

The function and origin of multi-copy tRNA genes in *Pseudomonas fluorescens*

Dissertation in fulfillment of the requirements for the degree
“Doctor rerum naturalium”
of the Faculty of Mathematics and Natural Sciences
at the Christian Albrechts University of Kiel

Submitted by

Zahra Khomarbaghi

Microbial Evolutionary Dynamics Group
Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology

Kiel, December 2019

First examiner: Dr. Jenna Gallie (Max Planck Institute for Evolutionary Biology)

Second examiner: Prof. Dr. Eva Stukenbrock (Christian-Albrecht University of Kiel, Max Planck Institute for Evolutionary Biology)

Third examiner: Prof. Dr. Hinrich Schulenburg (Christian-Albrecht University of Kiel, Max Planck Institute for Evolutionary Biology)

Chairperson: Prof. Dr. Stanislav Gorb (Christian-Albrecht University of Kiel)

Date of oral examination: 07.02.2020

“natural selection merely modified, while redundancy created”

Evolution by gene duplication, Ohno, 1970

This thesis is sincerely dedicated to my beloved parents; Tahereh and Asghar, both of whom providing the opportunity for me to follow my dreams in higher educations and enjoy the chance of learning new things in every moment of the academic life, the chance that they never had been granted in their own life despite of being so talented.

Abstract

Transfer RNAs play a central role in protein synthesis. Organisms contain sets of genes that code for different tRNA isotypes, and these sets differ among species and strains. An interesting question is how and why these differences come to exist. This study is focused on answering this question by applying the *Pseudomonas fluorescens* as a model system. First, in the introduction chapter, a review of the functions and life cycle of a tRNA molecule, and the genetic organization of genes encoding tRNAs in all domains of life, is provided. Next, the tRNA gene sets characteristics in several strains of *Pseudomonas fluorescens*, with particular focus on three strains with differing degrees of evolutionary divergence (SBW25, A506 and Pf0-1), are discussed for further analyses. By comparing the tRNA gene sets of these strains, the tRNAs present in all three genomes (core tRNAs) and those present in only a subset (accessory tRNAs) are identified.

In the next section, the fitness effects of removing accessory tRNAs from individual genomes are investigated. Unexpectedly, the results indicate no significant deleterious effects of removing accessory tRNAs from A506 and Pf0-1. Considering the general tolerance of A506 and Pf0-1 to the deletion of one copy of multi-copy tRNA genes, I conclude that there is a degree of redundancy in tRNA gene sets. This redundancy provides a certain level of robustness to tRNA mutations.

I next hypothesized that there is a threshold for this redundancy. By deleting more than one copy of the redundant tRNA-Glu^{UUC} and tRNA-Gly^{GCC} genes from the SBW25 background, the threshold was passed, and phenotyping showed a significant fitness defect. This strain was then used as a founder for five independent lineages of an evolution experiment along with five independent wild type control lineages. 1% of each population was passaged daily into fresh medium, for 21 days. The fitness defect was recovered after ~210 generations. To unravel the genetic bases of this fast recovery, the genomic DNA of a single isolate from day 20 of each lineage was subjected to whole genome sequencing. Based on the sequencing analyses, genomic rearrangements, which resulted in duplications as large as 1Mb, are the main molecular mechanisms behind the rapid adaptation. Interestingly, this duplicated region includes a copy of the tRNA-Gly^{GCC} gene in every case. Thus, the duplications have increased the tRNA-Gly^{GCC} gene dosage, likely resulting in a compensatory increase in mature tRNA-Gly^{GCC}. This study has provided a real time example of gene duplication as a mechanism in the evolution of tRNA gene sets.

Zusammenfassung

Transfer-RNAs spielen eine zentrale Rolle bei der Proteinsynthese. Organismen enthalten spezifische Gruppen von Genen, die für verschiedene tRNA-Isotypen kodieren, und diese Gruppen unterscheiden sich zwischen verschiedenen Arten und Stämmen. Eine interessante Frage ist, wie und warum diese Unterschiede entstehen. Die vorliegende Studie beschäftigt sich mit der Beantwortung dieser Frage am Beispiel des Modellorganismus *Pseudomonas fluorescens*. Zunächst wird im Einführungskapitel ein Überblick über die Funktionen und den Lebenszyklus eines tRNA-Moleküls sowie über die genetische Organisation von Genen, die für tRNAs in verschiedenen Lebewesen kodieren, gegeben. Anschließend werden die Eigenschaften der tRNA-Gene in mehreren *Pseudomonas fluorescens*-Stämmen untersucht, wobei drei Stämme mit unterschiedlichem Grad an evolutionärer Divergenz (SBW25, A506 und Pf0-1) im Mittelpunkt der weiteren Analysen stehen. Durch Vergleich der tRNA-Gensätze dieser Stämme werden die in allen drei Genomen vorhandenen tRNAs (Kern-tRNAs) und diejenigen, die nur in einer Teilmenge vorhanden sind (zusätzliche tRNAs), identifiziert.

Im nächsten Abschnitt werden die Effekte der Entfernung von zusätzlichen tRNAs aus einzelnen Genomen auf die Fitness untersucht. Unerwarteterweise zeigen die Ergebnisse keine signifikanten schädlichen Auswirkungen der Entfernung von zusätzlichen tRNAs in den Stämmen A506 und Pf0-1. In Anbetracht der Toleranz von A506 und Pf0-1 gegenüber der Entfernung eines Satzes von zusätzlichen tRNA-Genen komme ich zu dem Schluss, dass die tRNA-Gensätze einen gewissen Grad an Redundanz aufweisen. Diese Redundanz bietet ein Maß an Robustheit gegenüber Mutationen in tRNA-Genen.

Als nächstes stellte ich die Hypothese auf, dass diese Redundanz nur bis zu einer bestimmten Toleranzschwelle funktioniert. Durch Löschen von mehr als einer Kopie der redundanten tRNA-Glu^{UUC}- und tRNA-Gly^{GCC}-Gene aus dem SBW25-Genom wurde diese Schwelle überschritten, und der entsprechende Phänotyp zeigte eine signifikante verringerte Fitness. Dieser Mutant wurde dann als Gründer für fünf unabhängige Populationen eines Evolutionsexperiments verwendet, zusammen mit fünf unabhängigen Wildtyp-Kontrollpopulationen. Dabei wurde 1% jeder Population jeweils 21 Tage lang täglich in frisches Medium überführt. Der Fitnessnachteil der Mutanten wurde dabei nach ca. 210

Generationen behoben. Um die genetischen Grundlagen dieser schnellen Anpassung zu entschlüsseln, wurde die genomische DNA eines einzelnen Isolats vom 20. Tag jeder Population einer vollständigen Genomsequenzierung unterzogen. Basierend auf den Sequenzierungsanalysen stellen sich genomische Reorganisationen, die zu Duplikationen von bis zu 1 Mb führten, als der wichtigste molekulare Mechanismus für die schnelle Anpassung heraus. Interessanterweise enthält diese duplizierte Region in jedem beobachteten Fall eine Kopie des tRNA-Gly^{GCC}-Gens. Somit haben die Duplikationen die Dosierung des tRNA-Gly^{GCC}-Gens erhöht, was wahrscheinlich zu einer kompensatorischen Erhöhung der reifen tRNA-Gly^{GCC} führt. Diese Studie lieferte damit ein Echtzeitbeispiel für Gen-Duplikationen als ein Mechanismus bei der Evolution von tRNA-Genpools.

Table of Contents

Abstract	v
Zusammenfassung	vi
1. Introduction	1
1.1 The role of transfer RNA (tRNA) in cells	1
1.1.1 Translation of genetic information in organisms: from DNA to proteins	1
1.1.2 The structure of a tRNA molecule	3
1.1.3 Codon-anticodon interaction rules	5
1.1.4 The cellular life cycle of a tRNA molecule	6
1.1.5 tRNA gene structure and organization in genome.....	10
1.1.6 Factors shaping tRNA gene sets in genomes	11
1.1.7 tRNA repertoires in different domains	12
1.1.8 tRNA databases.....	13
1.2 Experimental Evolution.....	13
1.3 <i>Pseudomonas fluorescens</i> : a model organism for studying evolution.....	15
1.4 Aims	16
2. Materials and Methods	17
2.1 Materials	17
2.1.1 Plasmids	18
2.1.2 Enzymes, antibiotics, and reagents.....	19
2.1.3 Media and cultures	20
2.1.4 Primers.....	21
2.1.5 Molecular biology kits.....	23
2.2 Methods.....	24
2.2.1 Genomic DNA extraction	24
2.2.2 Polymerase chain reaction (PCR)	25
2.2.3 Gel electrophoresis	26
2.2.4 Preparing chemically competent cells	26

2.2.5 Cloning and transformation techniques	26
2.2.6 Conjugation.....	28
2.2.7 Sequencing.....	30
2.2.8 Biological assays.....	31
2.2.9 Statistical analyses	33
2.2.10 Experimental evolution.....	34
3. Results	36
3.1 Transfer RNA gene sets vary across <i>P. fluorescens</i> strains.....	36
3.2 Expanding tRNA gene sets in <i>P. fluorescens</i> can be reduced by gene deletion	41
3.3 Reduction of tRNA gene sets in <i>P. fluorescens</i> leads to limited phenotypic effects	44
3.3.1 Phenotypic analysis of the A506 tRNA deletion mutants	45
3.3.2 Phenotypic analysis of the Pf0-1 tRNA gene deletion mutants	51
3.4 Passing the level of redundancy in tRNA gene families with multiple copies results in fitness defect in SBW25	56
3.5 SBW25ΔGluGlyGluGly fitness defect compensates fast in experimental evolution	62
3.6 Genome sequencing of the evolved mutants shows large duplications in all lineages	66
3.7 Concluding statement.....	70
4. Discussion	71
References	80
Appendixes	88
AFFIDAVIT.....	103
Curriculum vitae	104

1. Introduction

1.1 The role of transfer RNA (tRNA) in cells

This section provides a review of the function of tRNA molecules, the genetic organization of genes encoding tRNAs, and the life cycle of a tRNA in a cell.

1.1.1 Translation of genetic information in organisms: from DNA to proteins

The genetic material of an organism (nucleic acid) contains the information that governs all cellular processes. Nucleic acids are made up of four different building blocks called nucleotides: adenine (A), thymine (T) in DNA or uracil (U) in RNA, cytosine (C), and guanine (G). The information of cellular life is arranged in the form of genetic “codes”. Genetic codes are triplets of nucleotides that dictate the sequence of a peptide chain (Table 1.1).

Table 1.1. There are 61 universal sense code (DNA)/codons (mRNA) for coding the 20 amino acid

		2nd base →				
1st base ↓		U	C	A	G	3rd base ↓
U	UUU } Phe	UCU } Asn UCC } UCA } UCG }	UAU } Tyr UAC } UAA } Stop UAG } Stop	UGU } Cys UGC } UGA } Stop UGG } Trp	U	
	UUC } UUA } Leu UUG }				C	
					A	
					G	
C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U	
					C	
					A	
					G	
A	AUU } AUC } Ile AUA } AUG } Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U	
					C	
					A	
					G	
G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U	
					C	
					A	
					G	

Red shadings are the nonsense (stop) codons. U is used in mRNA, and is replaced with T in DNA codes.

The “codes” in DNA are first copied into the “codons” in RNA *via* a process called transcription. The transcribed copy of RNA is called messenger RNA (mRNA). This mRNA

is transferred into the ribosomes where it is decoded into peptides (Alberts, 1983). The ribosome translates the mRNA codons into proteins with the help of tRNAs, which carry amino acids to the ribosome. Proteins consist of one or more peptide chains and catalyze cellular functions. Transfer RNAs have three consecutive bases called an “anticodon”, which has a complementary sequence to its corresponding “codon”. Codon-anticodon matching is key in the decoding of the information from DNA into functional proteins. A simplified diagram of the decoding process is provided in Figure 1.1.

The basic concept of the universal triplet codes was first proposed by Crick in 1968. He proposed that, as there are four distinct bases in DNA (A, T, C, and G) and 20 proteinogenic amino acids, the most plausible code would be a triplet code (Crick, 1968). Hence, some (most) amino acids are coded for by more than one codon.

