetase gene, α-subunit of glycine tRNA synthetase (*glyQ*) was incorporated into non-native backgrounds. This was achieved by a scar-free genetic engineering technique, whereby, *glyQ* gene was replaced by its alleles in the genome. Three strains of *P. fluorescens* were used for this study – SBW25, A506 and Pf0-1 – and *glyQ* in each of these strains was replaced by *glyQ* from the two other non-native backgrounds. The gene *glyQ* exclusively shows synonymous differences; in other words, the primary protein sequence of GlyQ is exactly the same in all three strains and differences are at the DNA level. The synonymous strains generated by engineering naturally evolved synonymous alleles of *glyQ* in non-native background, therefore, *only* differ at the *glyQ* locus.

4.4.2 Construction of synonymous mutants with the "designed" *glyQ* gene

Synonymous variants of *glyQ* were designed with the aim of altering codon composition along the gene such that different stages of the translation process might be targeted. The beginning of the gene is crucial to initiation of translation, the middle for elongation and the end for termination. Therefore slow and fast variants of the start, middle and end were designed. The approach was to drastically alter the codon composition by changing codon identities all along the gene – *faster glyQ*, *slow glyQ* and *flipped glyQ* were designed. The idea here was that different extents of global changes in gene output might be observed.

None of these constructs could be successfully isolated from culture, despite multiple attempts at constructing these strains. The presence of the mutants was detected in culture by using mutant-specific PCR reactions. This indicates that these constructions are highly deleterious and do not permit the cells to grow to a frequency such that these can be isolated as individual colonies. The beginning of the gene is exceptionally affected when changing codons due to deleterious 5' end mRNA secondary structures and due to formation of deleterious codon pairs (Kelsic *et al.*, 2016; Frumkin *et al*., 2017).

4.4.3 Growth and competitive fitness of the *glyQ* synonymous mutants

The constructed strains were all tested for growth rates and competitive fitness. Growth was tested in different media, with varying degrees of stress so as to tease out any (dis)advantages in growth (if any). Competitive fitness was tested out in KB only.

Neither in LB nor in KB was a significant difference in growth rate observed for any of the SBW25, A506 or Pf0-1 background mutants. Some differences that were seen between one of the Pf0-1 background mutants (A506 Pf0-1(1)) as compared to the cognate wild type is in all likelihood due to some background genomic mutation since the same behaviour of growth was not observed for the second biological replicate of the same strain (A506 Pf0-1(2)). Due to the lack of evidence in growth differences in LB and KB, conditions of stress were introduced for growth of the mutants by way of growing the strains in minimal medium, M9, supplemented with 20% glucose. Under these conditions, neither the SBW25 nor the A506 background mutants showed any differences in growth although the same Pf0-1 mutant that differed in growth in both KB and LB, was once again, differed in growth from the cognate wild type.

Under conditions of faster growth, with Terrific broth, none of the mutants showed significant differences in growth rates as compared to the corresponding wild type strains. The same pattern was observed for growth in Super broth as well. Therefore, although a number of changes have been made in the *glyQ* gene by swapping the alleles of *glyQ* in non-native backgrounds, it is possible that the changes do not have a large enough effect in order to translate to differences in growth rate.

Next, a competitive fitness of the mutants with respect to the wild type strain was carried out. This was done in KB, wherein A506 *glyQ* mutants in Pf0-1 showed significant

difference in competitive fitness relative to the wild type strain. A possible explanation is that the large number of synonymous differences (67 differences across the gene) between A506 *glyQ* and Pf0-1 *glyQ* are large enough to cause a fitness effect when the strains are competed against one another. However, why the reverse combination – Pf0-1 *glyQ* mutant in A506 background – is not true is confounding. It is possible that the reverse combination does not cause large enough effects on the translational output of the gene and is able to meet the translational demands and hence grows as well as the wild type strain. For the SBW25 background strains and A506 background strains, we do not observe any significant difference in competitive fitness as compared to the respective wild type strains, in KB. It is possible that conditions of faster growth might help to resolve differences (if any) between these strains better. Fast growth would demand faster protein translation and if the synonymous mutations in any manner affect protein translation then it would be discernible.

4.4.4 Quantification of *glyQ* expression in native and non-native backgrounds using qRT-PCR

Synonymous mutations may not have a large effect to be observable as differences in growth or competitive fitness but may be observed at the level of mRNA. Therefore, a quantification of *glyQ* mRNA in each of the mutants as well as the wild types was done. For each of the backgrounds, one mutant each showed a significant, albeit weakly significant difference as compared to the wild type. A506 *glyQ* mutant in SBW25 background and Pf0-1 *glyQ* mutant in A506 background showed a significantly lower mRNA quantity as compared to the wild type strains of the respective backgrounds. The A506 *glyQ* mutant in Pf0-1 background strain had a significantly higher amount of *glyQ* mRNA. Synonymous changes are known to affect the stability of mRNA and in the case of SBW5 and A506 mutants, it is possible that the synonymous alterations make the mRNA more prone to degradation and reduce the stability and possibly half-life of mRNA. While, in the case of Pf0-1 mutant, the synonymous alterations possibly make it more stable and hence the relative quantity of mRNA for the mutant is higher than that of the wild type. Also, the *glyQ* expression results are reflected in the competitive fitness results for A506 *glyQ* mutant in Pf0-1 background. The mutant has higher fitness as compared to the wild type and hence it is possible that the higher amounts of mRNA are being translated to higher amounts of protein. This is turn makes the mutant fitter than the wild type as it now possible has more number of mature glycine tRNA, thereby giving it an advantage during competition.

It is therefore evident, that in spite of the large number of synonymous changes effected in *glyQ* gene, all the changes do not translate into large differences in fitness. The synonymous changes made in this part of the study, are a result of evolution and may not necessarily have been due to selection but genetic drift. If these were brought about by genetic drift, then it is no surprise that introducing these does not have massive effects on the fitness. Moreover, it is also possible that the positions in the gene that have these synonymous changes have accumulated through evolution only at these regions as these do not affect the function of the gene; these positions are more "tolerant" to synonymous mutations at least, as these mutations may not directly affect the stability of mRNA, translation speed or accuracy and/or protein folding.

Lastly, a certain combination of number and position of synonymous alterations may translate into fitness differences that have not been achieved in this part of the study. Also, the right conditions for observing a functional effect of the synonymous mutations may not have been tested and as a result the observable effects of the change have not been perceived. As mentioned already, synonymous mutations brought about in *glyQ* through allele swap may not be large enough to translate to growth differences, but may be observed as differences at the mRNA level. Such differences at the cellular level (mRNA or tRNA level) may have arisen due to the synonymous alterations introduced through allele swap, but the techniques used in this study so far do not have the resolution for detecting changes at the level of cellular RNA pool. RNA sequencing or tRNA sequencing techniques will help resolve these small differences.

Chapter 5 : Analysis of routes to mat formation in SBW25, A506, PICF7 and Pf0-1

5.1 Introduction

5.1.1 *Pseudomonas fluorescens* experimental system

In order to manipulate the genome of the model system, *P. fluorescens,* it is imperative to understand the experimental system. *Pseudomonas* as a genus is well known for secretion of exopolysaccharides and leading to formation of structures such as mats and biofilms (Spiers *et al.*, 2003; Friedman & Kolter, 2004b; Chang *et al*., 2007; Ueda & Wood, 2009; Ueda & Saneoka, 2014; Cárcamo-Oyarce *et al*., 2015; Ghods *et al*., 2015; Díaz-Salazar *et al.*, 2017; Farias *et al*., 2019). Currently genome assemblies and annotation reports of 213 *Pseudomonas fluorescens* strains are available on the NCBI database (National Centre for Biotechnology Information) that have been isolated from plant and soil environments around the world. These strains are varied both genetically as well as phenotypically. This is a complex group of organisms that are frequently associated with both plants and animals and range between commensals and pathogens (Silby *et al.*, 2011). In this chapter, four strains of *P. fluorescens* – SBW25, A506, Pf0-1 and PICF7 – are being characterized and compared with respect to a specific phenotype: mat formation at the air-liquid interface of static cultures.

P. fluorescens SBW25 was first isolated in 1989 from the leaf of a sugar beet plant in Oxford, UK (Rainey & Bailey, 1996). It is commonly found associated with leaf surfaces (phyllosphere) and root surfaces (rhizosphere). *P. fluorescens* A506 was isolated from the phyllosphere of pear trees of an orchard from the West Coast of USA (Wilson & Lindow, 1993). This strain has also been shown to be associated with pear leaves and pear blossoms. *P. fluorescens* Pf0-1 on the other hand was isolated from loam soil of a farm from the East Coast of USA (Compeau *et al*., 1988). *P. fluorescens* PICF7 was isolated from the roots of olive plants and has been demonstrated to be an endophytic strain (Mercado-Blanco *et al*., 2004). All of the four strains are Gram-negative, rod shaped, aerobic bacteria that can grow in association with both the rhizosphere and phyllosphere.

5.1.2 Niche specialization

All of the *P. fluorescens* strains – SBW25, A506, Pf0-1 and PICF7 – when grown on King's medium B (KB), produce circular colonies with a defined edge; these are called "smooth" colonies. However, extensive work with *P. fluorescens* SBW25 has demonstrated that it can adapt to a heterogeneous environment, developing niche specialists. In other words, when cultured in a spatially structured environment, cells from smooth SBW25 colonies diversify to variant genotypes that are adapted to specific niches. The niche specialists are distinguishable at the colony morphology level and can be distinguished genetically. Based on tractable genetic assays, the diverse genotypes are categorized as smooth (SM) that colonize the broth, wrinkly spreader (WS) that colonize the air-liquid interface and fuzzy spreader (FS) that form weak raft at the air-liquid interface which later collapse to the bottom of the microcosm, respectively (see Figure 1.5, section 1.2.3.1; (Rainey & Travisano, 1998; Ferguson *et al*., 2013)).

5.1.3 Predicting mat-formation behavior

Mat formation behavior in *P. fluorescens* SBW25 has been described in a series of papers (Rainey & Travisano, 1998; Spiers *et al*., 2002; Spiers *et al*., 2003; Goymer *et al*., 2006; Bantinaki *et al*., 2007; McDonald *et al*., 2009; Ferguson *et al.*, 2013; Ardre *et al.*, 2019). The structural component of the mats in SBW25 was found to be composed of acetylated cellulosic polymer (ACP) (Spiers *et al.*, 2002; Spiers *et al*., 2003). Through genetic studies the *wss* operon was identified to be responsible for formation of ACP (Spiers *et al*., 2002). Mats in *P. fluorescens* SBW25 have been demonstrated to be result of overproduction of ACP due to increased amounts of the bacterial secondary messenger, c-di-GMP. Intracellular concentration of c-di-GMP is modulated through the opposing activities of diguanylate cyclases (DGCs) that increase c-di-GMP levels and phosphdiesterases (PDEs) that degrade c-di-GMP. Therefore, gain-of-function mutations in DGCs and loss-of-function mutations in PDEs would lead to overproduction of c-di-GMP and hence ACP (see Figure 1.4). A suppressor analysis of a WS genotype revealed that mutations leading to the wrinkly spreader phenotype are typically found in one of the three loci – *wsp, aws* or *mwsR* (McDonald *et al*., 2009).

A recent publication describes a possible way to predict similar evolutionary outcomes based on the presence of similar sets of genes in related organisms (Lind *et al.*, 2018). The

work compares six species of *Pseudomonas – P. fluorescens, P. protegens, P. putida, P. syringae, P. savastanoi* and *P. stutzeri*. The author demonstrates that the three main diguanylate cyclases (DGCs, *wspR, awsR* and *mwsR*) in SBW25 vary from being absent in the some strains (*P. stutzeri*) to being present in their entirety in some other strains (*P. protegens, P. putida* and *P. aeruginosa*). Based on the presence of different groups of DGCs, the author makes predictions for the possible evolutionary outcomes of mat formation in *P. protegens* Pf-5. Pf-5 and SBW25 share 33 DGCs, and additionally encode six unique DGCs each. Therefore, it seems logical to think that mat-forming mutations in Pf-5 might appear in a similar order and in similar positions, to those seen in SBW25. Genes for cellulose biosynthesis (*wssA-J,* (Spiers *et al*., 2002)) are absent in Pf-5. Mats in Pf-5 could be therefore be formed by the overproduction of some other exopolymeric substance (EPS) as a consequence of elevated c-di-GMP levels. Experimental evolution with Pf-5 under static conditions, followed by spreading diluted cultures on agar plates, revealed divergent colony morphologies had developed over the course of the experiment. Altogether, 43 independent mutants were isolated, of which 40 mutants were identified to have mutations in the negative regulators of the *wsp, aws* and *mwsR* pathways, as was predicted by the mathematical model.

Given the success of Lind (2018) in predicting the outcomes of evolution in *P. fluorescens* Pf-5, similar predictions regarding mat forming ability may be expected to hold true for other strains of *P. fluorescens*. Such predictions should be based on the presence or absence of genes responsible for the mat formation phenotype in SBW25 (*wsp, aws, mws*).

5.2 Aims

1. To repeat the mat formation experiment with *P. fluorescens* SBW25 in our laboratory including the isolation of ten independent mat forming genotypes, followed by whole genome sequencing and identification of the mutational causes.
2. To determine the presence/absence of genes of interest (*wss*, *wsp*, *aws*, *mws*) in the other *P. fluorescens* strains (A506, Pf0-1, PICF7), using bioinformatics techniques.
3. To experimentally demonstrate the presence/absence of mat forming abilities, in each strain of *P. fluorescens* (A506, Pf0-1 and PICF7).
4. To isolate and phenotypically characterise any colony morphs observed in aim 3.
5. To use whole genome sequencing (Illumina) to identify mutations in the morphotypes isolated in aim 4.

6. To describe the genetic pathways involved in mat formation in each of *P. fluorescens* A506, Pf0-1 and PICF7.

5.3 Results

5.3.1.1 *P. fluorescens* SBW25 forms mats at the air-liquid interface

The ability of *P. fluorescens* strains to form mats in KB static microcosms was determined by incubation of SBW25 cultures under static conditions. Ten independent microcosms (25 ml glass vials with 6 ml KB) were inoculated with individual SBW25 colonies and incubated under static conditions at 28ºC for 72 hours. During this time SBW25 forms a mat at the air-liquid interface and the ancestor diversifies to form different colony morphological types. Each of these morphotypes was identified by vortexing the 72 hours old culture, followed by dilution and spreading the diluted culture on KB agar plates (see section 2.2.11.2 for details).

Figure 5.1 shows the development of mat in SBW25 microcosms at 72 hours and the different colony morphologies that were observed from spreading the culture from 72 hours old microcosms on KB agar plates. Three different colony morphologies were observed – smooth (SM), wrinkly spreader (WS) and fuzzy spreader (FS). SM and WS types were seen on the plates from all ten microcosms, while the FS was observed from only one of ten microcosms. The wrinkly spreader and the fuzzy spreader colonies when purified by streaking on fresh KB agar plates produced exclusively wrinkly spreaders and fuzzy spreaders, respectively. These colony morphologies were therefore heritable. Moreover, the wrinkly spreader when inoculated into KB in a microcosm, formed mats within 24 hours. Ten individual WS types isolated from each of the ten microcosms were purified, grown in liquid culture and stored at -80ºC. Each of these stored WS was later sequenced in order to identify the mutation causes of the WS phenotype.

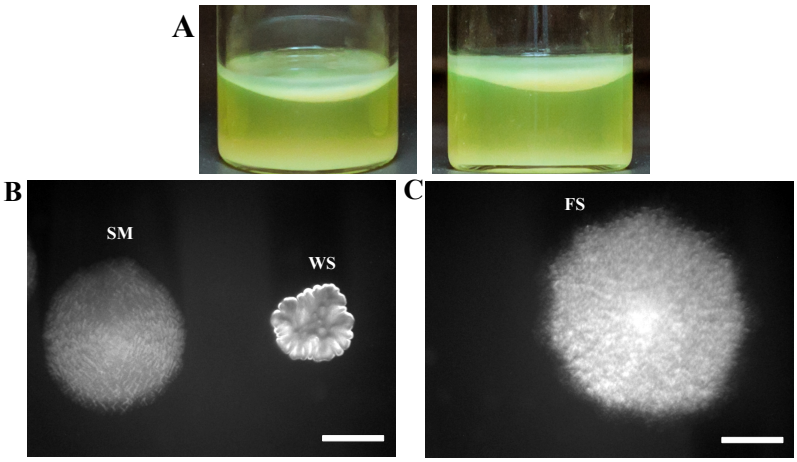


Figure 5.1: **Mat formation and colony morphologies of SBW25 show the expected phenotype. A.** SBW25 forms mats at the air broth interface of the microcosms. **B.** All microcosms showed the smooth (SM) and the wrinkly spreader (WS) colonies. C. Only microcosm 4, showed fuzzy spreader (FS) colonies in addition to the wrinkly spreader (WS) and smooth (SM) colonies. Scale bar is 5 mm.

5.3.1.2 Mutational causes of WS in SBW25 were found to be the *wsp*, *aws* and *mwsR* loci

Identification of the mutational causes of WS in SBW25 was performed with Illumina platform of whole genome sequencing. The ten WS samples isolated in the above section were purified from glycerol stocks, grown in liquid KB and genomic DNA was isolated using DNeasy® Blood & Tissue kit as described in section 2.2.1. Samples were sequenced at the MPI sequencing unit using the MiSeq System from Illumina® by Dr. Sven Künzel. The samples were sequenced as paired end with a run of 2x300 bp and an average of 1,000,000 reads per sample. Sequencing traces were analysed using *breseq* (Deatherage & Barrick, 2014) Geneious® version 10.1.3 (see section 2.2.5.2).

Table 5.1 shows a list of the predicted mat-forming genes that were identified to carry mutations in SBW25.

Strain	Gene	DNA change	Protein change
1	*mwsR*	G3244A	E1082K
2	*mwsR*	A2476T	S826C
3	*mwsR*	T2183C	V728A

4	*wspF*[*]	Δ151-165 (CTGATGGACCTGATC)	Δ51-55 (LMDLI)
5	*mwsR*[*]	G3095T	R1032L
6	*awsR*	A79C	T27P
7	*awsR*	G574A	A192T
8	*wspF*	A820G	T274A
9	*wspA*	C1313T	S438F
10	*wspE*[*]	G1912T	D638Y

Table 5.1: Mat forming genes of SBW25 that were identified to carry mutations using whole genome sequencing. WS colonies isolated from ten microcosms of SBW25 that underwent experimental evolution under static conditions, in parallel, were used for whole genome sequencing. [*]These mutations have been reported previously in (McDonald *et al*., 2009**;** Gallie *et al*., 2015) and (Gallie *et al*., 2019).

Four out of ten sequenced samples (lines 1, 2, 3 and 5) were found to have point mutations in *mwsR*. This gene has both the GGDEF and EAL domains that are responsible for synthesis and breakdown of c-di-GMP, respectively (McDonald *et al.*, 2009). Therefore, loss of function mutations in the EAL domain would lead to increased c-di-GMP levels, overproduction of ACP and hence WS. The mutations in lines 1 and 5 are in the EAL domain and probably are loss of functions mutations that lead to the WS phenotype. Mutations identified in *awsR* (lines 6 and 7), another DGC indicate that these mutations might be activating it, thereby leading to WS. Lines 4 and 8 showed mutations in *wspF*, which is a negative regulator of the methyl accepting chemotaxis protein (MCP) WspA and hence can control kinase activity (Bantinaki *et al.*, 2007). Any loss-of-function mutations in *wspF* therefore would constitutively keep the DGC of the *wsp* operon (*wspR*) active producing the WS phenotype (see section 1.2.3.3). Mutations in lines 9 and 10 were found to be in the *wspA* and *wspE* genes, respectively. These mutations lead to WS phenotype and hence may be causing activation of the genes leading to increased c-di-GMP levels. The mutations observed in lines 4, 5 and 10 have previously been observed in (McDonald *et al*., 2009; Gallie *et al*., 2015; Gallie *et al*., 2019), respectively. This, together with the identification of a mutation in *wsp*, *aws*, or *mwsR* in every isolate demonstrates a high degree of predictability in the evolutionary emergence of WS genotypes in *P. fluorescens* SBW25.

5.3.2 Several genes of mat formation are present in the four studied strains of *P. fluorescens*

In this work, I aim to explain the evolutionary outcomes in related strains of the same species. Based on the DGCs that may be present or absent in SBW25, A506, Pf0-1 and PICF7, one could make predictions regarding mat formation behaviour, as well as the pathways that might be involved in mat formation. The genes responsible for mat formation in *P. fluorescens* SBW25 were identified by transposon mutagenesis of a single wrinkly spreader phenotype (LSWS); transposon mutant colonies were screened to identify those that were defective in wrinkly spreader phenotype and mat formation. Two loci were found to have transposon insertions: the *wss* and *wsp* operons (Spiers *et al*., 2002). The genes contained in operons are listed in Table 5.2 and 5.3; presence (or absence) of the genes in all the four strains that are being studied for mat formation behaviour are also indicted in these tables. Extrapolating the forecasting of evolutionary outcomes from (Lind *et al.*, 2018), it is possible that due to the presence (or absence) or certain groups of genes in the strains of interest, mat formation behaviour in A506, Pf0-1 and PICF7 is similar to that in SBW25. The genes of the *wss* and *wsp* operons in A506, Pf0-1 and PICF7 were identified using gene similarities by BLASTP alignments. Percent identity of the *wsp* gene homologues of A506, Pf0-1 and PICF7 are also mentioned in the table. The percent identity is to the SBW25 *wsp* genes.

Gene	SBW25	A506	Pf0-1	PICF7
wssA	PFLU0300 (+)	-	-	-
wssB	PFLU0301 (+)	-	-	-
wssC	PFLU0302 (+)	-	-	-
wssD	PFLU0303 (+)	-	-	-
wssE	PFLU0304 (+)	-	-	-
wssF	PFLU0305 (+)	-	-	-
wssG	PFLU0306 (+)	-	-	-
wssH	PFLU0307 (+)	-	-	-
wssI	PFLU0308 (+)	-	-	-
wssJ	PFLU0309 (+)	-	-	-

Table 5.2: Presence of the genes of the *wss* operon that are involved in mat formation in SBW25. The table gives a list of genes that are described through previous work to be involved in the biosynthesis of ACP, the structural component of the SBW25 mats. Genes of this operon are present in SBW25 only. No homologues of these genes were identified in A506, Pf0-1 or PICF7.

Gene	SBW25	A506	Pf0-1	PICF7
wspA	PFLU1219	PflA506_1184 (87.41%)	Pfl01_1052 (83.52%)	PFLUOLIPICF7_02270 (95.93%)
wspB	PFLU1220	PflA506_1185 (88.82%)	Pfl01_1053 (76.22%)	PFLUOLIPICF7_02265 (87.06%)
wspC	PFLU1221	PflA506_1186 (86.63%)	Pfl01_1054 (74.29%)	PFLUOLIPICF7_02260 (89.23%)
wspD	PFLU1222	PflA506_1187 (91.42%)	Pfl01_1055 (82.81%)	PFLUOLIPICF7_02255 (91.85%)
wspE	PFLU1223	PflA506_1188 (93.79%)	Pfl01_1056 (84.82%)	PFLUOLIPICF7_02250 (96.29%)
wspF	PFLU1224	PflA506_1189 (95.83%)	Pfl01_1057 (86.13%)	PFLUOLIPICF7_02245 (97.02%)
wspR	PFLU1225	PflA506_1190 (96.10%)	Pfl01_1058 (89.79%)	PFLUOLIPICF7_02240 (98.20%)

Table 5.3: Presence of the genes of the *wsp* operon that are known to be involved in mat formation in SBW25. The table gives a list of genes of the *wsp* operon that are described through previous work to regulate expression of the *wss* genes in SBW25. Genes of this operon are present in all of the four strains. The numeric name of the genes in the four strains is also mentioned. The percentages indicate percent identity with *wsp* genes homologues in SBW25.

A third locus, the *aws* (alternate wrinkly spreader) was identified as involved in the wrinkly spreader phenotype through transposon suppressor analysis of the WS genotype of SBW25 with the *wsp* operon deleted (McDonald *et al.*, 2009). The *aws* operon is a ~2.3kb operon that consists of three genes – *awsX, awsO* and *awsR.* The operon is present in SBW25, and was found to be present in A506 and PICF7. No homologues of these genes were identified in Pf0-1 (Table 5.4). Presence of the genes of the *aws* operon in A506, Pf0-1 and PICF7 were identified using gene similarities by BLASTP alignments. Percent identity of the *wsp* gene homologues of A506 and PICF7 are also mentioned in the table. The percent identity is to the SBW25 *aws* genes.

Gene	SBW25	A506	Pf0-1	PICF7
awsX	PFLU5211	PflA506_4499 (84.21%)	-	PFLUOLIPICF7_12710 (91.58%)
awsR	PFLU5210	PflA506_4498 (89.05%)	-	PFLUOLIPICF7_12715 (93.81%)

awsO	PFLU5209	PflA506_4497 (81.60%)	-	PFLUOLIPICF7_12720 (92.64%)

Table 5.4: Presence of the genes of the *aws* operon that are known to be involved in mat formation in SBW25. The table gives a list of genes of the *aws* operon that are described through previous work to be involved in mat formation in SBW25. Genes of this operon are present in A506 and PICF7 but not in Pf0-1. The numeric name of the genes in the four strains is also mentioned. The percentages indicate percent identity with *aws* genes homologues in SBW25.

Deletion of both the *wsp* and the *aws* locus in SBW25 still did not remove the ability to produce mats and the wrinkly spreader phenotype. Transposon suppressor analysis of this double deletion WS genotype led to the identification of the *mws* (mike's wrinkly spreader) locus (McDonald *et al*., 2009). This is a ~3.8kb locus that consists of a single gene, *mwsR*. MwsR has multiple domains and is a large protein made of 1,283 amino acids. . Homologues of this gene are present in each of the strains (Table 3.5). The presence of *mwsR* in A506, Pf0-1 and PICF7 was identified using gene similarities by BLASTP alignments. Percent identity of the *wsp* gene homologues of A506, Pf0-1 and PICF7 are also mentioned in the table. The percent identity is to the SBW25 *mwsR* gene.

Gene	SBW25	A506	Pf0-1	PICF7
mwsR	PFLU5329	PflA506_4625 (94.15%)	Pfl01_4876 (83.16%)	PFLUOLIPICF7_12195 (96.41%)

Table 5.5: Presence of the *mws* locus that are involved in mat formation in SBW25. The table shows the presence of *mwsR* in the three other strains of interest. The gene is present in all of the strains. In A506, the gene is called *morA*. The numeric name of the gene in the four strains is also mentioned. The percentages indicate percent identity with *mwsR* homologue in SBW25.

Cellulose based mats that are the result of the activation of the *wss* operon, are more common in SBW25. However, a secondary exopolysaccharide, poly-beta-1,6-N-acetyl-D-glucosamine (PGA) can also form the structural basis of mats (Lind *et al*., 2017). Transposon suppressor analysis of the WSΔ*wss* (WS with *wss* operon deletion) revealed insertions in a cluster of four genes that were similar to the *pgaABCD* operon in *E. coli* and *Acenetobacter baumannii*. Even in the absence of the conventional *wss* pathway, SBW25 was still able to form mats through alternative routes. Homologues of the *pgaABCD* genes were found to be present in A506, Pf0-1 and PICF7 (Table 5.6). These were identified using gene similarities by BLASTP alignments. In addition to the phenotype due to the *pga* locus (wrinkly spreaders), another phenotype was observed – colonies like the wild type but with cell

chaining ability (CC) (Lind *et al*., 2017). The CC phenotype could not be attributed to a gene via transposon suppressor analysis since differences at the level of colony morphology were indistinguishable. However, Sanger sequencing provided the necessary information – the CC phenotype was due to mutation in a gene that was similar to the *nlpD* gene in *E. coli.* This gene codes for a lipoprotein, which in turn is an activator of a cell division protein. Loss of function of *nlpD* would therefore lead to improper cell division and hence the cell chaining phenotype. The presence of the *nlpD* locus in A506, Pf0-1 and PICF7 was confirmed and the gene was identified using gene similarities by BLASTP alignments (Table 5.6).

The fuzzy spreader locus (*fuz*) was identified as being involved for the fuzzy spreader (FS) phenotype (Ferguson *et al*., 2013) and in a previous study, it was found to be suppressing the WS phenotype in both the LSWS and the *aws* mutant (McDonald *et al*., 2009). A loss of function mutation in one of the genes of the operon was found to be responsible for lipopolysaccharide (LPS) modification that results in the fuzzy spreader phenotype. Homologues of the *fuz* genes were identified and found to be present in A506, Pf0-1 and PICF7, using gene similarities by BLASTP alignments (Table 5.6).

Besides the above-mentioned operons that have been shown through previous work to be involved in mat formation in SBW25, numerous other operons are known to be present not only in *P. fluorescens* but also in other pseudomonads and are known to be involved in the pathways that could be important to mat formation. One such operon is the *alg* operon that is involved in alginate production which helps *P. aeruginosa* to form thick biofilms (Boyd & Chakrabarty, 1995; Nivens *et al.*, 2001). The presence of the operon therefore would allow these strains of *P. fluorescens* to form surface mats even in the absence of the *wss* pathway. The genes of the *alg* operon are present in SBW25, A506, Pf0-1 and PICF7 and were identified using gene similarities by BLASTP alignments (Table 5.6). The alternative sigma factor AlgU controls the trasncription of the genes in the *alg* operon (Ertesvåg *et al*., 2017).

There are many more genes, in addition to the ones mentioned above, that have been reported to be involved in mat formation pathways, at least in *P. aeruginosa*. Some of these are the genes of the *pel* and *psl* operons (Friedman & Kolter, 2004a; Colvin *et al*., 2012; Jones & Wozniak, 2017; Marmont, *et al*., 2017b). The genomes of the four strains of *P. fluorescens* were screened for the presence for the genes of the *pel* and *psl* operons. The *pel* operon has seven genes (*pelA, pelB, pelC, pelD, pelE, pelF* and *pelG*) in *P. aeruginosa* (Friedman & Kolter, 2004a) and this operon was found to be present in one out of the four strains of *P. fluorescens* that are being studied – PICF7 (Table 5.6). The *psl* operon has eleven genes in *P. aeruginosa – pslA, pslB, pslC, pslD, pslE, pslF, pslG, pslH, pslI, pslJ* and *pslK* (Friedman &

Kolter, 2004b). Besides Pf0-1, all the other strains of *P. fluorescens* – SBW25, A506 and PICF7 show the presence of genes of the *psl* operon. Table 5.6 lists the genes of the *psl* operon and the presence of these genes in the strains of *P. fluorescens*. The *pel* and *psl* operons in each of the four strains of *P. fluorescens* were identified using gene similarities by BLASTP alignments.

Operon	SBW25	A506	Pf0-1	PICF7
pga	+	+	+	+
nlpD	+	+	+	+
fuz	+	+	+	+
alg	+	+	+	+
pel	-	-	-	+
psl	+	+	-	+

Table 5.6: Presence of the genes of other exopolymeric substances (EPS) known to be involved in mat formation. All genes of the *alg, pga, nlpD* and *fuz* loci are present in all of the strains. Genes of *psl* are present in all, except Pf0-1, while genes of *pel* are present exclusively in PICF7.

It is therefore obvious that while the ACP producing genes (*wss* locus) are present exclusively in SBW25 among the four *P. fluorescens* strains, there are a number of other loci that can be activated to form mats. This implies that mats of A506, Pf0-1 and PICF7, if these were to form, would likely develop thorough the overproduction of c-di-GMP. The c-di-GMP would in turn activate one or more of the loci among *pga, pel, psl, alg, nlpD* and other mat forming genes that may be present in these strains. Unlike SBW25, the structural basis of the mats in A506, Pf0-1 and PICF7 would therefore be composed of EPSs that are a product of or more of the genes mentioned above and not ACP.

5.3.3 Mat forming ability of A506, Pf0-1 and PICF7

As has already been demonstrated through previous work with SBW25 (see section 5.3.2), absence of the *wss* genes does not preclude mat formation under static conditions. Drawing on this observation, it therefore seems logical to think that each of A506, Pf0-1 and PICF7 would also form mats under static conditions, although these strains do not have the *wss* genes.

Ten independent colonies of each of *P. fluorescens* A506, Pf0-1 and PICF7 were inoculated into static KB microcosms, which were incubated at 28ºC for 72 hours (see section 2.2.11.2). The formation of mats at the air-liquid interface was closely observed, and at 72 hours the microcosms were harvested and colony morphology on KB agar investigated.

Figure 5.2 shows the development of a mat-like structure in A506 microcosms at 72 hours, and the different colony morphologies that were observed from plating the culture on KB agar plates. The mat-like structure of A506 was visually not as thick as the SBW25 mats. The smooth (SM) colony morphology was observed from each of the ten microcosms; the web-like (Web) colony morphologies were observed on plates from eight out of ten microcosms. On purification of the web-like colonies by streaking on fresh KB agar plates, both web-like colonies and SM colonies were obtained. From these, web-like and SM colony types were again sub-streaked onto KB agar. The web-like colony again gave a mixture of web-like and SM colonies, while the SM colony gave rise to only SM colonies. These observations are consistent with the web-like colony morphology being semi-heritable; that is, the web-like colony is maintained for some generations, but switches back to the SM morphology at higher frequency than is typically seen by mutation. Notably however, when seeded with the web-like colonies, microcosms produced mats within 24 hours of inoculation.

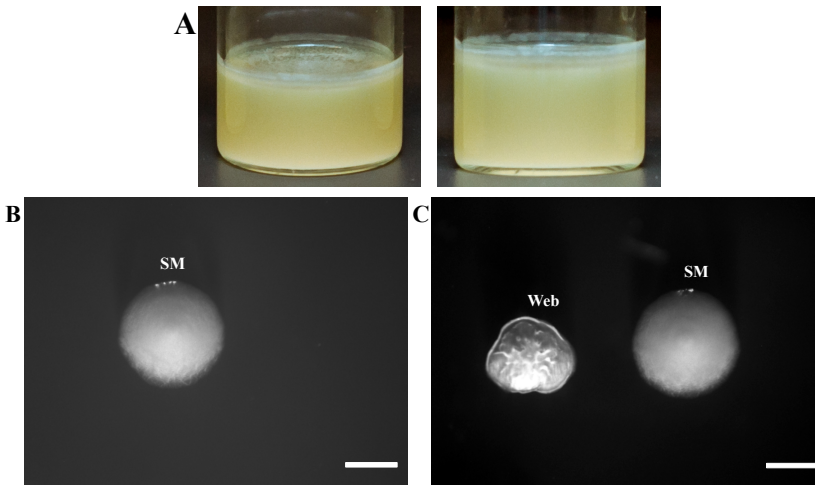


Figure 5.2: Mat formation is similar across all A506 microcosms but colony morphology differs slightly. A. Mat formation after 72 hours. Extent of mat formation is similar across all the microcosms. **B.** The smooth (SM) colony was observed for all microcosms. **C.** The web-like colonies (Web) were observed in all microcosms except microcosms 8 & 10. Scale bar is 5 mm.

P. fluorescens Pf0-1, as described in section 5.3.2, has a number of EPS synthesis genes such as the *alg* and *pga* operons, and biofilm formation by Pf0-1 has been described in

a number of studies (Monds *et al*., 2007; Newell *et al.*, 2009; Newell *et al.*, 2011a; Boyd *et al*., 2012). However, under static conditions, none of the ten microcosms founded with Pf0-1 formed any discernable mats at the air-liquid interface (Figure 5.3). Upon plating dilutions of the culture on agar plates, three different colony morphologies were observed - one colony was a large disc shaped colony (LD), the second was a smaller disc shaped colony (SD), while the third one was a smooth colony (SM). While SM colonies were observed on plates from all ten microcosms, LD were observed on plates from seven out of ten microcosms and SD were observed on plates from nine out of ten microcosms.