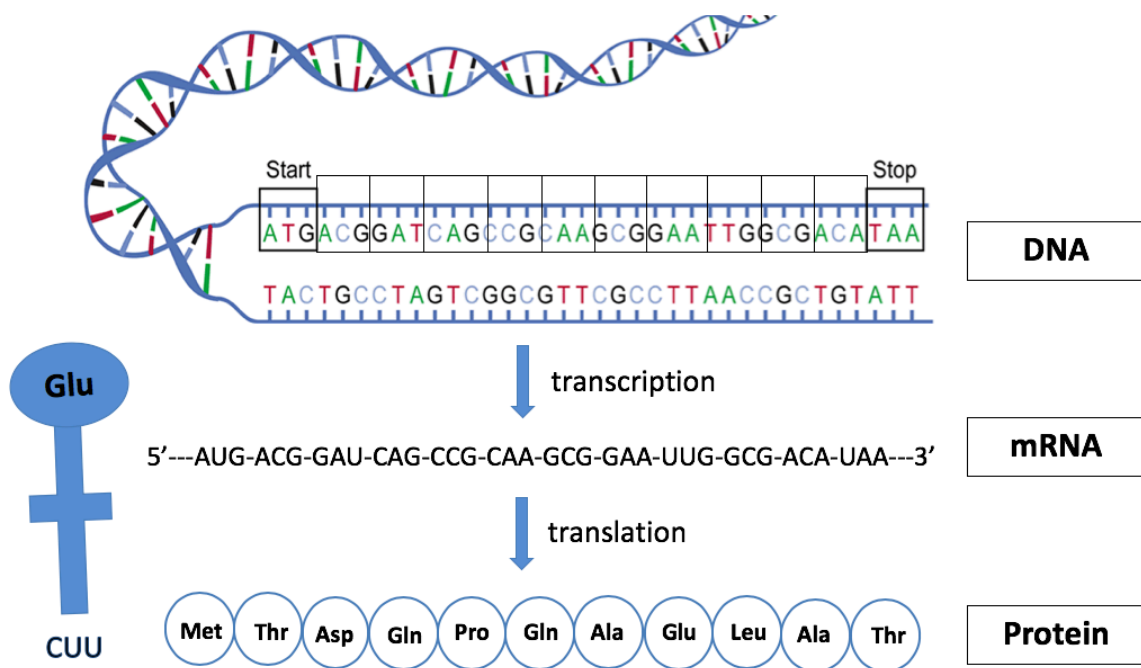


Figure 1. 1. The pathway from DNA to protein

1.1.2 The structure of a tRNA molecule

Transfer RNA (tRNA)s are small RNA molecules of 75 to 90 nucleotides. They are highly abundant in the cell during exponential growth (*e.g.* in each yeast cell, around three million tRNA molecules are functional at any given time; Phizicky and Hopper, 2010). Unlike mRNA, tRNA is not translated into protein. Instead, tRNAs fold into a defined, conserved secondary structure with a specific role; tRNAs mediate the translation of mRNA into protein. During translation, codons in mRNA are matched to the anticodon triplet located in the anticodon loop of the tRNA molecule, and the amino acid carried by the tRNA is incorporated into the nascent polypeptide chain. In 1970, Woese proposed that primitive tRNA molecules included only the anticodon stem-loop, and pairing with the putative messenger molecule would have occurred by changing the stacking interactions between the bases of this small molecule (Woese, 1970). The primary nucleotide sequence and two-dimensional structure of a modern tRNA molecule was first described by Holley (Holley *et al.*, 1965). Transfer RNAs have a distinctive cloverleaf shape that includes three arms composed of stems and loops, with each arm contributing in a specific way to tRNA function (Figure 1.2). As mentioned above, the anticodon loop is responsible for recognizing the codons inside the mRNA. The D-loop is involved in the recognition of the tRNA molecule by aminoacyl-tRNA synthetase (Smith and Yarus, 1989). The T-loop acts as a recognition site for the ribosome during translation (Rich and RajBhandary, 1976). The variable loop is not universal and its length differs among species (Pettersson, 2009). The tertiary L-shape structure is the final functional structure of the tRNA, with the anticodon loop and the amino acid acceptor positions exposed at the top of each arm (Figure 1.2).

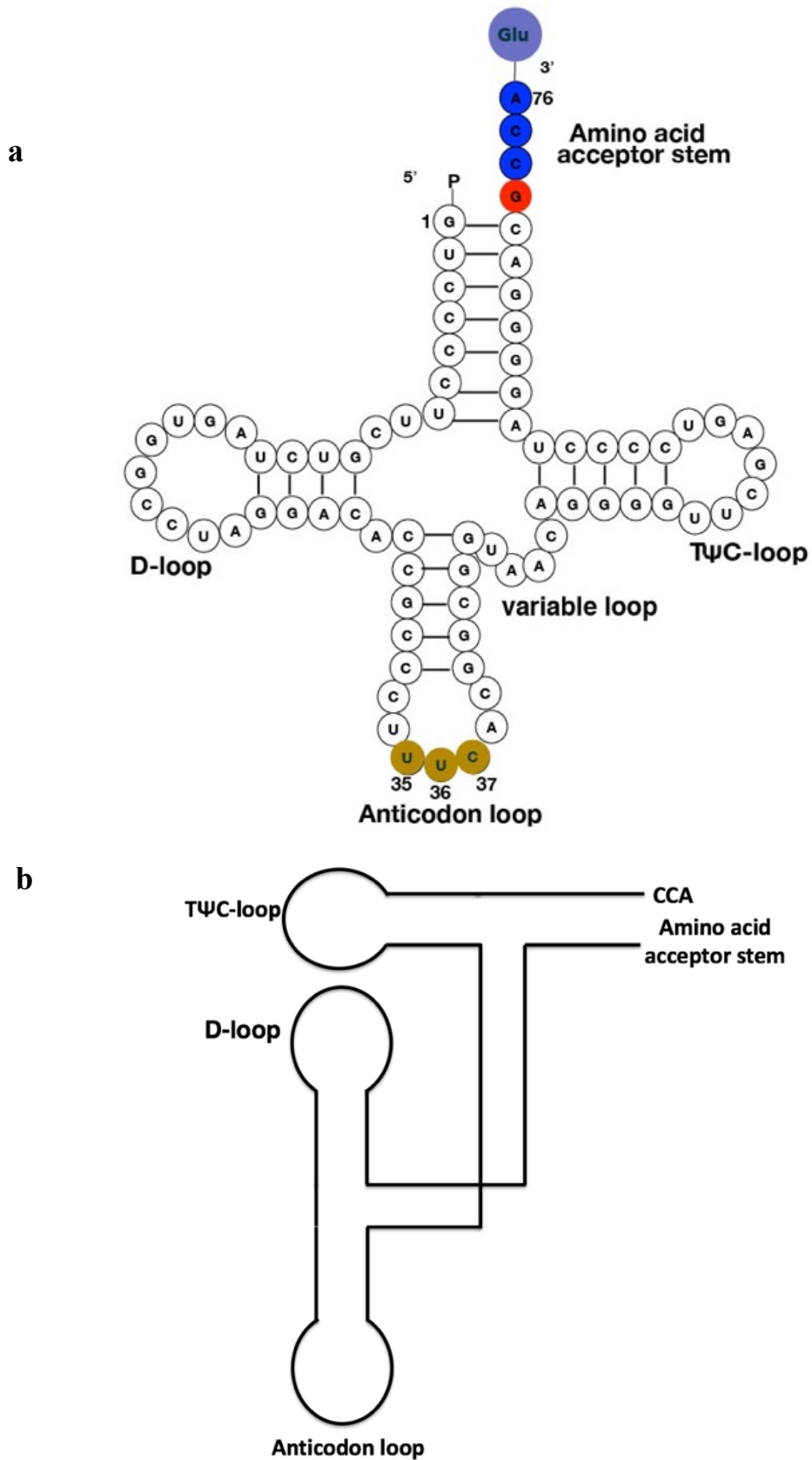


Figure 1.2. A schematic figure for glutamic acid-carrying tRNA-CUU isotype. a) Secondary structure. b) Functional 3D structure of a tRNA molecule.

1.1.3 Codon-anticodon interaction rules

During translation in the ribosome, the tRNA anticodon base pairs with the mRNA codon, bringing the cognate amino acid into the ribosome. This codon-anticodon pairing is based on Watson-Crick base pairing rules: A pairs with U, and C pairs with G. However, there are some deviations from this rule – the first base of the anticodon pairs with the third base of mRNA codon according to the “wobble pairing” rule first suggested by Francis Crick (Crick, 1966). In this position, due to conformation of the tRNA-ribosome complex, more relaxed pairing can occur (Nelson *et al.*, 2008), allowing G to pair with U (in addition to the usual C). Further, the 5' anticodon base is sometimes modified post-transcriptionally, extending the binding properties. An example of this is Inosine (I) – a modified A that can pair with A, C, or U (Murphy and Ramakrishnan, 2004). Therefore, in many cases, a single tRNA anticodon (a tRNA “isotype” or a tRNA “species”) can decode multiple mRNA codons. For example, tRNA isotype 5'-UUC-3' can theoretically decode codon 5'-GAA-3' (by Watson-Crick base pairing rules), and 5'-GAG-3' (by wobble pairing; Figure 1.3).

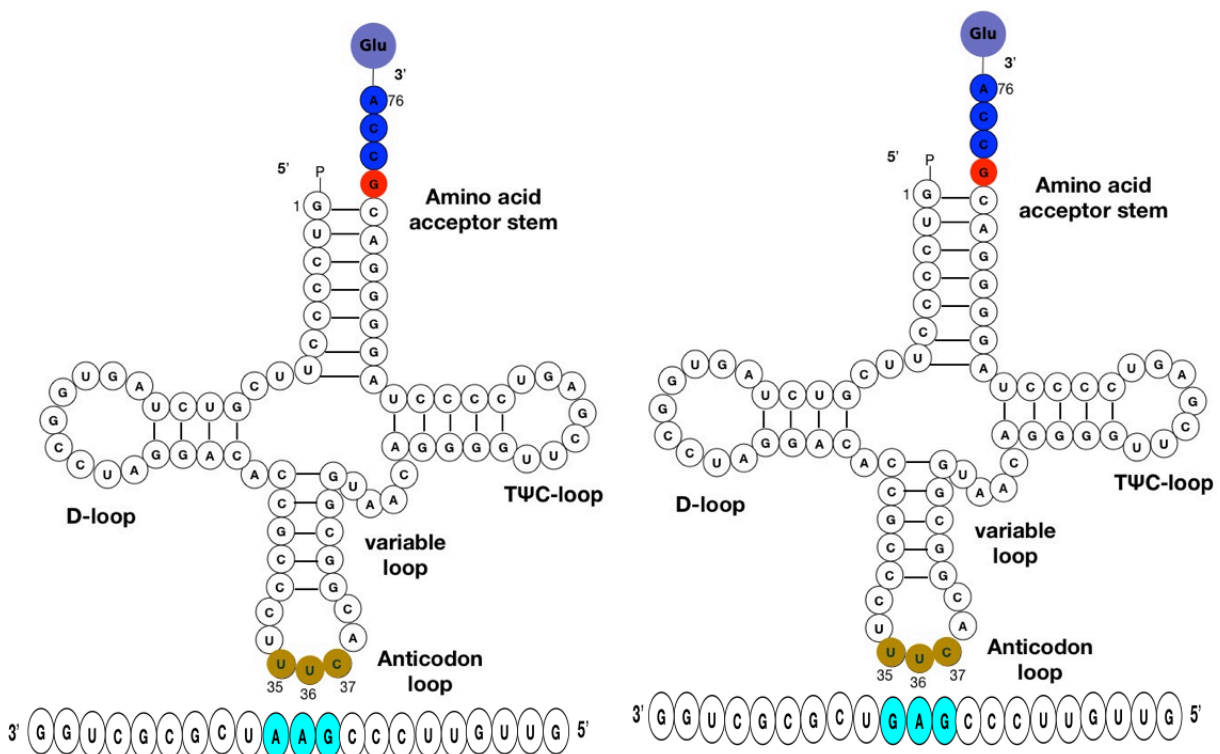


Figure 1.3. Codon-anticodon interactions are governed by Watson-Crick (left) and wobble pairing rules (right). The turquoise bases are two different codons coding for glutamic acid.

Wobble base pairing is an important factor in shaping the tRNA gene sets in organisms. Transfer RNAs that can decode more than one codon *via* wobble base pairing often appear in multiple copies, while those without this capability are more prone to be totally absent from the genome. For example, there are two theoretically possible tRNA isotypes that carry glutamic acid: tRNA-Glu^{UUC} and tRNA-Glu^{CUC}. The former isotype can decode both 5'-GAA-3' and 5'-GAG-3' *via* Watson-Crick and wobble pairing rules, respectively (Figure 1.3). In many genomes there are several copies of the gene coding for tRNA-Glu^{UUC}, while the gene for tRNA-Glu^{CUC} is absent or present as a single copy (Lowe and Eddy, 1997). By applying the wobble base pairing rule, all the 61 codons can theoretically be decoded by a minimum set of 32 tRNA isotypes (Rak *et al.*, 2018). Thus, extended base pairing at the 5' anticodon position explains why genomes typically encode only a subset (typically not more than 50; Grosjean *et al.*, 2010) of the 61 theoretically possible tRNA isotypes.

1.1.4 The cellular life cycle of a tRNA molecule

Transfer RNA was first discovered by Zamecnik and Hoagland in 1958 as an enzyme in rat liver extract that could be incorporated into activated amino acids in an ATP-dependent manner, and facilitate addition of amino acids into proteins (Hoagland *et al.*, 1958). The mature tRNA has a conserved cloverleaf structure with the stem-loop compartments that are made by the inverted repeats conserved among all tRNA species (Altman, 1980; Figure 1.2). Production of the mature tRNA involves several processing steps and enzymes that vary depending on the organism and tRNA isotype (Li and Deutscher, 1996). However, in prokaryotes the general steps are as follows:

1. The tRNA gene is first transcribed by RNA polymerase into a precursor RNA; this precursor is longer than the mature tRNA.
2. Next, the precursor is cleaved by several RNAases. First, RNase E and RNase III process the 3' end of the tRNA molecule, downstream of the conserved 5'-CCA-3' region (see section 1.1.5). If there is more than one tRNA sequence in the precursor, RNase E and RNase III cleave the RNA into constituent tRNAs. Then, RNase P removes bases from the

5' end of the molecule(s) (Altman, 1971). Finally extra nucleotides from the 3' end of the precursor are trimmed by exoribonucleases including RNase II, BN, D and PH (Sakano and Shimura, 1978).

3. The processed tRNA is next subjected to various tRNA modification enzymes, resulting in chemical changes on special bases. These modifications are necessary for maintaining the tRNA structure, efficient codon-anticodon base pairing and translation (Lorenz *et al.*, 2017). So far, around 100 post-transcriptional modifications have been discovered. The modifications result in different derivatives of all four common nucleotides. The modifications in positions 34 (first base of anticodon) and 37 of a tRNA molecule are among the most important ones; these modifications result in stabilization of codon-anticodon interactions during translation (Hagervall and Björk, 2005).

4. After the modifications are completed, the tRNA molecule passes through aminoacylation, which is a two-stage process catalyzed by aminoacyl tRNA synthetases. Aminoacyl tRNA synthetases constitute a large enzyme family that can be divided into two classes: class I and II. In *E.coli*, the class I synthetases recognize tRNAs carrying Val, Ile, Leu, Met, Cys, Tyr, Trp, Glu, Gln, and Arg, while class II synthetases aminoacylate tRNAs carrying the rest of amino acids (Ser, Thr, Pro, Gly, His, Asp, Lys, Asn, Phe, Ala; Eriani *et al.*, 1990). The enzyme first activates the amino acid by consuming an ATP molecule. The resulting aminoacyl adenylate is next transferred to the OH group at the 3' end of the tRNA molecule (Ibba and Söll, 2003). Each synthetase must recognize only its cognate corresponding tRNA(s), resulting in high substrate affinity (sub-optimal affinity results in mischarging of a tRNA molecule, and thus incorporation of incorrect amino acids in proteins). The tRNA structure forms a scaffold in the interaction with the synthetase enzyme; There are some conserved bases inside tRNA molecules that belong to a single amino acid family, which allows this family of tRNAs to be recognized by the same aminoacyl-tRNA synthetase (Westhof *et al.*, 1985; Ibba and Söll, 2003). Thus, each tRNA isoacceptor group has one specific aminoacyl synthetase enzyme and in total in most of organisms there are 20 types of these enzymes in the cell. There are some bases that play critical role in recognition of specific tRNAs by the cognate synthetases, which are called as “determinants”. The determinants are mostly located in the anticodon loop and the amino acid acceptor stem (including the discriminator base: N73). Furthermore, some of the crucial bases for correct recognition are positioned in the core structure of the tRNA (*e.g.* between nucleotides 8-31

and 39-65; Giegé *et al.*, 1998). All the aforementioned properties of the specific bases of the tRNA molecule reduce the probability of the synthetase recognizing a wrong tRNA type. However, mischarging of a tRNA molecule with the wrong amino acid may occur when two amino acids have a similar structure (*e.g.* the valine and the alanine structures differ only by a single methyl group, thus the synthetase enzyme may not distinguish well between them (Loftfield, 1963; Loftfield and Vanderjagt, 1972). Early experimental studies estimated that the error frequency of incorporation of a wrong amino acid is around 1 *per* 3,000, and 2.6 *per* 10,000 aminoacylation (Loftfield, 1963; Loftfield and Vanderjagt, 1972). Later in 1990, Söll proposed that there are certain positions in each tRNA type that act as an “identity element” for recognition by the correct aminoacyl-tRNA synthetase and the approximate error rate of aminoacylation is 10^{-6} *per* event (Söll, 1990; Schulman, 1991).

5. Finally, the mature, aminoacylated tRNA is ready to be recruited by the translation machinery or other cellular processes. During translation, the aminoacylated tRNA binds to the ribosome based on the anticodon matching with the corresponding codon in the mRNA. The amino acid on top of the tRNA molecule forms a peptide bond with the previous amino acid in the nascent polypeptide, in the P site of the ribosome. *Via* a translocation event, the discharged tRNA moves to the exit site of the ribosome (E Site) and is released from the ribosome (Alberts, 1983). The free, deacylated tRNA may then pass through more aminoacylation cycles.

Aminoacylated tRNA molecules can also be involved in other cellular processes including peptidoglycan synthesis, antibiotic resistance, antibiotic synthesis and membrane phospholipid modification pathways (Shepherd and Ibba, 2015; Raina and Ibba, 2014). Figure 1.4 summarizes the life cycle of a tRNA molecule in a prokaryotic cell.

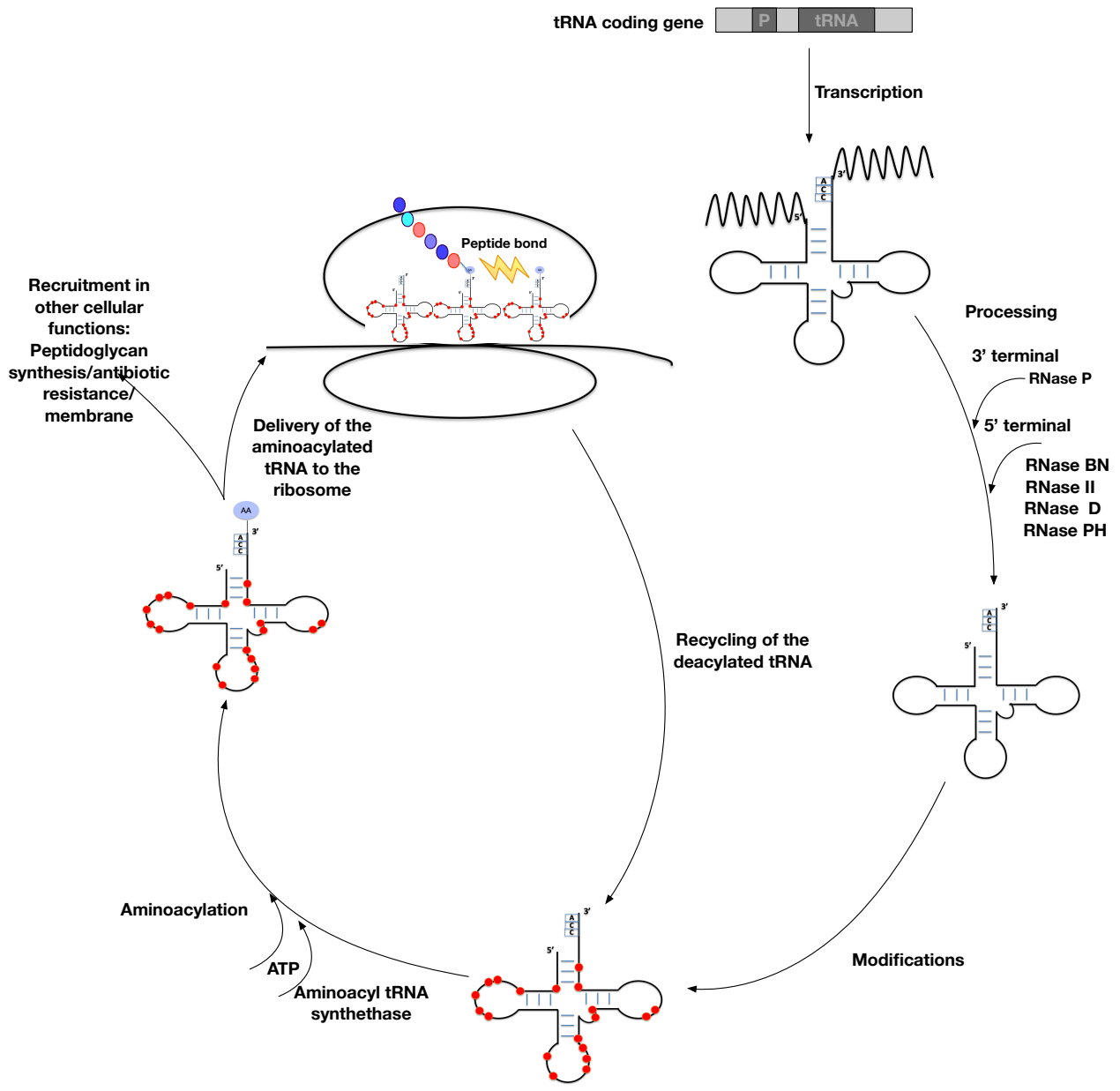


Figure 1.4. The life cycle of a bacterial tRNA molecule. The red circles show the modified bases.

1.1.5 tRNA gene structure and organization in genome

Transfer RNAs are coded by genes, which are scattered all over genomes. The structure of tRNA genes and their organization inside the genome is different between prokaryotes and eukaryotes. The promoter of a tRNA gene in eukaryotes is internal. That is, there are two highly conserved sequences inside the tRNA gene called A box and B box. A box starts after 8 nucleotides from the start point of the gene, and B box is positioned around 30 to 60 nucleotides downstream of the A box. Interaction of RNA polymerase III with these two sites promotes tRNA transcription (Paule *et al.*, 2000).

Prokaryotic tRNA genes have external promoters, which are usually conserved sequences at -10 and -35 of the transcription start of the gene, and are recognized by the σ subunit of the bacterial RNA polymerase (Helmann and Chamberlin 1988; Toh *et al.*, 2001). A further difference between prokaryotic and eukaryotic tRNA genes is that in prokaryotes the universal 5'-CCA-3' end of the tRNA – which is essential for aminoacylation by tRNA synthetases – is mostly present in the tRNA gene sequence itself, however in eukaryotes and archaea it is added post-transcriptionally to the amino acid acceptor stem (Hou, 2010).

The organisation of eukaryotic tRNA genes is also different from that in bacteria. In eukaryotes, tRNA genes are contained in monocistronic transcripts, which consist of only tRNA genes. These tRNA genes are usually organized as multigene families dispersed all over the genome. Each family comprises of several copies of the same or different tRNAs, which are called as homologous and heterologous clusters respectively (Santana *et al.*, 2010). In prokaryotes, the organization of the tRNA genes in the genome varies in each clade, and can be categorized into three groups (see Figure 1.5):

- 1-Transcription units that consist of one or more tRNA genes.
- 2-Transcription units that contain rRNA genes, with tRNA genes in the spacer regions.
- 3-Transcription units that contain tRNA genes in combination with some protein-coding genes (often genes coding for proteins involved in transcription or translation regulation; Toh *et al.*, 2001).

In the model system of this study, the bacterium *Pseudomonas fluorescens* SBW25, all three of the aforementioned categories of tRNA transcription units are present (data from UNAM online prokaryotic operon prediction tool; Taboada and Verde, 2010).