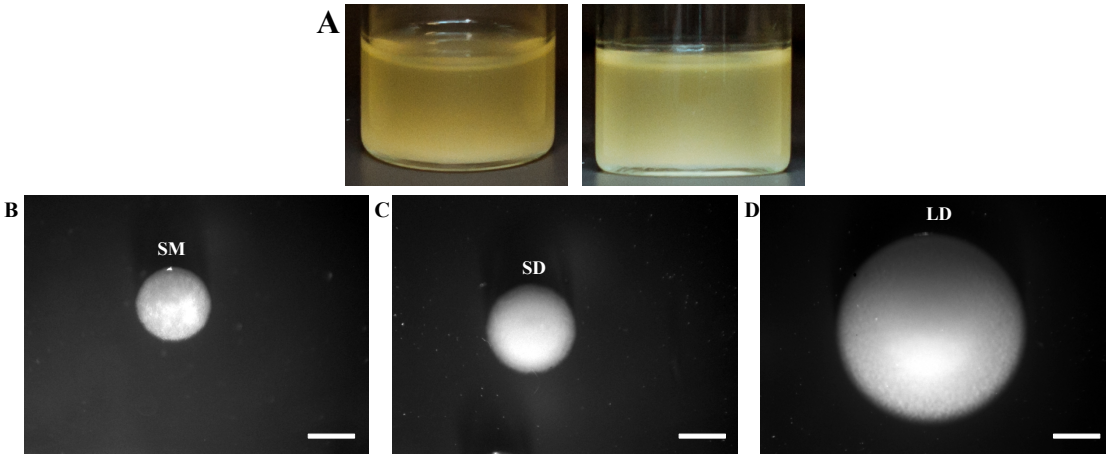


Figure 5.3: Mat formation and colony morphologies are similar across all Pf0-1 backgrounds. A. No mat-like structure was visible even at then of 72 hours. **B.** The smooth colony (SM) was observed for all microcosms. **C.** The smaller disc shaped colony (SD) was visible in all microcosms except microcosm 6. **D.** The large disc shaped colony (LD) was visible in microcosms 7, 8 & 9. Scale bar is 5 mm.

The ten microcosms founded with PICF7 formed mat-like structures that closely resembled those observed in SBW25, as can be seen in Figure 5.4. On KB agar plates, five different colony morphologies were observed. One colony was the smooth (SM) colony morphology, the second one looked similar to the WS from SBW25 microcosms. Both the SM and WS were observed on the plates from all ten microcosms. The third looked similar to the SBW25 fuzzy spreader (FS), and was observed on plates from six out of ten microcosms; the fourth was a disc shaped colony (as was seen for the Pf0-1 microcosms (SD)), observed on plate from only one of ten microcosms; the fifth morphotype was a larger variant of the disc shaped colony (LD), again observed on plate from only one of ten microcosms. The wrinkly spreader (WS) colonies of PICF7 upon purification exclusively produced wrinkly spreader colonies and these colonies gave rise to mats within 24 hours when inoculated into KB in microcosms.

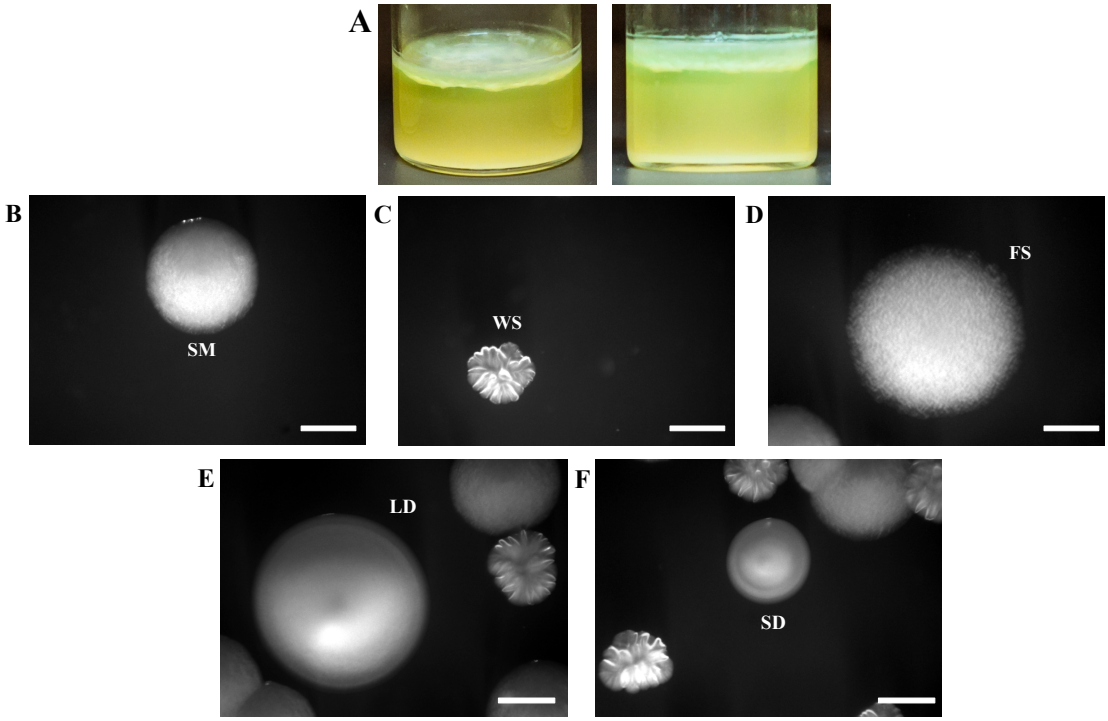


Figure 5.4: Mat formation is similar across all PICF7 backgrounds but colony differs across microcosms. A. Mat formation after 72 hours. **B.** Smooth colony was observed on plates from all microcosms. **C.** Wrinkly spreader like colonies observed across all microcosms. **D.** Fuzzy like colony was observed in some microcosms (microcosms 4-9). **E-F.** Disc shaped colonies were also observed but only in two microcosms. Microcosm 2 had the larger disc shaped colony (E), while microcosm 9 had the smaller disc shaped colony (F). Scale bar is 5 mm.

Altogether, the data suggests that A506 and PICF7 use different strategies to form mats, while Pf0-1 does not form mats in liquid culture. Both A506 and PICF7 form mats in microcosms, under static conditions, at the end of 72 hours. PICF7 forms mats that closely resemble SBW25 mats; A506 forms mats that look much different from the mats of both SBW25 and PICF7. The colony morphological types isolated from all ten microcosms of each of the strains also demonstrate a great degree of variation. On KB agar plates, A506 produced web-like and SM colonies; Pf0-1 produced disc shaped colonies, LD (large disc) and SD (small disc). Contrariwise, PICF7 produced varied colony morphotypes – SM, WS and FS that resemble the WS and FS from SBW25 microcosms and disc shaped colonies, LD (large disc) and SD (small disc). Additionally, the WS colonies of PICF7, upon purification exclusively produced WS colonies, while the web-like colonies generated both web-like and SM colonies. Further purification of the web-like colonies obtained post-purification, produced a mix of web-like and SM colonies. This is indicative that the WS of PICF7 are

generated *via* genetic mutations, but the web-like colonies are semi-heritable; the web-like colonies switch at a frequency much higher than that observed for mutations.

5.3.4 Identifying the mutational causes of the A506 and PICF7 colony morphotypes

In order to understand the pathways involved in mat formation phenotype, it is imperative to identify the genes and the interactions thereof. Mat formation genes and pathways in SBW25 are well studied and have been identified to a great extent (see section 5.1.3). Those in A506, Pf0-1 and PICF7 are yet to be explored thoroughly. However, Pf0-1 did not form any observable mats and hence the morphotypes isolated from Pf0-1 microcosms will not be used any further in the scope of this study.

Ten independent PICF7 WS colonies and ten independent A506 web-like colonies were purified and then grown in liquid culture. Genomic DNA was extracted as described in section 2.2.1 Whole genome sequencing was used to identify the mutations in each of the twenty strains. Results from the whole genome sequencing run were analysed using *breseq* (Deatherage & Barrick, 2014) and Geneious (see section 2.2.5.2). Table 5.7 lists the genes that were found to carry mutations and thus may be the proximate causes for mat formation.

Drawing on the parallels in SBW25 and PICF7 mat appearance, the mutational targets of PICF7 samples (21 – 30) appear to be obvious candidates. Five out of ten lines (22, 23, 28, 29 and 30) show mutations in the homologue of *wspF*, while the others show mutations in homologues of *wspA, awsX* and *awsO* (lines 21, 24 and 26, respectively). Given that in SBW25, WspF is a negative regulator of WspR and loss-of-function mutations are more common, this result is not surprising and as has been discussed in section 5.3.1.2, loss-of-function mutations in *wspF* would lead to constitutive expression of c-di-GMP. Lines 22, 23 and 30 show nonsense mutations (amino acids changed to stop codon), while in the remaining two (lines 28 and 29), one has a transversion and the other, an in-frame deletion. The mutations found in homologues of *mwsR* and *wspA* may increase c-di-GMP levels, leading to increased production of some polymer (polymer that is overproduced leading to WS in PICF7 is not yet known) that produces WS. In SBW25, *awsX* is the negative regulator of *awsR*, while *awsO* functions as an outer membrane porin (McDonald *et al*., 2009). In the inactive state, AwsX is bound to AwsO that dissociates when AwsO receives an as yet unidentified environmental signal leading to physical association between AwsX and two

domains of AwsR. Consequently c-di-GMP is overproduced causing ACP overproduction and WS. Drawing upon the parallels with SBW25 mat formation pathway, I conclude that similar pathways for activation of genes – DGCs and the negative regulators – ensues in PICF7.

Strain	Gene	DNA change	Protein change
PICF7			
21	*PFLUOLIPICF7_02270* (*wspA*)	C161T	S54L
22	*PFLUOLIPICF7_02245* (*wspF*)	C250T	R82(STOP+)
23	*PFLUOLIPICF7_02245* (*wspF*)	C747G	I248(STOP+)
24	*PFLUOLIPICF7_12710* (*awsX*)	Δ232-264	Δ78-88
25	**NGS fail**	**NGS fail**	**NGS fail**
26	*PFLUOLIPICF7_12720* (*awsO*)	T127A	F43I
27	*PFLUOLIPICF7_12195* (*mwsR*)	G2976T	M992I
28	*PFLUOLIPICF7_02245* (*wspF*)	A880C	T294P
29	*PFLUOLIPICF7_02245* (*wspF*)	Δ949-951(GCT)	ΔA317
30	*PFLUOLIPICF7_02245* (*wspF*)	C898T	Q300(STOP#)

Table 5.7: Mat forming genes that were identified to carry mutations following whole genome sequencing. Strain numbers 21-30 are of PICF7 mat forming samples.

Initial whole genome sequencing and comparative analysis of the A506 mat forming strains and the A506 wild type strain revealed possible mutations in two genes – *dppA2* and *dppA3*. Due to low quality coverage, the two genes were sequenced using Sanger sequencing technique. This revealed that *dppA2* and *dppA3* genes did not carry the mutations that were identified to be present in these genes from whole genome sequencing analysis. These mutations may have been an artefact of the sequencing method or analysis. It is therefore implied that mats in A506 were developing *via* an environmental signal and did not ensue through a genetic change.

In conclusion, the strategy used by PICF7 for mat formation closely mirrors the mat formation strategy of SBW25 – overproduction of c-di-GMP, which in turn activates some EPS producing genes. But, as has been mentioned before, the mat formation strategy of A506 is unlike that of SBW25; the mats in A506 are formed through an environment signal that may be activating EPS producing genes, which ultimately leads to mats at the air-liquid interface in microcosms.

5.3.5 Genetic pathways involved in mat formation in PICF7 and A506

Pathways involved in mat formation can be deduced from transposon mutagenesis screens. In this process, the strain of interest is subjected to random transposon mutagenesis and many mutants are screened for loss of the phenotype of interest. This technique is an informative tool in identifying genes and interactions thereof that might be contributing to the phenotype of interest; in this case the colony morphotypes that give rise to mats.

5.3.5.1 Transposon suppressor analysis in PICF7

The mat formation pathways in SBW25 have been studied extensively and still being investigated. The pathways leading up to mat formation in PICF7 and A506, on the other hand are not known. In this section, I use transposon mutagenesis to investigate the structural basis of PICF7 mats.

Four of the ten mat forming lines of PICF7 – line 23, 24, 27 and 28 – were used for transposon suppressor analysis (see section 2.2.10). These strains contained mutations in *wspF* (23 and 28), *awsX* (24) and *mwsR* (27). A total of around 7,875 colonies were screened from eight independent conjugations were screened for loss of the wrinkly spreader phenotype.

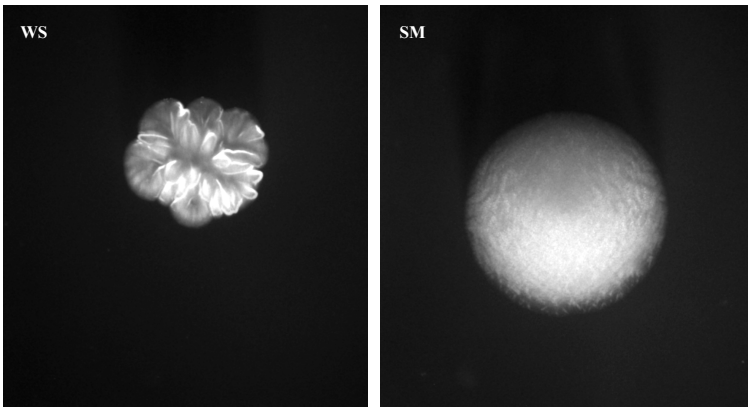


Figure 5.5: The wrinkly spreader phenotype (WS) of PICF7 was used to isolate transposon mutant derivatives (SM). Dissection microscope images of the wrinkly spreader (WS) phenotype of PICF7 and the derived transposon mutant that had lost the wrinkly spreader phenotype and reverted to the smooth (SM) state.

A total of 158 smooth transposon mutants were obtained. Each of these transposon mutants was screened for the loss of WS-mat forming ability under static conditions at 24 hours (see section 2.2.11.2). Those mutants that did not form mats at 24 hours were used for sequencing of the transposon insertion site. Of the 158 smooth (SM) transposon (Tn) mutants, 94 were transposon mutants that had los the ability to form the typical WS-mats. 51 of these were from mutant *wspF* transposon mutant (strain 23), six of these were from mutant *awsX* transposon mutant (strain 24), 16 of these were from mutant *mwsR* transposon mutant (strain 27) and 21 of these were from mutant *wspF* transposon mutant (strain 28). All 94 SM Tn mutants were analysed for the exact insertion site of the transposon by AP-PCR, a two-step semi-degenerate PCR technique with a first degenerate PCR step followed by a standard PCR step (see section 2.2.2.3).

In three out of the four starting genotypes, genes of the *pel* operon were found to have transposon insertions leading to loss of WS phenotype. The *pel* locus has seven genes – *pelA, pelB, pelC, pelD, pelE, pelF* and *pelG*. Protein domain search using IntroProScan from EMBL-EBI revealed that PelA is functionally a hydrolase while PelF is a glycosyltransferase; PelD and PelE are sugar transporters and PelG is a polysaccharide exporter. PelB is a protein that has tetratricopeptide domains that is present in transmembrane proteins involved in protein-protein interactions and assembly of multiprotein complexes. Lastly, PelC has a signal peptide domain (Figure 5.6). The sites of transposon insertions in the genes of the *pel* operon are marked in Figure 5.6.

Additionally, genes of the *wsp, aws* and *mws* operons were also found to have transposon insertions (Table 7.31 in Appendix 3 lists the 94 mutants that were sequenced and the sites where transposon insertions were found).

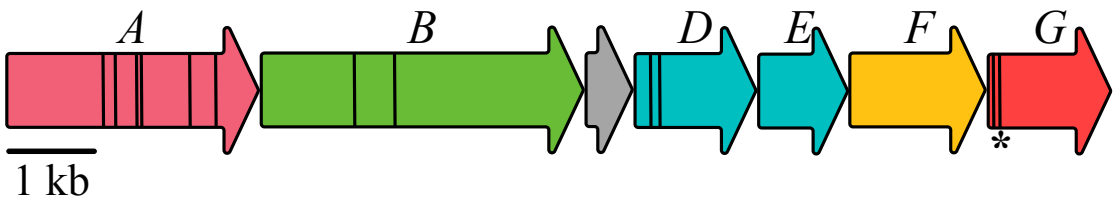


Figure 5.6: Genetic structure of the PICF7 *pel* operon. The grey arrow is *pelC*. Transposon insertion sites in the genes are marked by black lines. Two independent transposon mutants that were sequenced, had the exact same insertion site (indicated by the *).

Pel (N-acetylglucosamine with β(1-4) linkage), a product of the *pel* operon has been reported to be responsible for biofilm formation in *P. aeruginosa* (Vasseur *et al*, 2005; Colvin *et al*., 2012; Jennings *et al*., 2015; Marmont *et al*., 2017b),. Additionally, Pel has been established as the cause of thick pellicles at the air-liquid interface in *P. aeruginosa* biofilms (Friedman & Kolter, 2004b; Colvin *et al.*, 2012; Limoli, Jones, & Wozniak, 2015),. Furthermore, Pel production, like ACP, is dependent on elevated levels of c-di-GMP (Römling *et al*., 2013; Dae-Gon & O'Toole, 2015; Pérez-Mendoza & Sanjuán, 2016). This information indicates that in the absence of the *wss* operon, the *pel* operon is activated under elevated c-di-GMP (due to mutations in *wsp, aws, mwsR*), resulting in the WS phenotype in PICF7 (Figure 5.7).

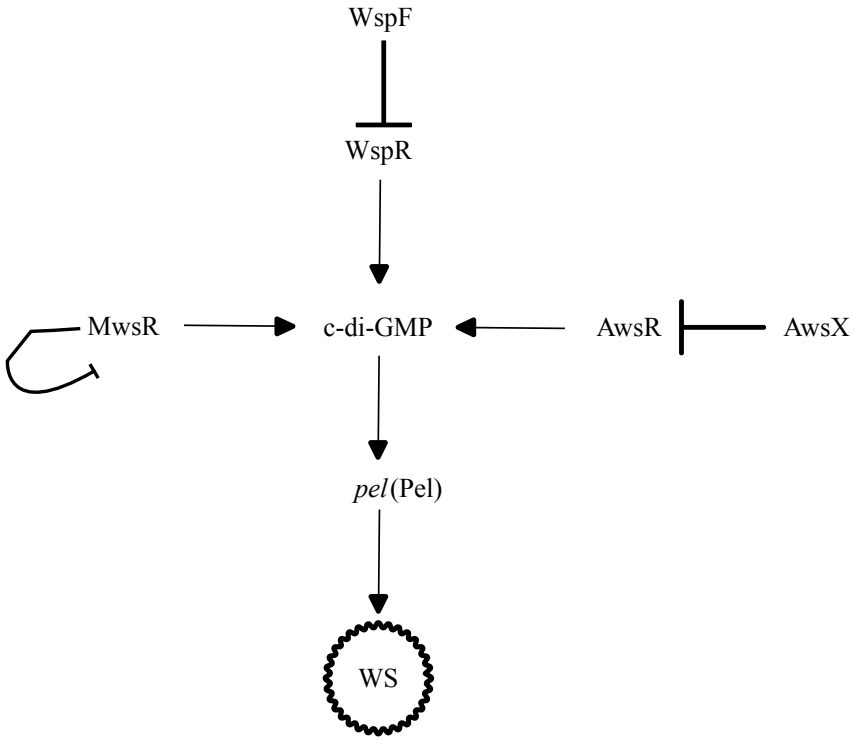


Figure 5.7: One h**ypothesised mat formation pathway in PICF7, based on the transposon suppressor screen.** PICF7 lacks the *wss* operon but is able to make WS colonies as a consequence of c-di-GMP overproduction and activation of the *pel* operon. The structural basis of PICF7 mats is produced by Pel proteins.

5.3.5.2 Mat formation pathway in A506

The web-like colonies in A506 appear to be semi-heritable; the web-like phenotype switches between smooth and web. Therefore, it is intractable to identify mat formation pathways in A506 using transposon suppressor screen, as this screen depends on heritable,

reliable colony morphologies. Instead, in this section, I construct targeted deletions in the A506 genome and assess the effect(s) on the emergent web-like phenotype.

A study with *Salmonella enterica* demonstrated the role of PGA (N-acetylglucosamine with β(1-6) linkage) in formation of a surface mat similar to that observed in my A506 microcosms (Echeverz *et al*., 2017). PGA has been shown to be the primary polymer in biofilms of *E. coli* as well, which occurs through the activation of the *pga* operon (Wang, Preston, & Romeo, 2004; Steiner *et al.,* 2013). PGA production in both *Salmonella* and *E. coli* ensues as a consequence of augmented levels of c-di-GMP. Additionally, the *psl* locus has been observed to be a contributor to biofilms in *P. aeruginosa* (Colvin *et al*., 2012; Dae-Gon & O'Toole, 2015; Limoli *et al*., 2015). In fact the colony morphology of A506 web-like colonies looks very similar to the rugose small colony variants (RSCV) isolated from cystic fibrosis (CF) patients (Irie *et al*., 2010). These RSCVs had elevated levels of Psl.

It is therefore possible that either *pga* or *psl* or a combination of the two loci is the structural basis of the web-like colony morphology of A506. Each of Δ*pga,* Δ*psl* and Δ*pga*Δ*psl* strains of A506 was constructed using the methods of SOE-PCR and two-step allelic exchange (see sections 2.2.2.2 and 2.2.7) in order to determine which of these loci were responsible for the web colony phenotype and hence surface mats. The double deletion strain was constructed through the sequential deletion of *pga* and *psl*. Five independent static KB microcosms were seeded with five individual purified colonies of each of A506, A506Δ*pga* and A506Δ*psl*. The microcosms were incubated at 28ºC for 72 hours and the formation of mats was closely observed. The microcosms were harvested at 72 hours and colony morphology was investigated on KB agar plates (see section 2.2.11.2). Figure 5.8 shows the mat formation in the Δ*pga,* Δ*psl* and wild type strains of A506 and the colony morphological types observed for each of the strains.

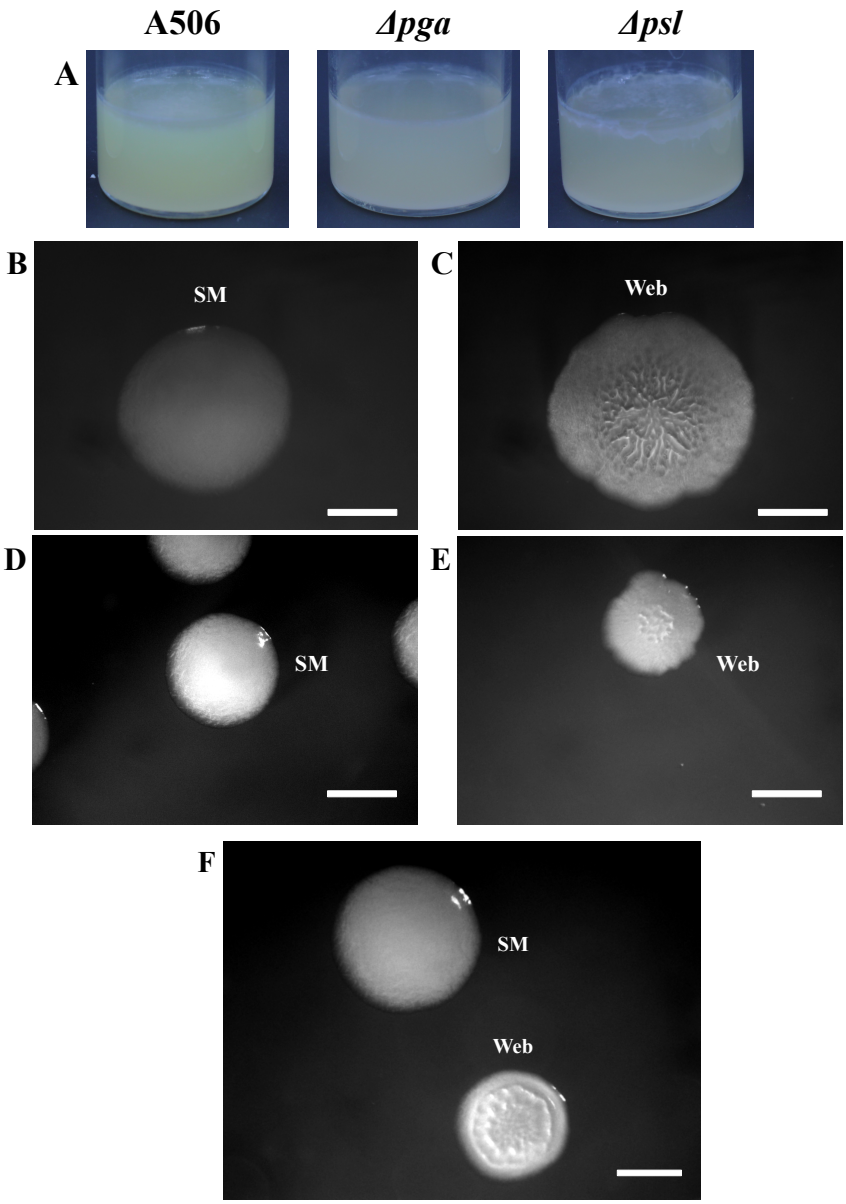


Figure 5.8: Mat formation observed in the wild type A506 and Δ*psl* strains of A506 but not in Δ*pga*. A. Mat formation after 72 hours in A506, Δ*pga* and Δ*psl* strains. **B.** Smooth colony was observed on plates from all microcosms of A506. **C.** Web like colonies observed on plates from all microcosms of A506. **D.** Smooth colony was observed on plates from all microcosms of Δ*pga* of A506. **E.** Web like colonies observed on plates from all microcosms of Δ*pga* of A506. **F.** Smooth and web-like colonies observed on plates from all microcosms of Δ*psl* of A506. Scale bar is 5 mm.

Mat formation in A506 is controlled at least partly by the *pga* operon, as the deletion of this operon led to almost no discernable mat at the air-liquid interface of A506 microcosms. However, some web-like colonies with a less pronounced web-like structure were still observed, indicating that there is a second component in A506 mats. The strategy

that results in mat formation, however, is still not clear – whether through increased amounts of c-di-GMP or some other path. In order to determine whether c-di-GMP overproduction leads to mat formation, a *wspF* loss-of-function mutation was constructed in A506. The construction was done using the methods of SOE-PCR and two-step allelic exchange (see sections 2.2.2.2 and 2.2.7). This particular mutation (15 bp deletion) was based on previous observation with SBW25 that was isolated in this work and published work (see Table 5.1) that leads to formation of a mat and WS colonies. The idea here is that c-di-GMP will be constitutively overexpressed (due to inactivation of the negative regulator of *wspR* DGC), leading to mats in A506 and web-like colonies.

The experiment was set up in the same manner as mentioned in the previous section – six independent purified colonies of each of the *wspF* loss-of-function mutants, reconstructed A506 wild type strain and the A506 wild type strain were seeded into six independent static KB microcosms. The microcosms were incubated at 28ºC for 24 hours and 72 hours, following which the culture was harvested by plating dilutions of the cultures on KB agar plates for visualization of colony morphologies at both time points. Figure 5.9 shows mat formation in two biological replicates of *wspF* loss-of-function mutants at 24 hours, and Figure 5.10 shows mat formation at the end of 72 hours as compared to mat formation in wild type A506 and engineered wild type strain of A506. The *wspF* loss-of-function mutants formed mats by the end of 24 hours, while the wild type strains did not. The wild type strains however formed mats only by 72 hours.

The web-like from the colonies from the *wspF* loss-of-function mutants were purified by streaking on KB agar plates. This generated stable and heritable web-like colonies. Contrariwise, upon purification streaking of the web-like colonies from the A506 wild type strains on KB agar plates, both web-like and smooth colonies were obtained.

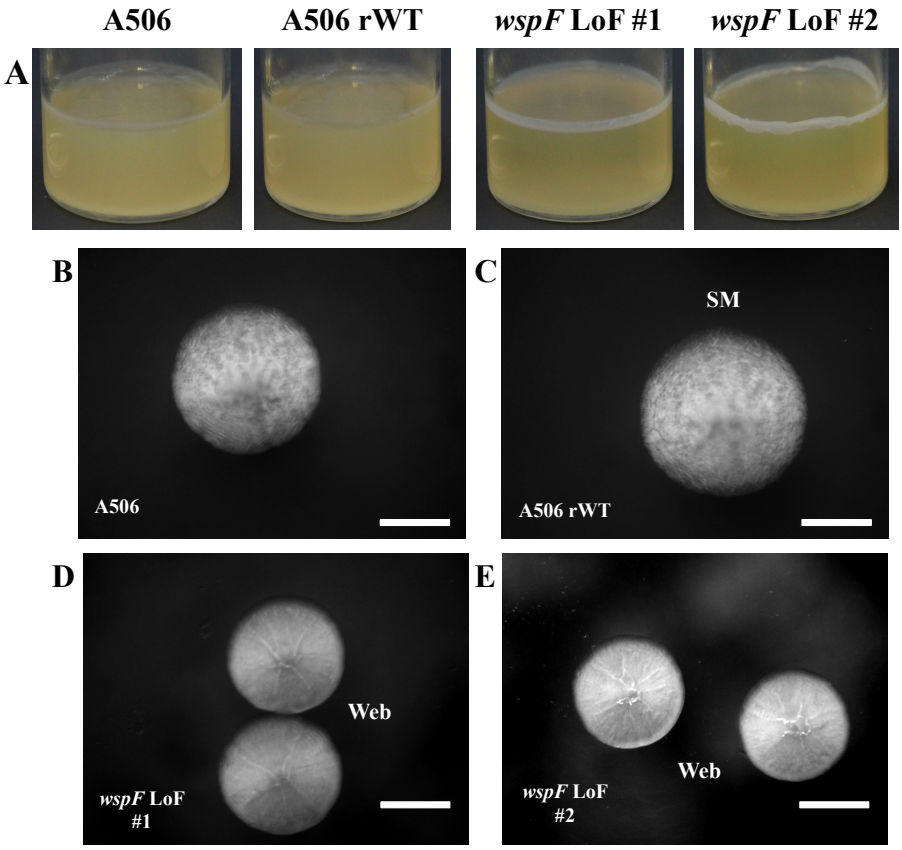
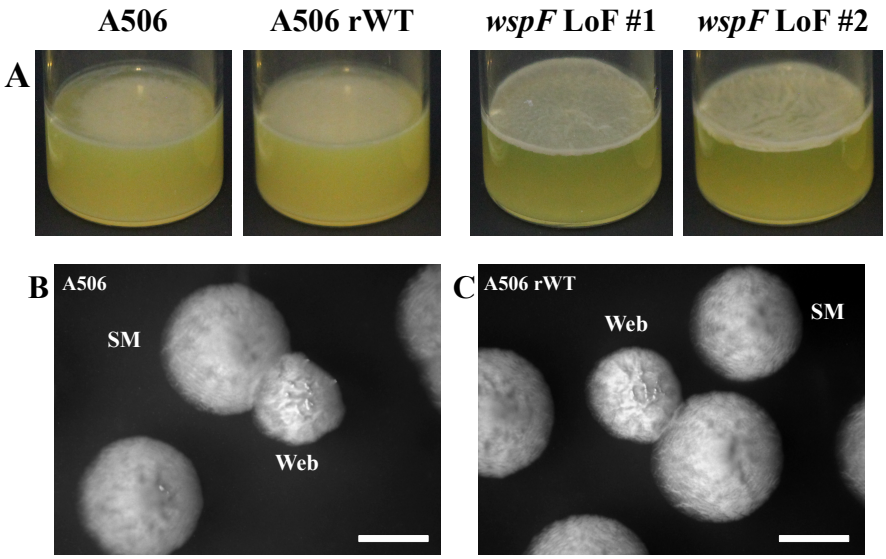


Figure 5.9: Mat formation observed in the *wspF* loss-of-function strains of A506 but not in the wild type strains after 24 hours. A. Mat formation after 24 hours in *wspF* loss-of-functions strains of A506 but no observable mats in the wild type strains. **B-C.** Smooth colony was observed on plates from all microcosms of A506 wild type and reconstructed wild type strains. **D-E.** Web like colonies observed on plates from all microcosms of *wspF* mutants of A506. Scale bar is 5 mm. A506 rWT is the reconstructed wild type strain of A506.



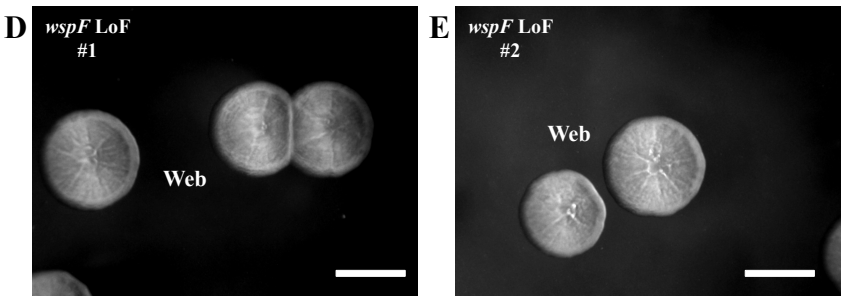


Figure 5.10: Mat formation observed in all of wild type and *wspF* strains after 72 hours. A-D. Mat formation after 72 hours in *wspF* loss-of-functions strains of A506 and in the wild type strains. **B-C.** Smooth colony and web-like colonies were observed on plates from all microcosms of A506 wild type strains. **D-E.** Web like colonies observed on plates from all microcosms of *wspF* mutants of A506. Scale bar is 5 mm. A506 rWT is the reconstructed wild type strain of A506.

Based on the results observed with Δ*pga* and Δ*psl* deletion mutants, it appears that the *pga* operon is involved in mat formation in A506. The web-like colonies that were observed with the *wspF* loss-of-function mutations were phenotypically similar to those generated by the A506 wild type strains. The *wspF* mutant web-like colonies in fact exhibited more pronounced web-like structures. The strategy to mat formation in A506 therefore appears to be *via* the Wsp pathway and likely through the activation and overproduction of the *pga* locus. Identification of the exopolymeric substance that is produced to generate mats in the *wspF* loss-of-function strains of A506, by techniques such as mass spectrometry, in addition to the genetic studies mentioned here, might shed more light on the pathway that A506 uses for mat formation in general and in the presence of elevated c-di-GMP levels. Altogether, the results indicate that the mat formation in A506 might be due to PGA overproduction that may be induced due to increased c-di-GMP. This is in turn is mediated by the Wsp chemosensory pathway (Figure 5.11).

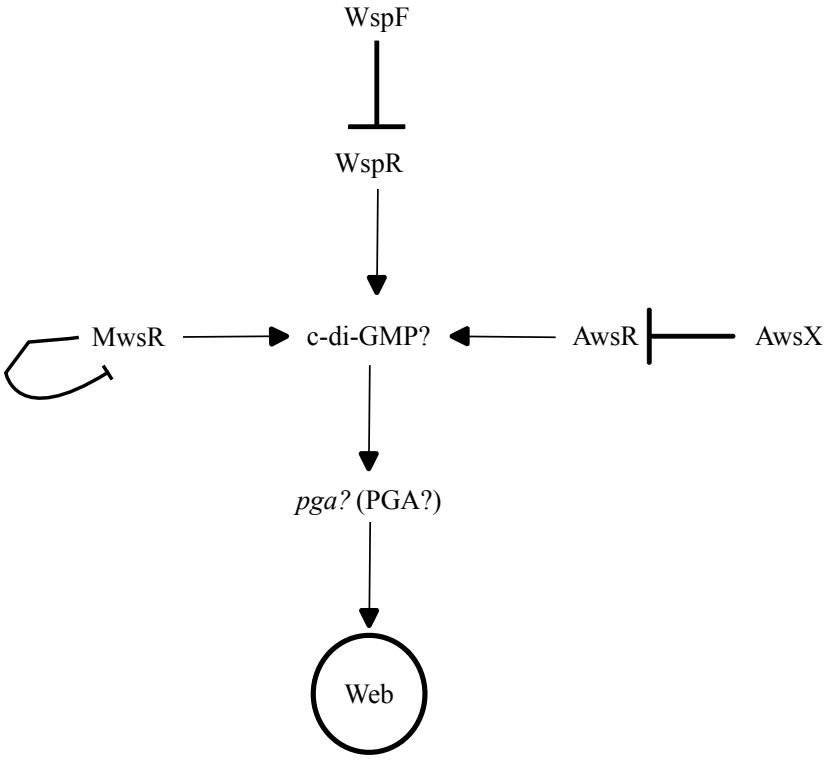


Figure 5.11: One hypothesised mat formation pathway in A506 based on mat formation by *pga* and *psl* deletion mutants and *wspF* loss-of-function mutants of A506. A506 lacks the *wss* operon but is able to make web-like colonies as a possible consequence of c-di-GMP overproduction and activation of the *pga* operon. The polymer in A506 mats in all likelihood is PGA.

5.4 Discussion

Altogether, the data demonstrates, that all the four *P. fluorescens* strains have a varied range of EPS producing genes and all strains employ different strategies to mat formation. Of the four strains, only three (SBW25, PICF7 and A506) demonstrated the ability to form mats in a spatially structured environment (Figure 5.12A). SBW25 has the *wss* locus and the structural basis of the SBW25 mats is ACP. On the other hand, PICF7, which lacks the *wss* genes, uses the *pel* genes to form mats; Pel proteins generate the structural basis of the PICF7 mats. A506, on the other hand has an entirely different strategy and mat formation in A506 ensues through an environmental signal that in all likelihood activates the *pga* locus, possibly *via* the overproduction of c-di-GMP. Pf0-1 does have some of the EPS producing genes, however, it was unable to form any discernible mats in liquid culture. Nevertheless, on plating

of the Pf0-1 microcosm cultures, it, in addition to the other three exhibited distinct colony morphologies that had diversified from the ancestral smooth colonies (Figures 5.12B, 5.12C, 5.12D and 5.12E).