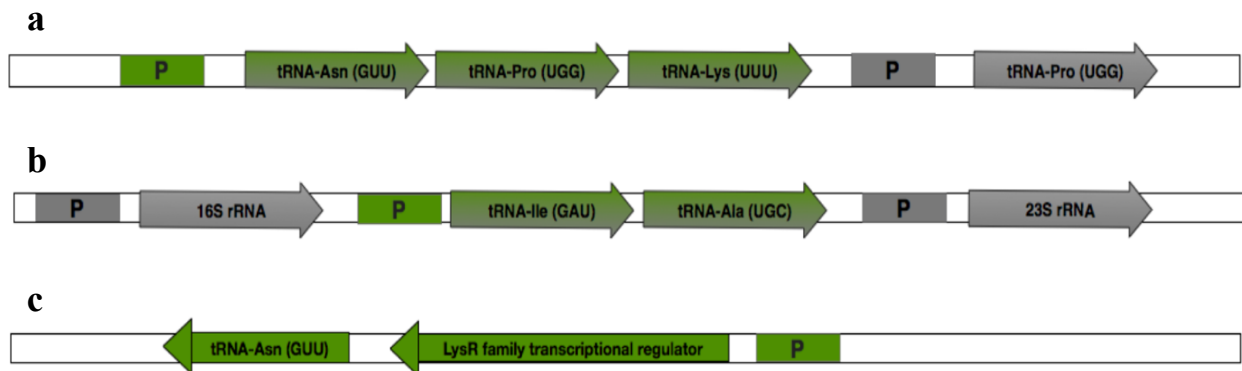


Figure 1.5. Transfer RNA genes are present in different types of transcriptional unit, all three of which have examples in the *P. fluorescens* SBW25 genome. a) Several tRNA genes share a common promoter (green P) and the tRNA-Pro^{UGG} gene in the c (grey P) has a separate promoter. b) tRNA-Ile^{GAU} – tRNA-Ala^{UGC} are located between the two rRNA promoters and both share a common promoter. c) The tRNA-Glu^{UUC} gene has a common promoter with a transcription factor for LysR family (P stands for promoter).

1.1.6 Factors shaping tRNA gene sets in genomes

Thus far, the structure and synthesis of a tRNA molecule, and its subsequent role in cellular processes have been discussed. Next, the factors involved in shaping the tRNA gene sets throughout the course of evolution will be discussed. Previous studies have shown that the cellular tRNA level is the rate-limiting factor during protein translation (Gingold *et al.*, 2012). Furthermore, it has been proposed that tRNA isotypes correlate with regard to gene copy number and proportion in the mature tRNA pool (Percudani *et al.*, 1997; Kanaya *et al.*, 1999). Thus, the tRNA gene set of genomes is under strong selection. This selection favours a balance between the tRNA genes as “supply” and the coding sequences as “demand” in the genomes.

Next, several factors that are involved in forming the supply and demand dynamics in genomes will be discussed; the demand can be changed both in evolutionary timescales and the lifespan of the organism. Codon usage in genomes is constantly changing during evolution (Sharp and Matassi, 1994). For instance, mutations and horizontal gene transfer can both result in changes in codon usage (Gogarten *et al.*, 2002). Moreover, during the life span of an organism (*e.g.* across different cellular phases, developmental stages, or diverse tissues of multicellular organism), many genes are subject to differential expression levels, which alters the codon usage in the mRNA pool of the organism. In addition to the internal

causes of differential codon usage, cells also have to adapt their codon usages in response to environmental stress, that is, in response to stress different sets of genes are expressed and the new gene set may have codon usage that differs from the genes in normal condition (Rak *et al.*, 2018).

All the aforementioned changes in codon usage shape the demand for tRNA isotypes. As the supply, the mature tRNA pool is expected to change in response. The supply is affected by multiple factors, including the gene copy number for each tRNA species, the expression level of tRNA genes (and all the components that affect the expression level, including promoter type and the genomic location of the tRNA gene), and the existence or absence of other types of tRNA (*e.g.* other tRNAs compete for translation machinery components such as translation factors and tRNA synthetases; Dong *et al.*, 1996). Thus, we can conclude that the efficiency and accuracy of translation, and ultimately the fitness of the cell, depends on a very intricate balance between tRNA supply and demand.

In this study, I focus on one component of this complex system: the tRNA gene copy number. Specifically, tRNA gene copy numbers will be perturbed in various *P. fluorescens* strains, and the downstream effects documented (see section 1.4).

1.1.7 tRNA repertoires in different domains

The number and type of genes coding for tRNAs (the “tRNA gene set”) varies among organisms. Across the tree of life, each domain shows unique tRNA gene set features. Firstly, archaea have the simplest tRNA gene sets with the lowest variation in gene copy numbers between its members (Mallick *et al.*, 2005). Bacterial tRNA gene sets tend to be more complex, but are still far simpler than those in eukaryotes (Novoa *et al.*, 2012). There are also many variations within domains, species and even strains. For example, among eukaryotes, humans have over 400 tRNA genes while *Drosophila melanogaster* has 295. Similar variations are observed in the bacterial domain as well. For instance, *Escherichia coli* has around 89 tRNA genes, while *Pseudomonas fluorescens* has around 70 (Lowe and Eddy, 1997). There is much debate about the origin of these differences; proposed mechanisms of tRNA gene evolution include gene duplication and horizontal gene transfer (Wald and Margalit, 2014).

1.1.8 tRNA databases

Among the many databases available that contain information regarding the tRNA repertoires of species is the GtRNAdb - Genomic tRNA database - which was created and maintained by Lowe lab. This database uses a program that identifies more than 99% of the tRNA genes in DNA sequences. This program uses some relaxed parameters for the primary tRNA detection, and then analyzes the candidates by a highly selective covariance model (Lowe and Eddy, 1997). This program also predicts the unusual selenocysteine tRNAs (or suppressor tRNAs) and tRNA pseudogenes. Thus far, 540 eukaryotic genomes, 217 archaeal genomes, and 4,037 bacterial genomes have been scanned by this algorithm, leading to the detection of around 400,000 tRNA genes. The GtRNAdb has the data for tRNA repertoires of eight *P. fluorescens* strains.

1.2 Experimental Evolution

During the course of evolution, adaptations occur in response to the environment. These adaptations are the result of the occurrence and accumulation of beneficial mutations in populations. In addition to natural environments, evolution occurs in laboratory populations (Lenski *et al.*, 1991; Riehle *et al.*, 2001; Meyer *et al.*, 2012; Perry *et al.*, 2016; Gallie *et al.*, 2019).

Experimental evolution, coupled with advances in genome sequencing technology in recent decades, has been widely used to study the genetic bases of adaptation (Heineman *et al.*, 2005; Turner *et al.*, 2011; Lindsay *et al.*, 2019; Manrique-Poyato *et al.*, 2019). This approach provides the opportunity to control several factors such as the selective environment, the bottleneck size between transfers, the ancestor genotypes, and the number of replicate populations.

Microorganisms are ideal for evolution experiments. Firstly, their short generation time (*E. coli* divides every 20 minutes, *P. fluorescens* SBW25 every 40 minutes) enables the researcher to conduct the evolution experiment in the time span of the laboratory. Secondly, the population sizes attained are large enough to sample a large set of mutations (*E. coli* and *P. fluorescens* grow to population sizes of $\sim 10^9$ individuals *per ml* in overnight cultures). Thirdly, the potential freeze indefinitely, and to subsequently resuscitate samples, makes the bacteria a very attractive evolutionary model for scientists. In essence, bacteria provide the

chance to access and directly compare all individuals across the complete time span of a genetic lineage. Thus, microbes have been used extensively for studying the evolutionary dynamics of mutations within populations (Lenski and Travisano, 1994; Travisano and Lenski, 1996). Whole genome sequencing of evolved populations for the first time showed surprising discordance between the rates of phenotypic and genotypic change in evolution (Lang and Desai, 2014). Some interesting reasons were proposed for these observations such as emergence of mutators in *E.coli* populations (these strains have acquired defects in the DNA repair system, and thus show elevated mutation rates; Barrick *et al.*, 2009), or occurrence of many synonymous mutations in yeast populations (Lang *et al.*, 2013).

During evolution, over generations, beneficial mutations will generally increase in populations, whereas neutral mutations will generally not. On the other hand, there are selective pressures against deleterious mutations in populations; hence, they are expected to be lost from the population (Bruger and Marx, 2018).

Mutations are very diverse, and include transitions, transversions, single base deletions or insertions. In addition to point mutations, genomic rearrangements such as large duplications, deletions and inversions can occur during evolution. For instance, the frequency of duplication in most sites of the bacterial chromosome is estimated to be 10^{-3} to 10^{-4} (Sonti and Roth, 1989). In an experimental evolution carried out on thermal adaptation of *E. coli*, Riehle *et al.* repeatedly found large duplications in some of lineages adapted to high temperatures. They proposed that the duplications increase the gene dosages of heat shock proteins and genes encoding the stress sigma factor (Riehle *et al.*, 2001).

The general design of experimental evolution is simple: a population of bacteria (containing one or more genotypes) is inoculated into a tube containing known medium. At specified intervals, a proportion of the population is transferred into fresh medium. This process may be continued for thousands of generations (*e.g.* over 70,000 generations in Lenski's long term evolution experiment; Lamrabet *et al.*, 2019). Finally, the phenotypic and genotypic characteristic of the evolved populations is compared to those of the ancestral founding population. Observed changes can be attributed to the evolutionary effect on the populations. In this study, I apply the same principals to design an evolution experiment with a tRNA gene set mutant of *P. fluorescens* SBW25 (see section 2.2.10).

1.3 *Pseudomonas fluorescens*: a model organism for studying evolution

Pseudomonas is a genus of diverse, Gram-negative Gammaproteobacteria, encompassing strains from the opportunistic human pathogen *P. aeruginosa* to the plant pathogen *P. syringae* (Silby *et al.*, 2011). The wide variety of habitats of *Pseudomonas* members is a reflection of their diverse metabolic capacity (Palleroni, 1992). This diverse metabolic capacity has made some members of this clade useful biotechnological tools. For example, some *P. fluorescens* strains interact with plants and antagonize plant pathogens, so they can play an important role as plant growth-promoting bacteria (Glick, 2012). Furthermore, some species of *Pseudomonas* can degrade hydrocarbons, and are used in bioremediation projects (*e.g.* some saprophytic strains of *P. putida* are capable of degrading petroleum contaminated soils (Shrivastava *et al.*, 2011). Some pseudomonads are involved in biogeochemical cycles (*e.g.* some *P. stutzeri* strains are able to fix nitrogen; Lalucat *et al.*, 2006).

In this study, the species of interest is *P. fluorescens*. Members of this species are found in water, soil, and on plant surfaces. In the scarcity of iron, the fluorescent *Pseudomonas* spp. including all the *P. fluorescens* strains, produce green-yellowish fluorescent pigments called pyoverdine. These chelate iron, which is then brought into the cell (Wendenbaum *et al.*, 1983). Many members of the species have been successfully cultured and used extensively in the laboratory (Table 1.2). There are common features in the genome of the sequenced strains of *P. fluorescens* (Silby *et al.*, 2009). In particular, repetitive extragenic palindromic sequences (REPs) are present in most of the sequenced genomes (Bertels and Rainey, 2011). These are thought to facilitate genomic rearrangements during evolution.

The wealth of phenotypic, genetic and evolutionary data available for *P. fluorescens* renders this species a useful and tractable model for performing evolutionary studies. Specifically, SBW25 has been widely used in experimental evolution (Rainey and Travisano, 1998; Beaumont *et al.*, 2009; Koza *et al.*, 2017; Gallie *et al.*, 2019). In this study, SBW25 and two other *P. fluorescens* strains (A506 and Pf0-1) were used as model strains for conducting genetic manipulation and experimental evolution (see next section).

Table 1.2. Some worldwide isolated strains of *P. fluorescens*

Strain	Isolation site	Natural habitat	Reference
A506	CA, USA	Pear tree leaf	(Lindow <i>et al.</i> , 1998)
PICF-7	Spain	Olive tree leaf	(Gómez-Lama Cabanás <i>et al.</i> , 2014)
Pf0-1	MA, USA	Loam soil	(Kim <i>et al.</i> , 2009)
SBW25	Oxford, UK	Sugar beet leaf	(Rainey and Bailey, 1996)

1.4 Aims

In this introduction, the fundamental role of tRNA molecules in translation, and the factors involved in balancing intracellular tRNA levels during evolution have been discussed. Among the most important factors in shaping mature tRNA pools are the copy number of tRNA genes. Studying the tRNA gene repertoire of organisms from all domains of life indicates that organisms typically contain a subset of the 61 available tRNA isotypes, with multiple copies. These copy numbers range from one to six in *P. fluorescens* strains.

The first aim of this thesis is to study the functional importance of multiple-copy tRNA genes in a selected set of three evolutionarily distant *P. fluorescens* strains: SBW25, A506, Pf0-1. The tRNA gene set of each genome shows that there are 66 tRNA genes that are shared among these three strains, which in this study are defined as “core” tRNA genes. There are three and seven additional tRNA genes in A506 and Pf0-1, respectively. These are defined as “accessory” tRNA genes. The aim is to systemically delete all the accessory tRNA genes from A506 and Pf0-1 genome using two-steps allelic replacement. Phenotyping of each mutant will give insights about contribution of each identical copies of tRNA gene to the fitness. In case of observing no phenotype after deletion of one copy of a specific tRNA gene, further copies of that gene will be deleted until a phenotype is observed. In next section of this study, a mutant with defected fitness will be evolved in the lab to study the evolution of tRNA genes in real time. The expectation is that the fitness defect will be compensated, and genome sequencing of compensated isolates could increase our understanding of tRNA gene sets evolution.

2. Materials and Methods

2.1 Materials

2.1.1 Bacterial strains

Bacterial strains used in this study are listed in Table 2.1. All strains were stored in glycerol saline stock 45% (v/v) in -80°C and are available in the Gallie lab.

Table 2.1. Characterization of bacterial strains used/constructed in this study

Strain	Characteristics	Reference
<i>Pseudomonas fluorescens</i> A506		
A506	Wild type <i>P. fluorescens</i> isolated from a pear leaf in USA	(Lindow <i>et al.</i> , 1998)
A506ΔAsn	<i>P. fluorescens</i> A506 with one of three tRNA-Asn ^{GUU} deleted	This study
A506 ΔAsn_EC	<i>P. fluorescens</i> A506 that has passed through the same engineering process for deletion of tRNA-Asn ^{GUU} , but still carries the wild type sequence	This study
A506ΔAlaIle	<i>P. fluorescens</i> A506 with one of six tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublets deleted	This study
A506ΔAlaIle_EC	<i>P. fluorescens</i> A506 that has passed through the same engineering process for deletion of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet, but still carries the wild type sequence	This study
A506ΔAsnAlaIle	<i>P. fluorescens</i> A506 with one out of three tRNA-Asn ^{GUU} and one of six tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublets deleted	This study
A506ΔAsnAlaIle_EC	<i>P. fluorescens</i> A506 that has passed through two successive engineering processes for deletion of tRNA-Ala ^{UGC} and tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet, but carries the wild type sequence	This study
<i>Pseudomonas fluorescens</i> Pf0-1		
Pf0-1	Wild type <i>P. fluorescens</i> isolated from loam soil in Sherborn, Massachusetts, USA.	Kim <i>et al.</i> , 2009
Pf0-1ΔGln	<i>P. fluorescens</i> Pf0-1 with one of two tRNA-Gln ^{UUG} deleted.	This study
Pf0-1ΔSecGln	<i>P. fluorescens</i> Pf0-1 with tRNA-Sec ^{UCA} and one of two Gln ^{UUG} have been deleted.	This study
Pf0-1ΔSecGlnGluGly	<i>P. fluorescens</i> Pf0-1 with tRNA-Sec ^{UCA} , one of two tRNA-Gln ^{UUG} , and one of three tRNA-Glu ^{UUC} / tRNA-Gly ^{GCC} doublets, deleted.	This study

Pf0-1ΔSecGlnGluGly_EC	<i>P. fluorescens</i> Pf0-1 that has passed through the same engineering processes for deletion of the tRNA-Glu ^{UUC} / tRNA-Gly ^{GCC} doublet, but still carries the wild type sequence.	This study
<i>Pseudomonas fluorescens</i> SBW25		
SBW25	Wild type <i>P. fluorescens</i> SBW25 isolated from sugar beet plant in Oxford, 1989	(Rainey and Bailey, 1996)
SBW25_lacZ	SBW25 with phenotypically neutral <i>lacZ</i> marker used for competition experiment	(Zhang and Rainey, 2007)
SBW25ΔGluGlyGluGly	<i>P. fluorescens</i> SBW25 with a simultaneous deletion of four tRNA genes (two tRNA-Glu ^{UUC} and two tRNA-Gly ^{GCC} ; 500bp)	This study
RWT	Wild type <i>P. fluorescens</i> SBW25 strain that has been reconstructed from SBW25ΔGluGlyGluGly	This study
SBW25ΔAsn1	A <i>P. fluorescens</i> SBW25 with one out of its two tRNA-Asn ^{GUU} has been deleted.	This study
ZK1	A <i>P. fluorescens</i> SBW25 isolated from the 210 th generation of the experimental evolution (line 1) with the SBW25ΔGluGlyGluGly as the founding population	This study
ZK2	A <i>P. fluorescens</i> SBW25 isolated from the 210 th generation of the experimental evolution (line 2) with the SBW25ΔGluGlyGluGly as the founding population	This study
ZK3	A <i>P. fluorescens</i> SBW25 isolated from the 210 th generation of the experimental evolution (line 2) with the SBW25ΔGluGlyGluGly as the founding population	This study
ZK4	A <i>P. fluorescens</i> SBW25 isolated from the 210 th generation of the experimental evolution (line 2) with the SBW25ΔGluGlyGluGly as the founding population	This study
ZK5	A <i>P. fluorescens</i> SBW25 isolated from the 210 th generation of the experimental evolution (line 5) with the SBW25ΔGluGlyGluGly as the founding population	This study
ZK6	A <i>P. fluorescens</i> SBW25 isolated from the 210 th generation of the experimental evolution (line 3) with the wild type SBW25 as the founding population	This study
<i>Escherichia coli</i> K-12		
DH5α_λpir	Strain used for cloning; contains chromosomally-encoded λpir, recA mutation, and endA mutation. It has a deletion in the lac operon that results in white colonies on X-gal	Invitrogen
S17-1_λpir	Strain for biparental conjugation that contains mobilization genes and λpir on its chromosome	Invitrogen

2.1.2 Plasmids

Plasmids used in this study are listed in the following Table.

Table 2.2. Characteristics of plasmids used for this study.

Name	Characteristics	Reference
pCR TM 8/GW/TOPO TM	Sequencing plasmid (2.8kb), Spe ^R , Puc <i>ori</i>	Invitrogen
pUIC3	allelic replacement plasmid, Tc ^R , <i>mob</i> , <i>oriR6K</i> , <i>bla</i> , with a promotorless <i>lacZY</i> operon, which makes the colony blue on X-gal	(Rainey, 1999)

2.1.3 Enzymes, antibiotics, and reagents

Most of the enzymes were purchased from Promega. The antibiotics and enzymes used in this study are listed in Table 2.3 and 2.4 respectively:

Table 2.3. Characteristics of antibiotic used for this study.

Name	Solvent	Final concentration	Function
Tetracycline (Tc)	50% ethanol	12.5 µg ml ⁻¹	Selective agent for pUIC3
Spectinomycin (Spe)	dH ₂ O	100 µg ml ⁻¹	Selective agent for pCR TM 8/GW/TOPO TM
N-[5-Nitro-2-furfurylidene]-1-aminofurofuran-2(1H)-one (NF)	Dimethyl sulfoxide (DMSO)	100 µg ml ⁻¹	Inhibits <i>E. coli</i> growth
Cycloserine	37°C dH ₂ O	800 µg ml ⁻¹	Kills growing bacterial cells

Table 2.4. Characteristics of enzymes used for molecular cloning.

Name	Function	Manufacturer
<i>Bgl</i> II	Restriction enzyme	Promega
T4 DNA ligase	Ligase	Promega
GoTaq DNA polymerase	DNA polymerase	Promega
Pfu polymerase	DNA polymerase (with proofreading ability)	Promega
TSAP	Thermosensitive Alkaline Phosphatase	Promega

Additional reagents and buffers were used for molecular biology. These are listed in Table 2.5.

Table 2.5. Ingredients list for reagents used in this study.

Reagent	Ingredients	Reference
TBFI (final autoclaved pH:5.8)	30 mM KOAc, 100 mM RbCl, 50 mM MnCl, 10 mM CaCl ₂	(Sambrook, 1989)
TBFII (final autoclaved pH:7)	10 mM MOPS, 10 mM RbCl, 75mM CaCl ₂ , 15 % (v/v) Glycerol	(Sambrook, 1989)
chromogenic indicator of β -galactosidase activity of <i>lacZ</i> operon (X-gal)	5-Bromo-4-chloto-3-indolyl- β -D-galactopyranoside Final concentration: 60 μ g ml ⁻¹	(Sambrook, 1989)
5xCES	2.7 M betaine (540 μ l), 6.7 mM didithiothreitol (200 μ l), 55 μ g/mlvBSA (1 μ l), 6.7% (v/v) DMSO (67 μ l), Distilled water (192 μ l)	(Ralser <i>et al</i> , 2006)
Ringer solution	2.25g NaCl, 0.105 g KCl, 0.12 g CaCl ₂ , 0.05 NaHCO ₃ , dissolved in 1L distilled water	(Sambrook, 1989)

2.1.4 Media and cultures

Unless otherwise stated, media components were purchased from Sigma-Aldrich. Bacterial culture details are listed in Table 2.6. For preparing the solid forms of each medium, 1.5% agar was dissolved into each bottle.

Table 2.6. Media used for culturing bacteria.

Medium	Ingredients	Reference
King's B (KB)	10 g glycerol, 20 g tryptone (for making plates) or Proteose peptone No3 (for the broth), 1.5 g K ₂ PO ₄ .3H ₂ O, 1.5g MgSO ₄ .7H ₂ O in 1L	(King <i>et al.</i> , 1954)
Lysogeny Broth (LB)	10 G NaCl, 10 g tryptone, 5 g yeast extract in 1L	(Bertani, 1951)
Minimal M9 medium	34 g Na ₂ HPO ₄ .3HPO ₄ .7H ₂ O, 15 g KH ₂ PO ₄ , 2.5 g NaCl, 5 g 1.8M NH ₄ Cl, 15 mg 1M CaCl ₂ .6H ₂ O, 1 ml 1M MgSO ₄ .7H ₂ O, 10ml 20% (w/v) glucose in 1L	(Sambrook <i>et al.</i> , 1989)
Terrific Broth (TB)	2.31 g KH ₂ PO ₄ , 12.54 g K ₂ HPO ₄ , 12 g tryptone, 24 g yeast extract, 4 ml glycerol in 1L	(Sambrook <i>et al.</i> , 1989)

2.1.5 Primers

Primers were designed manually, and their suitability checked using the Integrated DNA Technologies (IDT) online tool. Primers were synthesized by Metabion, and resuspended in deionized water to final concentration of 100pmol μl^{-1} . The 100 pmol μl^{-1} stocks were stored at -20°C. The working concentration was 5 pmol μl^{-1} with changes in concentration noted when necessary (Table 2.7).

Table 2.7. Sequence and target of primers designed to make *P. fluorescens* mutants.

Name	Sequence (5'→3')	Target
Primers for allelic replacement in A506		
A506_Asn3598_1	<u>GAAGATCTGGATGACAAT</u> GACCTCCATCGTAAAC	~400bp upstream of 5' end of tRNA-Asn ^{GUU} (at position 3988498)
A506_Asn3598_2	GATAGAGAAAAGCCTGTA GAGATACAG	5' end of tRNA-Asn ^{GUU} (at position 3988498)
A506_Asn3598_3	CTGTATCTCTACAGGCTTT TCTCTATCGCCACGAAAAA CGTGGTGGAGAC	3' end of tRNA-Asn ^{GUU} (at position 3988498)
A506_Asn3598_4	<u>GAAGATCTGTGCCGAATGA</u> GAGCAATGGATATG	~400bp downstream of 3' end of tRNA-Asn ^{GUU} (at position 3988498)
A506_Asn3598_F	CTCCAACCAGGAAAGCGA ACCTTG	~900bp upstream of 5' end of tRNA-Asn ^{GUU} (at position 3988498)
A506_Asn3598_R	CTTCTAATGATGCGATGC AGCGTC	~900bp downstream of 3' end of tRNA-Asn ^{GUU} (at position 3988498)