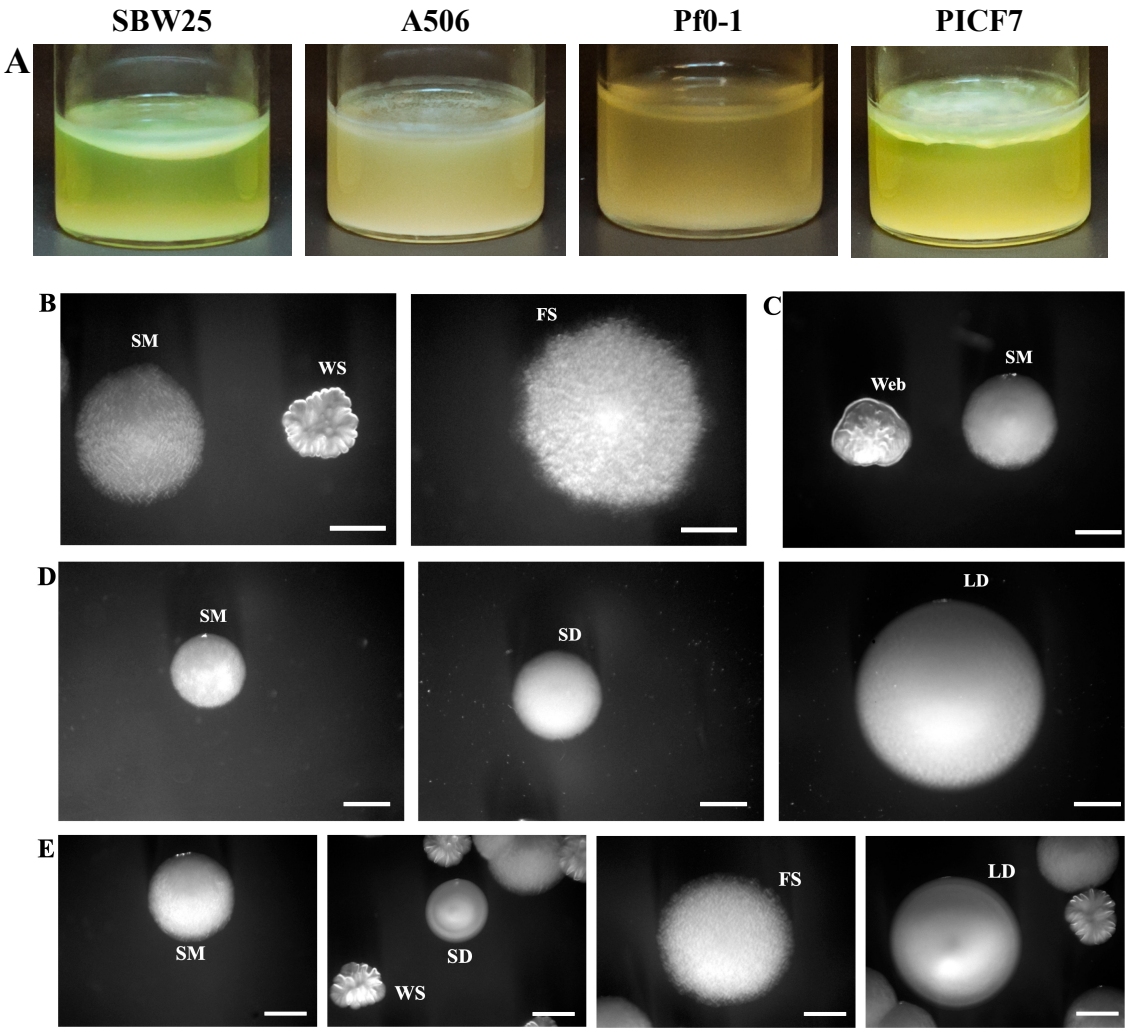


Figure 5.12: Mat formation and colony morphotypes isolated from SBW25, A506, Pf0-1 and PICF7 microcosms. A. Mat formation was observed for SBW25, PICF7 and A506 strains but not for Pf0-1. **B.** SBW25 produced smooth (SM), wrinkly spreaders (WS) and fuzzy spreaders (FS). **C.** A506 produced smooth (SM) and Web-like colonies. **D.** Pf0-1 produced smooth (SM), small disc shaped (SD) and large disc shaped (LD) colonies. **E.** PICF7 produced smooth (SM), wrinkly spreader (WS), fuzzy spreaders (FS), small disc shaped (SD) and large disc shaped (LD) colonies. Scale bar: 5mm

5.4.1 Mat formation experiment with *P. fluorescens* SBW25

In this chapter, mat formation in SBW25 was successfully repeated; SBW25 formed mats at the air liquid interface and produced three different colony morphotypes – smooth (SM), wrinkly spreader (WS) and fuzzy spreader (FS). On whole genome sequencing of WS, mutations were identified to be in *mwsR, wspF, awsR, wspA* and *wspE*. Three of ten mutations identified across ten WS populations have been reported in previously published work, indicating that particular mutations are commonly observed. Either these mutations are more likely to arise (*e.g.*, deletions in repetitive DNA), or provide a large fitness advantage and thus rapidly sweep through the population once they arise.

5.4.2 Genes of mat formation in *P. fluorescens* strains

Several genes and operons have been reported to be involved in mat formation in varied species of *Pseudomonas*. The *wss* loci is involved in mat formation in SBW25, however, even in the absence of ACP (product of *wss*), other exopolymeric substances (EPS) may produce mats. A number of genes that manufacture diverse EPSs are present in A506, Pf0-1 and PICF7. Therefore, it could be predicted that these strains, like SBW25, might form mats at the air-liquid interface upon induction of one or more of the loci that generate EPS.

5.4.3 Mat forming ability of A506, Pf0-1 and PICF7

Based on the repertoire of EPS producing genes present in A506, Pf0-1 and PICF7, it might be predicted that these strains would form mats under static conditions. These strains were tested for mat forming abilities – while A506 and PICF7 formed mats, Pf0-1 did not form any observable mats. The inability of Pf0-1 to form mats at the air-liquid is surprising as this strain has many of the EPS producing genes. Studies reporting biofilm formation by Pf0-1 demonstrate the ability of Pf0-1 to form biofilms via induction of *lap* genes in response to elevated c-di-GMP levels (Monds *et al.,* 2007; Newell *et al*., 2009; Newell *et al.*, 2011b). Therefore it is remarkable that under static conditions, Pf0-1 did not form any detectable mats.

The PICF7 mats resembled the SBW25 mats, and these also produced WS and FS colony morphotypes. However, PICF7 mats are achieved by c-di-GMP mediated activation of the *pel*

operon. A506 forms mats at the air-liquid interface, although these are unlike the mats of SBW25 and PICF7. The colony morphotypes of A506 mat forming variants are also distinctive. These are only semi-heritable, web-like colonies. The structural basis of these colonies and mats is at least in part the *pga* locus, and again is mediated through c-di-GMP. Interestingly, mat formation in A506 appears to occur by a readily reversible, less heritable mechanism as opposed to by a sub-population of stable mutant cells. The semi-heritable nature of A506 mats is possibly attributed to the activation of polymer biosynthetic genes that are associated with a positive feedback loop, in response to an environmental signal. For instance, in the study by (Remigi *et al*., 2019), the switching between capsulated and non-capsulated cells occurs through a balance between the translation and translation inhibition of a positive regulator for capsule synthesis. In the aforementioned study, translation of the positive regulator (*pflu3655*) establishes a positive feedback loop, wherein, PFLU3655 induces expression from its own promoter, thereby increasing its quantities and maintaining the capsulated stage. However, sequestration of *pflu3655* by its inhibitor abolishes capsulation. The balance between these two activities – translation and inhibition of *pflu3655* maintains capsulation bistability. A similar balancing act may be active in A506 mat forming strains. The positive feedback loop for the polymer biosynthetic genes (in response to environmental signal) maintains the web-like colonies. But when the signal subsides, the feedback loop subsides in response, increasing the number of cells that switch to the SM state.

5.4.4 Identifying mutations in A506 and PICF7 morphotypes

The mutations leading to WS and web-like colonies and consequently mats at the air-liquid interface were identified using whole genome sequencing. The mutations identified in PICF7 WS morphotypes were comparable to the ones that were observed in SBW25. Majority of mutations were identified to be in the gene that is a homologue of *wspF*. This result is not surprising as WspF is a negative regulator of WspR and loss-of-function mutations are more common. Any loss-of-function mutation in *wspF* would lead to constitutive expression of c-di-GMP, which in turn could switch on one or more EPS producing gene leading to mat formation in PICF7. The other mutations were identified in homologues of *wspA, mwsR, awsO* and *awsX*. AwsX is a negative regulator of AwsO in SBW25 (McDonald *et al*., 2009). Mutations inactivating AwsX, while activating AwsO, WspA and MwsR might lead to overproduction of c-di-GMP and consequently mats and WS.

No mutations were found in the A506 mat forming strains on whole genome sequencing. The web-like colonies isolated from the A506 microcosms were semi-heritable and altogether this information implies that A506 mats form through an environmental signal and not as a result of a mutation.

5.4.5 Genetic pathways involved in mat formation in A506 and PICF7

The potential pathways that lead to mat formation in PICF7 were identified using transposon suppressor screen. Multiple, independent transposon insertions were found to be in various genes of the *pel* operon. Pel is an exopolymeric substance and has been reported to be involved in biofilm formation in *P. aeruginosa*. Altogether, from the whole genome sequencing and transposon suppressor screen results, it seems that Pel overproduction leads to mats and WS morphotypes in PICF7, which in turn is a result of overexpression of c-di-GMP.

The mat formation pathway in A506, on the other hand does not appear to be as straightforward. Firstly, transposon suppressor screen could not be used for web-like colonies as these switch between the web-like phenotype and smooth phenotype and therefore a screen for loss of web-like phenotype would be challenging; it would be difficult to distinguish the transposon suppressors from those that spontaneously switched to the smooth phenotype. Potential EPS loci for A506 mat forming variants were identified based on similar mat morphologies and deletion mutants of some of these EPS loci showed promise. The *psl* deletion mutants still formed surface mats and web-like colonies; however, the *pga* deletion mutants demonstrated weak or nearly undetectable mats and the web-like colonies of these cultures had less distinctive web-like structures. Another set of mutants – *wspF* loss-of-function mutants – was constructed in order to identify the role of c-di-GMP, if any, in A506 mat formation. These mutants did form mats, and the web-like colony morphology of *wspF* mutants was like those of the A506 mat forming variants. It is possible that the *pga* locus has some role to play in mat and web-like colony formation in A506. However, it is likely that augmented c-di-GMP levels may be targeting a different EPS producing locus. The nature of the EPS that is produced in response to c-di-GMP may be identified using among other techniques, mass spectrometry. Identification of the EPS with other methods, in addition to the genetic techniques used in this study, may provide further insight into the mat formation strategy of A506.

Chapter 6 : Discussion

6.1 Overview

The genetic code is degenerate: there are 61 protein-coding codons in the standard genetic code, which translate to 20 proteinogenic amino acids. This leaves multiple codons for most amino acids; such multiple codons are referred to as synonymous codons. A multitude of genomic data and studies indicate that synonymous codons are not used equally. Across the tree of life, preferential use of synonymous codons has been observed – a phenomenon called codon usage bias (Sharp, Tuohy, & Mosurski, 1986; Plotkin *et al*., 2004; Camiolo, Farina, & Porceddu, 2012; Tello *et al*., 2013; Trotta, 2013; Goodman *et al*., 2013; Boël *et al*., 2016; Narula *et al*., 2019). There is a mounting set of studies that quantify the degree of preference for codons in different genomes (Sharp & Li, 1987; Reis *et al*., 2004; Xia, 2007; Sabi & Tuller, 2014). However, such quantification attempts do not shed light on the evolutionary forces that shape the skewed use of codons.

With the relative ease of use of sequencing technology, it is becoming apparent that many synonymous changes are not neutral, as was often assumed previously. Recently, studies have provided evidence that a number of synonymous mutations, either alone or in combination with other synonymous mutations, in fact have non-neutral effects (Agashe *et al*., 2013; Agashe *et al.*, 2016; Kristofich *et al*., 2018). Synonymous mutations can, in some instances, impair growth and competitive fitness to almost the same degree as non-synonymous mutations (Lind *et al*., 2010). In the study by Lind *et. al.,*(2010), the distribution of fitness effects of synonymous mutations in two ribosomal protein genes (*rpsT* and *rplA*) of *Salmonella typhimurium* were found to be similar to that of non-synonymous mutations. In fact a large number of synonymous mutations in the aforementioned genes had significant effects on fitness. The fitness effects were attributed to alteration in mRNA structure and stability. Other studies with bacteria have also demonstrated that synonymous substitutions can have varying degrees of non-neutral effects on the fitness of the bacteria (Agashe *et al*., 2013; Bailey *et al.,* 2014; Agashe *et al*., 2016; Kristofich *et al.*, 2018). Despite the increasing number of studies demonstrating the non-negligible impact of synonymous mutations on cellular physiology, extensive knowledge of the drivers that shape preferential codon use is lacking. However, some information is available about the role of different factors that might contribute to the evolution of codon usage bias. One of the factors that has been found to be significantly correlated to codon use is the tRNA gene set of the organism (Ikemura, 1981;

Ikemura, 1982; Duret, 2000; Higgs & Ran, 2008; Novoa *et al*., 2012; Yona *et al*., 2013; Frumkin *et al*., 2018).

A second factor that has been proposed to affect the evolution of codon usage is mRNA structure. The structure and stability of mRNA has been shown to be influenced by altered codon identities, not only in unicellular organisms but also in multicellular organisms (Presnyak *et al*., 2015; Bazzini *et al*., 2016; Harigaya & Parker, 2016a; Harigaya & Parker, 2016b; Radhakrishnan *et al*., 2016).

The idea that tRNA availability and mRNA structure and stability have a bearing on codon usage results from the fact that these factors affect translational speed and efficiency. Similar codons – either identical codons, or codons that can be decoded by the same tRNA through wobble base pairing – tend to cluster together in mRNA, thereby increasing speed of translation (Cannarozzi *et al*., 2010). Theoretical work also lends support to the role of tRNA availability and rate of codon translation (Rudorf & Lipowsky, 2015). Availability of mature tRNAs, ternary complexes (aminoacyl-tRNA+GTP+ribosome), and charging of discharged tRNA, all contribute to codon elongation rate. This rate would be specific to the identity of not only the codon of interest, but also the neighbouring codons. Accurate and efficient decoding of codons would depend on the local concentration of cognate tRNAs. Increased elongation rates due to misincorporation of amino acids can therefore be avoided if codons decoded by the same tRNA are present together in the mRNA.

Codon identity can affect not only translation elongation, but also initiation. Measurement of translation rate both *in vivo* (Morisaki *et al*., 2016; Yan *et al.*, 2016) and *in silico* simulation of translation initiation and elongation (Shah *et al.*, 2013) have shown that initiation is the rate-limiting step of translation. Rare codons at the initiation site aid efficient protein synthesis by eliminating the chance of ribosome crowding along the mRNA (Mitarai, Sneppen, & Pedersen, 2008; Zarai, Margaliot, & Tuller, 2016). However, as observed in yeast, beyond an optimal number of rare codons at the initiation site, rare codons at the 3' end of an mRNA transcript can significantly reduce the output of translated protein (Chu *et al*., 2014).

Altogether, codon identities can have an impact on structure of mRNA and its stability, which in turn can affect the output of protein. Not only the structure of mRNA but also protein synthesis rates and efficiencies may be affected by altering codon composition. Therefore it is not entirely surprising, that although altered codons do not change the primary sequence of proteins, they may still have a bearing on the fitness of an organism. Modified

mRNA structure and/or stability can also affect mRNA quantities, which can lead to decreased (or increased) protein amounts.

6.2 Review of the findings

6.2.1 Bioinformatics screening for codon biased of *P. fluorescens* genomes

The genome wide codon usage is overall similar in all four strains tested of *P. fluorescens* – SBW25, A506, Pf0-1 and PICF7 – however, the proportions of the most used codons in all of these four strains were different. The differences in codon usage are found to increase with increasing evolutionary distance between the strains. The differences increase when compared with another species of *Pseudomonas* (PAO1) and most pronounced when compared with an entirely divergent genus of bacteria (*E. coli* B REL606).

A screen for genes that showed only synonymous differences (and no non-synonymous differences) across *P. fluorescens* SBW25, A506 and Pf0-1 was performed at the start of this project in conjunction with. Dr. Frederic Bertels. Three genes are found to have these properties – *glyQ, acpP* and *rpsJ* – genes involved in charging tRNAs for glycine with the cognate amino acid, a lipid metabolism gene and the ribosomal protein gene S10, respectively. Additionally, genes involved in two-component signal transduction (TCST) pathway were also analysed for codon usage biases. The genes of the TCST were central to this study as mat formation is a characteristic feature of many species of *Pseudomonas* (Spiers *et al*., 2002; Friedman & Kolter, 2004b; Chang *et al*., 2007; Ueda & Wood, 2009; Ghods *et al*., 2015; Farias *et al*., 2019). Moreover, mat formation behaviour of the four strains of *P. fluorescens* has been extensively investigated in the scope of this study.

The TCST integral to mat formation in *P. fluorescens* involves the histidine kinases (HK) and diguanylate cyclases (DGC). Some of the more important HK and DGCs identified to date, to be involved in mat formation in SBW25 are genes of the *wsp* and *aws* operons and the *mwsR* gene. The codon use of HK and DGC of each of these operons was compared to the genome wide codon use for each of the four strains. While the codon use of *wspE*, the HK of the *wsp* operon mirrors that of the genome in SBW25 in terms of preference for most preferred codon, *awsO* (HK of *aws* operon), *wspR* and *awsR* (DGC of the *wsp* and *aws,* respectively) varies in preference for codons. The proportions of the codons, however, are

different in the genes as compared to the genomes, irrespective of preference. For A506, Pf0-1 and PICF7, the preference as well as proportion for codons in each of the genes and genomes is dissimilar. In addition, the negative regulator of the Wsp pathway, *wspF,* was also screened for codon use (dis)similarities. Again, not surprisingly, the codon use pattern of *wspF* alleles from each of the four strains varied. Genes of the TCST were screened for codon use pattern as (dis)similarities between these genes might be indicative of the potential distinctiveness in mat formation pathways among the strains. It is important to bear in mind that the differences in codon use might only be *indicative of potential* differences and may not be realised. Codon adaptation index of *wspF* was computed as well. This gives a measure of how well adapted the gene is to the genome in terms of codon use and translational output and a higher value is indicative of adaption to faster translation. This is true especially in unicellular organisms such as bacteria since bacteria have shorter generation times and hence faster translation would provide some fitness advantage. The *wspF* gene may not be the most highly expressed gene as it is involved in a pathway that may not be the most extensively used for survival of a bacterium. CAI of *wspF* in the four strains was found to range between 0.67 and 0.69, which is in the moderate range and is indicative that modest amounts of *wspF* are probably produced for regulation of *wsp* pathway.

Genes that exclusively show differences in synonymous codon use – *glyQ, acpP* and *rpsJ* – were screened as well. For SBW25, A506 and Pf0-1, the codon use of *acpP* and *rpsJ* is different with respect to the genome wide codon usage for both codon preference and proportion, while that of *glyQ* is more similar to the genome wide codon usage. The codon preference in *glyQ* in the three strains is identical except for serine codon use. Next, the codon adaptation index (CAI) of the genomes of all four strains of *P. fluorescens* was computed, as was the CAI of the three genes. Comparison of genome CAI revealed the degree of relatedness of the strains – with respect to codon use, SBW25 and A506 are comparable, while PICF7 and Pf0-1 differ in codon use from both SBW25 and A506. In all of the strains, *glyQ* has the highest CAI, which indicates that the codon usage pattern of *glyQ* might be quite similar to that of the genome. CAI of *acpP* is intermediate of the other two genes, corresponding to the intermediate level of variation in codon use of *acpP* as compared to the genome. Across all of the strains, *rpsJ* has the lowest CAI, which is indicative that the codon usage pattern of the gene differs extensively from the genome wide codon usage.

Altogether, the codon usage pattern of the genomes differs from each other and this might be reflective of the varied evolutionary forces that might have independently shaped the genomes, as the strains were isolated from different environmental niches. The variation in codon usage pattern is also reflected in the codon adaptation of the genomes and the genes.

This is further indicative of the relatedness of the strains and adaptation of the genes to their associated genomes.

6.2.2 Effects of synonymous codon changes in *glyQ*

To test the effect of altering codon usage patterns, *glyQ* alleles were swapped into non-native backgrounds. For this study *glyQ* was selected, as this gene has the maximum number of synonymous differences between the three strains (SBW25, A506 and Pf0-1). This gene codes for the enzyme glycine-tRNA synthetase (also known as glycine tRNA ligase). Therefore if altering the synonymous codons of the gene affects mRNA structure and/or stability and/or protein translation (initiation, elongation or termination) in any way then the output of the gene would be affected. Consequently, the cellular pool of glycine tRNAs would be altered that might result in modified global (protein) output of glycine rich genes. This could potentially affect the physiology and hence the fitness of the bacteria.

The "natural" synonymous variants of *glyQ* were constructed using the scar-free technique of SOE-PCR followed by two-step allelic exchange (see sections 2.2.2.2 and 2.2.7). In total, six mutants of *glyQ* – *glyQ* from each strain in each of the two non-native backgrounds were constructed, in addition to three partial mutants that were isolated from the engineering process and engineered wild type strain for each background (see Figure 4.1 and Table 4.1). Furthermore, an attempt was made to construct the "designed" synonymous variants of the SBW52 *glyQ* (see section 4.3.2). These could not be constructed successfully. It is possible that the synonymous mutations engineered for these "designed" constructs, for instance, in the case of the slow and flip allele, were too many such that it became extremely deleterious and hence lethal for the organism. In the case of the faster allele, synonymous mutations were engineered in the beginning of the gene, which perhaps is not tolerant to changes. The beginning of the gene is usually conserved and any mutation culminates in being deleterious to the organism. Following construction of the "natural" synonymous variants of *glyQ*, the mutants and wild type strains were subjected to phenotypic characterization; this would provide insight into how changing synonymous codons in *glyQ* might affect the fitness of *P. fluorescens* strains.

Firstly, the strains were used for measuring growth rates and competitive fitness. Growth assay was performed in Lysogeny broth (LB), King's medium B (KB), minimal medium M9 and enriched media – Terrific broth (TB) and Super broth (SB). None of the mutants show

any significant differences in growth rate with respect to the corresponding wild type strains in either LB or KB. Next, the mutants and the corresponding wild type strains were grown in minimal medium conditions, in M9 minimal medium. Since this is a minimal medium with a single carbon source, this environment creates stress and hence strains that have any (dis)advantage due to synonymous codon changes would exhibit a lower or higher growth rate. But even under minimal medium conditions none of the mutants show any significant differences in growth rate as compared to the corresponding wild type strains. As no differences in growth rate are observed in either LB or KB or M9, strains were grown in enriched media to generate the stress of higher growth demands. This would be a stress as an enriched medium would cause elevated translational demands and the "fit" strains would cope with conditions of higher translational stress as opposed to the less fit strains. Both TB and SB were used for this purpose but again, significant differences in growth rate are not observed. The lack of significant differences in growth rate under regular or stressful growth conditions is unexpected; the synonymous changes effected in *glyQ* might be at positions that are tolerant to changes and hence the synonymous codon changes do not translate into fitness differences. It is likely that the synonymous differences across *glyQ* gene from SBW25, A506 and Pf0-1 have evolved over time due to interactions that are specific to the environment from where the strains were isolated. The uniqueness of the niche and specificity of interactions has led to this divergent codon usage of *glyQ* in the three strains. The growth environments used in the laboratory do not imitate the conditions in the wild and hence the fitness effects (due to allele swap) that might have been observable in nature are not discernible in a laboratory setting.

Next, the competitive fitness of the strains was tested against each of the cognate wild type strains (see section 4.3.3.6). Both biological replicates of the A506 *glyQ* in Pf0-1 background mutant show significant differences in competitive fitness as compared to the cognate wild type strain. One possible explanation for this is that the there are 67 synonymous differences between A506 and Pf0-1 *glyQ*, maximum for the six gene genome combinations. The mutant in fact is seen to be fitter than the wild type strain. The codon order and identity of A506 *glyQ* in Pf0-1 background may be in a way "adapted" to the Pf0-1 genome wide codon use and/or tRNA availability. This allows A506 *glyQ* to be translated faster and hence display higher competitive growth while in Pf0-1 background, as compared to the native combination. However, the reverse combination – Pf0-1 *glyQ* in A506 background – does not display a similar result. It is possible that the codon order in this case is not able to create (dis)advantages for translational output. This may be due to the fact that the codon changes effected with swap of Pf0-1 *glyQ* into A506 background do not alter either the mRNA structure and/or stability or the protein synthesis rate and/or efficiency or protein folding, or

some combination thereof. For the other remaining mutants, competitive fitness was observed to be similar to the cognate wild type strains, hinting that the positions where synonymous changes have been made, are possibly not crucial to either the structure or stability of mRNA or translational process.

In order to test out the effect of the synonymous mutations on mRNA, *glyQ* mRNA transcripts were quantified using quantitative PCR (see section 4.3.4). This would shed light on the changes at the mRNA level that might ensue as a result of synonymous mutations. It turns out that the mRNA levels of the Pf0-1 *glyQ* mutant in SBW25 background and that of Pf0-1 *glyQ* mutant in A506 background was lower than that of the corresponding wild type strains – SBW25 and A506, respectively. For the Pf0-1 background strains, however, the A506 *glyQ* mutant in Pf0-1 background has significantly higher levels of *glyQ* mRNA. This result echoes the competitive fitness results and is indicative that the higher relative mRNA expression might in fact be translated to higher protein levels. This would mean that there might be more number of mature glycine tRNAs in A506 *glyQ* mutant in Pf0-1 background. This in turn would imply that it higher glycine tRNA provide translational advantage that gets reflected in competitive growth. It is therefore evident that although all the synonymous mutations brought about by allele swap do not translate into significant differences in growth rate or competitive fitness, there definitely is some non-neutral effect of synonymous changes – altered mRNA levels of *glyQ*. There could be some potential explanations for this observation – firstly, the elevated *glyQ* mRNA levels in the SBW52 and A506 background mutants might be indicative that either the mRNA structure is very stable or that it is translated very slowly in these mutants. Secondly, the lower levels of *glyQ* mRNA in the Pf0-1 background might imply that in this mutant the mRNA is translated faster and hence the amount of mRNA is reduced or that the combination of synonymous codons in this mRNA transcript makes it prone to degradation.

In addition to characterizing growth patterns, competitive fitness and mRNA amounts in the mutants as well as the wild types, the *glyQ* mutants were also tested for their ability to form mats at the air liquid interface.

6.2.3 Mat formation ability of the strains of *P. fluorescens*

Mat formation at the air liquid interface has been extensively studied in the *P. fluorescens* SBW25 experimental system (Rainey & Travisano, 1998; Spiers *et al*., 2002; Spiers *et al*.,

2003; Goymer *et al.*, 2006; Bantinaki *et al*., 2007; McDonald *et al.*, 2009; Ferguson *et al*., 2013; Ardre *et al.*, 2019). Since altering the codon usage pattern of *glyQ* may affect the expression of glycine rich genes due to potential changes in the cellular glycine tRNA pool, it was hypothesised that the mat forming ability of *P. fluorescens* strains under static conditions might also be affected. Glycine is one most used amino acids and hence changes in gycine tRNA pool might alter activity of many genes, including those that are required for mat formation. Moreover, mat formation pathways have been extensively studied and some level of understanding of molecular routes is available for SBW25 but lacking for the other strains – A506, Pf0-1 and PICF7. It would therefore be add more to our current understanding of mat formation under static conditions if the mat formation patterns in the other strains were explored as well.

To examine the mat formation ability of *P. fluorescens* strains, all *glyQ* mutants and the cognate wild type strains were incubated under static conditions, in microcosms. As has been reported previously for SBW25, these form mats after incubation of 72 hours. On plating culture from these microcosms on KB agar plates, smooth (SM) colonies, wrinkly spreader (WS) and fuzzy spreader (FS) are observed. The A506 background strains also form mats after 72 hours, though these looked different from the SBW25 mats. On KB agar plates, two kinds of colony morphologies are observed for A506 microcosm cultures – smooth colonies (SM) and web-like colonies (Web). Contrariwise, Pf0-1 does not produce any observable mats, although it still produces different colony morphologies – smooth colonies (SM), small disc shaped (SD) and large disc shaped colonies (LD). Since these observations have not reported previously, the mat formation ability of A506 and Pf0-1 was explored further. PICF7 was also examined for its mat forming abilities along with SBW25, A506 and Pf0-1.

Ten independent microcosms were set up for each of the four strains and inoculated microcosms were incubated under static conditions. SBW25 cultures form mats at the air liquid interface by the end of 72 hours and produce smooth colonies (SM), wrinkly spreaders (WS) and fuzzy spreaders (FS) on KB agar plates (see Figure 5.1). Previous studies have already demonstrated that the WS are responsible for SBW25 mats and hence WS from ten independent colonies were subjected to whole genome sequencing. Of the ten WS populations, four lines show mutations in *mwsR*, two lines show mutations in *wspF*, two lines show mutations in *awsR* and one each show mutations in *wspA* and *wspE*. One of the *mwsR* mutations (G3095T), one of the *wspF* mutations (Δ151-165, ΔCTGATGGACCTGATC) and the *wspE* mutation (G1912T) have been reported in (McDonald *et al*., 2009; Gallie *et al*., 2015; Gallie *et al*., 2019), respectively. This, together with the identification of a mutation in

wsp, *aws*, or *mwsR* in every isolate demonstrates a high degree of predictability in the evolutionary emergence of WS genotypes in *P. fluorescens* SBW25.

PICF7 cultures form mats at the end of 72 hours and the mats at the air-liquid interface of these cultures look similar to the SBW25 mats. Five colony morphotypes were observed on plating out the PICF7 cultures on KB agar plates – smooth colonies (SM), wrinkly spreaders (WS), fuzzy spreaders (FS), small disc shaped (SD) and large disc shaped (LD) (see Figure 5.4). Drawing on the parallels in mat formation between SBW25 and PICF7, the wrinkly spreader populations generated from ten independent lines of mat forming microcosms were used for whole genome sequencing. Of the ten lines, five show mutations in the *wspF* homologue, and one each show mutations in homologues of *wspA, awsX, awsO* and *mwsR*. In SBW25, WspF is the negative regulator of the Wsp pathway, which is the diguanylate cyclase responsible for production of c-di-GMP. C-di-GMP in turn leads to over production of acetylated cellulosic polymer and hence WS phenotype. Seeing that about 50% of the mat forming lines show mutations in the homologue of *wspF*, and based on similar mat formation patterns of SBW25 and PICF7, it was hypothesized that c-di-GMP is constitutively expressed in these PICF7 lines that leads to overproduction of some exopolymeric substance (EPS), thereby leading to mat formation.

In order to identify the EPS that is being overproduced in PICF7 the WS of PICF7 were subjected to transposon suppressor analysis. Smooth transposon mutants were then used for Sanger sequencing in order to determine the transposon insertion site. In five out of eight independent conjugation, genes of the *pel* operon were found to have transposon insertions leading to loss of WS phenotype. Pel (N-acetylglucosamine with β(1-4) linkage), a product of the *pel* operon has been reported to be responsible for biofilm formation in *P. aeruginosa* (Vasseur *et al*., 2005; Colvin *et al.*, 2012; Jennings *et al.*, 2015; Marmont *et al.*, 2017b). It has also been reported to form pellicles at the air liquid interface in *P. aeruginosa* cultures (Friedman & Kolter, 2004b; Colvin *et al*., 2012; Limoli *et al.*, 2015). Together, the whole genome sequencing and transposon suppressor analysis results indicate that in the absence of the *wss* operon, the *pel* operon can be triggered due to elevated levels of c-di-GMP, thereby leading to WS and hence mat formation in PICF7. Multiple downstream targets of c-di-GMP are available and subject to the presence of one or many such targets, different organisms may use different strategies for mat formation.

A506 mats were distinct from both SBW25 and PICF7 mats. Although the mats did form at the end of 72 hours, nevertheless the mats looked much different from that of SBW25 and PICF7. The colony morphotypes isolated from A506 mat forming culture are smooth colonies

(SM) and web-like colonies (Web) (see Figure 5.2). The Web colonies were used for whole genome sequencing, which revealed mutations in the *dppA2* and *dppA3* genes. However, upon Sanger sequencing of the *dppA2* and *dppA3* genes from the A506 web-like colonies, no mutations were found in the genes. It is therefore implied that mats in A506 were developing *via* an environmental signal and did not ensue through a genetic change. What might the structural basis for A506 mats be was still not clear form the above result. Mats in *Salmonella enterica* (Echeverz *et al*., 2017) are visually very similar to the A506 mats. These *S. enterica* mats are due to overexpression of PGA (product of the *pga* operon). Another study wherein rugose small colony variants (RSCV) were isolated from cystic fibrosis (CF) patients, show colony morphology not unlike the Web colonies of A506 (Irie *et al*., 2010). The RSCV were found to have elevated levels of Psl (product of *psl* operon). Therefore, both the *pga* and *psl* operons were deleted independently in A506 wild type strain. A double deletion mutant, wherein the *psl* operon was deleted in the *pga* deletion background was also generated. The double deletion mutant and the Δ*pga* mutant do not form clearly observable mats, while the Δ*psl* strain still forms mat at the end of 72 hours. All of the deletion mutants produce both SM and Web colonies, albeit the double deletion mutant and the Δ*pga* mutant produce the Web colonies in lower frequencies. Based on this observation, I conclude that the *pga* operon might contribute to some extent to mat formation in A506.

What is not evident from the above is whether c-di-GMP has a role in mat production and Web colony formation in A506. To test this, a *wspF* loss-of-function mutation was constructed (as a route to generating high c-di-GMP levels). The 15 bp deletion that has been observed in this study and others, was engineered into the *wspF* homologue of A506. As has been mentioned before, an active *wspF* in SBW25 inhibits c-di-GMP production. Therefore inactivation of the gene leads to constitutive synthesis of c-di-GMP, keeping EPS production switched on. The *wspF* loss-of-function mutant forms mat by 24 hours, unlike the wild type A506, which forms mats at the end of 72 hours. The Web colonies of wild type A506 are semi heritable – these switch between the Web and smooth state, while those of *wspF* loss-of-function mutant was heritable. The mat formation is therefore constitutively switched and probably the mat is through the overproduction of PGA but substantial evidence to this direction is still lacking. Based on the observations made, the current hypothesis for mat formation in A506 is as follows – c-di-GMP is probably being overproduced in A506 under static conditions, as a consequence of which PGA expression is triggered and elevated levels of PGA lead to mats.

The lack of observable mats in Pf0-1 is unexpected. The strain has some of the EPS producing genes and is also known to form biofilm through the activity of the *lap* operon

(Monds *et al*., 2007; Newell *et al.*, 2009; Newell *et al.*, 2011b; Boyd *et al*., 2012). These studies have demonstrated the ability of Pf0-1 to form biofilms but on solid surfaces. One possible explanation could be that the *lap* operon can form biofilms exclusively via adhesion to solid surfaces and hence in the absence of an appropriate attachment area, Pf0-1 cannot commence mat formation. Activity of *lap* operon is mediated by inorganic phosphate (P_i) sufficient conditions; it is likely that in KB, the medium that has been used for setting up *P. fluorescens* cultures for mat formation, P_i concentrations are insufficient and hence mat formation cannot be triggered via the *lap* operon. One might argue that either the *pga* or *alg* operons could have produced the respective EPSs – PGA and alginate – and developed mats in Pf0-1; these are among the EPS manufacturing operons that are present in Pf0-1. But currently, there is no clear indication of why these operons did not get activated. To induce mat formation in Pf0-1 at the air-liquid interface, one might try growing Pf0-1 under static conditions, but under phosphate sufficient conditions. Such an environment might activate the *lap* operon and generate mats in Pf0-1 microcosms.

6.3 Future direction

The work reported in this thesis has provided an insight into the functional implications of synonymous mutations, and demonstrated that closely related organisms when subjected to identical selective forces, can utilize different strategies to arrive at similar phenotypic solutions. The latter specifically is demonstrative of the plasticity of biological systems – organisms that may have varying genotypic networks but converge at comparable phenotypic outputs. Nevertheless, further research into both the above aspects would provide a deeper understanding.