A506_4047_ampF	CACAGCGTTCAGTGCTAGT AGAAGG	~3150bp upstream of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet (at position 4478125)
A506_4047_ampR	CTGTGCAGGCTTTGGAAGT TGG	~3150bp downstream of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet (at position 4478125)
A506_4047_1	<u>GAAGATCTCTAGAGATAG</u> ATTGGTGCCTTCGG	~650bp upstream of 5' end of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet (at position 4478125)
A506_4047_2	AATTTTCGAGCGCGTAACTG ATTAGCTAAG	5' end of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet (at position 4478125)
A506_4047_3	CTTAGCTAATCAGTTACGC GCTCGAAATTTAACTGCTT CTGCTGTTAGAGTTTAG	3' end of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet (at position 4478125)
A506_4047_4	<u>GAAGATCTGGTACTAGTTC</u> ACTATCGGTCAGTCAG	~650bp downstream of 3' end of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet (at position 4478125)
A506_4047_F	GCAGTCATCAGTGCCAACT GAG	~3250bp upstream of 5' end of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet (at position 4478125)
A506_4047_R	GTCTGTGAGGTACATAACC GTGGAG	~3250bp downstream of 3' end of the tRNA-Ala ^{UGC} / tRNA-Ile ^{GAU} doublet (at position 4478125)
Primers for allelic replacement in Pf0-1		
PFL01_RS11885_1	<u>GAAGATCTCTGAGAATTCC</u> TCTGTGCTGATCG	~400bp upstream of 5' end of the tRNA-Gln ^{UUG} (at position 2710914)
PFL01_RS11885_2	TTTTTTCCTCGTTTTACCCT CCTC	5' end of the tRNA-Gln ^{UUG} (at position 2710914)
PFL01_RS11885_3	GAGGAGGGTGAAACGAGG AAAAAAGTAGCTATTGGTT TCAGTCCCGG	3' end of the tRNA-Gln ^{UUG} (at position 2710914)
PFL01_RS11885_4	<u>GAAGATCTCGACTGCCTGC</u> AATACATCCACC	~400bp downstream of 3' end of the tRNA-Gln ^{UUG} (at position 2710914)
PFL01_RS11885_F	CAGGAGTGCCTGTAACGCC TGAC	~600bp upstream of 5' end of the tRNA-Gln ^{UUG} (at position 2710914)
PFL01_RS11885_R	CTGATCACTGAACGCATCG ACAC	~600bp downstream of 3' end of the tRNA-Gln ^{UUG} (at position 2710914)
PFL01_RS13650_1	<u>GAAGATCTCATGAACCTGGA</u> ACCCGATCAGATGG	~550bp upstream of 5' end of the tRNA-Sec ^{UCA} (at position 3119094)
PFL01_RS13650_2	TTGAACTCAGTCCGCGCCG	5' end of the tRNA-Sec ^{UCA} (at position 3119094)
PFL01_RS13650_3	CGGCGCGGACTGAGTTCAA ATTTCCTCAAAGCCCAA CGGGTAC	3' end of the tRNA-Sec ^{UCA} (at position 3119094)
PFL01_RS13650_4	<u>GAAGATCTCTGAACGCGGT</u> GTTGTCGATGTC	~550bp downstream of 3' end of the tRNA-Sec ^{UCA} (at position 3119094)
PFL01_RS13650_F	CATGAACTGGAACCCGATC AGATGG	~700bp upstream of 5' end of the tRNA-Sec ^{UCA} (at position 3119094)
PFL01_RS13650_R	GGCACCGATCTGTTTCAGCG ACAG	~700bp downstream of 3' end of the tRNA-Sec ^{UCA} (at position 3119094)

PFL01_RS21055_60_1	<u>GAAGATCTGCAAGTCGCCG</u> GAAGTGCTC	~500bp upstream of 5' end of tRNA-Glu ^{UUC} (at position 4729574)
PFL01_RS21055_60_2	AATTCACAAAAACGCCGCT CAG	5' end of tRNA-Glu ^{UUC} (at position 4729574)
PFL01_RS21055_60_3	CTGAGCGGCGTTTTTGTGA ATTAGGCTACAAC TTTCAC TAATAAAAACGC	3' end of tRNA-Gly ^{GCC} (at position 4729452)
PFL01_RS21055_60_4	<u>GAAGATCTGGCAGGGAGT</u> GAATGTTGTAGTG	~500bp downstream of 3' end of tRNA-Gly ^{GCC} (at position 4729452)
PFL01_RS21055_60_F	CAAGCACAAACCACATCACC GACG	~650bp upstream of 5' end of tRNA-Glu ^{UUC} (at position 4729574)
PFL01_RS21055_60_R	CGAGTGTCGGAATATAGGC GTCAATC	~650bp downstream of 3' end of tRNA-Gly ^{GCC} (at position 4729452)
Primers for allelic replacement in SBW25		
SBW25_Asn4745_1	<u>GAAGATCTCGGACATCGAC</u> ATCGAACTCAC	~600bp upstream of 5' end of tRNA-Asn ^{GUU} (at position 4745918)
SBW25_Asn4745_2	GGAATGCGCCGTATGTTAC TG	5' end of tRNA-Asn ^{GUU} (at position 4745918)
SBW25_Asn4745_3	CAGTAACATACGGCGCATT CCGATAGAGAAAGCCTGC AGAGATGC	3' end of tRNA-Asn ^{GUU} (at position 4745918)
SBW25_Asn4745_4	<u>GAAGATCTGCACTGCGCTC</u> TACCTCATC	~600bp downstream of 3' end of tRNA-Asn ^{GUU} (at position 4745918)
SBW25_Asn4745_F	CCGCTGCTCAATCGTCTTG GC	~800bp upstream of 5' end of tRNA-Asn ^{GUU} (at position 4745918)
SBW25_Asn4745_R	CGGCGGATCTCTTCGGTTG AG	~800bp downstream of 3' end of tRNA-Asn ^{GUU} (at position 4745918)
SBW25_Glu202861ST	<u>GAAGATCTGTTGACGTTGA</u> GTTCCAGTCGGGTAAG	~500bp upstream of 5' end of tRNA-Glu ^{UUC} (at position 20288653)
SBW25_Glu202862ST	CACGTACAACATTCACTCC CTACCG	5' end of tRNA-Glu ^{UUC} (at position 20288653)
SBW25_Glu20286_3T	CGGTAGGGAGTGAATGTTG TACGTGCAACGAAAACGC CGATCAGTGATGATC	3' end of tRNA-Gly ^{GCC} (at position 2029067)
SBW25_Glu20286_4T	<u>GAAGATCTGATTACCTGGG</u> AGGCAAAC TGGGTATTG	~500bp downstream of 3' end of tRNA-Gly ^{GCC} (at position 2029067)
SBW25_Glu20286FST	GTGACCTGACCATGTAGGG CTTCGAG	~6500bp upstream of 5' end of tRNA-Glu ^{UUC} (at position 20288653)
SBW25_Glu20286RT	GCAAGGAAGTCCAAC TGA CCGTTAC	~650bp downstream of 3' end of tRNA-Gly ^{GCC} (at position 2029067)

2.1.6 Molecular biology kits

The commercial kits used for genetic work are listed in Table 2.8.

Table 2.8. Kits used for molecular biology techniques.

Kit Name	Application	Manufacturer
DNeasy® Blood & Tissue Kit	Kit for genomic DNA extraction	Qiagen
QIAprep® Spin Miniprep Kit	Kit for plasmid DNA extraction	Qiagen
pCR8™/GW/TOPO® TA Cloning® Kit	Kit for TA cloning	Invitrogen
QIAquick® Gel Extraction Kit	Kit for extraction of DNA from agarose gel	Qiagen
QIAquick® PCR Purification Kit	Kit for purification of PCR products	Qiagen
BigDye XTerminator® Purification Kit	Kit for purification of DNA sequencing reaction	Thermofisher scientific

2.2 Methods

2.2.1 Genomic DNA extraction

For genomic DNA extraction, 1ml of overnight KB culture grown at 28°C (200 rpm) was pelleted by centrifugation (5000 x rpm, 10 minutes). The pellet was resuspended in 180 ml ATL buffer and 20 µl of proteinase K was added and mixed thoroughly by inversion. The mixture was then incubated for 1.5 hours at 56°C in a shaking thermomixer. Next, 4 µl of RNase A was added and mixed by inversion followed by two minutes of incubation at room temperature. Then 400 µl of AL buffer and ethanol mix was added to the vial followed by immediate, thorough inversion. The filter columns from the DNeasy® Blood & Tissue Kit were used to purify the extracted DNA based on the manufacturer's instructions. This step included passing the mixture through the mini spin column by centrifugation at 8000 x rpm for 1 minute followed by washing the column sequentially by 500 µl buffer AW1 and AW2. Finally, the DNA was eluted by 50 µl sterile HPLC water heated to 56°C. The quality of the extracted DNA was checked by gel electrophoresis (See section 2.2.3) and Nanodrop microvolume spectrophotometry (checking for a 260:280 ratio, the ideal DNA has the value of ~ 1.8 to 2.0). When genomic DNA was being extracted for next generation sequencing, the mixing and inversion steps were done very gently to avoid damaging the DNA.

2.2.2 Polymerase chain reaction (PCR)

2.2.2.1 Standard PCR

Polymerase chain reactions were done using a Biometra T3000 96-Well thermal cycler (Biolab). The standard reaction volume was 25 μl including: 5 μl 5x flexi PCR buffer, 3 μl 25 mM MgCl_2 , 1 μl 10mM dNTP mix, 2 μl of each 5 pmol μl^{-1} primer, and 0.5 μl 5 unit/ μl high-performance Gotaq® G2 flexi DNA polymerase. Finally, the mixture was made up to 25 μl with deionized water. The cycling program started with 2 minutes of pre-warming at 94 °C followed by 30 cycles of the following program:

- 1- Denaturation at 94°C for 15 seconds
- 2- Annealing at 55 to 62°C (depending on the primer sequences) for 30 seconds
- 3- Extension at 72°C for 1 minute per kilobase (kb)

In case of TA cloning of the PCR products (into pCRTM8/GW/TOPOTM), a final extension step at 72°C for 10 minutes was applied to ensure poly-A tails were added to PCR products. Finally, the sample was cooled down to 4°C.

2.2.2.2 Strand overlap extension (SOE-PCR)

SOE-PCR (Ho *et al.* 1989) was used for constructing tRNA deletion fragments. In SOE-PCR, two sets of primers are designed for upstream and downstream of the region to be deleted. These primers have around 25 bp complementary sequences (see Table 2.7). This PCR includes two separate sets of PCRs; the products from two separate round 1 PCRs overlap. These overlapping products are mixed in a 1:1 molar ratio, and used as template for round 2 PCR, which ends with the two round 1 products sewn together in a single, continuous sequence. To avoid the introduction of unwanted mutations to the deletion fragments, Pfu polymerase with proofreading ability was used for these PCRs. The reaction mixture included 5 μl 10x Pfu polymerase buffer, 1 μl 10mM dNTP mix, 4 μl of each 5 pmol μl^{-1} primer, and 2 μl 5 unit/ μl Pfu DNA polymerase. Finally, the mixture was made up to 50 μl with deionized water. Thermocycling was carried out in the same thermocycler and same cycling conditions as standard PCR, except using a denaturation temperature of 95°C, and 2 minutes of extension time per kilobase of product.

2.2.3 Gel electrophoresis

Agarose was dissolved into TBE buffer to the final concentration of 1 % w/v. 5 μ l SYBR Safe DNA Gel stain (ThermoFisher Scientific) was dissolved to the agarose solution. DNA samples were mixed with the 6X loading dye (Promega) in 6:1 ratio before being loaded to the gel wells. Agarose gels were run at 90 V for approximately 45 minutes. The separated DNA bands were visualized by an UV Transilluminator and gel documentation apparatus. Excision of DNA bands from agarose gels was performed under a blue light using a sterile Qiagen gel extraction tool. DNA was then extracted using the QIAquick® Gel Extraction Kit, following the manufacturer's instructions.

2.2.4 Preparing chemically competent cells

For cloning, two types of *E. coli* chemically competent cells were used: DH5 α _ λ pir and S17-1_ λ pir. To prepare chemically competent cells, glycerol stocks were streaked onto LB agar plates (37°C, 16-24 h) and overnight cultures inoculated from single colonies (37°C, 200 rpm). 500 μ l of the pre-culture was inoculated to 200 ml fresh LB and the cells were grown to mid-logarithmic phase. Four 50ml aliquots of this culture were pelleted by centrifugation in 3000 x rpm for 10 minutes. The pellets were re-suspended in 1 ml cold TBFI and were combined together with the total volume made up to 15 ml by more TBFI. The mixture was incubated on ice for 1 hour. Then the cells were pelleted (centrifugation in 300 rpm, 10 minutes) and re-suspended in 4 ml cold TBFII (Table 2.5). Finally, 50 μ l aliquots were made on ice and the vials were frozen at -80°C immediately.