The effects of synonymous mutations that have been engineered in this study could be further investigated by examining protein levels. Quantities of mRNA are affected by swapping *glyQ* alleles in no-native backgrounds. However, if the differences observed in mRNA quantities are due to altered translation rates or due to altered mRNA secondary structure is not clear. Measurement of GlyQ levels might provide more information – if the mRNA levels are low, but protein levels are the same in both wild type and mutant, the protein translation is probably faster in the mutant. However, if the protein levels are proportionate to the mRNA levels, then that is indicative of the effect being on the mRNA secondary structure. As mRNA quantities have affected by the synonymous mutations, measurement of global protein expression could suggest if function of the gene has indeed

been affected. The gene *glyQ* codes for glycine tRNA synthetase, an enzyme responsible for generating mature glycine tRNAs. In the absence of an optimally functional *glyQ*, the pool of glycine tRNAs would be affected that might potentially lead to changes in expression of genes that are enriched in glycine. This could be achieved by RNA sequencing.

A deeper understanding of the functional implications of synonymous mutation in *glyQ*, could be provided by determining the number of mature glycine tRNAs in the cellular tRNA pool by sequencing the cellular pool of mature tRNAs. The absolute number of mature glycine tRNAs in the mutants with respect to the corresponding wild type strains would demonstrate the functional consequence of synonymous mutations.

With regards to the mat formation phenotype, current findings indicate that the mat formation pathway in PICF7 is evidently through the activity of increased amounts of Pel. What might be intriguing is to determine which EPS producing genes would be induced to produce mats, in the absence of *pel* (deletion of *pel* genes). For A506, the mat formation strategy is not obvious yet. There seems a possibility that PGA overproduction is leading to mat and Web colony formation and perhaps through the c-di-GMP mediated route. However convincing evidence is still lacking. A number of approaches may be undertaken to determine what the A506 mats are composed of. If the polymer is PGA, it may be difficult to identify this by dye-binding methods as bacterial cell wall, like PGA is composed of *N*-acetylglucosamine units. Other methods that could be employed are immunofluorescence technique; in other words, antibody targeting PGA that could be visualized with the aid of fluorescence microscopy as has been done in previous studies (Itoh *et al*., 2008; Roux *et al*., 2015). The other approach would be to use mass spectrometry, as has been done in previous work by (Itoh *et al.*, 2005; Fazekas, Kandra, & Gyémánt, 2012). Pf0-1 did not form any observable mats in this study, but one could attempt to induce mat formation by growing Pf0-1 under static conditions under P_i rich conditions. This approach might generate mats in Pf0-1, conceivable through the *lap* pathway.

6.4 Final comment

This work demonstrates a range of weak to strong fitness effects of introducing synonymous mutations. While the strong deleterious effects are implied from the inability to successfully isolate the mutants from culture, the weaker effects are clearly detected through the fitness assays and relative *glyQ* mRNA expression levels. Significantly altered

competitive fitness and mRNA levels were observed for some of the mutants, when compared with cognate wild type strains. It is evident that certain gene and environment interactions make the effects of synonymous mutations more apparent, while in other environments the effects may not be perceivable. Synonymous mutations might have non-neutral effects but the strength of those effects may differ depending on the environmental context. This work is the first to report on the mat formation phenotype of PICF7 and A506. The outcome, that individual strategies to mat formation may range from using different polymers, such as acetylated cellulosic polymer in SBW25 to Pel in PICF7 to using completely different schemes, such as from being a heritable phenotype, such as in both SBW25 and PICF7 to being semi-heritable in A506, is telling of the exceptional capacity of plasticity demonstrated by genomes. The final and quite perplexing option is to not form any mats at all, such as in Pf0-1, in spite of the presence of some EPS producing genes. Altogether, this work provides insight into how closely related genomes may respond differently to identical environmental and evolutionary stimuli.

Reference List

Agashe, D., Martinez-Gomez, N. C., Drummond, D. A., & Marx, C. J. (2013). Good codons, bad transcript: large reductions in gene expression and fitness arising from synonymous mutations in a key enzyme. *Molecular Biology and Evolution*, *30*(3), 549–560. http://doi.org/10.1093/molbev/mss273

Agashe, D., Sane, M., Phalnikar, K., Diwan, G. D., Habibullah, A., Martinez-Gomez, N. C., et al. (2016). Large-Effect Beneficial Synonymous Mutations Mediate Rapid and Parallel Adaptation in a Bacterium. *Molecular Biology and Evolution*, *33*(6), 1542–1553. http://doi.org/10.1093/molbev/msw035

Ardre, M., Dufour, D., Rainey, P. B., & O'Toole, G. (2019). Causes and Biophysical Consequences of Cellulose Production by *Pseudomonas fluorescens* SBW25 at the Air-Liquid Interface. *Journal of Bacteriology*, *201*(18), e00110–19. http://doi.org/10.1128/JB.00110-19

Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., et al. (2006). Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Molecular Systems Biology*, *2*(1), 2006.0008. http://doi.org/10.1038/msb4100050

Bailey, S. F., Hinz, A., & Kassen, R. (2014). Adaptive synonymous mutations in an experimentally evolved *Pseudomonas fluorescens* population. *Nature Communications*, *5*(1), 1–7. http://doi.org/10.1038/ncomms5076

Bantinaki, E., Kassen, R., Knight, C. G., Robinson, Z., Spiers, A. J., & Rainey, P. B. (2007). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. III. Mutational origins of wrinkly spreader diversity. *Genetics*, *176*(1), 441–453. http://doi.org/10.1534/genetics.106.069906

Bazzini, A. A., Del Viso, F., Moreno-Mateos, M. A., Johnstone, T. G., Vejnar, C. E., Qin, Y., et al. (2016). Codon identity regulates mRNA stability and translation efficiency during the maternal-to-zygotic transition. *The EMBO Journal*, *35*(19), 2087–2103. http://doi.org/10.15252/embj.201694699

Berg, O. (1997). Codon bias in *Escherichia coli*: the influence of codon context on mutation and selection. *Nucleic Acids Research*, *25*(7), 1397–1404. http://doi.org/10.1093/nar/25.7.1397

Bertani, G. (1951). Studies on lysogenesis. I. The mode of phage liberation by lysogenic *Escherichia coli*. *Journal of Bacteriology*, *62*(3), 293–300.

Blanchet, S., Cornu, D., Hatin, I., Grosjean, H., Bertin, P., & Namy, O. (2018). Deciphering the reading of the genetic code by near-cognate tRNA. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(12), 3018–3023. http://doi.org/10.1073/pnas.1715578115

Boël, G., Letso, R., Neely, H., Price, W. N., Wong, K.-H., Su, M., et al. (2016). Codon influence on protein expression in E. coli correlates with mRNA levels. *Nature*, *529*(7586), 358–363. http://doi.org/10.1038/nature16509

Boyd, A., & Chakrabarty, A. M. (1995). *Pseudomonas aeruginosa* biofilms: role of the alginate exopolysaccharide. *Journal of Industrial Microbiology*, *15*(3), 162–168. http://doi.org/10.1007/BF01569821

Boyd, C. D., Chatterjee, D., Sondermann, H., & O'Toole, G. A. (2012). LapG, required for modulating biofilm formation by *Pseudomonas fluorescens* Pf0-1, is a calcium-dependent protease. *Journal of Bacteriology*, *194*(16), 4406–4414. http://doi.org/10.1128/JB.00642-12

Brandis, G., & Hughes, D. (2016). The Selective Advantage of Synonymous Codon Usage Bias in Salmonella. *PLoS Genetics*, *12*(3), e1005926–16. http://doi.org/10.1371/journal.pgen.1005926

Buhr, F., Jha, S., Thommen, M., Mittelstaet, J., Kutz, F., Schwalbe, H., et al. (2016). Synonymous Codons Direct Cotranslational Folding toward Different Protein

Conformations. *Molecular Cell*, *61*(3), 341–351. http://doi.org/10.1016/j.molcel.2016.01.008
Bulmer, M. (1991). The selection-mutation-drift theory of synonymous codon usage. *Genetics*, *129*(3), 897–907.
Burmann, B. M., Schweimer, K., Luo, X., Wahl, M. C., Stitt, B. L., Gottesman, M. E., & Rösch, P. (2010). A NusE:NusG Complex Links Transcription and Translation. *Science*, *328*(5977), 501–504. http://doi.org/10.1126/science.1184953
Camiolo, S., Farina, L., & Porceddu, A. (2012). The Relation of Codon Bias to Tissue-Specific Gene Expression in *Arabidopsis thaliana*. *Genetics*, *192*(2), 641–649. http://doi.org/10.1534/genetics.112.143677
Camiolo, S., Melito, S., & Porceddu, A. (2015). New insights into the interplay between codon bias determinants in plants. *DNA Research : an International Journal for Rapid Publication of Reports on Genes and Genomes*, *22*(6), 461–470. http://doi.org/10.1093/dnares/dsv027
Cannarozzi, G., Schraudolph, N. N., Faty, M., Rohr, von, P., Friberg, M. T., Roth, A. C., et al. (2010). A Role for Codon Order in Translation Dynamics. *Cell*, *141*(2), 355–367. http://doi.org/10.1016/j.cell.2010.02.036
Carbone, A., Zinovyev, A., & Kepes, F. (2003). Codon adaptation index as a measure of dominating codon bias. *Bioinformatics*, *19*(16), 2005–2015. http://doi.org/10.1093/bioinformatics/btg272
Cárcamo-Oyarce, G., Lumjiaktase, P., Kümmerli, R., & Eberl, L. (2015). Quorum sensing triggers the stochastic escape of individual cells from *Pseudomonas putida* biofilms. *Nature Communications*, *6*(1), 1–9. http://doi.org/10.1038/ncomms6945
Chamary, J. V., Parmley, J. L., & Hurst, L. D. (2006). Hearing silence: non-neutral evolution at synonymous sites in mammals. *Nature Reviews Genetics*, *7*(2), 98–108. http://doi.org/10.1038/nrg1770
Chamary, J.-V., & Hurst, L. D. (2005). Biased codon usage near intron-exon junctions: selection on splicing enhancers, splice-site recognition or something else? *Trends in Genetics*, *21*(5), 256–259. http://doi.org/10.1016/j.tig.2005.03.001
Chan, C., Paul, R., Samoray, D., Amiot, N. C., Giese, B., Jenal, U., & Schirmer, T. (2004). Structural basis of activity and allosteric control of diguanylate cyclase. *Proceedings of the National Academy of Sciences*, *101*(49), 17084–17089. http://doi.org/10.1073/pnas.0406134101
Chang, W.-S., van de Mortel, M., Nielsen, L., de Guzman, G. N., Li, X., & Halverson, L. J. (2007). Alginate Production by *Pseudomonas putida* creates a Hydrated Microenvironment and Contributes to Biofilm Architecture and Stress Tolerance under Water-Limiting Conditions. *Journal of Bacteriology*, *189*(22), 8290–8299. http://doi.org/10.1128/JB.00727-07
Chantawannakul, P., & Cutler, R. W. (2008). Convergent host-parasite codon usage between honeybee and bee associated viral genomes. *Journal of Invertebrate Pathology*, *98*(2), 206–210. http://doi.org/10.1016/j.jip.2008.02.016
Chithambaram, S., Prabhakaran, R., & Xia, X. (2014a). Differential codon adaptation between dsDNA and ssDNA phages in *Escherichia coli*. *Molecular Biology and Evolution*, *31*(6), 1606–1617. http://doi.org/10.1093/molbev/msu087
Chithambaram, S., Prabhakaran, R., & Xia, X. (2014b). The Effect of Mutation and Selection on Codon Adaptation in *Escherichia coli* Bacteriophage. *Genetics*, *197*(1), 301–315. http://doi.org/10.1534/genetics.114.162842
Chu, D., Kazana, E., Bellanger, N., Singh, T., Tuite, M. F., & Haar, von der, T. (2014). Translation elongation can control translation initiation on eukaryotic mRNAs. *The EMBO Journal*, *33*(1), 21–34. http://doi.org/10.1002/embj.201385651
Colvin, K. M., Irie, Y., Tart, C. S., Urbano, R., Whitney, J. C., Ryder, C., et al. (2012). The Pel and Psl polysaccharides provide *Pseudomonas aeruginosa* structural redundancy within the biofilm matrix. *Environmental Microbiology*, *14*(8), 1913–1928. http://doi.org/10.1111/j.1462-2920.2011.02657.x
Compeau, G., Al-Achi, B. J., Platsouka, E., & Levy, S. B. (1988). Survival of rifampin-

resistant mutants of *Pseudomonas fluorescens* and *Pseudomonas putida* in soil systems. *Applied and Environmental Microbiology*, *54*(10), 2432–2438.
Cortazzo, P., Cerveñansky, C., Marín, M., Reiss, C., Ehrlich, R., & Deana, A. (2002). Silent mutations affect in vivo protein folding in *Escherichia coli*. *Biochemical and Biophysical Research Communications*, *293*(1), 537–541. http://doi.org/10.1016/S0006-291X(02)00226-7
Dae-Gon, H., & O'Toole, G. A. (2015). c-di-GMP and its effects on biofilm formation and dispersion: a *Pseudomonas aeruginosa* review. *Microbiology Spectrum*, *3*(2), MB–0003–2014. http://doi.org/10.1128/microbiolspec.MB-0003-2014
Davey, M. E., & O'Toole, G. A. (2000). Microbial Biofilms: from Ecology to Molecular Genetics. *Microbiology and Molecular Biology Reviews*, *64*(4), 847–867. http://doi.org/10.1128/MMBR.64.4.847-867.2000
de Lorenzo, V., Cases, I., Herrero, M., & Timmis, K. N. (1993). Early and late responses of TOL promoters to pathway inducers: identification of postexponential promoters in *Pseudomonas putida* with lacZ-tet bicistronic reporters. *Journal of Bacteriology*, *175*(21), 6902–6907. http://doi.org/10.1128/jb.175.21.6902-6907.1993
Deatherage, D. E., & Barrick, J. E. (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in Molecular Biology (Clifton, N.J.)*, *1151*(Chapter 12), 165–188. http://doi.org/10.1007/978-1-4939-0554-6_12
Dittmar, K. A., Goodenbour, J. M., & Pan, T. (2006). Tissue-specific differences in human transfer RNA expression. *PLoS Genetics*, *2*(12), e221. http://doi.org/10.1371/journal.pgen.0020221
Dittmar, K. A., Sørensen, M. A., Elf, J., Ehrenberg, M., & Pan, T. (2005). Selective charging of tRNA isoacceptors induced by amino-acid starvation. *EMBO Reports*, *6*(2), 151–157. http://doi.org/10.1038/sj.embor.7400341
Díaz-Salazar, C., Calero, P., Espinosa-Portero, R., Jiménez-Fernández, A., Wirebrand, L., Velasco-Domínguez, M. G., et al. (2017). The stringent response promotes biofilm dispersal in *Pseudomonas putida*. *Scientific Reports*, *7*(1), 1–13. http://doi.org/10.1038/s41598-017-18518-0
Duret, L. (2000). tRNA gene number and codon usage in the *C. elegans* genome are co-adapted for optimal translation of highly expressed genes. *Trends in Genetics : TIG*, *16*(7), 287–289. http://doi.org/10.1016/s0168-9525(00)02041-2
Echeverz, M., García, B., Sabalza, A., Valle, J., Gabaldón, T., Solano, C., & Lasa, I. (2017). Lack of the PGA exopolysaccharide in *Salmonella* as an adaptive trait for survival in the host. *PLoS Genetics*, *13*(5), e1006816. http://doi.org/10.1371/journal.pgen.1006816
Ertesvåg, H., Sletta, H., Senneset, M., Sun, Y.-Q., Klinkenberg, G., Konradsen, T. A., et al. (2017). Identification of genes affecting alginate biosynthesis in *Pseudomonas fluorescens* by screening a transposon insertion library. *BMC Genomics*, *18*(1), 227. http://doi.org/10.1186/s12864-016-3467-7
Farias, G. A., Olmedilla, A., & Gallegos, M. T. (2019). Visualization and characterization of *Pseudomonas syringae* pv. *tomato* DC3000 pellicles. *Microbial Biotechnology*, *12*(4), 688–702. http://doi.org/10.1111/1751-7915.13385
Fazekas, E., Kandra, L., & Gyémánt, G. (2012). Model for β-1,6-N-acetylglucosamine oligomer hydrolysis catalysed by DispersinB, a biofilm degrading enzyme. *Carbohydrate Research*, *363*, 7–13. http://doi.org/10.1016/j.carres.2012.09.016
Ferguson, G. C., Bertels, F., & Rainey, P. B. (2013). Adaptive divergence in experimental populations of Pseudomonas fluorescens. V. Insight into the niche specialist fuzzy spreader compels revision of the model *Pseudomonas* radiation. *Genetics*, *195*(4), 1319–1335. http://doi.org/10.1534/genetics.113.154948
Freist, W., Logan, D. T., & Gauss, D. H. (1996). Glycyl-tRNA Synthetase. *Biological Chemistry Hoppe-Seyler*, 343–356.
Friedman, L., & Kolter, R. (2004a). Genes involved in matrix formation in *Pseudomonas aeruginosa* PA14 biofilms. *Molecular Microbiology*, *51*(3), 675–690. http://doi.org/10.1046/j.1365-2958.2003.03877.x

Friedman, L., & Kolter, R. (2004b). Two Genetic Loci Produce Distinct Carbohydrate-Rich Structural Components of the *Pseudomonas aeruginosa* Biofilm Matrix. *Journal of Bacteriology*, *186*(14), 4457–4465. http://doi.org/10.1128/JB.186.14.4457-4465.2004

Frumkin, I., Lajoie, M. J., Gregg, C. J., Hornung, G., Church, G. M., & Pilpel, Y. (2018). Codon usage of highly expressed genes affects proteome-wide translation efficiency. *Proceedings of the National Academy of Sciences of the United States of America*. http://doi.org/10.1073/pnas.1719375115

Frumkin, I., Schirman, D., Rotman, A., Li, F., Zahavi, L., Mordret, E., et al. (2017). Gene Architectures that Minimize Cost of Gene Expression. *Molecular Cell*, *65*(1), 142–153. http://doi.org/10.1016/j.molcel.2016.11.007

Fu, J., Murphy, K. A., Zhou, M., Li, Y. H., Lam, V. H., Tabuloc, C. A., et al. (2016). Codon usage affects the structure and function of the *Drosophila* circadian clock protein PERIOD. *Genes & Development*, *30*(15), 1761–1775. http://doi.org/10.1101/gad.281030.116

Gallie, J., Bertels, F., Remigi, P., Ferguson, G. C., Nestmann, S., & Rainey, P. B. (2019). Repeated Phenotypic Evolution by Different Genetic Routes in Pseudomonas fluorescens SBW25. *Molecular Biology and Evolution*. http://doi.org/10.1093/molbev/msz040",

Gallie, J., Libby, E., Bertels, F., Remigi, P., Jendresen, C. B., Ferguson, G. C., et al. (2015). Bistability in a Metabolic Network Underpins the De Novo Evolution of Colony Switching in *Pseudomonas fluorescens*. *PLoS Biol*, *13*(3), e1002109. http://doi.org/10.1371/journal.pbio.1002109

Gerdes, S. Y., Scholle, M. D., Campbell, J. W., Balázsi, G., Ravasz, E., Daugherty, M. D., et al. (2003). Experimental determination and system level analysis of essential genes in *Escherichia coli* MG1655. *Journal of Bacteriology*, *185*(19), 5673–5684. http://doi.org/10.1128/jb.185.19.5673-5684.2003

Ghods, S., Sims, I. M., Moradali, M. F., Rehm, B. H. A., & Schottel, J. L. (2015). Bactericidal Compounds Controlling Growth of the Plant Pathogen *Pseudomonas syringae* pv. *actinidiae*, Which Forms Biofilms Composed of a Novel Exopolysaccharide. *Applied and Environmental Microbiology*, *81*(12), 4026–4036. http://doi.org/10.1128/AEM.00194-15

Giddens, S. R., Jackson, R. W., Moon, C. D., Jacobs, M. A., Zhang, X.-X., Gehrig, S. M., & Rainey, P. B. (2007). Mutational activation of niche-specific genes provides insight into regulatory networks and bacterial function in a complex environment. *Proceedings of the National Academy of Sciences of the United States of America*, *104*(46), 18247–18252. http://doi.org/10.1073/pnas.0706739104

Goodall, E. C. A., Robinson, A., Johnston, I. G., Jabbari, S., Turner, K. A., Cunningham, A. F., et al. (2018). The essential genome of *Escherichia coli* K-12. *mBio*, *9*(1), 385. http://doi.org/10.1128/mBio.02096-17

Goodman, D. B., Church, G. M., & Kosuri, S. (2013). Causes and effects of N-terminal codon bias in bacterial genes. *Science*, *342*(6157), 475–479. http://doi.org/10.1126/science.1241934

Goymer, P., Kahn, S. G., Malone, J. G., Gehrig, S. M., Spiers, A. J., & Rainey, P. B. (2006). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. II. Role of the GGDEF regulator WspR in evolution and development of the wrinkly spreader phenotype. *Genetics*, *173*(2), 515–526. http://doi.org/10.1534/genetics.106.055863

Gu, W., Zhou, T., & Wilke, C. O. (2010). A universal trend of reduced mRNA stability near the translation-initiation site in prokaryotes and eukaryotes. *PLoS Computational Biology*, *6*(2), e1000664. http://doi.org/10.1371/journal.pcbi.1000664

Hanson, G., & Coller, J. (2017). Codon optimality, bias and usage in translation and mRNA decay. *Nature Reviews. Molecular Cell Biology*, *151*(1), 389–30. http://doi.org/10.1038/nrm.2017.91

Harigaya, Y., & Parker, R. (2016a). Analysis of the association between codon optimality and mRNA stability in *Schizosaccharomyces pombe*. *BMC Genomics*, *17*(1), 895–16. http://doi.org/10.1186/s12864-016-3237-6

Harigaya, Y., & Parker, R. (2016b). Codon optimality and mRNA decay. *Cell Research*,

26(12), 1269–1270. http://doi.org/10.1038/cr.2016.127
Hengge, R. (2009). Principles of c-di-GMP signalling in bacteria. *Nature Reviews Microbiology*, *7*(4), 263–273. http://doi.org/10.1038/nrmicro2109
Hia, F., Yang, S. F., Shichino, Y., Yoshinaga, M., Murakawa, Y., Vandenbon, A., et al. (2019). Codon bias confers stability to human mRNAs. *EMBO Reports*, *33*, e48220. http://doi.org/10.15252/embr.201948220
Higgs, P. G., & Ran, W. (2008). Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Molecular Biology and Evolution*, *25*(11), 2279–2291. http://doi.org/10.1093/molbev/msn173
Ho, S. N., Hunt, H. D., Horton, R. M., Pullen, J. K., & Pease, L. R. (1989). Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene*, *77*(1), 51–59. http://doi.org/10.1016/0378-1119(89)90358-2
Ikemura, T. (1981). Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes. *Journal of Molecular Biology*, *146*(1), 1–21.
Ikemura, T. (1982). Correlation between the abundance of yeast transfer RNAs and the occurrence of the respective codons in protein genes. Differences in synonymous codon choice patterns of yeast and *Escherichia coli* with reference to the abundance of isoaccepting transfer RNAs. *Journal of Molecular Biology*, *158*(4), 573–597. http://doi.org/10.1016/0022-2836(82)90250-9
Ikemura, T. (1985). Codon usage and tRNA content in unicellular and multicellular organisms. *Molecular Biology and Evolution*, *2*(1), 13–34. http://doi.org/10.1093/oxfordjournals.molbev.a040335
Ikemura, T. (1999). Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. *Gene*, *238*(1), 143–155. http://doi.org/10.1016/s0378-1119(99)00225-5
Irie, Y., Starkey, M., Edwards, A. N., Wozniak, D. J., Romeo, T., & Parsek, M. R. (2010). *Pseudomonas aeruginosa* biofilm matrix polysaccharide Psl is regulated transcriptionally by RpoS and post-transcriptionally by RsmA. *Molecular Microbiology*, *78*(1), 158–172. http://doi.org/10.1111/j.1365-2958.2010.07320.x
Itoh, Y., Rice, J. D., Goller, C., Pannuri, A., Taylor, J., Meisner, J., et al. (2008). Roles of pgaABCD genes in synthesis, modification, and export of the *Escherichia coli* biofilm adhesin poly-beta-1,6-N-acetyl-D-glucosamine. *Journal of Bacteriology*, *190*(10), 3670–3680. http://doi.org/10.1128/JB.01920-07
Itoh, Y., Wang, X., Hinnebusch, B. J., Preston, J. F., & Romeo, T. (2005). Depolymerization of beta-1,6-N-acetyl-D-glucosamine disrupts the integrity of diverse bacterial biofilms. *Journal of Bacteriology*, *187*(1), 382–387. http://doi.org/10.1128/JB.187.1.382-387.2005
Jacobs, M. A., Alwood, A., Thaipisuttikul, I., Spencer, D., Haugen, E., Ernst, S., et al. (2003). Comprehensive transposon mutant library of *Pseudomonas aeruginosa*. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(SUPPL. 2), 14339–14344. http://doi.org/10.1073/pnas.2036282100
Jennings, L. K., Storek, K. M., Ledvina, H. E., Coulon, C., Marmont, L. S., Sadovskaya, I., et al. (2015). Pel is a cationic exopolysaccharide that cross-links extracellular DNA in the *Pseudomonas aeruginosa* biofilm matrix. *Proceedings of the National Academy of Sciences*, *112*(36), 11353–11358. http://doi.org/10.1073/pnas.1503058112
Jones, C. J., & Wozniak, D. J. (2017). Psl Produced by Mucoid *Pseudomonas* aeruginosaContributes to the Establishment of Biofilms and Immune Evasion. *mBio*, *8*(3), e00864–17. http://doi.org/10.1128/mBio.00864-17
Kelsic, E. D., Chung, H., Cohen, N., Park, J., Wang, H. H., & Kishony, R. (2016). RNA Structural Determinants of Optimal Codons Revealed by MAGE-Seq. *Cell Systems*, *3*(6), 563–571.e6. http://doi.org/10.1016/j.cels.2016.11.004
Keng, T., Webster, T. A., Sauer, R. T., & Schimmel, P. (1982). Gene for Escherichia coli glycyl-tRNA synthetase has tandem subunit coding regions in the same reading frame. *Journal of Biological Chemistry*, *257*(21), 12503–12508.

Khandia, R., Singhal, S., Kumar, U., Ansari, A., Tiwari, R., Dhama, K., et al. (2019). Analysis of nipah virus codon usage and adaptation to hosts. *Frontiers in Microbiology*, *10*(MAY), 92. http://doi.org/10.3389/fmicb.2019.00886

Kimchi-Sarfaty, C., Oh, J. M., Kim, I.-W., Sauna, Z. E., Calcagno, A. M., Ambudkar, S. V., & Gottesman, M. M. (2007). A "silent" polymorphism in the MDR1 gene changes substrate specificity. *Science*, *315*(5811), 525–528. http://doi.org/10.1126/science.1135308

King, E. O., Ward, M. K., & Raney, D. E. (1954). Two simple media for the demonstration of pyocyanin and fluorescin. *The Journal of Laboratory and Clinical Medicine*, *44*(2), 301–307.

Kitten, T., Kinscherf, T. G., McEvoy, J. L., & Willis, D. K. (1998). A newly identified regulator is required for virulence and toxin production in Pseudomonas syringae. *Molecular Microbiology*, *28*(5), 917–929.

Kristofich, J. C., Morgenthaler, A. B., Kinney, W. R., Ebmeier, C. C., Snyder, D. J., Old, W. M., et al. (2018). Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLoS Genetics*, *14*(8), e1007615. http://doi.org/10.1371/journal.pgen.1007615

LaBella, A. L., Opulente, D. A., Steenwyk, J. L., Hittinger, C. T., & Rokas, A. (2019). Variation and selection on codon usage bias across an entire subphylum. *PLoS Genetics*, *15*(7), e1008304–25. http://doi.org/10.1371/journal.pgen.1008304

Lebeuf-Taylor, E., McCloskey, N., Bailey, S. F., Hinz, A., & Kassen, R. (2019). The distribution of fitness effects among synonymous mutations in a gene under directional selection. *Elife*, *8*, 1542. http://doi.org/10.7554/eLife.45952

Lenski, R. E., Rose, M. R., Simpson, S. C., & Tadler, S. C. (1991). Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist*, *138*(6), 1315–1341. http://doi.org/10.1086/285289

Limoli, D. H., Jones, C. J., & Wozniak, D. J. (2015). Bacterial Extracellular Polysaccharides in Biofilm Formation and Function. *Microbiology Spectrum*, *3*(3). http://doi.org/10.1128/microbiolspec.MB-0011-2014

Lind, P. A., Berg, O. G., & Andersson, D. I. (2010). Mutational Robustness of Ribosomal Protein Genes. *Science*, *330*(6005), 825–827. http://doi.org/10.1126/science.1194617

Lind, P. A., Farr, A. D., & Rainey, P. B. (2017). Evolutionary convergence in experimental *Pseudomonas* populations. *The ISME Journal*, *11*(3), 589–600. http://doi.org/10.1038/ismej.2016.157

Lind, P. A., Libby, E., Herzog, J., & Rainey, P. B. (2018). Disentangling the effects of genetic architecture, mutational bias and selection on evolutionary forecasting. *bioRxiv*, 335711. http://doi.org/10.1101/335711

Loper, J. E., Hassan, K. A., Mavrodi, D. V., Davis, E. W., II, Lim, C. K., Shaffer, B. T., et al. (2012). Comparative Genomics of Plant-Associated *Pseudomonas* spp.: Insights into Diversity and Inheritance of Traits Involved in Multitrophic Interactions. *PLoS Genetics*, *8*(7), e1002784. http://doi.org/10.1371/journal.pgen.1002784

Luo, X., Hsiao, H.-H., Bubunenko, M., Weber, G., Court, D. L., Gottesman, M. E., et al. (2008). Structural and Functional Analysis of the E. coli NusB-S10 Transcription Antitermination Complex. *Molecular Cell*, *32*(6), 791–802. http://doi.org/10.1016/j.molcel.2008.10.028

Malone, J. G., Williams, R., Christen, M., Jenal, U., Spiers, A. J., & Rainey, P. B. (2007). The structure-function relationship of WspR, a *Pseudomonas fluorescens* response regulator with a GGDEF output domain. *Microbiology (Reading, England)*, *153*(Pt 4), 980–994. http://doi.org/10.1099/mic.0.2006/002824-0

Manoil, C. (2000). **Tagging exported proteins using *Escherichia coli* alkaline phosphatase gene fusions.** (pp. 1–13). Elsevier Inc.

Marmont, L. S., Rich, J. D., Whitney, J. C., Whitfield, G. B., Almblad, H., Robinson, H., et al. (2017a). Oligomeric lipoprotein PelC guides Pel polysaccharide export across the outer membrane of Pseudomonas aeruginosa. *Proceedings of the National Academy of Sciences*, *114*(11), 2892–2897. http://doi.org/10.1073/pnas.1613606114

Marmont, L. S., Whitfield, G. B., Rich, J. D., Yip, P., Giesbrecht, L. B., Stremick, C. A., et al. (2017b). PelA and PelB proteins form a modification and secretion complex essential for Pel polysaccharide-dependent biofilm formation in Pseudomonas aeruginosa. *The Journal of Biological Chemistry*, *292*(47), 19411–19422. http://doi.org/10.1074/jbc.M117.812842

Martínez-García, P., Ruano-Rosa, D., Schilirò, E., Prieto, P., Ramos, C., Rodríguez-Palenzuela, P., & Mercado-Blanco, J. (2015). Complete genome sequence of *Pseudomonas fluorescens* strain PICF7, an indigenous root endophyte from olive (Olea europaea L.) and effective biocontrol agent against Verticillium dahliae. *Standards in Genomic Sciences*, *10*(1), 10. http://doi.org/10.1186/1944-3277-10-10

Mascher, T., Helmann, J. D., & Unden, G. (2006). Stimulus perception in bacterial signal-transducing histidine kinases. *Microbiology and Molecular Biology Reviews*, *70*(4), 910–938. http://doi.org/10.1128/MMBR.00020-06

McDonald, M. J., Chou, C.-H., Swamy, K. B. S., Huang, H.-D., & Leu, J.-Y. (2015). The evolutionary dynamics of tRNA-gene copy number and codon-use in E. coli. *BMC Evolutionary Biology*, *15*(1), 163. http://doi.org/10.1186/s12862-015-0441-y

McDonald, M. J., Gehrig, S. M., Meintjes, P. L., Zhang, X.-X., & Rainey, P. B. (2009). Adaptive divergence in experimental populations of *Pseudomonas fluorescens*. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation. *Genetics*, *183*(3), 1041–1053. http://doi.org/10.1534/genetics.109.107110

Mercado-Blanco, J., Rodríguez-Jurado, D., Hervás, A., & Jiménez-Díaz, R. M. (2004). Suppression of Verticillium wilt in olive planting stocks by root-associated fluorescent *Pseudomonas* spp. *Biological Control*, *30*(2), 474–486. http://doi.org/10.1016/j.biocontrol.2004.02.002

Mitarai, N., Sneppen, K., & Pedersen, S. (2008). Ribosome Collisions and Translation Efficiency: Optimization by Codon Usage and mRNA Destabilization. *Journal of Molecular Biology*, *382*(1), 236–245. http://doi.org/10.1016/j.jmb.2008.06.068

Mittal, P., Brindle, J., Stephen, J., Plotkin, J. B., & Kudla, G. (2018). Codon usage influences fitness through RNA toxicity. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(34), 8639–8644. http://doi.org/10.1073/pnas.1810022115

Monds, R. D., Newell, P. D., Gross, R. H., & O'Toole, G. A. (2007). Phosphate-dependent modulation of c-di-GMP levels regulates *Pseudomonas fluorescens* Pf0-1 biofilm formation by controlling secretion of the adhesin LapA. *Molecular Microbiology*, *63*(3), 656–679. http://doi.org/10.1111/j.1365-2958.2006.05539.x

Morgan, J. L. W., McNamara, J. T., & Zimmer, J. (2014). Mechanism of activation of bacterial cellulose synthase by cyclic di-GMP. *Nature Structural & Molecular Biology*, *21*(5), 489–496. http://doi.org/10.1038/nsmb.2803

Morisaki, T., Lyon, K., DeLuca, K. F., DeLuca, J. G., English, B. P., Zhang, Z., et al. (2016). Real-time quantification of single RNA translation dynamics in living cells. *Science*, *352*(6292), 1425–1429. http://doi.org/10.1126/science.aaf0899

Nagel, G. M., Cumberledge, S., Johnson, M. S., Petrella, E., & Weber, B. H. (1984). The beta subunit of *E. coli* glycyl-tRNA synthetase plays a major role in tRNA recognition. *Nucleic Acids Research*, *12*(10), 4377–4384. http://doi.org/10.1093/nar/12.10.4377

Narula, A., Ellis, J., Taliaferro, J. M., & Rissland, O. S. (2019). Coding regions affect mRNA stability in human cells. *Rna*, rna.073239.119. http://doi.org/10.1261/rna.073239.119

Newell, P. D., Boyd, C. D., Sondermann, H., & O'Toole, G. A. (2011a). A c-di-GMP Effector System Controls Cell Adhesion by Inside-Out Signaling and Surface Protein Cleavage. *PLoS Biol*, *9*(2), e1000587. http://doi.org/10.1371/journal.pbio.1000587

Newell, P. D., Monds, R. D., & O'Toole, G. A. (2009). LapD is a bis-(3′,5′)-cyclic dimeric GMP-binding protein that regulates surface attachment by *Pseudomonas fluorescens* Pf0-1. *Proceedings of the National Academy of Sciences of the United States of America*, *106*(9), 3461–3466. http://doi.org/10.1073/pnas.0808933106

Newell, P. D., Yoshioka, S., Hvorecny, K. L., Monds, R. D., & O'Toole, G. A. (2011b). Systematic analysis of diguanylate cyclases that promote biofilm formation by *Pseudomonas fluorescens* Pf0-1. *Journal of Bacteriology*, *193*(18), 4685–4698.

http://doi.org/10.1128/JB.05483-11
Nivens, D. E., Ohman, D. E., Williams, J., & Franklin, M. J. (2001). Role of alginate and its O acetylation in formation of *Pseudomonas aeruginosa* microcolonies and biofilms. *Journal of Bacteriology*, *183*(3), 1047–1057. http://doi.org/10.1128/JB.183.3.1047-1057.2001
Nixon, B. T., Ronson, C. W., & Ausubel, F. M. (1986). Two-component regulatory systems responsive to environmental stimuli share strongly conserved domains with the nitrogen assimilation regulatory genes *ntrB* and *ntrC*. *Proceedings of the National Academy of Sciences*, *83*(20), 7850–7854. http://doi.org/10.1073/pnas.83.20.7850
Novoa, E. M., Pavon-Eternod, M., Pan, T., & Ribas de Pouplana, L. (2012). A Role for tRNA Modifications in Genome Structure and Codon Usage. *Cell*, *149*(1), 202–213. http://doi.org/10.1016/j.cell.2012.01.050
Ostrem, D. L., Biochemistry, P. B., 1974. (n.d.). Glycyl transfer ribonucleic acid synthetase from Escherichia coli. Purification, properties, and substrate binding. *ACS Publications*.
Parkinson, J. S. (2003). Bacterial chemotaxis: a new player in response regulator dephosphorylation. *Journal of Bacteriology*, *185*(5), 1492–1494. http://doi.org/10.1128/jb.185.5.1492-1494.2003
Paul, R., Weiser, S., Amiot, N. C., Chan, C., Schirmer, T., Giese, B., & Jenal, U. (2004). Cell cycle-dependent dynamic localization of a bacterial response regulator with a novel di-guanylate cyclase output domain. *Genes & Development*, *18*(6), 715–727. http://doi.org/10.1101/gad.289504
Peris, J. B., Davis, P., Cuevas, J. M., Nebot, M. R., & Sanjuan, R. (2010). Distribution of Fitness Effects Caused by Single-Nucleotide Substitutions in Bacteriophage f1. *Genetics*, *185*(2), 603–609. http://doi.org/10.1534/genetics.110.115162
Petersen, E., Mills, E., & Miller, S. I. (2019). Cyclic-di-GMP regulation promotes survival of a slow-replicating subpopulation of intracellular *Salmonella typhimurium*. *Proceedings of the National Academy of Sciences of the United States of America*, *116*(13), 6335–6340. http://doi.org/10.1073/pnas.1901051116
Pérez-Mendoza, D., & Sanjuán, J. (2016). Exploiting the commons: cyclic diguanylate regulation of bacterial exopolysaccharide production. *Current Opinion in Microbiology*, *30*, 36–43. http://doi.org/10.1016/j.mib.2015.12.004
Plotkin, J. B., & Kudla, G. (2011). Synonymous but not the same: the causes and consequences of codon bias. *Nature Reviews Genetics*, *12*(1), 32–42. http://doi.org/10.1038/nrg2899
Plotkin, J. B., Robins, H., & Levine, A. J. (2004). Tissue-specific codon usage and the expression of human genes. *Proceedings of the National Academy of Sciences*, *101*(34), 12588–12591. http://doi.org/10.1073/pnas.0404957101
Presnyak, V., Alhusaini, N., Chen, Y.-H., Martin, S., Morris, N., Kline, N., et al. (2015). Codon Optimality Is a Major Determinant of mRNA Stability. *Cell*, *160*(6), 1111–1124. http://doi.org/10.1016/j.cell.2015.02.029
Qin, H., Wu, W. B., Comeron, J. M., Kreitman, M., & Li, W.-H. (2004). Intragenic spatial patterns of codon usage bias in prokaryotic and eukaryotic genomes. *Genetics*, *168*(4), 2245–2260. http://doi.org/10.1534/genetics.104.030866
Qiu, S., Zeng, K., Slotte, T., Wright, S., & Charlesworth, D. (2011). Reduced efficacy of natural selection on codon usage bias in selfing *Arabidopsis* and *Capsella* species. *Genome Biology and Evolution*, *3*(0), 868–880. http://doi.org/10.1093/gbe/evr085
Quax, T. E. F., Claassens, N. J., Söll, D., & van der Oost, J. (2015). Codon Bias as a Means to Fine-Tune Gene Expression. *Molecular Cell*, *59*(2), 149–161. http://doi.org/10.1016/j.molcel.2015.05.035
Radhakrishnan, A., Chen, Y.-H., Martin, S., Alhusaini, N., Green, R., & Coller, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell*, *167*(1), 122–132.e9. http://doi.org/10.1016/j.cell.2016.08.053
Rainey, P. B. (1999). Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. *Environmental Microbiology*, *1*(3), 243–257. http://doi.org/10.1046/j.1462-2920.1999.00040.x

Rainey, P. B., & Bailey, M. J. (1996). Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Molecular Microbiology*, *19*(3), 521–533. http://doi.org/10.1046/j.1365-2958.1996.391926.x

Rainey, P. B., & Travisano, M. (1998). Adaptive radiation in a heterogeneous environment. *Nature*, *394*(6688), 69–72. http://doi.org/10.1038/27900

Ralser, M., Querfurth, R., Warnatz, H.-J., Lehrach, H., Yaspo, M.-L., & Krobitsch, S. (2006). An efficient and economic enhancer mix for PCR. *Biochemical and Biophysical Research Communications*, *347*(3), 747–751. http://doi.org/10.1016/j.bbrc.2006.06.151

Ran, W., & Higgs, P. G. (2010). The influence of anticodon-codon interactions and modified bases on codon usage bias in bacteria. *Molecular Biology and Evolution*, *27*(9), 2129–2140. http://doi.org/10.1093/molbev/msq102

Reis, dos, M., Savva, R., & Wernisch, L. (2004). Solving the riddle of codon usage preferences: a test for translational selection. *Nucleic Acids Research*, *32*(17), 5036–5044. http://doi.org/10.1093/nar/gkh834

Remigi, P., Ferguson, G. C., McConnell, E., De Monte, S., Rogers, D. W., & Rainey, P. B. (2019). Ribosome Provisioning Activates a Bistable Switch Coupled to Fast Exit from Stationary Phase. *Molecular Biology and Evolution*, *36*(5), 1056–1070. http://doi.org/10.1093/molbev/msz041

Riehl, N., Remy, P., Ebel, J. P., & Ehresmann, B. (1982). Crosslinking of N-acetyl-phenylalanyl [s4U]tRNAPhe to protein S10 in the ribosomal P site. *European Journal of Biochemistry*, *128*(2-3), 427–433. http://doi.org/10.1111/j.1432-1033.1982.tb06982.x

Rocha, Eduardo, P.C. (2004). Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization, *14*, 2279–2286.

Ross, P., Weinhouse, H., Aloni, Y., Michaeli, D., Weinberger-Ohana, P., Mayer, R., et al. (1987). Regulation of cellulose synthesis in *Acetobacter xylinum* by cyclic diguanylic acid. *Nature*, *325*(6101), 279–281. http://doi.org/10.1038/325279a0

Roux, D., Cywes-Bentley, C., Zhang, Y.-F., Pons, S., Konkol, M., Kearns, D. B., et al. (2015). Identification of Poly-N-acetylglucosamine as a Major Polysaccharide Component of the *Bacillus subtilis* Biofilm Matrix. *The Journal of Biological Chemistry*, *290*(31), 19261–19272. http://doi.org/10.1074/jbc.M115.648709

Römling, U., Galperin, M. Y., & Gomelsky, M. (2013). Cyclic di-GMP: the First 25 Years of a Universal Bacterial Second Messenger. *Microbiology and Molecular Biology Reviews*, *77*(1), 1–52. http://doi.org/10.1128/MMBR.00043-12

Rudorf, S., & Lipowsky, R. (2015). Protein Synthesis in E. coli: Dependence of Codon-Specific Elongation on tRNA Concentration and Codon Usage. *PLoS ONE*, *10*(8), e0134994. http://doi.org/10.1371/journal.pone.0134994

Rutherford, K., Parkhill, J., Crook, J., Horsnell, T., 2000. (n.d.). Artemis: sequence visualization and annotation. *Molecular Biology and Evolution*. http://doi.org/10.1093/bioinformatics/16.10.944","inLanguage":"en","copyrightHolder“:” Oxford

Sabi, R., & Tuller, T. (2014). Modelling the efficiency of codon-tRNA interactions based on codon usage bias. *DNA Research : an International Journal for Rapid Publication of Reports on Genes and Genomes*, *21*(5), 511–526. http://doi.org/10.1093/dnares/dsu017

Sabi, R., Daniel, R. V., Tuller, T., 2016. (n.d.). stAIcalc: tRNA adaptation index calculator based on species-specific weights. *Molecular Biology and Evolution*. http://doi.org/10.1093/bioinformatics/btw647,Oxford

Schindelin, J., Arganda-Carreras, I., Frise, E., Kaynig, V., Longair, M., Pietzsch, T., et al. (2012). Fiji: an open-source platform for biological-image analysis. *Nature Methods*, *9*(7), 676–682. http://doi.org/10.1038/nmeth.2019

Schmidt, A. J., Ryjenkov, D. A., & Gomelsky, M. (2005). The ubiquitous protein domain EAL is a cyclic diguanylate-specific phosphodiesterase: enzymatically active and inactive EAL domains. *Journal of Bacteriology*, *187*(14), 4774–4781. http://doi.org/10.1128/JB.187.14.4774-4781.2005

Seward, E. A., & Kelly, S. (2016). Dietary nitrogen alters codon bias and genome composition in parasitic microorganisms. *Genome Biology*, *17*(1), 226.

http://doi.org/10.1186/s13059-016-1087-9
Shah, P., Ding, Y., Niemczyk, M., Kudla, G., & Plotkin, J. B. (2013). Rate-limiting steps in yeast protein translation. *Cell*, *153*(7), 1589–1601. http://doi.org/10.1016/j.cell.2013.05.049
Sharp, P. M., & Li, W.-H. (1986). Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for "rare" codons. *Nucleic Acids Research*, *14*(19), 7737–7749. http://doi.org/10.1093/nar/14.19.7737
Sharp, P. M., & Li, W.-H. (1987). The codon adaptation index-a measure of directional synonymous codon usage bias, and its potential applications. *Nucleic Acids Research*, *15*(3), 1281–1295. http://doi.org/10.1093/nar/15.3.1281
Sharp, P. M., Tuohy, T. M., & Mosurski, K. R. (1986). Codon usage in yeast: cluster analysis clearly differentiates highly and lowly expressed genes. *Nucleic Acids Research*, *14*(13), 5125–5143.
Silby, M. W., Cerdeño-Tárraga, A. M., Vernikos, G. S., Giddens, S. R., Jackson, R. W., Preston, G. M., et al. (2009). Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biology*, *10*(5), R51. http://doi.org/10.1186/gb-2009-10-5-r51
Silby, M. W., Winstanley, C., Godfrey, S. A. C., Levy, S. B., & Jackson, R. W. (2011). Pseudomonas genomes: diverse and adaptable. *Molecular Biology and Evolution*.
Spiers, A. J., Bohannon, J., Gehrig, S. M., & Rainey, P. B. (2003). Biofilm formation at the air-liquid interface by the *Pseudomonas fluorescens* SBW25 wrinkly spreader requires an acetylated form of cellulose. *Molecular Microbiology*, *50*(1), 15–27. http://doi.org/10.1046/j.1365-2958.2003.03670.x
Spiers, A. J., Folorunso, A. O., medical, K. Z. C., 2013. (n.d.). Cellulose expression in *Pseudomonas fluorescens* SBW25 and other environmental pseudomonads. *Rke.Abertay.Ac.Uk*.
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*, 33–46.
Sprouffske, K., & Wagner, A. (2016). Growthcurver: an R package for obtaining interpretable metrics from microbial growth curves. *BMC Bioinformatics*, *17*(1), 172. http://doi.org/10.1186/s12859-016-1016-7
Steiner, S., Lori, C., Boehm, A., & Jenal, U. (2013). Allosteric activation of exopolysaccharide synthesis through cyclic di-GMP-stimulated protein–protein interaction. *The EMBO Journal*, *32*(3), 354–368. http://doi.org/10.1038/emboj.2012.315
Tello, M., Avalos, F., & Orellana, O. (2018). Codon usage and modular interactions between messenger RNA coding regions and small RNAs in *Escherichia coli*. *BMC Genomics*, *19*(1), 657. http://doi.org/10.1186/s12864-018-5038-6
Tello, M., Saavedra, J. M., & Spencer, E. (2013). Analysis of the use of codon pairs in the HE gene of the ISA virus shows a correlation between bias in HPR codon-pair use and mortality rates caused by the virus. *Virology Journal*, *10*(1), 180. http://doi.org/10.1186/1743-422X-10-180
Trotta, E. (2013). Selection on codon bias in yeast: a transcriptional hypothesis. *Nucleic Acids Research*, *41*(20), 9382–9395. http://doi.org/10.1093/nar/gkt740
Tuller, T., Carmi, A., Vestsigian, K., Navon, S., Dorfan, Y., Zaborske, J., et al. (2010a). An Evolutionarily Conserved Mechanism for Controlling the Efficiency of Protein Translation. *Cell*, *141*(2), 344–354. http://doi.org/10.1016/j.cell.2010.03.031
Tuller, T., Waldman, Y. Y., Kupiec, M., & Ruppin, E. (2010b). Translation efficiency is determined by both codon bias and folding energy. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(8), 3645–3650. http://doi.org/10.1073/pnas.0909910107
Ueda, A., & Saneoka, H. (2014). Characterization of the Ability to Form Biofilms by Plant-Associated *Pseudomonas* Species. *Current Microbiology*, *70*(4), 506–513. http://doi.org/10.1007/s00284-014-0749-7
Ueda, A., & Wood, T. K. (2009). Connecting Quorum Sensing, c-di-GMP, Pel

Polysaccharide, and Biofilm Formation in *Pseudomonas aeruginosa* through Tyrosine Phosphatase TpbA (PA3885). *PLOS Pathogens*, *5*(6), e1000483. http://doi.org/10.1371/journal.ppat.1000483

Valencia-Sánchez, M. I., Rodríguez-Hernández, A., Ferreira, R., Santamaría-Suárez, H. A., Arciniega, M., Dock-Bregeon, A.-C., et al. (2016). Structural Insights into the Polyphyletic Origins of Glycyl tRNA Synthetases. *The Journal of Biological Chemistry*, *291*(28), 14430–14446. http://doi.org/10.1074/jbc.M116.730382

Vasseur, P., Vallet-Gely, I., Soscia, C., Genin, S., & Filloux, A. (2005). The pel genes of the *Pseudomonas aeruginosa* PAK strain are involved at early and late stages of biofilm formation. *Microbiology (Reading, England)*, *151*(3), 985–997. http://doi.org/10.1099/mic.0.27410-0

Wang, X., Preston, J. F., & Romeo, T. (2004). The pgaABCD Locus of *Escherichia coli* Promotes the Synthesis of a Polysaccharide Adhesin Required for Biofilm Formation. *Journal of Bacteriology*, *186*(9), 2724–2734. http://doi.org/10.1128/JB.186.9.2724-2734.2004

Webster, T. A., Gibson, B. W., Keng, T., Biemann, K., & Schimmel, P. (1983). Primary structures of both subunits of *Escherichia coli* glycyl-tRNA synthetase. *Journal of Biological Chemistry*, *258*(17), 10637–10641.

Wilson, M., & Lindow, S. E. (1993). Interactions Between the Biological-Control Agent *Pseudomonas fluorescens* A506 and Erwinia-Amylovora in Pear Blossoms. *Phytopathology*, *83*(1), 117–123. http://doi.org/10.1094/Phyto-83-117

Wong, E. H. M., Smith, D. K., Rabadan, R., Peiris, M., & Poon, L. L. M. (2010). Codon usage bias and the evolution of influenza A viruses. Codon Usage Biases of Influenza Virus. *BMC Evolutionary Biology*, *10*(1), 253. http://doi.org/10.1186/1471-2148-10-253

Xia, X. (2007). An improved implementation of codon adaptation index. *Evolutionary Bioinformatics Online*, *3*, 53–58.

Yan, X., Hoek, T. A., Vale, R. D., & Tanenbaum, M. E. (2016). Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell*, *165*(4), 976–989. http://doi.org/10.1016/j.cell.2016.04.034

Yona, A. H., Bloom-Ackermann, Z., Frumkin, I., Hanson-Smith, V., Charpak-Amikam, Y., Feng, Q., et al. (2013). tRNA genes rapidly change in evolution to meet novel translational demands. *Elife*, *2*, e01339. http://doi.org/10.7554/eLife.01339

Zarai, Y., Margaliot, M., & Tuller, T. (2016). On the Ribosomal Density that Maximizes Protein Translation Rate. *PLoS ONE*, *11*(11), e0166481. http://doi.org/10.1371/journal.pone.0166481

Zhang, X.-X., & Rainey, P. B. (2007). Construction and validation of a neutrally-marked strain of *Pseudomonas fluorescens* SBW25. *Journal of Microbiological Methods*, *71*(1), 78–81. http://doi.org/10.1016/j.mimet.2007.07.001

Zhou, M., Guo, J., Cha, J., Chae, M., Chen, S., Barral, J. M., et al. (2013). Non-optimal codon usage affects expression, structure and function of clock protein FRQ. *Nature*, *495*(7439), 111–115. http://doi.org/10.1038/nature11833

Zorraquino, V., García, B., Latasa, C., Echeverz, M., Toledo-Arana, A., Valle, J., et al. (2013). Coordinated cyclic-di-GMP repression of *Salmonella* motility through YcgR and cellulose. *Journal of Bacteriology*, *195*(3), 417–428. http://doi.org/10.1128/JB.01789-12

Chapter 7 : Appendices

Appendix 1. Appendix for Chapter 3

Appendix 1.1 Codon usage for each amino acid across bacterial strains

The table below lists the most preferred codon and its proportion across the four different strains of *P. fluorescens* – SBW25, A506, Pf0-1 and PICF7. Two outgroups were used for comparison – *P. aeruginosa* PAO1 and *E. coli* B REL606. See section 3.3.1 for explanation and graphical representation of data.

Amino Acid	SBW25		PICF7		A506		Pf0-1		PAO1		REL606	
	Co	Pr	Co	Pr	Co	Pr	Co	Pr	Co	Pr	Co	Pr
A	GCC	0.5	GCC	0.495	GCC	0.506	GCC	0.481	GCC	0.584	GCG	0.352
C	TGC	0.812	TGC	0.796	TGC	0.802	TGC	0.814	TGC	0.901	TGC	0.548
D	GAC	0.680	GAC	0.670	GAC	0.666	GAC	0.635	GAC	0.803	GAT	0.628
E	GAA	0.551	GAA	0.564	GAA	0.550	GAA	0.592	GAG	0.615	GAA	0.691
F	TTC	0.705	TTC	0.709	TTC	0.681	TTC	0.784	TTC	0.952	TTT	0.576
G	GGC	0.630	GGC	0.662	GGC	0.636	GGC	0.608	GGC	0.736	GGC	0.403
H	CAC	0.658	CAC	0.643	CAC	0.625	CAC	0.661	CAC	0.710	CAT	0.568
I	ATC	0.757	ATC	0.759	ATC	0.73	ATC	0.762	ATC	0.909	ATC	0.415
K	AAG	0.628	AAG	0.635	AAG	0.632	AAG	0.591	AAG	0.875	AAA	0.765
L	CTG	0.579	CTG	0.580	CTG	0.570	CTG	0.621	CTG	0.668	CTG	0.490
M	ATG	1	ATG	1	ATG	1	ATG	1	ATG	1	ATG	1
N	AAC	0.762	AAC	0.759	AAC	0.749	AAC	0.768	AAC	0.859	AAC	0.546

P	CCG	0.542	CCG	0.527	CCG	0.495	CCG	0.660	CCG	0.658	CCG	0.521
Q	CAG	0.675	CAG	0.679	CAG	0.662	CAG	0.713	CAG	0.853	CAG	0.651
R	CGC	0.565	CGC	0.556	CGC	0.563	CGC	0.516	CGC	0.648	CGC	0.396
S	AGC	0.401	AGC	0.394	AGC	0.405	AGC	0.383	AGC	0.472	AGC	0.274
T	ACC	0.649	ACC	0.644	ACC	0.651	ACC	0.647	ACC	0.790	ACC	0.427
V	GTG	0.559	GTG	0.558	GTG	0.553	GTG	0.504	GTG	0.485	GTG	0.370
W	TGG	1	TGG	1	TGG	1	TGG	1	TGG	1	TGG	1
Y	TAC	0.717	TAC	0.706	TAC	0.691	TAC	0.691	TAC	0.793	TAT	0.572
*** (STOP)**	TGA	0.562	TGA	0.551	TGA	0.561	TGA	0.628	TGA	0.793	TAA	0.585

Table 7.1: Most preferred codon for each amino acid/species. The proportion of usage for the most preferred codon for each amino acid in *P. fluorescens* SBW25, PICF7, A506, Pf0-1; *P. aeruginosa* PAO1; *E. coli* B REL606. As expected, *E. coli* B REL606 shows more divergence in codon usage with respect to the most preferred codon as well as the proportions of individual codons (differences highlighted in red). The most preferred codon for all four *P. fluorescens* strains is the same for each amino acid, but the frequency of these codons differs across the three strains.

Appendix 1.2 Codon usage in histidin kinase and diguanylate cyclase of the *wsp* and *aws* loci in four strains of *P. fluorescens*

The tables below list the most preferred codon and its proportion for the histidine kinase and diguanylate cyclase of the *wsp* and *aws* loci. *P. fluorescens* Pf0-1 does not have the *aws* locus homologue and hence for this strain the codon usage of only the *wsp* genes are shown. See section 3.3.2 for explanation and graphical representation of data.

Amino acid	*Genome*		Histidine Kinase				Diguanylate cyclase			
			wspE		*awsO*		*wspR*		*awsR*	
	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.
A	GCC	0.5	GCC	0.535	GCG	0.571	GCG	0.471	GCC	0.672
C	TGC	0.812	TGC	0.750	TGC	1	TGC	0.5	TGC	1
D	GAC	0.680	GAC	0.561	GAC	0.625	GAC	0.520	GAC	0.8
E	GAA	0.551	GAA	0.613	GAA	0.667	GAG	0.556	GAA	0.640

F	TTC	0.705	TTC	0.750	TTC	0.5	TTC	0.625	TTC	0.9
G	GGC	0.630	GGC	0.569	GGC	0.765	GGG	0.632	GGC	0.759
H	CAC	0.658	CAC	0.6	CAC	1	CAC	0.556	CAC	0.750
I	ATC	0.757	ATC	0.667	ATC	0.714	ATC	0.636	ATC	0.885
K	AAG	0.628	AAG	0.611	AAG	0.571	AAG	0.909	AAG	0.7
L	CTG	0.579	CTG	0.583	CTG	0.474	CTG	0.641	CTG	0.492
M	ATG	1	ATG	1	ATG	1	ATG	1	ATG	1
N	AAC	0.762	AAC	0.636	AAC	0.6	AAC	0.643	AAC	0.769
P	CCG	0.542	CCG	0.686	CCG	0.4	CCG	0.714	CCG	0.571
Q	CAG	0.675	CAG	0.625	CAA	0.556	CAG	0.882	CAA	0.6
R	CGC	0.565	CGC	0.638	CGT	0.7	CGC	0.440	CGC	0.643
S	AGC	0.401	AGC	0.415	AGC	0.286	TCG	0.389	AGC	0.576
T	ACC	0.649	ACC	0.533	ACC	0.727	ACC	0.556	ACC	0.824
V	GTG	0.559	GTG	0.548	GTG	0.6	GTG	0.650	GTG	0.630
W	TGG	1	TGG	1	TGG	1	TGG	1	TGG	1
Y	TAC	0.717	TAC	0.5	TAC	1	TAT	0.556	TAC	0.8
* (STOP)	TGA	0.562	TGA	1	TGA	1	TAG	1	TGA	1

Table 7.2: Codon usage of the histidine kinases and diguanylate cyclases of *wsp* & *aws* operons in *P. fluorescens* SBW25. The table shows the most preferred codon of the genes *wspE* and *awsO* (histidine kinases) and of *wspR* and *awsR* (diaguanylate cyclases) in *P. fluorescens* SBW25. There is a difference in the most preferred codon for some of the amino acids (where the most preferred codon differs from that of the genome, this is highlighted in red). For the other amino acids, though the most preferred codon remains the same in all three genes, the proportion of codon usage differs.

Amino acid	*Genome*		Histidine Kinase				Diaguanylate cyclase			
			wspE		*awsO*		*wspR*		*awsR*	
	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.
A	GCC	0.495	GCC	0.569	GCG	0.538	GCC	0.457	GCC	0.540
C	TGC	0.796	TGC	1	TGC	1	TGC	0.5	TGC	1
D	GAC	0.670	GAC	0.655	GAC	0.857	GAC	0.720	GAC	0.737
E	GAA	0.564	GAA	0.633	GAA	0.625	GAA	0.667	GAA	0.577
F	TTC	0.709	TTC	0.667	TTC	0.833	TTC	0.625	TTC	0.8
G	GGC	0.662	GGC	0.592	GGC	0.833	GGC	0.421	GGC	0.759
H	CAC	0.643	CAT	0.6	CAC	1	CAT	0.556	CAC	0.706
I	ATC	0.759	ATC	0.667	ATC	0.714	ATC	0.905	ATC	0.875
K	AAG	0.635	AAG	0.550	AAA	0.714	AAG	0.818	AAA	0.556
L	CTG	0.580	CTG	0.6	CTG	0.632	CTG	0.634	CTG	0.492
M	ATG	1	ATG	1	ATG	1	ATG	1	ATG	1
N	AAC	0.759	AAC	0.615	AAC	0.8	AAC	0.533	AAC	0.750

P	CCG	0.527	CCG	0.667	CCA	0.5	CCG	0.786	CCG	0.5
Q	CAG	0.679	CAG	0.718	CAG	0.7	CAG	0.750	CAA	0.625
R	CGC	0.556	CGC	0.623	CGC	0.5	CGC	0.480	CGC	0.462
S	AGC	0.394	AGC	0.4	AGT	0.273	TCG	0.389	AGC	0.531
T	ACC	0.644	ACC	0.552	ACC	0.692	ACC	0.647	ACC	0.737
V	GTG	0.558	GTG	0.563	GTG	0.706	GTG	0.632	GTG	0.483
W	TGG	1	TGG	1	TGG	1	TGG	1	TGG	1
Y	TAC	0.706	TAT	0.6	TAC	1	TAC	0.667	TAC	0.8
* (STOP)	TGA	0.551	TGA	1	TGA	1	TGA	1	TGA	1

Table 7.3: Codon usage of the histidine kinases and diguanylate cyclases of *wsp* & *aws* operons in *P. fluorescens* PICF7. The table shows the most preferred codon of the genes *wspE* and *awsO* (histidine kinases) and of *wspR* and *awsR* (diaguanylate cyclases) in *P. fluorescens* PICF7. There is a difference in the most preferred codon for some of the amino acids (where the most preferred codon differs from that of the genome, this is highlighted in red). For the other amino acids, though the most preferred codon remains the same in all three genes, the proportion of codon usage differs.

Amino acid	*Genome*		Histidine Kinase				Diaguanylate cyclase			
			wspE		*awsO*		*wspR*		*awsR*	
	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.
A	GCC	0.506	GCC	0.581	GCC	0.429	GCC	0.514	GCC	0.610
C	TGC	0.802	TGC	1	TGC	1	TGT	1	TGC	1
D	GAC	0.666	GAC	0.617	GAC	0.833	GAC	0.760	GAC	0.727
E	GAA	0.55	GAA	0.633	GAA	0.6	GAA	0.722	GAA	0.565
F	TTC	0.681	TTC	0.750	TTC	0.667	TTC	0.8	TTC	0.818
G	GGC	0.636	GGC	0.640	GGC	0.941	GGC	0.579	GGC	0.615
H	CAC	0.625	CAT	0.692	CAC	1	CAC	0.875	CAC	0.765
I	ATC	0.73	ATC	0.522	ATC	0.714	ATC	0.850	ATC	0.652
K	AAG	0.632	AAA	0.526	AAA	0.875	AAG	0.727	AAA	0.571
L	CTG	0.57	CTG	0.571	CTG	0.611	CTG	0.595	CTG	0.586
M	ATG	1	ATG	1	ATG	1	ATG	1	ATG	1
N	AAC	0.749	AAC	0.643	AAC	0.667	AAC	0.625	AAC	0.833
P	CCG	0.495	CCG	0.545	CCG	0.429	CCG	0.786	CCC	0.571
Q	CAG	0.662	CAG	0.842	CAA	0.667	CAG	0.8	CAG	0.529
R	CGC	0.563	CGC	0.7	CGC	0.727	CGC	0.8	CGC	0.613
S	AGC	0.405	AGC	0.4	AGC	0.455	AGC	0.375	AGC	0.588
T	ACC	0.651	ACC	0.586	ACC	0.583	ACC	0.706	ACC	0.765
V	GTG	0.553	GTG	0.657	GTG	0.5	GTG	0.526	GTG	0.5
W	TGG	1	TGG	1	TGG	1	TGG	1	TGG	1
Y	TAC	0.691	TAC	0.6	TAC	1	TAC	1	TAC	0.750

* (STOP)	TGA	0.561	TGA	1	TAA	1	TAA	1	TGA	1

Table 7.4: Codon usage of the histidine kinases and diguanylate cyclases of *wsp* & *aws* operons in *P. fluorescens* A506. The table shows the most preferred codon of the genes *wspE* and *awsO* (histidine kinases) and of *wspR* and *awsR* (diaguanylate cyclases) in *P. fluorescens* A506. There is a difference in the most preferred codon for some of the amino acids (where the most preferred codon differs from that of the genome, this is highlighted in red). For the other amino acids, though the most preferred codon remains the same in all three genes, the proportion of codon usage differs.

Amino acid	*Genome*		Histidine kinase		Diaguanylate cyclase	
			wspE		*wspR*	
	Codon	Prop.	Codon	Prop.	Codon	Prop.
A	GCC	0.481	GCC	0.487	GCG	0.545
C	TGC	0.814	TGC	1	TGC	1
D	GAC	0.635	GAC	0.725	GAC	0.808
E	GAA	0.592	GAG	0.507	GAA	0.737
F	TTC	0.784	TTC	0.714	TTC	0.857
G	GGC	0.608	GGC	0.612	GGC	0.526
H	CAC	0.661	CAC	0.688	CAC	0.625
I	ATC	0.762	ATC	0.818	ATC	0.789
K	AAG	0.591	AAG	0.591	AAG	0.727
L	CTG	0.621	CTG	0.619	CTG	0.682
M	ATG	1	ATG	1	ATG	1
N	AAC	0.768	AAC	0.667	AAC	0.625
P	CCG	0.66	CCG	0.719	CCG	0.786
Q	CAG	0.713	CAG	0.821	CAG	0.833
R	CGC	0.516	CGC	0.642	CGC	0.615
S	AGC	0.383	AGC	0.349	TCG	0.333
T	ACC	0.647	ACC	0.8	ACC	0.692
V	GTG	0.504	GTG	0.589	GTG	0.714
W	TGG	1	TGG	1	TGG	1
Y	TAC	0.691	TAT	0.667	TAC	0.7
* (STOP)	TGA	0.628	TGA	1	TGA	1

Table 7.5: Codon usage of the histidine kinases and diguanylate cyclases of *wsp* operon in *P. fluorescens* Pf0-1. The table shows the most preferred codon of the genes *wspE* and *wspR* in *P. fluorescens* Pf0-1. Pf0-1 does not have the *aws* operon. There is a difference in the most preferred codon for some of the amino acids (where the most preferred codon differs from that of the genome,

this is highlighted in red). For the other amino acids, though the most preferred codon remains the same in all three genes, the proportion of codon usage differs.

Amino acid	*wspF* (SBW25)		*wspF* (PICF7)		*wspF* (A506)		*wspF* (Pf0-1)	
A	GCC	0.511	GCC	0.542	GCC	0.533	GCC	0.491
C	TGC	1	TGC	1	TGC	0.667	TGC	1
D	GAC	0.625	GAC	0.688	GAC	0.706	GAC	0.714
E	GAA	0.556	GAA	0.588	GAA	0.5	GAA	0.632
F	TTC	0.571	TTC	0.714	TTC	0.5	TTC	0.714
G	GGC	0.714	GGC	0.786	GGC	0.714	GGC	0.630
H	CAC	0.833	CAC	0.667	CAC	0.667	CAC	0.571
I	ATC	0.750	ATC	0.714	ATC	0.579	ATC	0.857
K	AAG	0.5	AAG	0.625	AAG	0.625	AAG	0.667
L	CTG	0.605	CTG	0.590	CTG	0.528	CTG	0.676
M	ATG	1	ATG	1	ATG	1	ATG	1
N	AAT	0.556	AAC	0.667	AAC	0.556	AAC	0.727
P	CCG	0.571	CCG	0.450	CCG	0.478	CCG	0.476
Q	CAG	0.5	CAG	0.733	CAG	0.714	CAG	0.786
R	CGC	0.652	CGC	0.609	CGC	0.391	CGC	0.5
S	AGC	0.625	AGC	0.667	AGC	0.462	AGC	0.429
T	ACC	0.583	ACC	0.833	ACC	0.615	ACC	0.818
V	GTG	0.788	GTG	0.656	GTG	0.6	GTG	0.567
W	TGG	1	TGG	1	TGG	1	TGG	1
Y	TAT	0.750	TAT	0.750	TAT	0.6	TAC	0.8
* (STOP)	TGA	0.551	TGA	1	TGA	1	TGA	1

Table 7.6: Codon usage of *wspF*, the negative regulator of the the Wsp pathway across the four strains of *P. fluorescens*. The table shows the most preferred codon of the gene *wspF* in strains of *P. fluorescens*. There is a difference in the proportion of the most preferred codon for the amino acids.

Appendix 1.3 CAI values of *wspF* of all four *P. fluorescens* strains

The table below lists the codon adaptation indices (CAI) of *wspF* across all four strains of *P. fluorescens.* See section 3.3.2 for explanation.

Strains	CAI of *wspF*
SBW25	0.687

A506	0.694
Pf0-1	0.668
PICF7	0.686

Table 7.7: Codon Adaptation Index of *glyQ*, *acpP* and *rpsJ* in SBW25, A506, Pf0-1 and PICF7.

Appendix 1.4 Codon usage of *glyQ*, *acpP* and *rpsJ* across SBW25, A506 and Pf0-1

The tables below list the most preferred codon and its proportion for *glyQ, acpP* and *rpsJ*. See section 3.3.3 for explanation and graphical representation of data.

Amino acid	*Genome*		*glyQ*		*acpP*		*rpsJ*	
	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.
A	GCC	0.5	GCC	0.414	GCT	0.625	GCG	0.429
C	TGC	0.812	TGC	0.571	TGC	0	TGC	0
D	GAC	0.680	GAC	0.857	GAC	0.8	GAC	0.625
E	GAA	0.551	GAA	0.857	GAA	0.688	GAA	0.6
F	TTC	0.705	TTC	0.833	TTC	1	TTC	0.5
G	GGC	0.630	GGC	0.667	GGC	0.5	GGT	0.75
H	CAC	0.658	CAC	0.571	CAC	1	CAT	0.667
I	ATC	0.757	ATC	0.6	ATC	1	ATC	0.875
K	AAG	0.628	AAG	0.846	AAA	0.75	AAA	0.571
L	CTG	0.579	CTG	0.5	CTG	0.8	CTG	0.556
M	ATG	1	ATG	1	ATG	1	ATG	1
N	AAC	0.762	AAC	0.818	AAC	1	AAC	0.5
P	CCG	0.542	CCG	0.31	CCG	1	CCA	0.6
Q	CAG	0.675	CAG	0.667	CAA	0.667	CAG	0.556
R	CGC	0.565	CGC	0.269	CGC	1	CGT	0.6
S	AGC	0.401	TCG	0.289	AGC	0.5	TCC	0.667
T	ACC	0.649	ACC	0.375	ACT	0.571	ACC	0.625
V	GTG	0.559	GTG	0.562	GTT	0.4	GTG	0.3
W	TGG	1	TGG	1	TGG	0	TGG	0
Y	TAC	0.717	TAC	0.667	TAC	1	TAC	1
* (STOP)	TGA	0.562	TGA	1	TAA	1	TAA	1

Table 7.8: Codon usage of *glyQ, acpP* and *rpsJ* in *P. fluorescens* SBW25. The table shows the most preferred codon of the genes *glyQ, acpP and rpsJ* in *P. fluorescens* SBW25. There is a difference in the most preferred codon for some of the amino acids (where the most preferred codon differs from that of the genome, this is highlighted in red). For the other amino acids, though the most preferred codon remains the same in all three genes, the proportion of codon usage differs.