2.2.5 Cloning and transformation techniques

The commercially available pCRTM8/GW/TOPOTM vector was used for cloning purified SOE-PCR fragments. This vector is provided as a linear fragment with poly-T ends, so the fragment to be ligated into the vector needs to have a poly-A tails. This tail was added to the purified SOE-PCR fragment using a non-proofreading Taq polymerase. The polyadenylation reaction mixture included 6 μ l 5x flexi PCR buffer, 3 μ l 25 mM MgCl₂, 1 μ l 10mM dNTP mix, and 0.1 μ l 5 unit/ μ l high-performance Gotaq® G2 flexi DNA

polymerase. The reaction mixture was incubated at 72°C for 20 minutes followed by 5 minutes of cooling to 4°C. 4 µl of the poly-A PCR product was used for setting up the ligation reaction, which included 1 µl of salt solution and 1 µl of pCRTM8/GW/TOPOTM vector. After 10 minutes of incubation at room temperature, 4 µl of the ligation mixture was gently pipetted into a thawed vial of chemically competent *E. coli* DH5α_λpir chemically competent cells (see section 2.2.4). After 10 minutes' incubation on ice, the vial of competent cells was subjected to a heat shock (42°C for 30 seconds) followed by 2 minutes of incubation on ice. After the heat shock, 250 µl of room temperature LB was added to the vial. The vial was incubated at 37°C in shaken incubator (200 rpm) for 1 hour to allow the growth and expression of the spectinomycin resistance gene. Appropriate volumes of the cells were spread on LB + Spe plates followed by incubation at 37°C overnight to screen for the transformant colonies.

Transformant colonies were checked for the insert with the correct size by applying a PCR approach. The pCRTM8/GW/TOPOTM TA Cloning[®] Kit contains a set of primers that bind outside the poly-T cloning site in the vector (GW1/GW2). transformant colonies were suspended into 50 µl Ringer solution and were used as PCR template. The PCR reaction mixture contained 5 µl 5x PCR buffer, 5 µl 5x CES PCR enhancer buffer (Table 2.5), 3 µl 25 mM MgCl₂, 1 µl 10mM dNTP mix, and 0.25 µl 5 unit/µl high-performance Gotaq[®] G2 flexi DNA polymerase, 2 µl of each reverse and forward primers (5 pmol/ µl), and 1 µl colony resuspension. The cycling condition was as described in section 2.2.2.1. The size of the insert was checked by agarose gel electrophoresis. Transformant colonies with the correct insert size were selected for making overnight cultures for extracting the plasmid. 5 µl of each colony suspension was inoculated into 4ml LB broth containing 100 µg/ml spectinomycin and the tubes were incubated at 37°C, 200 rpm overnight. 2 ml of the overnight culture was used for plasmid DNA extraction. The plasmid extraction was done using the QIAprep[®] Spin Miniprep Kit based on the manufacturer instruction. The purified vectors were sequenced by Sanger sequencing with GW1/GW2 and/or primers that targeted the insert sequence (see section 2.2.7.1).

Plasmids carrying the desired insert were digested using the restriction enzyme *Bg*III. The inserts were then ligated into *Bg*III digested, TSAP treated pUIC3 vector using T4 DNA ligase (Rainey, 1999). Ligation reactions were carried out at room temperature overnight. 4 µl of the ligation mixture was pipetted into a thawed vial of *E. coli* S17-1_λpir chemically competent cells. All the process of transformation was similar to that described in the start of this section, except that here tetracycline was used to select for transformant colonies. The

transformant colonies were checked for the correct sized insert by plasmid extraction followed by digestion with *Bgl*III. Colonies containing the desired insert were frozen at -80°C with 45% v/v glycerol. These colonies were subsequently used as the donor strains for conjugation (see next section).

2.2.6 Conjugation

The cloned SOE-PCR fragments were used in two-step allelic exchange process to incorporate the deletion fragment to the same region of the genome of the wild type *P. fluorescens* strains. The two-steps allelic exchange process is done via bi-parental conjugation. Use of this method is well established in SBW25 (e.g. Gallie *et al.*, 2015, Gallie *et al.*, 2019), however it has not been previously used in A506 or Pf0-1. There are few differences between the conjugation process in the SBW25 and A506/Pf0-1. Each type of conjugation is explained below:

2.2.6.1 Bi-parental conjugation in A506 and Pf0-1

The conjugation started with an overnight culture of A506 or Pf0-1 as the recipient, and *E. coli* S17-1_λpir + pUIC3-insert as the donor. Cultures were grown in LB containing the appropriate antibiotics. 1 ml of the recipient and 300 µl of the donor were pelleted by centrifugation (13,000 x rpm, 1 minute). The resulting pellets were re-suspended in 200 µl fresh LB and then mixed together. The mixture was pelleted with the same centrifugation condition and re-suspended in 30 µl LB. The re-suspension was spread into a circle approximately the size of a 2-cent coin in the center of a pre-warmed LB agar plate. After the drop had dried, the plate was incubated at 28°C overnight. At the end of the incubation, the inoculum was re-suspended in 1 ml LB, and a dilution series to 10⁻⁴ made in Ringers Solution. All dilutions were plated on LB plates containing Tc, NF and X-gal (Table 2.3). This step screens for transformant colonies: recipients that have received pUIC3+insert from the donor. The pUIC3 plasmid is unable to replicate in SBW25, A506 or Pf0-1 (*i.e.*, is a suicide vector), and so the only way in which these *P. fluorescens* strains can gain resistance to Tc is to incorporate pUIC3+insert into the chromosome. This is done *via* a *recA*-dependent recombination event at a region of homology between the *P. fluorescens*-derived insert and

the recipient chromosome. NF selects against *E. coli* (*P. fluorescens* is resistant), and the Tc selects for the donor plasmid and against the recipient.

After 48 hours of incubation at 28°C, 6 to 10 blue transconjugant colonies were selected and grown in LB medium without any antibiotic to allow individuals that have passed through a second homologous recombination event - resulting in loss of the plasmid - to grow. Notably, these homologous recombination events are rare, naturally occurring events. Next, 25 µl and 100 µl of the specific dilutions (10^{-5} and 10^{-6}) of the overnight LB culture were plated on LB+X-gal plates. After 48 hours of incubation at 28°C, white colonies were identified amongst many blue colonies. These colonies were tested for Tc sensitivity. White, Tc sensitive colonies have lost the pUIC3 vector, taking with it either the original insert, or the homologous region of the chromosome. This leaves behind either the wild type genome (used for controls for effects of the engineering process), or the desired mutant, respectively. Sensitive clones were subjected to PCR with primers outside the region of manipulation, followed by Sanger sequencing across the entire region of manipulation to ensure scar-free deletion of the desired region (see Table 2.7).

2.2.6.2 Bi-parental conjugation in SBW25

The SBW25 needs to be passed through the heat shock before starting the conjugation. Exposing the bacteria to 45°C for 20 minutes destroys the enzymes which cleaves the foreign DNA. The other steps are similar to what was explained for A506 and Pf0-1. However, for SBW25, an enriched protocol can be applied (Kitten *et al.*, 1998). In this protocol the overnight culture of 6 to 10 transconjugant colonies (the outcome of the first recombination event; See last section), were combined together and 10 µl of this mixture was inoculated into 200 ml LB in a 1 L flask. After 16 hours of incubation at 28°C, 200rpm, 400 µl of this culture was transferred into a 250 ml flask containing 20 ml fresh LB. The flask was incubated with the same mentioned condition and after 30 minutes, 12.5 µg ml⁻¹ fresh tetracycline was added to the flask followed by 2 more hours of incubation in the similar condition. Tc selects for the bacteria that the pUIC3 (and the *Tc^R* genes) has been incorporated into their genome. After 2 hours, 800 µg ml⁻¹ cycloserine was added to the flask. Cycloserine kills the Tc Resistant *Pseudomonas* cells, selecting for the cells that had undergone a second round of homologous recombination. During the second round of recombination, the *Tc^R* and the *lacZ* genes are lost, thus the colonies become white. After 5 hours of growth in presence of the cycloserine, 1 ml of the flask was centrifuged to harvest

the cells followed by washing. Appropriate dilutions of the cells were spread onto LB+X-gal plates and incubated at 28°C for 48 hours. The white colonies were purified, checked for Tc sensitivity, and the accuracy of the manipulate region of their genome was checked by PCR and Sanger sequencing.

2.2.7 Sequencing

2.2.7.1 Sanger sequencing

Sanger sequencing was used to check the accuracy of the desired fragments (*e.g.* checking the insert sequence after cloning, and checking the manipulated genomic region of the mutants after the homologues recombinations). Before sequencing, the samples were purified with the appropriate kit (*i.e.* QIAquick® PCR Purification Kit for PCR products, and QIAprep® Spin Miniprep Kit for vectors that carried the insert). The quality of the purified DNA was checked by spectrophotometry prior to submitting the samples into sequencing unit of Max Planck institute for evolutionary biology (Plön, Germany). The sequencing reaction mixture was set up in an Applied Biosystems Microplate with 0.1 ml wells and it contained 0.5 µl Big Dye Termonator, 1 µl DMSO, 2 µl Big Dye buffer, 1 µl primer (5 pmol/ µl), 1 µl of the purified DNA to be sequenced, made up to 10 µl by sterile HPLC water. The cycling reaction was done in Applied Biosystems Veriti Thermal Cycler machine started by 1 minutes at 96°C, followed by 30 cycles of denaturation for 10 seconds at 96°C, annealing for 15 seconds at 56°C, and amplification for 4 minutes at 60°C. Finally, the reaction was stopped by cooling to 4°C for 5 minutes. The samples were cleaned up using the BigDye XTerminator® Purification Kit prior to the sequencing. Sanger sequencing was done using a 16- capillary 3130xl Genetic Analyzer (ThermoFisher Scientific). Resulting chromatograms were analyzed and aligned to the corresponding reference sequences using Geneious version 10.0.7.

2.2.7.2 Whole-genome sequencing

The Miseq v2 sequencer system from Illumina Solexa was used for whole-genome sequencing of the evolved mutants from experimental evolution (Section 2.2.10). Genomic DNA from purified isolates grown up in overnight cultures (KB; 28°C, 200 rpm) was extracted using DNeasy® Blood & Tissue Kit. After checking the quality of the extracted DNA by spectrophotometry and gel electrophoresis, the samples were submitted at the sequencing unit of Max Planck Institute for Evolutionary Biology for Illumina sequencing. Illumina MiSeq (500 bp, paired-end libraries, resulting in a mean coverage of 50 *per* genome base) was carried out. The output of sequencing was aligned to the *P. fluorescens* SBW25 genome sequence (Silby *et al.*, 2009) using a combination of breseq version 0.33.2 (Deatherage and Barrick, 2014) and Geneious version 10.0.7.

2.2.8 Biological assays

2.2.8.1 Colony size measurement

Colony size of particular strains on KB agar plates was measured for each mutant; single colonies from glycerol stocks of the wild types and the mutants were grown in liquid KB overnight. A serial dilution was made from each culture, and 50ul and 100ul from 10^{-5} and 10^{-6} were plated. After incubation for 48h at 28°C, plates that had roughly equal numbers of colonies (approximately 10 colonies) were selected for each strain. 5 to 10 colonies from each strain were photographed under a Leica MS5 high-performance stereomicroscope. All photographs were taken at the same magnification (x12.5). The diameter of each colony in the photograph series was measured with a ruler directly on the computer screen and recorded. This means that colony size measurements should be interpreted relatively, rather than as a quantitative measurement of actual colony size. Two-sample *t*-tests were run to test for a difference in mean colony diameter between strains.

2.2.8.2 Growth curve analysis

In order to analyze the growth of strains under various media and temperature conditions, a microtitre plate reader (BioTek® from Epoch) was used to record optical density as a measure of bacterial growth. That is, at least five replicate single colonies of each strain were grown in overnight pre-cultures. 2 µl of these pre-cultures were used to inoculate 198 µl of fresh media in a well of a 96-well plate. OD₆₀₀ was measured every five minutes for 72 hours (at 28°C, with 5 seconds of shaking before each read). The mean and standard error of OD₆₀₀ over time was calculated for each strain and condition, and plotted as a growth curve using Microsoft Excel 2011. Gen5 analysis software (Epoch) was used to calculate the maximum growth rate (V_{max}), lag time and time to V_{max} of each replicate, using a sliding window of 9 time points. These data are used for testing for differences in growth. Also, of each OD₆₀₀ graph. The mean and standard error of the replicates was calculated for each read.

2.2.8.3 Competition assays

In competition assays, the growth of two bacterial strains is compared directly by growing the strains in the same tube (starting from roughly equal numbers of each). The change in the starting and finishing ratios of the two strains is used as a measure of relative fitness. In cases where the two competitors form colonies that are morphologically indistinguishable, then each strain should be competed against SBW25-*lacZ*, and the competition results compared indirectly. SBW25-*lacZ* is *P. fluorescens* SBW25 carrying a neutral *lacZ* marker (Zhang and Rainey, 2007). This marker means that SBW25-*lacZ* colonies are blue on LB+X-gal plates, while SBW25 colonies are white. Thus, we can reliably use it as a morphologically distinguishable competitor.