Amino acid	*Genome*		*glyQ*		*acpP*		*rpsJ*	
	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.
A	GCC	0.506	GCC	0.630	GCT	0.5	GCG	0.429
C	TGC	0.802	TGC	0.5	TGC	0	TGC	0
D	GAC	0.666	GAC	0.765	GAC	0.8	GAC	0.625
E	GAA	0.55	GAA	0.636	GAA	0.688	GAA	0.6
F	TTC	0.681	TTC	1	TTC	1	TTC	0.5
G	GGC	0.636	GGC	0.952	GGC	0.5	GGT	0.75
H	CAC	0.625	CAC	0.714	CAC	1	CAT	0.667
I	ATC	0.73	ATC	0.750	ATC	1	ATC	0.875
K	AAG	0.632	AAG	0.750	AAG	0.5	AAA	0.571
L	CTG	0.57	CTG	0.667	CTG	0.8	CTG	0.556
M	ATG	1	ATG	1	ATG	1	ATG	1
N	AAC	0.749	AAC	0.909	AAC	1	AAC	0.5
P	CCG	0.495	CCG	0.556	CCG	1	CCA	0.6
Q	CAG	0.662	CAG	0.636	CAA	0.667	CAG	0.556
R	CGC	0.563	CGC	0.471	CGC	1	CGT	0.6
S	AGC	0.405	TCG	0.364	AGC	0.5	TCC	0.667
T	ACC	0.651	ACC	0.556	ACT	0.571	ACC	0.625
V	GTG	0.553	GTG	0.538	GTT	0.4	GTG	0.3
W	TGG	1	TGG	1	TGG	0	TGG	0
Y	TAC	0.691	TAC	0.778	TAC	1	TAC	1
* (STOP)	TGA	0.561	TGA	1	TAA	1	TAA	1

Table 7.9: Codon usage of *glyQ, acpP* and *rpsJ* in *P. fluorescens* A506. The table shows the most preferred codon of the genes *glyQ, acpP and rpsJ* in *P. fluorescens* A506. There is a difference in the most preferred codon for some of the amino acids (where the most preferred codon differs from that of the genome, this is highlighted in red). For the other amino acids, though the most preferred codon remains the same in all three genes, the proportion of codon usage differs.

Amino acid	*Genome*		*glyQ*		*acpP*		*rpsJ*	
	Codon	Prop.	Codon	Prop.	Codon	Prop.	Codon	Prop.
A	GCC	0.481	GCC	0.593	GCT	0.5	GCG	0.429
C	TGC	0.814	TGC	0.5	TGC	0	TGC	0
D	GAC	0.635	GAC	0.882	GAC	1	GAC	0.625
E	GAA	0.592	GAA	0.636	GAA	0.625	GAA	0.6
F	TTC	0.784	TTC	1	TTC	1	TTC	0.5

G	GGC	0.608	GGC	0.810	GGT	1	GGT	0.75
H	CAC	0.661	CAC	0.857	CAC	1	CAT	0.667
I	ATC	0.762	ATC	0.875	ATC	1	ATC	0.875
K	AAG	0.591	AAG	0.875	AAG	0.5	AAA	0.571
L	CTG	0.621	CTG	0.778	CTG	0.8	CTG	0.444
M	ATG	1	ATG	1	ATG	1	ATG	1
N	AAC	0.768	AAC	1	AAC	1	AAC	0.5
P	CCG	0.66	CCG	0.722	CCT	1	CCA	0.6
Q	CAG	0.713	CAG	0.727	CAA	0.667	CAG	0.556
R	CGC	0.516	CGC	0.471	CGC	1	CGT	0.6
S	AGC	0.383	AGC	0.364	AGC	0.5	AGC	0.333
T	ACC	0.647	ACC	0.667	ACC	0.571	ACC	0.625
V	GTG	0.504	GTG	0.423	GTT	0.4	GTG	0.4
W	TGG	1	TGG	1	TGG	0	TGG	0
Y	TAC	0.691	TAC	0.944	TAC	1	TAC	1
* (STOP)	TGA	0.628	TGA	1	TAA	1	TAA	1

Table 7.10: Codon usage of *glyQ, acpP* and *rpsJ* in *P. fluorescens* Pf0-1. The table shows the most preferred codon of the genes *glyQ, acpP and rpsJ* in *P. fluorescens* Pf0-1. There is a difference in the most preferred codon for some of the amino acids (where the most preferred codon differs from that of the genome, this is highlighted in red). For the other amino acids, though the most preferred codon remains the same in all three genes, the proportion of codon usage differs.

Amino acid	*glyQ* (SBW25)		*glyQ* (A506)		*glyQ* (Pf0-1)	
A	GCC	0.414	GCC	0.630	GCC	0.593
C	TGC	0.571	TGC	0.5	TGC	0.5
D	GAC	0.857	GAC	0.765	GAC	0.882
E	GAA	0.857	GAA	0.636	GAA	0.636
F	TTC	0.833	TTC	1	TTC	1
G	GGC	0.667	GGC	0.952	GGC	0.810
H	CAC	0.571	CAC	0.714	CAC	0.857
I	ATC	0.6	ATC	0.750	ATC	0.875
K	AAG	0.846	AAG	0.750	AAG	0.875
L	CTG	0.5	CTG	0.667	CTG	0.778
M	ATG	1	ATG	1	ATG	1
N	AAC	0.818	AAC	0.909	AAC	1
P	CCG	0.31	CCG	0.556	CCG	0.722
Q	CAG	0.667	CAG	0.636	CAG	0.727
R	CGC	0.269	CGC	0.471	CGC	0.471

S	TCG	0.289	TCG	0.364	AGC	0.364
T	ACC	0.375	ACC	0.556	ACC	0.667
V	GTG	0.562	GTG	0.538	GTG	0.423
W	TGG	1	TGG	1	TGG	1
Y	TAC	0.667	TAC	0.778	TAC	0.944
* (STOP)	TGA	1	TGA	1	TGA	1

Table 7.11: Codon usage of *glyQ,* in strains of *P. fluorescens*. The table shows the most preferred codons and the proportions of the codons for *glyQ* in *P. fluorescens* SBW25, A506 and Pf0-1, respectively. Most of the preferred codons are the same between the strains, but the proportions of the codons in the three strains differ.

Appendix 1.5 CAI values of *glyQ, acpP* and *rpsJ* of all four *P. fluorescens* strains

The table below lists the codon adaptation indices (CAI) of *glyQ, acpP* and *rpsJ* across all four strains of *P. fluorescens.* See section 3.3.4 for explanation

Strains	Genes		
	glyQ	*acpP*	*rpsJ*
SBW25	0.72	0.6	0.546
A506	0.719	0.631	0.56
Pf0-1	0.767	0.6	0.528
PICF7	0.725	0.606	0.554

Table 7.12: Codon Adaptation Index of *glyQ*, *acpP* and *rpsJ* in SBW25, A506, Pf0-1 and PICF7.

Appendix 1.6 tAI values for *glyQ, acpP, rpsJ, wsp* and *aws* genes across all four *P. fluorescens* strains

The table below lists the tRNA adaptation indices (tAI) of *glyQ, acpP, rpsJ, wspE, wspF, wspR, awsO* and *awsR* across all four strains of *P. fluorescens.* See section 3.3.5 for explanation.

Strains	Genes

	glyQ	*acpP*	*rpsJ*	*wspE*	*wspF*	*wspR*	*awsO*	*awsR*
SBW25	0.336	0.330	0.296	0.331	0.307	0.322	0.339	0.321
A506	0.279	0.270	0.275	0.263	0.263	0.211	0.235	0.266
Pf0-1	0.266	0.270	0.252	0.248	0.278	0.250	-	-
PICF7	0.272	0.250	0.231	0.249	0.290	0.259	0.275	0.202

Table 7.13: tRNA Adaptation Index of *glyQ, acpP, rpsJ, wspE, wspF, wspR, awsO* and *awsR* in SBW24, A506, Pf0-1 and PICF7.

Appendix 2. Appendix for Chapter 4

This section provides the statistical methods that were used for calculating differences in growth rates, competitive fitness and relative expression of *glyQ* between the mutants and wild type strains of the same background.

Shapiro-Wilk test for normality was used to determine if data collected for each strain came from a normal distribution. For comparison of two or more strains, the Levene's test was used to test for equality of variances between the strains. The test for normality and equality of variances were considered violated if $P<0.05$ for Shapiro-Wilk test for normality and for Levene's test for equality of variances, respectively. In such cases where the test for normality was met, *t*-test was performed to test differences between population means; if the requirement for normality were not met (violated) a Mann-Whitney-Wilcoxon would be used. For cases where the test for equality of variances was met a parametric two-sample *t*-test or ANOVA (comparison of more than two groups) was used; if the assumption for equal variances was violated then a non-parametric type of two-sample *t*-test or non-parametric form of ANOVA (Kruskal-Wallis test) was used, respectively.

Appendix 2.1 Raw data and statistical test for growth rates

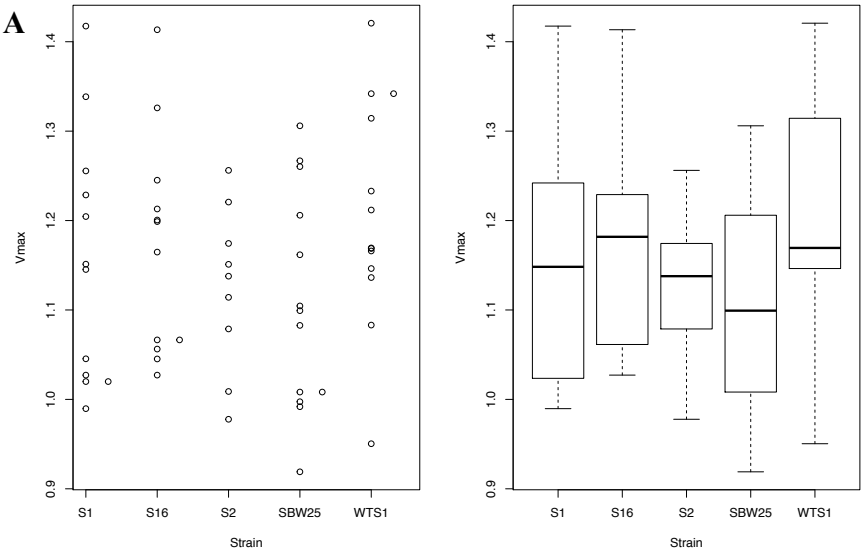
Data for growth rate in LB for SBW25, A506 and Pf0-1 background strains. See section 4.3.3.1 for explanation.

Strain	Growth rate	Mean ± S.E.*	Shapiro-Wilk P-value
SBW25			
A506_*glyQ*_SBW25	1.019, 1.019, 0.989, 1.045, 1.417, 1.229, 1.027, 1.145, 1.204, 1.151, 1.338, 1.225	1.15 ± 0.0402	0.2799
Pf0-1_*glyQ*_SBW25	1.009, 0.978, 1.221, 1.151, 1.079, 1.138, 1.174, 1.114, 1.256	1.12 ± 0.0305	0.8935
Part A506_*glyQ*_SBW25	1.066, 1.066, 1.213, 1.045, 1.245, 1.056, 1.326, 1.027, 1.165, 1.199, 1.413, 1.201	1.17 ± 0.0352	0.2279
SBW25_*glyQ*_SBW25	1.342, 1.342, 1.166, 1.169, 1.136, 1.146,	1.21 ± 0.0348	0.7798

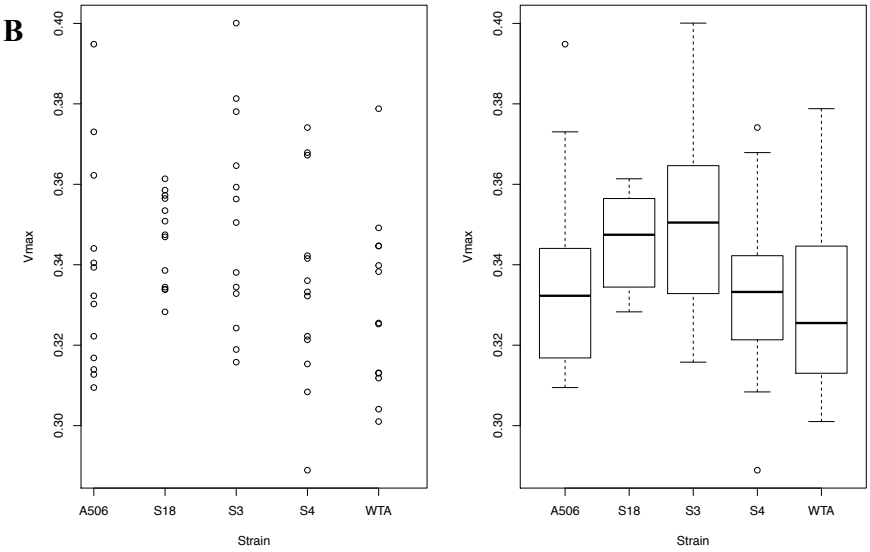
	0.950, 1.083, 1.169, 1.314, .1.212, 1.421, 1.233		
SBW25	1.008, 1.008, 0.997, 0.992, 0.919, 1.104, 1.099, 1.083, 1.260, 1.306, 1.206, 1.267, 1.162	1.11 ± 0.0342	0.4678
A506			
SBW25_*glyQ*_A506	0.324, 0.338, 0.333, 0.381, 0.350, 0.359, 0.378, 0.364, 0.334, 0.356, 0.319, 0.400, 0.316	0.3503 ± 0.0072	0.7166
Pf0-1_*glyQ*_A506	0.342 0.321, 0.374, 0.289, 0.332, 0.308, 0.342, 0.333, 0.315, 0.368, 0.336, 0.322, 0.367	0.3347 ± 0.0069	0.7507
Part Pf0-1_*glyQ*_A506	0.361, 0.359, 0.347, 0.351, 0.356, 0.328, 0.334, 0.334, 0.334, 0.353, 0.339, 0.347, 0.357	0.346 ± 0.0031	0.2308
A506_*glyQ*_A506	0.345, 0.338, 0.379, 0.340, 0.325, 0.349, 0.313, 0.345, 0.313, 0.312, 0.326, 0.304, 0.301	0.329 ± 0.0061	0.3918
A506	0.395, 0.344, 0.332, 0.373, 0.339, 0.322, 0.330, 0.309, 0.317, 0.313, 0.314, 0.340, 0.362	0.338 ± 0.0071	0.1540
Pf0-1			
SBW25_*glyQ*_Pf0-1	0.799, 0.691, 0.761, 0.837, 0.655, 0.706, 0.646, 0.683	0.722 ± 0.0246	0.4035
A506_*glyQ*_Pf0-1 (1)	0.951, 0.972, 1.063, 1.002, 0.959, 1.017, 0.982, 0.999	0.993 ± 0.0128	0.6071
A506_*glyQ*_Pf0-1 (2)	0.633, 0.810, 0.721, 0.702, 0.670, 0.649, 0.653, 0.636, 0.687, 0.664	0.685 ± 0.0167	0.0751
Part A506_*glyQ*_Pf0-1 (1)	0.741, 0.752, 0.775, 0.746, 0.662, 0.646, 0.649, 0.769, 0.650, 0.704	0.709 ± 0.0168	0.0584
Part A506_*glyQ*_Pf0-1 (2)	0.650, 0.723, 0.789, 0.678, 0.689, 0.745, 0.681, 0.661, 0.651, 0.709	0.697 ± 0.0141	0.3378
Pf0-1_*glyQ*_Pf0-1	0.549, 0.582, 0.658, 0.729, 0.781, 0.712, 0.837, 0.639, 0.628, 0.614	0.673 ± 0.0286	0.8122
Pf0-1	0.743, 0.697, 0.670, 0.813, 0.848, 0.678, 0.770, 0.646, 0.627, 0.664	0.719 ± 0.0231	0.5733

Table 7.14: Raw data and test for normality on growth rates in LB. S.E.*= standard error

A


Vmax
0.9
1.0
1.1
1.2
1.3
1.4
S1
S16
S2
SBW25
WTS1
Strain

B


Vmax
0.30
0.32
0.34
0.36
0.38
0.40
A506
S18
S3
S4
WTA
Strain

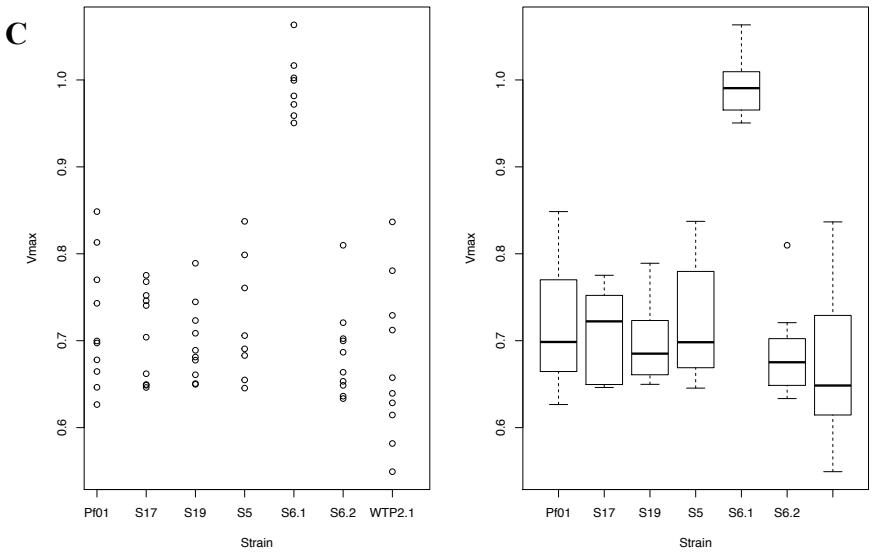


Figure 7.1: Scatter plots (left) and boxplots (right) of the data in Table 7.14. S1= A506_*glyQ*_SBW25, S2= Pf0-1_*glyQ*_SBW25, S16= Part A506_*glyQ*_SBW25, WTS1= SBW25_*glyQ*_SBW25, S3=SBW25_*glyQ*_A506, S4=Pf0-1_*glyQ*_A506, S18= Part Pf0-1_*glyQ*_A506, WTA= A506_*glyQ*_A506, S5= SBW25_*glyQ*_Pf0-1, S6.1 A506_*glyQ*_Pf0-1 (1), S6.2= A506_*glyQ*_Pf0-1 (2), S17= Part A506_*glyQ*_Pf0-1 (1), S19= Part A506_*glyQ*_Pf0-1 (2), WTP2.1= Pf0-1_*glyQ*_Pf0-1.

Samples	Levene test		ANOVA	
	F-stat	*P*-value	*F*-stat	*P*-value
SBW25				
A506_*glyQ*_SBW25, Pf0-1_glyQ_SBW25, Part A506_glyQ_SBW25, SBW25_*glyQ*_SBW25, SBW25	0.473	0.755	1.211	0.317
A506				
SBW25_*glyQ*_A506, Pf0-1_glyQ_A506, Part Pf0-1_glyQ_A506, A506_*glyQ*_A506, A506	1.438	0.232	1.784	0.144
Pf0-1				
SBW25_*glyQ*_Pf0-1, A506_glyQ_Pf0-1, Part A506_glyQ_Pf0-1 (1), Part A506_*glyQ*_Pf0-1 (2), Pf0-1_*glyQ*_Pf0-1, Pf0-1	1.40715	0.217	24.919	0

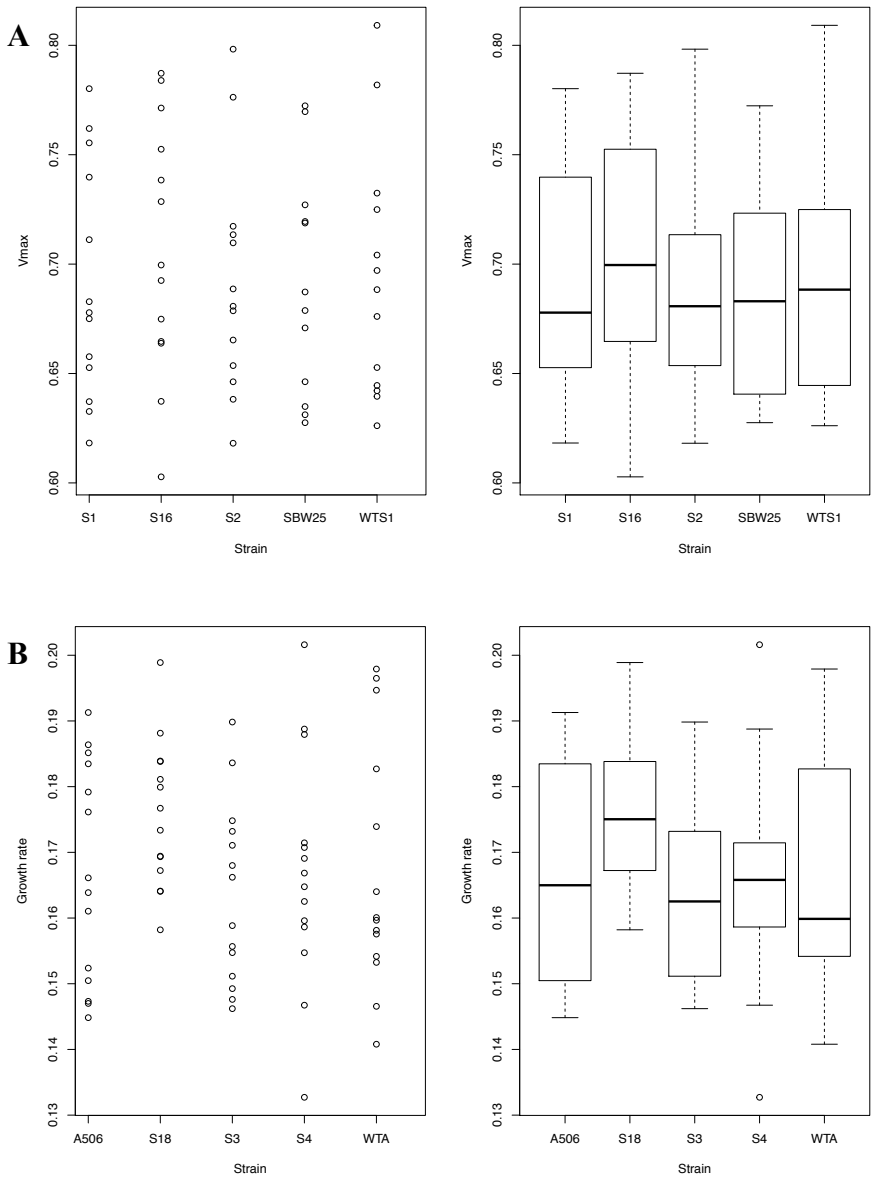
Table 7.15: Statistics for differences in growth rates between mutants and corresponding wild type strains when grown in LB. *F*-stat (*F* statistic) and *P*-value of Levene test and ANOVA provided in the table.

Data for growth rate in KB for SBW25, A506 and Pf0-1 background strains. See section 4.3.3.2 for explanation.

Strain	Growth rate	Mean ± S.E.*	Shapiro-Wilk P-value
SBW25			
A506_*glyQ*_SBW25	0.762, 0.780, 0.740, 0.633, 0.653, 0.683, 0.658, 0.678, 0.675, 0.711, 0.637, 0.618, 0.755	0.691 ± 0.0149	0.3512
Pf0-1_*glyQ*_SBW25	0.798, 0.717, 0.638, 0.681, 0.713, 0.679, 0.776, 0.665, 0.689, 0.646, 0.618, 0.654, 0.710	0.691 ± 0.0145	0.4263
Part A506_*glyQ*_SBW25	0.771, 0.665, 0.753, 0.664, 0.675, 0.738, 0.700, 0.693, 0.787, 0.637, 0.603, 0.729, 0.784	0.707 ± 0.0162	0.7263
SBW25_*glyQ*_SBW25	0.725, 0.704, 0.676, 0.697, 0.645, 0.640, 0.688, 0.653, 0.642, 0.626, 0.732, 0.809, 0.782	0.694 ± 0.0157	0.2507
SBW25	0.772, 0.770, 0.646, 0.679, 0.719, 0.631, 0.687, 0.635, 0.719, 0.628, 0.727, 0.671	0.690 ± 0.0148	0.2769
A506			
SBW25_*glyQ*_A506	0.171, 0.159, 0.148, 0.184, 0.151, 0.168, 0.166, 0.156, 0.173, 0.190, 0.155, 0.149, 0.175, 0.146,	0.164 ± 0.0037	0.4521
Pf0-1_*glyQ*_A506	0.165, 0.189, 0.155, 0.171, 0.188, 0.169, 0.162, 0.167, 0.171, 0.159, 0.147, 0.160, 0.133, 0.202	0.167 ± 0.0047	0.8531
Part Pf0-1_*glyQ*_A506	0.164, 0.167, 0.164, 0.180, 0.188, 0.184, 0.199, 0.177, 0.169, 0.158, 0.181, 0.173, 0.184, 0.169	0.176 ± 0.0030	0.8996
A506_*glyQ*_A506	0.160, 0.141, 0.196, 0.183, 0.164, 0.153, 0.156, 0.154, 0.147, 0.198, 0.174, 0.195, 0.160, 0.158	0.167 ± 0.0050	0.0986
A506	0.185, 0.145, 0.186, 0.176, 0.147, 0.179, 0.152, 0.161, 0.150, 0.147, 0.166, 0.191, 0.184, 0.164	0.167 ± 0.0044	0.1384
Pf0-1			
SBW25_*glyQ*_Pf0-1	0.128, 0.116, 0.202, 0.119, 0.124, 0.119, 0.125, 0.132	0.14 ± 0.0043	0.0617
A506_*glyQ*_Pf0-1 (1)	0.113, 0.105, 0.107, 0.108, 0.110, 0.108, 0.108, 0.108	0.106 ± 0.0025	0.9970

A506_*glyQ*_Pf0-1 (2)	0.132, 0.129, 0.131, 0.123, 0.229, 0.131, 0.138, 0.129, 0.128	0.151 ± 0.0035	0.6966
Part A506_*glyQ*_Pf0-1 (1)	0.126, 0.120, 0.119, 0.128, 0.188, 0.117, 0.128, 0.133, 0.222	0.133 ± 0.0102	0.0594
Part A506_*glyQ*_Pf0-1 (2)	0.127, 0.121, 0.133, 0.126, 0.119, 0.126, 0.128, 0.131, 0.132	0.139 ± 0.0033	0.2249
Pf0-1_*glyQ*_Pf0-1	0.124, 0.124, 0.339, 0.113, 0.197, 0.118, 0.121, 0.119	0.134 ± 0.0048	0.6394
Pf0-1	0.122, 0.137, 0.182, 0.203, 0.124, 0.126, 0.190, 0.116, 0.200	0.17 ± 0.0105	0.0597

Table 7.16: Raw data and test for normality on growth rates in KB. S.E.*= standard error



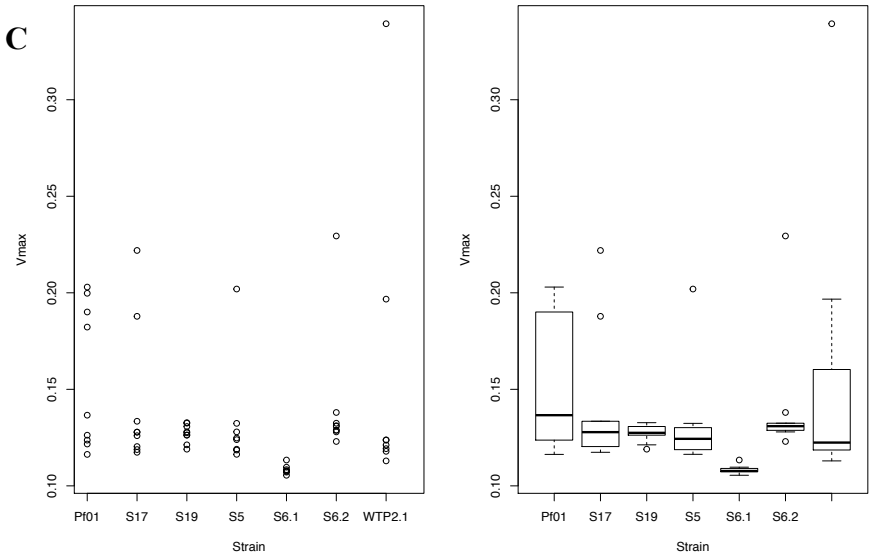


Figure 7.2: Scatter plots (left) and boxplots (right) of the data in Table 7.16. S1= A506_*glyQ*_SBW25, S2= Pf0-1_*glyQ*_SBW25, S16= Part A506_*glyQ*_SBW25, WTS1= SBW25_*glyQ*_SBW25, S3=SBW25_*glyQ*_A506, S4=Pf0-1_*glyQ*_A506, S18= Part Pf0-1_*glyQ*_A506, WTA= A506_*glyQ*_A506, S5= SBW25_*glyQ*_Pf0-1, S6.1 A506_*glyQ*_Pf0-1 (1), S6.2= A506_*glyQ*_Pf0-1 (2), S17= Part A506_*glyQ*_Pf0-1 (1), S19= Part A506_*glyQ*_Pf0-1 (2), WTP2.1= Pf0-1_*glyQ*_Pf0-1.

Samples	Levene test		ANOVA	
	F-stat	*P*-value	*F*-stat	*P*-value
SBW25				
A506_*glyQ*_SBW25, Pf0-1_glyQ_SBW25, Part A506_glyQ_SBW25, SBW25_*glyQ*_SBW25, SBW25	0.357	0.876	0.741	0.595
A506				
SBW25_*glyQ*_A506, Pf0-1_glyQ_A506, Part Pf0-1_glyQ_A506, A506_*glyQ*_A506, A506	0.704	0.592	1.122	0.354
Pf0-1				
SBW25_*glyQ*_Pf0-1, A506_glyQ_Pf0-1, Part A506_glyQ_Pf0-1 (1), Part A506_*glyQ*_Pf0-1 (2), Pf0-1_*glyQ*_Pf0-1, Pf0-1	1.282	0.281	1.565	0.176

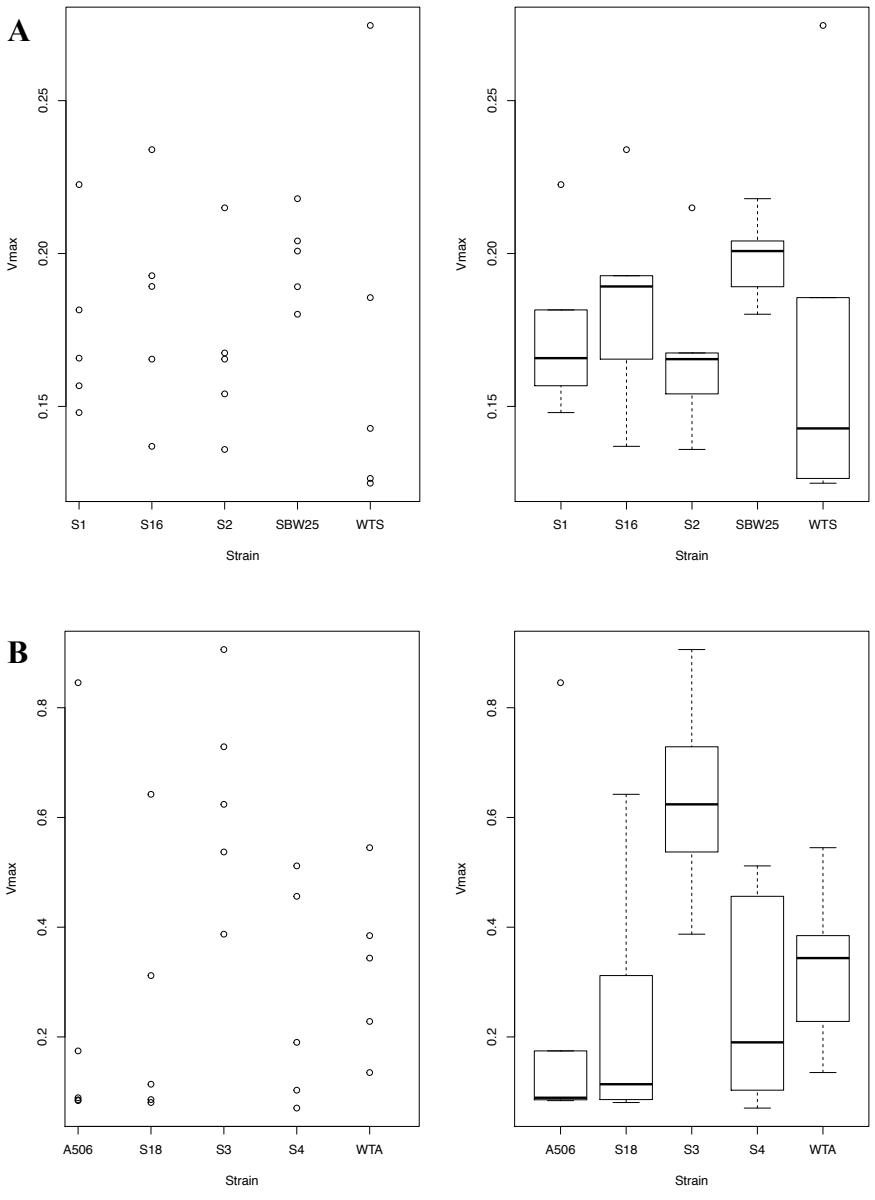
Table 7.17: Statistics for differences in growth rates between mutants and corresponding wild type strains when grown in KB. *F*-stat (*F* statistic) and *P*-value of Levene test and ANOVA provided in the table.

Data for growth rate in M9+glucose (20%) for SBW25, A506 and Pf0-1 background strains. See section 4.3.3.3 for explanation.

Strain	Growth rate	Mean ± S.E.*	Shapiro-Wilk P-value
SBW25			
A506_*glyQ*_SBW25	0.762, 0.780, 0.740, 0.633, 0.653, 0.683, 0.658, 0.678, 0.675, 0.711, 0.637, 0.618, 0.755	0.175 ± 0.0131	0.894
Pf0-1_*glyQ*_SBW25	0.798, 0.717, 0.638, 0.681, 0.713, 0.679, 0.776, 0.665, 0.689, 0.646, 0.618, 0.654, 0.710	0.168 ± 0. 0131	0.906
Part A506_*glyQ*_SBW25	0.771, 0.665, 0.753, 0.664, 0.675, 0.738, 0.700, 0.693, 0.787, 0.637, 0.603, 0.729, 0.784	0.184 ± 0. 0161	0.979
SBW25_*glyQ*_SBW25	0.725, 0.704, 0.676, 0.697, 0.645, 0.640, 0.688, 0.653, 0.642, 0.626, 0.732, 0.809, 0.782	0.171 ± 0.0282	0.817
SBW25	0.772, 0.770, 0.646, 0.679, 0.719, 0.631, 0.687, 0.635, 0.719, 0.628, 0.727, 0.671	0.198 ± 0.0065	0.984
A506			
SBW25_*glyQ*_A506	0.171, 0.159, 0.148, 0.184, 0.151, 0.168, 0.166, 0.156, 0.173, 0.190, 0.155, 0.149, 0.175, 0.146,	0.637 ± 0.0876	0.9981
Pf0-1_*glyQ*_A506	0.165, 0.189, 0.155, 0.171, 0.188, 0.169, 0.162, 0.167, 0.171, 0.159, 0.147, 0.160, 0.133, 0.202	0.266 ± 0.0914	0.2192
Part Pf0-1_*glyQ*_A506	0.164, 0.167, 0.164, 0.180, 0.188, 0.184, 0.199, 0.177, 0.169, 0.158, 0.181, 0.173, 0.184, 0.169	0.247 ± 0.108	0.0668
A506_*glyQ*_A506	0.160, 0.141, 0.196, 0.183, 0.164, 0.153, 0.156, 0.154, 0.147, 0.198, 0.174, 0.195, 0.160, 0.158	0.327 ± 0.0699	0.9585
A506	0.185, 0.145, 0.186, 0.176, 0.147, 0.179, 0.152, 0.161, 0.150, 0.147, 0.166, 0.191, 0.184, 0.164	0.256 ± 0.148	0.0016#
Pf0-1			
SBW25_*glyQ*_Pf0-1	0.128, 0.116, 0.202, 0.119, 0.124, 0.119, 0.125, 0.132	0.139 ± 0.0122	0.4351
A506_*glyQ*_Pf0-1 (1)	0.113, 0.105, 0.107, 0.108, 0.110, 0.108, 0.108, 0.108	0.863 ± 0.227	0.8873

A506_*glyQ*_Pf0-1 (2)	0.132, 0.129, 0.131, 0.123, 0.229, 0.131, 0.138, 0.129, 0.128	0.113 ± 0.0068	0.2707
Part A506_*glyQ*_Pf0-1 (1)	0.126, 0.120, 0.119, 0.128, 0.188, 0.117, 0.128, 0.133, 0.222	0.107 ± 0.0097	0.8356
Part A506_*glyQ*_Pf0-1 (2)	0.127, 0.121, 0.133, 0.126, 0.119, 0.126, 0.128, 0.131, 0.132	0.13 ± 0.0172	0.1882
Pf0-1_*glyQ*_Pf0-1	0.124, 0.124, 0.339, 0.113, 0.197, 0.118, 0.121, 0.119	0.149 ± 0.013	0.6971
Pf0-1	0.122, 0.137, 0.182, 0.203, 0.124, 0.126, 0.190, 0.116, 0.200	0.116 ± 0.0138	0.1484

Table 7.18: Raw data and test for normality on growth rates in M9+glucose (20%). S.E.*= standard error. # Violation of test for normality.



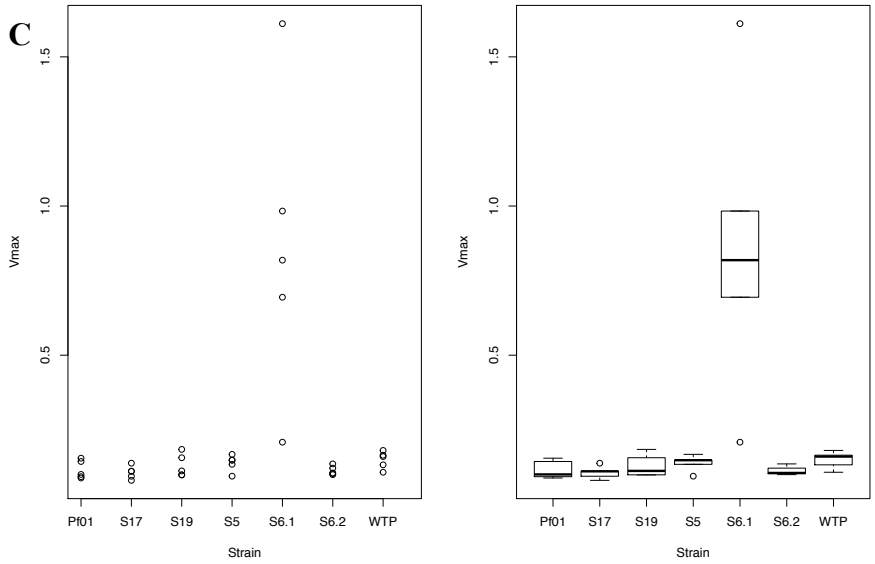


Figure 7.3: Scatter plots (left) and boxplots (right) of the data in Table 7.18. S1= A506_*glyQ*_SBW25, S2= Pf0-1_*glyQ*_SBW25, S16= Part A506_*glyQ*_SBW25, WTS1= SBW25_*glyQ*_SBW25, S3=SBW25_*glyQ*_A506, S4=Pf0-1_*glyQ*_A506, S18= Part Pf0-1_*glyQ*_A506, WTA= A506_*glyQ*_A506, S5= SBW25_*glyQ*_Pf0-1, S6.1 A506_*glyQ*_Pf0-1 (1), S6.2= A506_*glyQ*_Pf0-1 (2), S17= Part A506_*glyQ*_Pf0-1 (1), S19= Part A506_*glyQ*_Pf0-1 (2), WTP2.1= Pf0-1_*glyQ*_Pf0-1.

Samples	Levene test		ANOVA	
	F-stat	*P*-value	*F*-stat	*P*-value
SBW25				
A506_*glyQ*_SBW25, Pf0-1_glyQ_SBW25, Part A506_glyQ_SBW25, SBW25_*glyQ*_SBW25, SBW25	0.784	0.549	0.533	0.713
A506				
SBW25_*glyQ*_A506, Pf0-1_glyQ_A506, Part Pf0-1_glyQ_A506, A506_*glyQ*_A506, A506	0.062	0.992	-	0.085[#]
Pf0-1				
SBW25_*glyQ*_Pf0-1, A506_glyQ_Pf0-1, Part A506_glyQ_Pf0-1 (1), Part A506_*glyQ*_Pf0-1 (2), Pf0-1_*glyQ*_Pf0-1, Pf0-1	4.234	0.004	-	0.007[#]

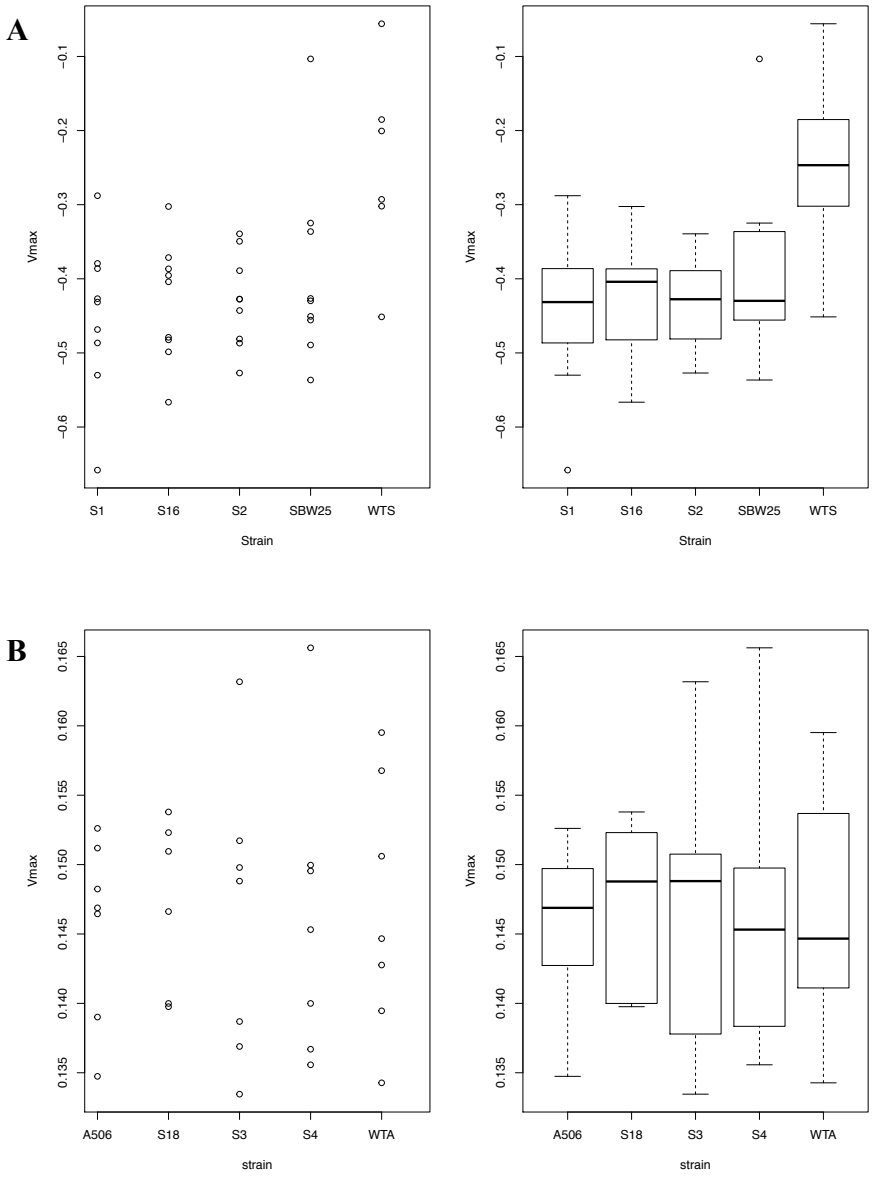
Table 7.19: Statistics for differences in growth rates between mutants and corresponding wild type strains when grown in M9+glucose (20%). *F*-stat (*F* statistic) and *P*-value of Levene test and

ANOVA provided in the table. # These are non-parametric ANOVA (Kruskal-Wallis test) since the assumption for normality and equality of variance were not met for A506 and Pf0-1, respectively.

Data for growth rate in TB for SBW25, A506 and Pf0-1 background strains. See section 4.3.3.4 for explanation.

Strain	Growth rate	Mean ± S.E.*	Shapiro-Wilk P-value
SBW25			
A506_*glyQ*_SBW25	0.515, 0.418, 0.220, 0.326, 0.370, 0.340, 0.295, 0.374, 0.411	0.363 ± 0.0278	0.9716
Pf0-1_*glyQ*_SBW25	0.408, 0.458, 0.361, 0.374, 0.297, 0.447, 0.330, 0.326, 0.374	0.375 ± 0. 0182	0.6993
Part A506_*glyQ*_SBW25	0.425, 0.317, 0.394, 0.332, 0.411, 0.402, 0.498, 0.271, 0.329	0.376 ± 0. 023	0.8070
SBW25_*glyQ*_SBW25	0.879, 0.354, 0.499, 0.510, 0.630, 0.653	0.587 ± 0.073	0.8242
SBW25	0.461, 0.473, 0.372, 0.324, 0.350, 0.354, 0.375, 0.291	0.375 ± 0.0223	0.3496
A506			
SBW25_*glyQ*_A506	0.133, 0.163, 0.137, 0.152, 0.150, 0.149, 0.139	0.146 ± 0.0039	0.6252
Pf0-1_*glyQ*_A506	0.150, 0.145, 0.166, 0.136, 0.140, 0.150, 0.137	0.146 ± 0.0039	0.3298
Part Pf0-1_*glyQ*_A506	0.154, 0.152, 0.147, 0.151, 0.140, 0.140	0.147 ± 0.0025	0.2073
A506_*glyQ*_A506	0.160, 0.151, 0.139, 0.157, 0.145, 0.143, 0.134	0.147 ± 0.0035	0.8301
A506	0.153, 0.135, 0.147, 0.146, 0.148, 0.151, 0.139	0.146 ± 0.0024	0.3657
Pf0-1			
SBW25_*glyQ*_Pf0-1	0.270, 0.311, 0.218, 0.318, 0.296, 0.340	0.292 ± 0.0176	0.5931
A506_*glyQ*_Pf0-1 (1)	0.302, 0.360, 0.391, 0.523, 0.443, 0.385	0.401 ± 0.0308	0.8673
A506_*glyQ*_Pf0-1 (2)	0.240, 0.264, 0.320, 0.288, 0.305, 0.216	0.272 ± 0.0163	0.8697
Part A506_*glyQ*_Pf0-1 (1)	0.266, 0.396, 0.301, 0.463, 0.198, 0.326	0.325 ± 0.0384	0.9801
Part A506_*glyQ*_Pf0-1 (2)	0.278, 0.279, 0.196, 0.287, 0.229, 0.360	0.272 ± 0.0228	0.7066
Pf0-1_*glyQ*_Pf0-1	0.180, 0.282, 0.223, 0.207, 0.294, 0.378	0.247 ± 0.0285	0.5584
Pf0-1	0.287, 0.307, 0.255, 0.300, 0.263, 0.336	0.292 ± 0.0121	0.8639

Table 7.20: Raw data and test for normality on growth rates in TB. S.E.*= standard error.



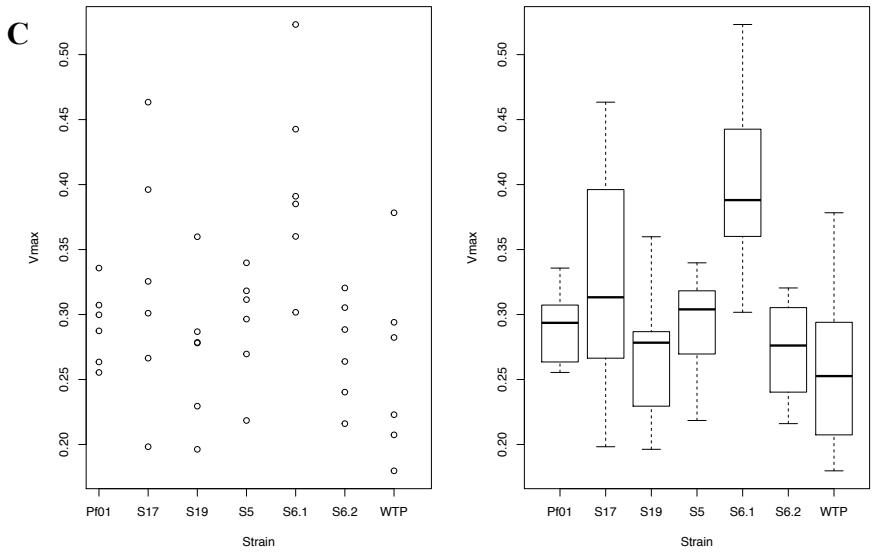


Figure 7.4: Scatter plots (left) and boxplots (right) of the data in Table 7.20. S1= A506_*glyQ*_SBW25, S2= Pf0-1_*glyQ*_SBW25, S16= Part A506_*glyQ*_SBW25, WTS1= SBW25_*glyQ*_SBW25, S3=SBW25_*glyQ*_A506, S4=Pf0-1_*glyQ*_A506, S18= Part Pf0-1_*glyQ*_A506, WTA= A506_*glyQ*_A506, S5= SBW25_*glyQ*_Pf0-1, S6.1 A506_*glyQ*_Pf0-1 (1), S6.2= A506_*glyQ*_Pf0-1 (2), S17= Part A506_*glyQ*_Pf0-1 (1), S19= Part A506_*glyQ*_Pf0-1 (2), WTP2.1= Pf0-1_*glyQ*_Pf0-1.

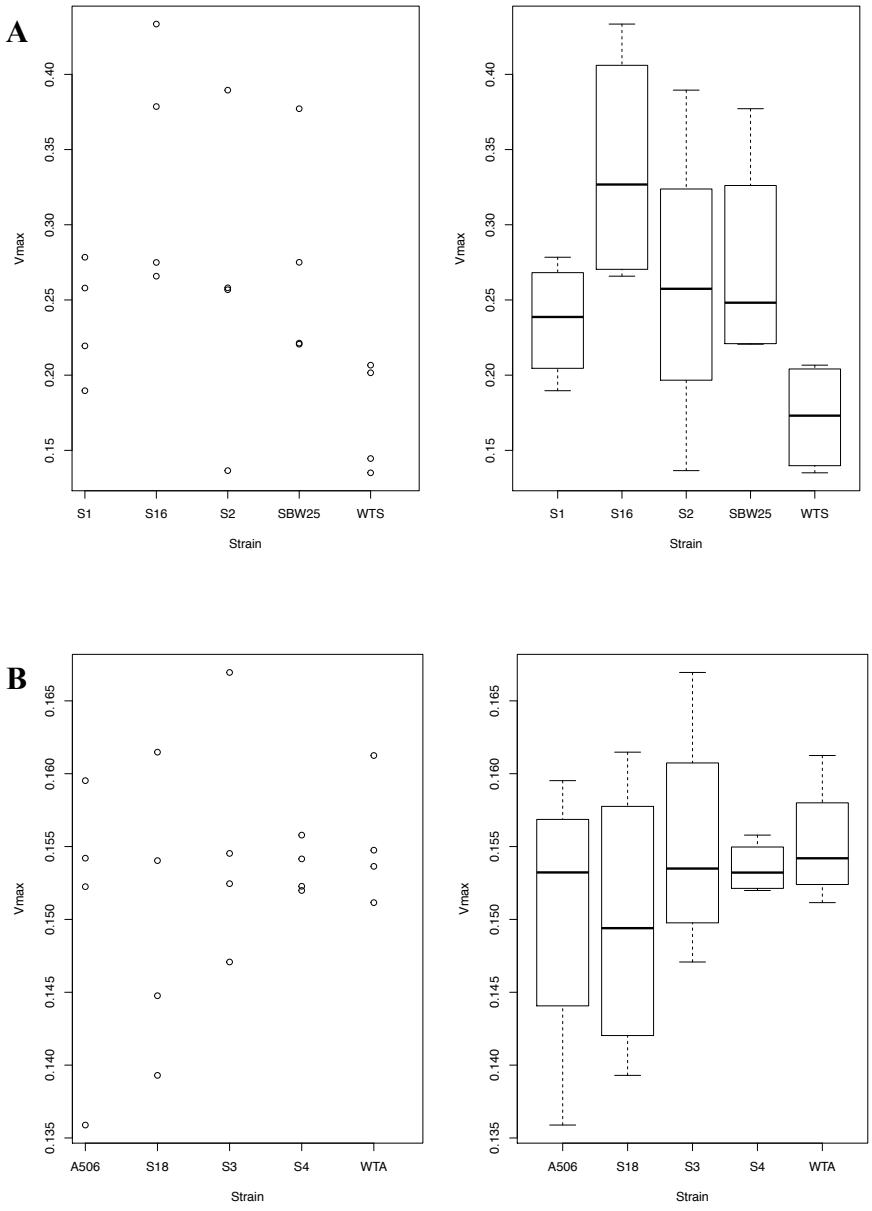
Samples	Levene test		ANOVA	
	F-stat	*P*-value	*F*-stat	*P*-value
SBW25				
A506_*glyQ*_SBW25, Pf0-1_glyQ_SBW25, Part A506_glyQ_SBW25, SBW25_*glyQ*_SBW25, SBW25	0.587	0.674	4.224	0.006
A506				
SBW25_*glyQ*_A506, Pf0-1_glyQ_A506, Part Pf0-1_glyQ_A506, A506_*glyQ*_A506, A506	0.548	0.702	0.0375	0.997
Pf0-1				
SBW25_*glyQ*_Pf0-1, A506_glyQ_Pf0-1, Part A506_glyQ_Pf0-1 (1), Part A506_*glyQ*_Pf0-1 (2), Pf0-1_*glyQ*_Pf0-1, Pf0-1	1.172	0.344	3.590	0.007

Table 7.21: Statistics for differences in growth rates between mutants and corresponding wild type strains when grown in TB. *F*-stat (*F* statistic) and *P*-value of Levene test and ANOVA provided in the table.

Data for growth rate in SB for SBW25, A506 and Pf0-1 background strains. See section 4.3.3.5 for explanation.

Strain	Growth rate	Mean ± S.E.*	Shapiro-Wilk P-value
SBW25			
A506_*glyQ*_SBW25	0.515, 0.418, 0.220, 0.326, 0.370, 0.340, 0.295, 0.374, 0.411	0.236 ± 0.0198	0.8189
Pf0-1_*glyQ*_SBW25	0.408, 0.458, 0.361, 0.374, 0.297, 0.447, 0.330, 0.326, 0.374	0.260 ± 0. 0517	0.6912
Part A506_*glyQ*_SBW25	0.425, 0.317, 0.394, 0.332, 0.411, 0.402, 0.498, 0.271, 0.329	0.338 ± 0. 0408	0.3368
SBW25_*glyQ*_SBW25	0.879, 0.354, 0.499, 0.510, 0.630, 0.653	0.172 ± 0.0187	0.1465
SBW25	0.461, 0.473, 0.372, 0.324, 0.350, 0.354, 0.375, 0.291	0.274 ± 0.0368	0.1807
A506			
SBW25_*glyQ*_A506	0.133, 0.163, 0.137, 0.152, 0.150, 0.149, 0.139	0.155 ± 0.0042	0.5644
Pf0-1_*glyQ*_A506	0.150, 0.145, 0.166, 0.136, 0.140, 0.150, 0.137	0.154 ± 0.0009	0.4628
Part Pf0-1_*glyQ*_A506	0.154, 0.152, 0.147, 0.151, 0.140, 0.140	0.150 ± 0.0049	0.8527
A506_*glyQ*_A506	0.160, 0.151, 0.139, 0.157, 0.145, 0.143, 0.134	0.155 ± 0.0021	0.4968
A506	0.153, 0.135, 0.147, 0.146, 0.148, 0.151, 0.139	0.150 ± 0.0051	0.3421
Pf0-1			
SBW25_*glyQ*_Pf0-1	0.270, 0.311, 0.218, 0.318, 0.296, 0.340	0.378 ± 0.0364	0.1505
A506_*glyQ*_Pf0-1 (1)	0.302, 0.360, 0.391, 0.523, 0.443, 0.385	0.505 ± 0.0189	0.6494
A506_*glyQ*_Pf0-1 (2)	0.240, 0.264, 0.320, 0.288, 0.305, 0.216	0.342 ± 0.0355	0.3658
Part A506_*glyQ*_Pf0-1 (1)	0.266, 0.396, 0.301, 0.463, 0.198, 0.326	0.417 ± 0.0233	0.4194
Part A506_*glyQ*_Pf0-1 (2)	0.278, 0.279, 0.196, 0.287, 0.229, 0.360	0.376 ± 0.0254	0.9549
Pf0-1_*glyQ*_Pf0-1	0.180, 0.282, 0.223, 0.207, 0.294, 0.378	0.365 ± 0.0238	0.5868
Pf0-1	0.287, 0.307, 0.255, 0.300, 0.263, 0.336	0.377 ± 0.0201	0.3347

Table 7.22: Raw data and test for normality on growth rates in TB. S.E.*= standard error.



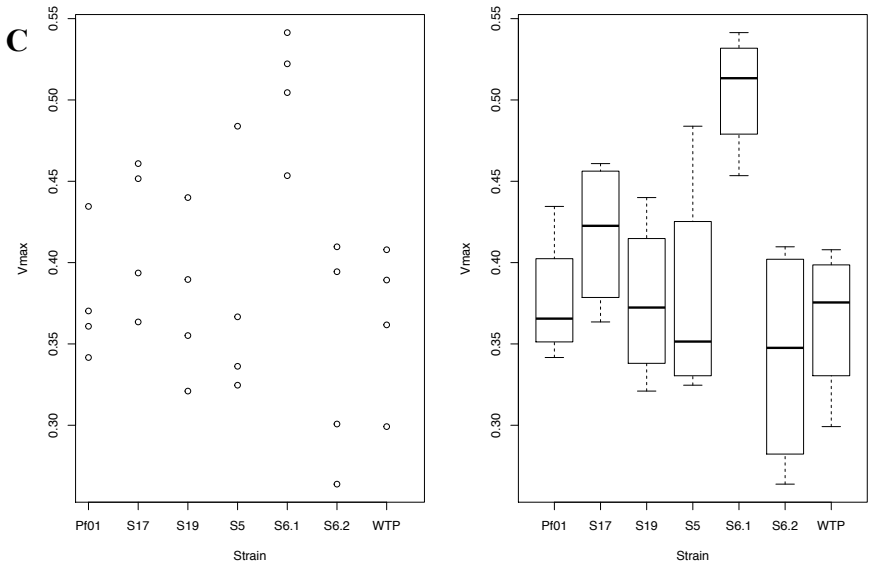


Figure 7.5: Scatter plots (left) and boxplots (right) of the data in Table 7.22. S1= A506_*glyQ*_SBW25, S2= Pf0-1_*glyQ*_SBW25, S16= Part A506_*glyQ*_SBW25, WTS1= SBW25_*glyQ*_SBW25, S3=SBW25_*glyQ*_A506, S4=Pf0-1_*glyQ*_A506, S18= Part Pf0-1_*glyQ*_A506, WTA= A506_*glyQ*_A506, S5= SBW25_*glyQ*_Pf0-1, S6.1= A506_*glyQ*_Pf0-1 (1), S6.2= A506_*glyQ*_Pf0-1 (2), S17= Part A506_*glyQ*_Pf0-1 (1), S19= Part A506_*glyQ*_Pf0-1 (2), WTP2.1= Pf0-1_*glyQ*_Pf0-1.

Samples	Levene test		ANOVA	
	F-stat	*P*-value	*F*-stat	*P*-value
SBW25				
A506_*glyQ*_SBW25, Pf0-1_glyQ_SBW25, Part A506_glyQ_SBW25, SBW25_*glyQ*_SBW25, SBW25	0.655	0.632	2.825	0.063
A506				
SBW25_*glyQ*_A506, Pf0-1_glyQ_A506, Part Pf0-1_glyQ_A506, A506_*glyQ*_A506, A506	1.172	0.362	0.447	0.773
Pf0-1				
SBW25_*glyQ*_Pf0-1, A506_glyQ_Pf0-1, Part A506_glyQ_Pf0-1 (1), Part A506_*glyQ*_Pf0-1 (2), Pf0-1_*glyQ*_Pf0-1, Pf0-1	0.601	0.727	3.971	0.008

Table 7.23: Statistics for differences in growth rates between mutants and corresponding wild type strains when grown in SB. *F*-stat (*F* statistic) and *P*-value of Levene test and ANOVA provided in the table.

Appendix 2.2 Raw data for relative fitness experiments.

Data for competitive fitness assay in KB of SBW25, A506 and Pf0-1 background strains. See section 4.3.3.6 for explanation.

For SBW25 background

Competition 1.1: SBW25 *lacZ* (a) and SBW25_*glyQ*_SBW25 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	32	41	100	52	62	7.393	7.321	1.010
2	100	21	25	100	40	53	7.552	7.659	0.986
3	100	27	31	100	16	17	6.384	6.307	1.012
4	100	34	33	100	25	39	6.600	7.075	0.933
5	100	30	28	100	19	33	6.451	7.072	0.912
6	100	23	23	100	12	23	6.257	6.908	0.906
7	100	21	21	100	76	83	8.194	8.282	0.989
8	100	40	43	100	19	17	6.163	5.980	1.031
Mean Relative fitness ± S.E. : 0.972 ± 0.0172									

Competition 1.2: SBW25 *lacZ* (a) and SBW25_*glyQ*_SBW25 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	32	41	50	25	41	7.354	7.601	0.968
2	100	21	25	50	70	61	8.805	8.493	1.037
3	100	27	31	50	60	59	8.399	8.244	1.019
4	100	34	33	50	59	80	8.152	8.486	0.961
5	100	30	28	50	53	56	8.170	8.294	0.985
6	100	23	23	50	77	73	8.809	8.756	1.006
7	100	21	21	50	51	68	8.488	8.776	0.967
8	100	40	43	50	54	68	7.901	8.059	0.980

Mean Relative fitness ± S.E. : 0.990 ± 0.001

Competition 2.1: SBW25 *lacZ* (a) and A506_*glyQ*_SBW25 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	25	23	100	41	35	7.402	7.328	1.010
2	100	39	30	100	46	54	7.073	7.496	0.944
3	100	13	10	100	57	46	8.386	8.434	0.994
4	100	14	16	100	36	41	7.852	7.849	1.000
5	100	90	73	100	25	34	5.627	6.144	0.916
6	100	18	11	100	40	45	7.706	8.317	0.927
7	100	17	19	100	55	75	8.082	8.281	0.976
8	100	27	48	100	50	45	7.524	6.843	1.099
Mean Relative fitness ± S.E. : 0.983 ± 0.021									

Competition 2.2: SBW25 *lacZ* (a) and A506_*glyQ*_SBW25 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	25	23	50	40	37	8.071	8.076	0.999
2	100	39	30	50	56	51	7.963	8.132	0.979
3	100	13	10	50	56	35	9.061	8.854	1.023
4	100	14	16	50	60	72	9.056	9.105	0.995
5	100	90	73	50	47	33	6.951	6.807	1.021
6	100	18	11	50	43	59	8.472	9.281	0.913
7	100	17	19	50	72	64	9.044	8.815	1.026
8	100	27	48	50	66	59	8.495	7.807	1.088
Mean Relative fitness ± S.E. : 1.005 ± 0.0191									

Competition 3.1: SBW25 *lacZ* (a) and SBW25_*glyQ*_SBW25 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	82	95	50	72	77	7.471	6.828	1.094
2	100	93	53	50	39	43	6.732	7.568	0.890
3	100	40	46	50	26	33	7.170	7.006	1.023
4	100	33	44	50	24	34	7.282	6.968	1.045
5	100	30	34	50	43	40	7.961	6.710	1.186
6	100	34	37	100	48	60	7.253	7.046	1.029
7	100	41	59	100	79	104	7.564	6.819	1.109
8	100	65	30	50	85	101	7.869	7.853	1.002
Mean Relative fitness ± S.E. : 1.047 ± 0.031									

Competition 3.2: SBW25 *lacZ* (a) and SBW25_*glyQ*_SBW25 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	82	95	50	59	68	7.272	7.267	1.001
2	100	93	53	50	59	61	7.146	7.741	0.923
3	100	40	46	50	46	38	7.741	7.410	1.045
4	100	33	44	50	71	61	8.367	7.928	1.055
5	100	30	34	50	66	56	8.389	8.100	1.036
6	100	34	37	50	44	70	7.859	8.238	0.954
7	100	41	59	50	47	67	7.737	7.728	1.001
8	100	65	30	50	61	76	7.537	8.530	0.884
Mean Relative fitness ± S.E. : 0.987 ± 0.022									

Competition 4.1: SBW25 *lacZ* (a) and Pf0-1_*glyQ*_SBW25 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	16	19	100	49	64	8.027	8.122	0.988
2	100	24	19	100	172	175	8.877	9.128	0.973
3	100	12	17	100	32	33	7.889	7.571	1.042
4	100	22	24	100	21	13	6.861	6.295	1.090
5	100	45	53	100	58	51	7.162	6.869	1.043
6	100	36	32	100	34	38	6.851	7.080	0.968
7	100	47	29	100	28	44	6.390	7.325	0.872
8	100	63	40	100	43	51	6.526	7.151	0.913
Mean Relative fitness ± S.E. : 986 ± 0.025									

Competition 4.2: SBW25 *lacZ* (a) and Pf0-1_*glyQ*_SBW25 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	16	19	Yeast contamination			n/a	n/a	n/a
2	100	24	19	50	70	69	8.671	8.891	0.975
3	100	12	17	50	86	81	9.570	9.162	1.045
4	100	22	24	50	51	74	8.442	8.727	0.967
5	100	45	53	50	62	66	7.921	7.820	1.013
6	100	36	32	50	52	60	7.969	8.230	0.968
7	100	47	29	50	62	58	7.878	8.294	0.950
8	100	63	40	50	52	56	7.409	7.937	0.933
Mean Relative fitness ± S.E. : 0.979 ± 0.014									

Tables 7.24: Colony counts, volume plated, Malthusian parameter and Relative fitness of competitors. Colony counts were obtained by plating either 50 µl or 100 µl of 10^{-5} dilution for each replicate. The tables are for competitions 1.1 to 4.2

For A506 background

Competition 1.1: SBW25 *lacZ* (a) and A506 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	24	11	100	9	14	9.010	8.671	1.039
2	100	39	43	100	10	20	7.752	8.543	0.907
3	100	27	15	100	3	14	7.601	8.554	0.889
4	100	27	30	100	10	24	8.112	9.093	0.892
5	100	19	33	100	2	28	6.407	9.598	0.668
6	100	21	23	100	21	18	9.119	9.056	1.007
7	100	27	26	100	4	28	7.339	9.247	0.794
8	100	27	32	100	6	39	7.536	9.578	0.787
Mean Relative fitness ± S.E. : 0.873 ± 0.043									

Competition 1.2: SBW25 *lacZ* (a) and A506 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	24	11	100	7	41	7.978	10.526	0.758
2	100	39	43	100	9	110	7.744	10.150	0.763
3	100	27	15	100	2	55	6.608	10.510	0.629
4	100	27	30	100	Yeast contamination	61	n/a	9.920	n/a
5	100	19	33	100	6	38	8.058	9.351	0.862
6	100	21	23	100	4	53	7.552	10.045	0.752
7	100	27	26	100	8	100	7.994	10.557	0.757
8	100	27	32	100	18	114	8.805	10.481	0.840
Mean Relative fitness ± S.E. : 0.990 ± 0.001									

Competition 2.1: SBW25 *lacZ* (a) and A506_*glyQ*_A506 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	50	53	69	100	16	14	7.319	6.922	1.057
2	50	76	84	100	26	45	7.445	7.893	0.943
3	50	38	54	100	10	22	7.182	7.619	0.943
4	50	36	27	100	14	44	7.573	9.006	0.841
5	50	33	46	100	12	24	7.506	7.867	0.954
6	50	25	37	100	8	26	7.378	8.164	0.904
7	50	30	46	100	12	18	7.601	7.579	1.003
8	50	41	48	100	11	27	7.202	7.942	0.907
Mean Relative fitness ± S.E. : 0.944 ± 0.023									

Competition 2.2: SBW25 *lacZ* (a) and A506_*glyQ*_A506 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	50	53	69	100	1	33	4.547	7.780	0.584
2	50	76	84	100	4	89	5.573	8.575	0.650
3	50	38	54	100	3	46	5.978	8.357	0.715
4	50	36	27	100	2	57	5.627	9.264	0.607
5	50	33	46	100	4	100	6.407	9.294	0.689
6	50	25	37	100	4	55	6.685	8.914	0.750
7	50	30	46	100	2	35	5.809	8.244	0.705
8	50	41	48	100	6	46	6.595	8.475	0.778
Mean Relative fitness ± S.E. : 0.685 ± 0.024									

Competition 3.1: SBW25 *lacZ* (a) and SBW25_*glyQ*_A506 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	50	26	64	100	9	60	7.456	8.453	0.882
2	50	112	47	100	22	22	6.890	7.758	0.888
3	50	23	41	100	4	32	6.768	8.269	0.818
4	50	21	28	100	10	34	7.775	8.711	0.893
5	50	41	86	100	12	39	7.289	7.726	0.943
6	50	50	53	100	6	33	6.397	8.043	0.795
7	50	24	36	100	12	39	7.824	8.597	0.910
8	50	49	69	100	11	21	7.023	7.328	0.958
Mean Relative fitness ± S.E. : 0.886 ± 0.02									

Competition 3.2: SBW25 *lacZ* (a) and SBW25_*glyQ*_A506 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	50	26	64	100	2	111	5.952	9.068	0.656
2	50	112	47	100	9	86	5.996	9.121	0.657
3	50	23	41	100	9	83	7.579	9.222	0.822
4	50	21	28	100	8	60	7.552	9.279	0.814
5	50	41	86	100	3	109	5.902	8.754	0.674
6	50	50	53	100	6	87	6.397	9.013	0.710
7	50	24	36	100	2	6	6.032	6.725	0.897
8	50	49	69	100	8	72	6.705	8.560	0.783
Mean Relative fitness ± S.E. : 0.752 ± 0.033									

Competition 4.1: SBW25 *lacZ* (a) and Pf0-1_*glyQ*_A506 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	50	28	45	100	16	20	7.958	7.706	1.033
2	50	46	62	100	24	16	7.867	7.163	1.098
3	50	34	76	100	20	24	7.987	7.365	1.084
4	50	70	58	100	9	36	6.466	8.040	0.804
5	50	81	77	100	12	22	6.608	7.264	0.910
6	50	275	93	100	16	59	5.673	8.062	0.704
7	50	61	51	100	12	28	6.891	7.918	0.870
8	50	64	84	100	14	16	6.997	6.859	1.020
Mean Relative fitness ± S.E. : 0.94 ± 0.05									

Competition 4.2: SBW25 *lacZ* (a) and Pf0-1_*glyQ*_A506 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	50	28	45	100	11	109	5.280	7.099	0.744
2	50	46	62	100	7	29	4.332	5.455	0.794
3	50	34	76	100	9	59	4.885	5.961	0.820
4	50	70	58	100	9	78	4.163	6.511	0.639
5	50	81	77	100	8	48	3.900	5.742	0.679
6	50	275	93	100	4	82	1.984	6.089	0.326
7	50	61	51	100	3	104	3.202	6.927	0.462
8	50	64	84	100	11	75	4.454	6.101	0.730
Mean Relative fitness ± S.E. : 0.636 ± 0.068									

Tables 7.25: Colony counts, volume plated, Malthusian parameter and Relative fitness of competitors. Colony counts were obtained by plating either 50 µl or 100 µl of 10^{-5} dilution for each replicate. The tables are for competitions 1.1 to 4.2.

For Pf0-1 background

Competition 1.1: SBW25 *lacZ* (a) and Pf0-1 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	10	14	100	32	81	8.071	7.500	1.076
2	100	9	28	100	10	63	7.013	7.613	0.921
3	100	16	16	100	43	69	7.896	7.381	1.070
4	100	18	24	100	36	86	7.601	7.491	1.015
5	100	18	24	100	13	49	6.582	7.947	0.828
6	100	20	16	100	17	65	6.745	8.472	0.796
7	100	24	28	100	13	73	6.295	8.479	0.742
8	100	24	21	100	23	71	6.865	8.168	0.840
Mean Relative fitness ± S.E. : 0.911 ± 0.046									

Competition 1.2: SBW25 *lacZ* (a) and Pf0-1 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	10	14	100	1	28	4.605	9.903	0.465
2	100	9	28	100	2	35	5.404	8.635	0.626
3	100	16	16	100	Yeast contamination		n/a	n/a	n/a
4	100	18	24	100	3	55	5.116	9.529	0.537
5	100	18	24	100	2	43	4.711	9.688	0.486
6	100	20	16	100	2	71	4.605	10.700	0.430
7	100	24	28	100	3	96	4.828	10.219	0.472
8	100	24	21	100	3	102	4.828	10.568	0.457
Mean Relative fitness ± S.E. : 0.496 ± 0.023									

Competition 2.1: SBW25 *lacZ* (a) and Pf0-1_*glyQ*_Pf0-1 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	14	32	100	25	35	7.488	6.418	1.167
2	100	20	16	100	61	58	8.023	7.080	1.133
3	100	18	18	100	39	60	7.681	7.339	1.047
4	100	18	15	100	25	64	7.236	8.030	0.901
5	100	20	8	100	101	51	8.527	7.141	1.194
6	100	32	24	100	27	60	6.738	7.994	0.843
7	100	12	12	100	14	43	7.062	8.030	0.879
8	100	16	30	100	6	44	5.927	8.272	0.717
Mean Relative fitness ± S.E. : 0.985 ± 0.062									

Competition 2.2: SBW25 *lacZ* (a) and Pf0-1_*glyQ*_Pf0-1 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	14	32	100	5	51	5.878	8.403	0.699
2	100	20	16	100	7	66	5.858	9.375	0.625
3	100	18	18	100	4	88	5.404	9.999	0.540
4	100	18	15	100	3	62	5.116	10.119	0.506
5	100	20	8	100	1	118	3.912	12.595	0.311
6	100	32	24	100	4	43	4.828	9.570	0.505
7	100	12	12	100	16	200	7.195	9.433	0.763
8	100	16	30	100	2	111	4.828	10.296	0.469
Mean Relative fitness ± S.E. : 0.552 ± 0.05									

Competition 3.1: SBW25 *lacZ* (a) and SBW25_*glyQ*_Pf0-1 (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	12	18	100	15	37	7.131	7.628	0.935
2	100	18	12	100	11	45	6.415	8.230	0.780
3	100	24	18	100	18	51	6.620	7.949	0.833
4	100	33	9	100	36	56	6.995	8.736	0.801
5	100	28	8	100	20	29	6.571	8.196	0.802
6	100	20	32	100	26	23	7.170	6.578	1.090
7	100	18	27	100	17	29	6.851	6.979	0.982
8	100	18	27	100	28	46	7.350	7.441	0.988
Mean Relative fitness ± S.E. : 0.901 ± 0.04									

Competition 3.2: SBW25 *lacZ* (a) and SBW25_*glyQ*_Pf0-1 (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	100	12	18	50	1	71	5.116	8.973	0.570
2	100	18	12	50	3	90	5.809	9.616	0.604
3	100	24	18	50	1	49	4.423	8.602	0.514
4	100	33	9	50	3	64	5.203	9.563	0.544
5	100	28	8	50	1	38	4.269	9.159	0.466
6	100	20	32	50	4	78	5.991	8.492	0.706
7	100	18	27	50	2	67	5.404	8.510	0.635
Mean Relative fitness ± S.E. : 0.577 ± 0.03									

Competition 4.1: SBW25 *lacZ* (a) and A506_*glyQ*_Pf0-1 (1) (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	150	44	106	100	16	16	6.302	5.422	1.162
2	150	12	51	100	15	16	7.536	6.154	1.225
3	150	13	34	100	21	10	7.793	6.089	1.280
4	150	3	28	100	18	21	9.105	7.026	1.296
5	150	10	60	100	18	18	7.901	6.109	1.293
6	150	17	59	100	24	16	7.658	6.008	1.275
Mean Relative fitness ± S.E. : 1.26 ± 0.021									

Competition 4.2: SBW25 *lacZ* (a) and A506_*glyQ*_Pf0-1 (1) (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	150	44	106	100	10	106	5.832	8.389	0.695
2	150	12	51	100	19	139	7.773	7.451	1.043
3	150	30	85	100	5	115	5.521	9.002	0.613
4	150	90	131	100	18	133	5.704	8.532	0.668
5	150	13	34	100	21	106	7.793	7.565	1.030
6	150	3	28	100	10	148	8.517	7.369	1.156
7	150	10	60	100	9	128	7.208	7.771	0.928
8	150	17	59	100	18	139	7.370	7.708	0.956
Mean Relative fitness ± S.E. : 0.886 ± 0.071									

Competition 5.1: SBW25 *lacZ* (a) and A506_*glyQ*_Pf0-1 (2) (b), 24 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	150	12	22	100	5	4	6.438	5.608	1.148
2	150	13	25	100	3	2	5.847	4.787	1.221
3	150	76	72	100	5	4	4.592	4.423	1.038
4	150	31	29	100	9	6	6.076	5.738	1.059
5	150	12	26	100	9	10	7.026	6.358	1.105
6	150	26	38	100	18	10	6.945	5.978	1.162
Mean Relative fitness ± S.E. : 1.12 ± 0.028									

Competition 5.2: SBW25 *lacZ* (a) and A506_*glyQ*_Pf0-1 (2) (b), 48 hours competition

Reps.	Counts at 0 hour			Counts at 24 hours			Malthusian parameter		Relative fitness
	Vol.	a	b	Vol.	a	b	a	b	
1	150	6	11	100	12	51	8.006	7.749	1.033
2	150	7	11	100	35	55	8.923	6.908	1.292
3	150	13	25	100	8	51	6.828	8.106	0.842
4	150	76	72	100	5	86	4.592	9.807	0.468
5	150	31	29	100	9	76	6.076	9.108	0.667
6	150	6	13	100	9	57	7.719	7.980	0.967
7	150	8	19	100	9	214	7.431	9.212	0.807
8	150	26	38	100	15	94	6.763	8.364	0.809
Mean Relative fitness ± S.E. : 0.861 ± 0.087									

Tables 7.26: Colony counts, volume plated, Malthusian parameter and Relative fitness of competitors. Colony counts were obtained by plating either 100 µl or 150 µl of 10^{-6} dilution for each replicate. The tables are for competitions 1.1 to 5.2

Appendix 2.3 Raw reads and statistical tests for relative *glyQ* mRNA expression

Data for relative *glyQ* mRNA expression for SBW25, A506 and Pf0-1 background strains. See section 4.3.4 for explanation.

Reps.	*glyQ* quantity	*recA* quantity	*rpoD* quantity	Mean *recA*, *rpoD*	Rel. quantity*	Fold change#
SBW25						
1.1	166.55	134.51	209.36	171.94	0.971	1.063
1.2	233.28	230.29	280.01	255.154	0.917	0.940
1.3	187.24	161.39	227.20	194.30	0.964	0.983
2.1	156.24	188.81	234.45	211.63	0.743	0.813
2.2	147.87	186.59	225.35	205.97	0.718	0.736
2.3	144.57	203.18	274.34	238.76	0.606	0.617
3.1	177.51	171.56	223.96	197.76	0.899	0.984
3.2	154.60	207.01	282.59	244.80	0.627	0.643
3.3	180.99	157.01	226.14	191.57	0.945	0.964
4.1	106.57	92.70	141.01	116.85	0.914	1.000
4.2	129.62	104.68	161.18	132.93	0.975	1.000
4.3	150.50	119.85	186.95	153.40	0.981	1.000
A506						
5.1	132.41	56.06	95.80	75.93	1.744	1.370
5.2	134.01	51.80	158.46	105.13	1.275	1.052
6.1	104.76	57.64	155.64	106.64	0.982	0.772
6.2	87.46	58.04	161.39	109.72	0.797	0.658
6.3	103.08	54.14	143.62	98.88	1.042	0.834
7.1	98.67	55.98	93.76	74.87	1.318	1.035
7.2	80.37	47.96	88.03	68.00	1.182	0.975
7.3	85.81	62.44	93.19	77.81	1.103	0.882
8.1	80.38	50.79	75.48	63.13	1.273	1.000
8.2	103.55	59.08	111.83	85.46	1.212	1.000
8.3	145.53	90.84	142.02	116.43	1.250	1.000
Pf0-1						
9.1	58.69	50.87	79.60	65.23	0.900	0.659
9.2	51.17	63.70	53.64	58.67	0.872	0.683
9.3	73.64	53.79	43.78	48.79	1.510	1.400
10.1	183.94	104.08	31.70	67.89	2.709	1.984
10.2	79.51	42.09	50.95	46.52	1.709	1.338

10.3	128.04	57.54	56.95	57.25	2.237	2.074
11.1	129.38	46.61	72.02	59.32	2.181	1.597
11.2	180.85	63.85	79.76	71.81	2.519	1.971
11.3	160.64	70.55	87.03	78.79	2.039	1.891
12.1	114.82	94.09	112.37	103.23	1.112	0.815
12.2	141.79	78.70	86.24	82.47	1.719	1.345
12.3	128.95	58.94	67.92	63.43	2.033	1.886
13.1	90.66	69.61	63.18	66.40	1.365	1.000
13.2	103.46	77.70	84.24	80.97	1.278	1.000
13.3	81.80	72.07	79.67	75.87	1.078	1.000

Table 7.27: Raw reads for *glyQ, recA* and *rpoD* quantities. Rel. quantity* (Relative quantity) is quantity of *glyQ* relative to mean of *recA* and *rpoD*. Fold change[#] is relative *glyQ* quantity normalised to the wild type quantities. 1= A506_*glyQ*_SBW25, 2= Pf0-1_*glyQ*_SBW25, 3= SBW25_*glyQ*_SBW25, 4= SBW25, 5=SBW25_*glyQ*_A506, 6=Pf0-1_*glyQ*_A506, 7= A506_*glyQ*_A506, 8= A506, 9= SBW25_*glyQ*_Pf0-1, 10= A506_*glyQ*_Pf0-1 (1), 11= A506_*glyQ*_Pf0-1 (2), 12= Pf0-1_*glyQ*_Pf0-1, 13= Pf0-1

Strain	Growth rate	Mean ± S.E.*	Shapiro-Wilk *P*-value
SBW25			
A506_*glyQ*_SBW25	0.971, 0.917, 0.964	0.967 ± 0.0223	0.4138
Pf0-1_*glyQ*_SBW25	0.743, 0.718, 0.606	0.260 ± 0. 0416	0.3250
SBW25_*glyQ*_SBW25	0.899, 0.627, 0.945	0.338 ± 0. 102	0.3009
SBW25	0.914, 0.975, 0.981	0.172 ± 0.0223	0.1802
A506			
SBW25_*glyQ*_A506	1.744, 1.275	1.51 ± 0.235	N too small (2 data points)
Pf0-1_*glyQ*_A506	0.982, 0.797, 1.042	0.941 ± 0.0738	0.4526
A506_*glyQ*_A506	1.318, 1.182, 1.103	1.2 ± 0.0628	0.7127
A506	1.273, 1.212, 1.250	1.24 ± 0.0179	0.7348
Pf0-1			
SBW25_*glyQ*_Pf0-1	0.900, 0.872, 1.510	1.09 ± 0.208	0.0729
A506_*glyQ*_Pf0-1 (1)	2.709, 1.709, 2.237	2.22 ± 0.289	0.9401
A506_*glyQ*_Pf0-1 (2)	2.181, 2.519, 2.039	2.25 ± 0.142	0.5595
Pf0-1_*glyQ*_Pf0-1	1.112, 1.719, 2.033	1.63 ± 0.27	0.6597
Pf0-1	1.365, 1.278, 1.078	1.24 ± 0.085	0.5771

Table 7.28: Raw data and test for normality on relative *glyQ* expression. S.E.*= standard error.

Samples	Levene test		ANOVA	
	F-stat	*P*-value	*F*-stat	*P*-value
SBW25				
A506_*glyQ*_SBW25, Pf0-1_glyQ_SBW25, SBW25_*glyQ*_SBW25, SBW25	0.749	0.553	5.426	0.0249*
A506				
SBW25_*glyQ*_A506, Pf0-1_glyQ_A506, A506_*glyQ*_A506, A506	4.806	0.04	-	0.055#
Pf0-1				
SBW25_*glyQ*_Pf0-1, A506_glyQ_Pf0-1, Part A506_glyQ_Pf0-1 (1), Part A506_*glyQ*_Pf0-1 (2), Pf0-1_*glyQ*_Pf0-1, Pf0-1	0.445	0.774	6.327	.0083*

Table 7.29: Statistics for differences in relative *glyQ* expression between mutants and corresponding wild type strains. *F*-stat (*F* statistic) and *P*-value of Levene test and ANOVA provided in the table. # This is a non-parametric form of ANOVA (Kruskal Wallis test) since the assumption for equality of variance was violated.

Appendix 3. Appendix for Chapter 5

This section provides the results of the whole genome sequencing analysis of both SBW25 and PICF7 mat forming strains and transposon mutagenesis screen done with the PICF7 mat forming strains. See sections 5.3.1.2, 5.3.4 and 5.3.5.1 for explanation.

Strain	Gene	DNA change	Protein change
SBW25			
1	*mwsR*	G3244A	E1082K
2	*mwsR*	A2476T	S826C
3	*mwsR*	T2183C	V728A
4	*wspF**	Δ151-165 (CTGATGGAC CTGATC)	Δ51-55 (LMDLI)
5	*mwsR**	G3095T	R1032L
6	*awsR*	A79C	T27P
7	*awsR*	G574A	A192T
8	*wspF*	A820G	T274A
9	*wspA*	C1313T	S438F
10	*wspE**	G1912T	D638Y
1-10	*PFLU3154* → / ← *PFLU3155*	$(C)_{5\rightarrow3}$	intergenic (+55 /+23)
1, 3-5, 7-10	*PFLU0045* → / ← *PFLU0046*	+G	intergenic (+65 /+20)
2-5, 7-10	*PFLU0872* → / → *PFLU0873*	+C	intergenic (+17 /-136)
A506			
11-20	*lapA*	C→T	E1278E
11, 13, 15-16, 18	*lapA*	C→A	P1261P
11-13, 15-16, 18, 20	*lapA*	C→T	V1256V
11-15, 19	*dppA2*	2 bp→CT	coding (1091-1 092/1605 nt)
11-12,	*dppA3*	G→C	N414K

20			
13, 15, 19	*dppA2*	2 bp→GT	coding (1080-1081/1605 nt)
13, 15, 19	*dppA2*	G→C	Q360E
11- 16, 20	*dppA3*	C→G	L411L
11-12, 14 - 16, 20	*dppA3*	G→A	G410G
11-12, 14-16, 20	*dppA3*	4 bp→TGGC	coding (1219-1222/1629 nt)
11, 13-14, 18-19	*PflA506_1705* → / → *rfbA_2*	$(C)_{10\rightarrow11}$	intergenic (+334/-80)
11-12	*PflA506_2350* → / ← *ntaA_1*	T→C	intergenic (+370/+295)
11-12	*PflA506_2350* → / ← *ntaA_1*	T→C	intergenic (+375/+290)
13, 16, 19	*PflA506_3053* ← / ← *glsA*	C→T	intergenic (-393/+902)
13	*PflA506_0050*	C→T	A64A
13	*dppA3*	T→C	K407E
15	*dppA2*	2 bp→GG	coding (1067-1068/1605 nt)
15	*dppA2*	C→A	P355P
PICF7			
21	*PFLUOLIPICF7_02270* (*wspA*)	C161T	S54L
22	*PFLUOLIPICF7_02245* (*wspF*)	C250T	R82(STOP+)
23	*PFLUOLIPICF7_02245* (*wspF*)	C747G	I248(STOP+)
24	*PFLUOLIPICF7_12710* (*awsX*)	Δ232-264	Δ78-88
25	**NGS fail**	**NGS fail**	**NGS fail**
26	*PFLUOLIPICF7_12720* (*awsO*)	T127A	F43I
27	*PFLUOLIPICF7_12195* (*mwsR*)	G2976T	M992I
28	*PFLUOLIPICF7_02245* (*wspF*)	A880C	T294P
29	*PFLUOLIPICF7_02245* (*wspF*)	Δ949-951(GCT)	ΔA317

Strain	Gene	DNA change	Protein change
30	*PFLUOLIPICF7_02245* (*wspF*)	C898T	Q300(STOP#)
21-30	*PFLUOLIPICF7_00605* ← / ← *PFLUOLIPICF7_00615*	3 bp→50 bp	intergenic (-2458/+478)
21-30	*PFLUOLIPICF7_01230*	+51 bp	coding (208/1113 nt)
21-30	*PFLUOLIPICF7_01250*	6 bp→57 bp	coding (515-520/717 nt)
21-30	*PFLUOLIPICF7_02370* → / → *PFLUOLIPICF7_02380*	C→G	intergenic (+379/-562)
21-30	*PFLUOLIPICF7_02370* → / → *PFLUOLIPICF7_02380*	G→T	intergenic (+384/-557)
21-30	*PFLUOLIPICF7_02370* → / → *PFLUOLIPICF7_02380*	G→T	intergenic (+390/-551)
21-30	*PFLUOLIPICF7_02370* → / → *PFLUOLIPICF7_02380*	2 bp→52 bp	intergenic (+395/-545)
21-30	*PFLUOLIPICF7_03040*	G→C	A208G
21-30	*PFLUOLIPICF7_03040*	C→A	R207M
21-30	*algL* → / → *PFLUOLIPICF7_03430*	Δ88 bp	intergenic (+19/-121)
21-30	*PFLUOLIPICF7_04830* → / ← *PFLUOLIPICF7_04835*	Δ26 bp	intergenic (+299/+443)
21-30	*PFLUOLIPICF7_05150*	G→C	E23Q
21-30	*PFLUOLIPICF7_05150*	G→C	G24G
21-30	*PFLUOLIPICF7_05150*	G→C	G25A
21-30	*PFLUOLIPICF7_05150*	C→T	P26L
21-30	*PFLUOLIPICF7_05480* ← / ← *PFLUOLIPICF7_05485*	Δ40 bp	intergenic (-808/+836)
21-30	*PFLUOLIPICF7_05495* ← / ← *PFLUOLIPICF7_05500*	Δ28 bp	intergenic (-621/+344)
21-30	*PFLUOLIPICF7_07025* → / → *PFLUOLIPICF7_07030*	5 bp→45 bp	intergenic (+339/-110)
21-30	*PFLUOLIPICF7_07660* ← / ← *PFLUOLIPICF7_07665*	C→T	intergenic (-352/+221)
21-30	*trpA* → / ← *PFLUOLIPICF7_08075*	+53 bp	intergenic (+449/+141)

Strain	Gene	DNA change	Protein change
21-30	*PFLUOLIPICF7_08520* → / ← *PFLUOLIPICF7_08525*	16 bp→66 bp	intergenic (+762/+174)
21-30	*PFLUOLIPICF7_08790*	G→T	P38T
21-30	*PFLUOLIPICF7_08790*	G→C	P37A
21-30	*PFLUOLIPICF7_08790*	+48 bp	coding (94/4347 nt)
21-30	*PFLUOLIPICF7_09250*	3 bp→48 bp	coding (1101-1103/1350 nt)
21-30	*PFLUOLIPICF7_09320*	G→T	G461V
21-30	*PFLUOLIPICF7_09680*	C→A	P220T
21-30	*PFLUOLIPICF7_11175*	A→G	noncoding (114/116 nt)
21-30	*PFLUOLIPICF7_11985*	G→A	R90C
21-30	*PFLUOLIPICF7_11985*	G→C	L88L
21-30	*PFLUOLIPICF7_13470*	Δ30 bp	coding (442-471/2541 nt)
21-30	*PFLUOLIPICF7_13810* → / ← *PFLUOLIPICF7_13820*	C→G	intergenic (+221/+600)
21-30	*PFLUOLIPICF7_13810* → / ← *PFLUOLIPICF7_13820*	G→C	intergenic (+226/+595)
21-30	*PFLUOLIPICF7_13810* → / ← *PFLUOLIPICF7_13820*	G→C	intergenic (+229/+592)
21-30	*PFLUOLIPICF7_13810* → / ← *PFLUOLIPICF7_13820*	4 bp→42 bp	intergenic (+232/+586)
21-30	*PFLUOLIPICF7_15060*	$(AGCTGCTGCGGC)_{1\rightarrow2}$	coding (3100/3156 nt)
21-30	*PFLUOLIPICF7_15285*	G→A	G31E
21-30	*PFLUOLIPICF7_15285*	C→T	R33C
21-30	*PFLUOLIPICF7_15285*	+39 bp	coding (103/216 nt)
21-30	*PFLUOLIPICF7_16175*	C→T	L202L
21-30	*PFLUOLIPICF7_17640*	7 bp→51 bp	coding (214-220/237 nt)
21-30	*PFLUOLIPICF7_18065* → / → *PFLUOLIPICF7_18070*	6 bp→53 bp	intergenic (+558/-867)
21-30	*PFLUOLIPICF7_18575* → / ← *PFLUOLIPICF7_18585*	$(14\text{-bp})_{2\rightarrow1}$	intergenic (+71

			4/+356)
21-30	*PFLUOLIPICF7_21170* ← / → *PFLUOLIPICF7_21175*	+41 bp	intergenic (-394/-1328)
21-30	*PFLUOLIPICF7_21555*	10 bp→64 bp	coding (133-142/714 nt)
21-30	*PFLUOLIPICF7_21745* → / ← *PFLUOLIPICF7_21750*	C→T	intergenic (+655/+337)
21-30	*PFLUOLIPICF7_21935* → / → *PFLUOLIPICF7_21940*	G→T	intergenic (+230/-1159)
21-30	*PFLUOLIPICF7_21935* → / → *PFLUOLIPICF7_21940*	+43 bp	intergenic (+233/-1156)
21-30	*PFLUOLIPICF7_21980*	$(AGACTGCCTGGC)_{2\rightarrow1}$	coding (639-650/825 nt)
21-30	*PFLUOLIPICF7_22140*	Δ30 bp	coding (872-901/942 nt)
21-30	*PFLUOLIPICF7_23105* → / ← *PFLUOLIPICF7_23110*	+53 bp	intergenic (+551/+958)
21-30	*PFLUOLIPICF7_23545*	G→T	D80E
21-30	*PFLUOLIPICF7_23545*	C→A	R78M
21-30	*PFLUOLIPICF7_23755*	G→C	P218A
21-30	*PFLUOLIPICF7_24815*	G→A	D39N
21-30	*PFLUOLIPICF7_25790*	C→A	G96W
21-30	*PFLUOLIPICF7_25790*	C→T	G95E
21-30	*PFLUOLIPICF7_25790*	C→A	G94V
21-30	*PFLUOLIPICF7_26075* → / ← *PFLUOLIPICF7_26085*	17 bp→61 bp	intergenic (+3701/+9077)
21-30	*PFLUOLIPICF7_26285* ← / → *PFLUOLIPICF7_26290*	+79 bp	intergenic (-13/-17)
21-30	*PFLUOLIPICF7_26920* → / ← *PFLUOLIPICF7_26925*	+54 bp	intergenic (+10/+6)
21-30	*PFLUOLIPICF7_27025* → / → *PFLUOLIPICF7_27030*	T→G	intergenic (+25/-722)
21-30	*PFLUOLIPICF7_27580*	G→T	G32V
21-30	*PFLUOLIPICF7_27580*	G→C	G33G
21-30	*PFLUOLIPICF7_27580*	C→T	P34L
21-30	*PFLUOLIPICF7_27760*	Δ1 bp	noncoding (76/76 nt)
21-30	*PFLUOLIPICF7_27770*	+T	noncoding (76/

			76 nt)
21	*PFLUOLIPICF7_09320*	G→T	G462W
22	*PFLUOLIPICF7_21745* → / ← *PFLUOLIPICF7_21750*	$(A)_{6\rightarrow5}$	intergenic (+882/+110)
22	*PFLUOLIPICF7_21935* → / → *PFLUOLIPICF7_21940*	+43 bp	intergenic (+236/-1153)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	1 bp→TA	intergenic (+1/-242)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	+A	intergenic (+3/-240)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	4 bp→C	intergenic (+13/-227)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	T→G	intergenic (+18/-225)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	Δ1 bp	intergenic (+21/-222)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	A→G	intergenic (+23/-220)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	C→T	intergenic (+26/-217)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	2 bp→AT	intergenic (+37/-205)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	Δ7 bp	intergenic (+45/-192)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	Δ1 bp	intergenic (+53/-190)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	Δ1 bp	intergenic (+56/-187)
23	*PFLUOLIPICF7_12680* → / → *PFLUOLIPICF7_12685*	G→A	intergenic (+62/-181)
23	*PFLUOLIPICF7_21230*	C→T	R42H
23	*PFLUOLIPICF7_25600*	8 bp→92 bp	coding (18-25/1137 nt)
23	*PFLUOLIPICF7_27025*	+66 bp	coding (137/198 nt)
24	*PFLUOLIPICF7_10055* ← / → *PFLUOLIPICF7_10060*	+115 bp	intergenic (-191/-473)
26	*PFLUOLIPICF7_08790*	+48 bp	coding (90/4347 nt)

Strain	Gene	DNA change	Protein change
30	*PFLUOLIPICF7_17040*	+68 bp	coding (234/324 nt)

Table 7.30: Genes identified to have mutations in SBW25, A506 and PICF7 mat forming strains after whole genome sequencing analysis. Strains 1-10 are of SBW25, strain 11-20 are of A506 and sStrains 21-30 are of PICF7.

Mutant	PFLUOLIPICF7[1]	Gene[2]	Predicted protein product	Insertion[3]
PICF7 mat forming strain with mutation in *wspF* (strain 23)				
P23_100_1	09900	-	Dihydroxy-acid dehydratase	2168744
P23_100_2	02250	*wspE*	Hybrid histidine kinase response regulator	482359
P23_100_3	06820	*pelA*	PbsX family transcriptional regulator	1498807
P23_100_4	02240	*wspR*	Diguanylate cyclase response regulator	478469
P23_100_5	02240	*wspR*	Diguanylate cyclase response regulator	478365
P23_100_6	09900	-	Dihydroxy-acid dehydratase	2168744
P23_100_7	11615	-	Hypothetical protein	2515740
P23_100_8	06790	*pelG*	Putative polysaccharide transporter	1489071
P23_100_9	06790	*pelG*	Putative polysaccharide transporter	1489279
P23_100_11	02260	*wspC*	Putative chemotaxis-related methyltransferase	483829
P23_100_12	00770	-	Hypothetical protein	168524
P23_100_13	02240	*wspR*	Diguanylate cyclase response regulator	478469
P23_100_14	02240	*wspR*	Diguanylate cyclase response regulator	478469
P23_100_15	02250	*wspE*	Hybrid histidine kinase response regulator	481839
P23_100_16	01380	-	Serine hydrolase	292354
P23_100_17	06805	*pelD*	Putative polysaccharide auxiliary transport protein	1492100
P23_100_18	05490	-	Phosphate starvation inducible protein	1186461
P23_100_19	06820	*pelA*	PbsX family transcriptional regulator	1498390

Mutant	PFLUOLIPICF7[1]	Gene[2]	Predicted protein product	Insertion[3]
P23_100_20	03540	-	Acriflavin resistance protein	768932
P23_100_21	06820	*pelA*	PbsX family transcriptional regulator	1498529
P23_100_22	00715	-	Hypothetical protein	159301
P23_150_2	06790	*pelG*	Putative polysaccharide transporter	1489071
P23_150_4	26410	-	TonB-dependent siderophore receptor	5711989
P23_150_6	01080	-	Sigma factor negative regulatory protein	229172
P23_150_7	00105	-	Ethanolamine permease	26663
P23_150_8	02250	*wspE*	Hybrid histidine kinase response regulator	481121
P23_150_9	02250	*wspE*	Hybrid histidine kinase response regulator	481098
P23_150_10	06235	-	Glutamate synthase subunit beta	1363316
P23_150_11	RS28700	-	Hypothetical protein	5021663
P23_150_12	06795	-	Glycosyl transferase family 1	1490140
P23_150_13	02270	*wspA*	Methyl-accepting chemotaxis protein	485487
P23_150_14	21560	-	Phosphotransferase family protein	4625815
P23_150_15	06820	*pelA*	PbsX family transcriptional regulator	1499650
P23_150_18	06805	*pelD*	Putative polysaccharide auxiliary transport protein	1491957
P23_200_1	10515	-	AraC family trasncriptional regulator	2296422
P23_200_2	02270	*wspA*	Methyl-accepting chemotaxis protein	486518
P23_200_3	04240	-	PKHD-type hydroxylase	914179
P23_200_5	02270	*wspA*	Methyl-accepting chemotaxis protein	486400
P23_200_6	02260	*wspC*	Putative chemotaxis-related methyltransferase	483387
P23_200_7	02270	*wspA*	Methyl-accepting chemotaxis protein	486400
P23_200_8	05195	-	Histidine kinase	1112775
P23_200_9	24695	-	ABC transporter ATP-binding protein	5302028
P23_200_10	14095	-	Crossover junction endodeoxyribonuclease RuvC	3061902
P23_200_11	02860	-	Hypothetical protein	615793
P23_200_12	06820	*pelA*	PbsX family transcriptional regulator	1499330
P23_200_13	09600	-	Porin	2106837
P23_200_14	01155	-	Autotranspoter beta domain	248422
P23_200_15	IG	-	Ribonuclease III	659994
P23_200_17	IG	-	Citrate synthase	3336514

Mutant	PFLUOLIPICF7[1]	Gene[2]	Predicted protein product	Insertion[3]
P23_200_18	02270	*wspA*	Methyl-accepting chemotaxis protein	486400
P23_200_19	06820	*pelA*	PbsX family transcriptional regulator	1498749
PICF7 mat forming strain with mutation in *awsX* (strain 24)				
P24_1	12710	*awsX*	Hypothetical protein	2759482
P24_2	06795	-	Glycosyl transferase family 1	1489996
P24_3	12715	*awsR*	Diguanylate cyclase	2760666
P24_4	06815	*pelB*	Biofilm formation protein	1494846
P24_5	12710	*awsX*	Hypothetical protein	2759319
P24_6	01150	-	Hypothetical protein	243842
PICF7 mat forming strain with mutation in *mwsR* (strain 27)				
P27_1	01795	-	ABC transporter ATP-binding protein	382631
P27_2	09605	-	Peroxidase	2108320
P27_3	18235	-	TonB-dependent siderophore receptor	3930086
P27_4	00040	-	Type IV secretion protein Rhs	11511
P27_5	IG	-		4923848
P27_6	03710	-	16S rRNA methyltransferase	812909
P27_7	12195	*mwsR*	GGDEF and EAL domain containing protein	2641199
P27_8	22730	-	Cytochrome oxidase	4867337
P27_9	00485	-	PilZ domain containing protein	106310
P27_10	05435	-	Urea ABC transporter permease	1175676
P27_11	27460	-	3-isopropylmalate dehydratase large subunit	5934618
P27_12	18945	-	Membrane protein	4073137
P27_13	27735	-	Fe-S binding protein	5991783
P27_14	06685	-	MFS transporter	1466583
P27_15	12195	*mwsR*	GGDEF and EAL domain containing protein	2639301
P27_16	04100	-	Acyl-CoA dehydrogenase	881644
PICF7 mat forming strain with mutation in *wspF* (strain 28)				
P28_100_2	00080	-	Sensor histidine kinase	17921
P28_100_3	02270	*wspA*	Methyl-accepting chemotaxis protein	486184
P28_100_6	24340	*pnuC*	nicotinamide mononucleotide transporter	5212540
P28_100_7	02270	*wspA*	Methyl-accepting chemotaxis protein	486391
P28_100_8	27905	-	Pirin family protein	6023700

Mutant	PFLUOLIPICF7[1]	Gene[2]	Predicted protein product	Insertion[3]
P28_100_9	24960	*wcaF*	TonB-dependent siderophore receptor	5363072
P28_100_10	02240	*wspR*	Diguanylate cyclase response regulator	478248
P28_100_11	RS07105	-	ABC transporter	1570054
P28_100_12	02250	*wspE*	Hybrid histidine kinase response regulator	482241
P28_150_1	14960	*flgM*	flagellar biosynthesis anti-sigma factor	3236871
P28_150_2	02250	*wspE*	Hybrid histidine kinase response regulator	482149
P28_200_1	02260	*wspC*	chemotaxis protein	483211
P28_200_2	02270	*wspA*	Methyl-accepting chemotaxis protein	486107
P28_200_3	02270	*wspA*	Methyl-accepting chemotaxis protein	485727
P28_200_4	02270	*wspA*	Methyl-accepting chemotaxis protein	485057
P28_200_5	02240	*wspR*	Diguanylate cyclase response regulator	478807
P28_200_6	06480	-	Imidazolonepropionase	1420307
P28_200_9	02270	*wspA*	Methyl-accepting chemotaxis protein	485569
P28_200_10	05045	-	Adenosine deaminase	1082912
P28_200_11	IG	-	Hypothetical protein	1514700
P28_200_12	06815	*pelB*	Biofilm formation protein	1495324

Table 7.31: Point of insertion of transposon in the mutants derived from the transposon suppressor screen. PFLUOLIPICF7[1] number, the designated number of the gene in PICF7 genome; IG = intergenic. Gene[2], gene name assigned on the basis of BLASTP; dash = unnamed gene. Insertion[3], precise insertion point in the 5'➔3' direction in the PICF7 genome.

Author contributions

Thesis title: Functional importance of synonymous mutation in the bacterial experimental system *Pseudomonas fluorescens*

Chapter 3 – Bioinformatics characterization of codon use in *P. fluorescens:*
JG conceived the study. **AM** performed the codon usage analyses with input from CW.

Chapter 4 – Altering codon usage in the *glyQ* gene of *P. fluorescens*:
JG conceived the study. FB identified *glyQ* as a candidate gene for the study (containing only synonymous mutations across all strains). **AM**, JG and ZK designed the protocol for making gene replacements in A506 and Pf0-1. **AM** performed the allelic replacements of mutant *glyQ* alleles in SBW25, A506 and Pf0-1. **AM** collected and analysed growth and fitness data. **AM** designed, performed and analysed the qRT-PCR experiments. **AM** interpreted the results of the study, with input from JG.

Chapter 5 – Analysis of routes to mat formation in SBW25, A506, PICF7 and Pf0-1:
AM and JG conceived the study. **AM** performed all lab work, including the evolution experiment in static microcosms, whole genome sequencing, transposon mutagenesis, genetic engineering, and phenotypic assays. **AM** analyzed all results, including NGS and statistical analyses. **AM** and JG interpreted the results.

Manuscripts for publication:
AM and JG are currently writing a manuscript that incorporates the results in Chapter 5.

Author names in alphabetical order
AM: Anuradha Mukherjee
CW: Christian Wöhle (from Prof. Tal Dagan's lab at CAU Kiel)
FB: Frederic Bertels
JG: Jenna Gallie
ZK: Zahra Khomarbaghi

Plön, 28.11.2019

_____ _____

Anuradha Mukherjee Jenna Gallie

Acknowledgements

Firstly, I would like to thank my supervisor Dr. Jenna Gallie for her guidance and for always encouraging me, pushing me to do my best and for the ideas and discussions. Thank you for giving me the opportunity to pursue some fascinating projects under your supervision. I am also grateful to Dr. David Rogers for his support, patience, ideas and mentorship. I thank him for always answering my barrage of questions, for his feedback on my work and ideas and for his constant encouragement. I thank my Thesis Advisory Committee for their ideas, input and insightful comments on my project.

The one person who has been with me through thick and thin at this institute is Devika Bhave. Thank you Devika for always having the time to discuss about science and buoying me during the times that I was feeling low. Always up for a tea and chat about science, life, politics and most importantly food! You have been a pillar of support and I could not have sailed smoothly without you. The success of every experiment in the lab would not have been possible without the backbone of our lab, Gunda Dechow-Seligmann. Thank you Gunda for all the technical help and assistance with experiments! Thank you Gallie Lab for all the help.

Without friends it is always difficult to cope with challenges, be at work or outside. I thank all of the friends that I have made at the institute. Thank you Cecilia, Hyejin and Alejandro for being wonderful friends. I thank Hanna, Maria, Johana, Gillian, Juan, Malavi, Gökçe and Luka. I thank Dr. Michael Sieber and Nico Fuhrmann for helping me with the German translation of the abstract of this thesis. I would also like to thank Ellen McConnell, Michael Schwarz and Dr. Sven Künzel for their help with experiments, documentation and whole genome sequencing. I thank the Sequencing team, Derk Wachsmuth and the IT department, Iben Martinsen and the Library team and the Administration team for their support and making both work and life smooth in Plön. I thank the IMPRS for providing me with the opportunity and the financial assistance to work at this fine institute

Finally, I would like to thank my family and friends from back home, who also played a big role in my success here. It would not have been possible for me to make it this far if not for the long discussions and plans to science with Gokul, Ilina, Farhan, Saroj and Som. I would like to especially thank my partner, Gokul, for his constant support in work and life and his understanding. None of this would have been possible without the perpetual and unconditional love and encouragement of my parents and my sister. They have always believed in me and encouraged me to do my best. Thank you Ma, Baba and Bulu!

Affidavit

I hereby declare that:

i. Apart from my supervisor's guidance, the content and design of this thesis is the product of my own work. The co-authors contributions are listed in the dedicated section.

ii. This thesis has not been already submitted either partially or wholly as part of doctoral degree to another examination body, and no other materials are published or submitted for publication than indicated in the thesis.

iii. The preparation of the thesis has been subjected to the Rules of Good Scientific Practice of the German Research Foundation.

iv. Prior to this thesis, I have not attempted and failed to obtain a doctoral degree.

Plön, 28.11.2019

_____

Anuradha Mukherjee