To compare each tRNA deletion mutant to its wild type strain, three types of competition experiments were set up:

- A. SBW25-*lacZ* vs wild type
- B. SBW25-*lacZ* vs engineering control wild type
- C. SBW25-*lacZ* vs tRNA deletion mutant

Each competition contained five to eight replicates. From the glycerol stocks of each competitor, eight single colonies were isolated on KB agar plates. An overnight culture was

made from each single colony in KB broth (16 hours, 28°C, 200 rpm). Each tube was vortexed for 1 minute before mixing each competitor in a 1:1 ratio. Generally, this was achieved by mixing 1:1 volumes of the competitors; in cases where this did not result in a 1:1 ratio of colony types, the colonies per unit volume were determined for each competitor and mixing volumes adjusted accordingly. A 10^{-3} dilution of the 1:1 mixture was inoculated into fresh KB cultures (these are the T0 tubes). Aliquots of these tubes were dilution plated on LB+X-gal plates to determine the starting ratios (T0 ratios), and then the competition started by incubation of the T0 tubes at 28°C, 200 rpm for 24 hours. To determine the T24 ratio, the overnight competition tubes were vortexed for 1 minute and appropriate aliquots were plated on LB+X-gal plates (T24). The tubes were replaced at 28°C, 200 rpm, and the incubation continued until 48 hours. The plating process was then repeated to obtain the T48 colony counts. After 48 hours of incubation of LB+X-gal plates from each time point, blue and white colonies were counted for each competition tube at each time point. These raw counts were used to calculate changes in the relative frequencies and the Malthusian fitness of each genotype (See the statistical analyses section; Lenski *et al.*, 1991).

2.2.9 Statistical analyses

2.2.9.1 General statistical tests

For all the results with multiple replicates, the means, standard deviation, and standard errors were calculated using Microsoft Office Excel version 16.9 or the R studio version 1.2.1335 (using *s20x* and *coin* packages; R core team, 2017). For comparing data from two different sets, two-sample *t*-tests were used. This test assumes independent data points, equal variances, and normal distribution of the data set. In case of violation of any of the mentioned assumptions, less conservative non-parametric tests were used (*i.e.* Welch test and Mann-Whitney-Wilcoxon tests were used in cases of violation of the equal variances and normal distribution respectively). For comparing the data across three or more groups, one-way ANOVA was applied. This test carries the same assumptions as the *t*-test. If the equal variance assumption was not met, a Kruskal-Wallis one-way ANOVA was used.

2.2.9.2 Calculation of relative fitness in competition assays

The Malthusian parameter formula (Lenski *et al.*,1991) was applied to calculate the relative fitness of each strain to the SBW25-*lacZ* based on the following equations:

The exponential equation of population growth is shown in equation 1 where N_0 is the initial population size and the N_t is the population size after time t . The constant value of r is called as Malthusian parameter, which in turn can be calculated based on equation 2

$$N_t = N_0 e^{rt} \quad (\text{Equation 1})$$

$$R = \text{Ln} \left(\frac{(\text{Number of counted colonies} \times \text{dilution factor} \times \text{volume}) \text{ for } T_n}{(\text{Number of counted colonies} \times \text{dilution factor} \times \text{volume}) \text{ for } T_0} \right) \quad (\text{Equation 2})$$

After counting the blue and white colonies in each set of competition, the Malthusian parameters were calculated based on equation 2 and then the relative fitness of each strain was calculated based on equation 3:

$$\text{Relative fitness(RF)} = \frac{R_{\text{strain of interest}}}{R_{\text{SBW25-LacZ}}} \quad (\text{Equation 3})$$

Finally, the appropriate statistical analyses were used to calculate the significance between the relative fitness of the different groups and the results were plotted using R Studio.

2.2.10 Experimental evolution

Experimental evolution was carried out for five independent lineages of each of *P. fluorescens* SBW25 (wild type) and the quadruple tRNA deletion mutant SBW25 Δ GluGlyGluGly (see Table 2.1). The experiment was run in sterile 50 ml FalconTM tubes under conditions that promote rapid growth (10 ml KB, 28°C, 200 rpm). Every 24 hours, the cultures were vortexed for 1 minute, and 100 μ l (1%) of each population was transferred to 10 ml of fresh medium (Figure 2.1). Every day, a 1 ml sample of each population was stored at -80°C for future reference. At regular intervals, growth characteristics and colony morphology on KB agar were checked for each population. This was to (i) check for obvious contamination, and (ii) monitor phenotypic changes. The OD₆₀₀

of each populations was measured every day at the time of transfer, 8 hours after the transfer, and 24 hours after the transfer.

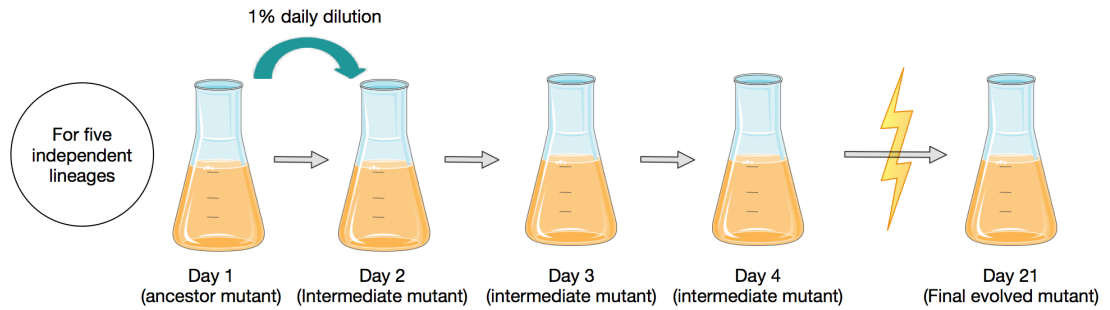


Figure 2.1. Schematic diagram of experimental evolution.

In order to calculate the approximate number of generations experienced in a 24 h growth cycle, the mean CFU (colony forming unit) of three populations of the wild type plus three populations of the mutant from T₀ and T₂₄ time points was calculated and the values were applied to equation 4, where the X is the generation time and N₀ and N_t are the average CFU for T₀ and T₂₄ respectively (Todar, 2008).

$$2^x \times n_0 = n_t \quad \text{or} \quad X = \frac{\log N_t - \log N_0}{\log 2} \quad \text{(equation 4)}$$

3. Results

3.1 Transfer RNA gene sets vary across *P. fluorescens* strains

The genome sequences of eight *P. fluorescens* strains were obtained from NCBI on April 2017. These eight strains are those for which tRNA gene set annotations are currently available in the genomic tRNA database (GtRNAdb; see below). The genome sequences were aligned together using *andi*, which is an algorithm that is designed for calculating the pairwise distances between closely related genomes (Haubold *et al.*, 2015), and a phylogenetic tree was constructed by applying PHYLIP neighbor joining program (Figure 3.1). This tree is used as a reference for degree of strain relatedness in subsequent sections.



Figure 3.1. Phylogenetic tree of eight *P. fluorescens* species. The tree was constructed using whole genome sequences and the neighbor joining method.

All available tRNA annotations in the genomic tRNA database (GtRNAdb) for *P. fluorescens* strains were downloaded (April 2017); this consisted of a total of eight strains

(those shown in Figure 3.1). Additionally, one set of *E. coli* and of *P. aeruginosa* tRNA gene set sequences were included in the analysis as outgroups (Lowe and Eddy, 1997). The genomic features of these strains indicate that *P. fluorescens* strains have various genome sizes (5.96 to 6.99 Mb), but all with similar GC contents compared to the outgroup strain from the same genus (*P. aeruginosa* PAO1) or from the outgroup strain out of the genus (*E. coli* MG1655; Table 3.1).

Table 3.3. Genome features across the ten bacterial strains included in the tRNA gene set analysis.

Strain	Genome size (Mb)	GC content (%)
PCL1751	6.14	60.4
PICF7	6.13	60.4
SBW25	6.72	60.5
A506	5.96	60
UK4	6.06	60.1
NCIMB	6.99	59
Pf0-1	6.43	60.5
F113	6.84	60.8
<i>P. aeruginosa</i> PAO1	6.26	66.6
<i>E. coli</i> MG1655	4.64	50.8

The first eight strains in the list belong to the species *P. fluorescens*. The strains are ordered based on their phylogenetic order in figure 3.1

Next, analysis of the GtRNAdb-annotated set of tRNA genes present in each of the eight *P. fluorescens* strains revealed that the total number of tRNA genes varies between 66 and 74 (Figure 3.2).

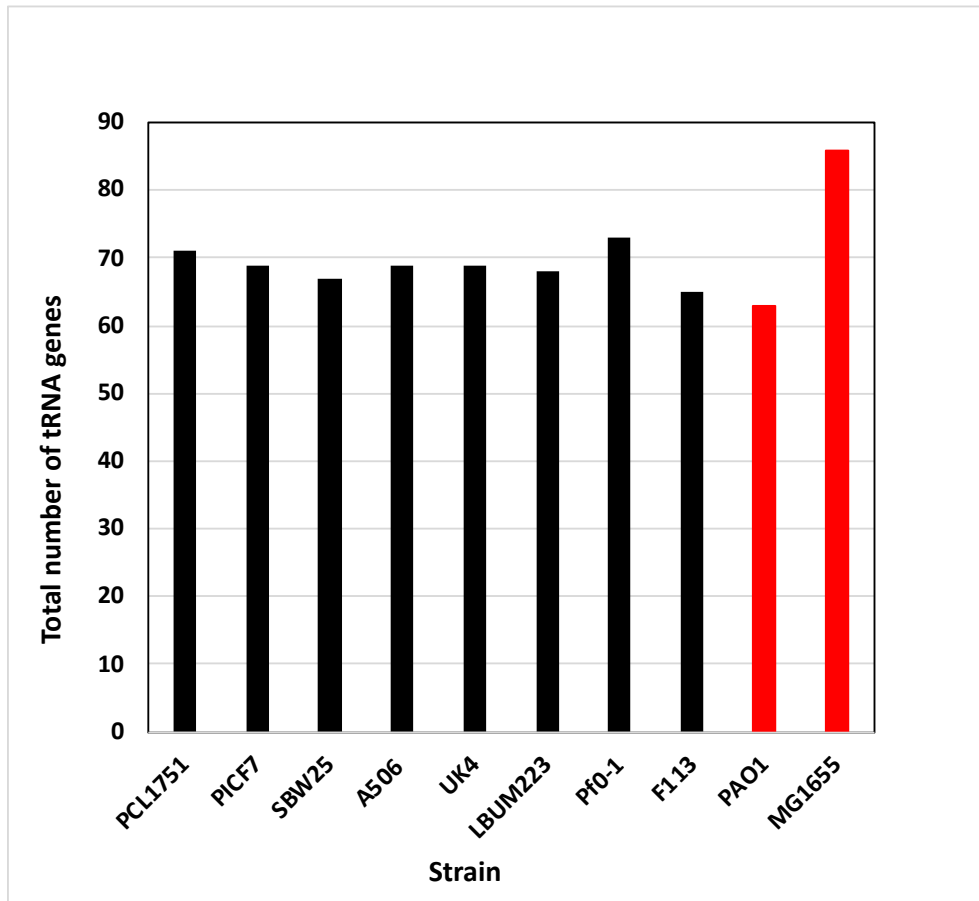


Figure 3.2. The total number of tRNA genes varies between 66 and 74 among *P. fluorescens* strains. Strains are ordered from left to right based on their phylogenetic order. The two outgroups are colored in red.

The number of every unique tRNA isotype was counted in each genome; the following barplots show the number of each tRNA isotypes in the eight selected *P. fluorescens* strains (Figure 3.3). From the barplots it can be concluded that 21 tRNA species are absent from all eight of the *P. fluorescens* genomes. For example, none of the eight genomes encode tRNA-Leu^{AAG}. Each genome has up to 40 unique tRNA isotype with one to six copies in the genome. Among these 40 tRNA isotypes, 39 of them are present in all eight strains. There are two isotypes that are present in only a subset of the *P. fluorescens* strains: tRNA-Lys^{CUU} and tRNA-Sec^{UCA}; tRNA-Lys^{CUU}, is only present in two closely related strains (PICF7 and PCL1751; Figure 3.1), while tRNA-Sec^{UCA} is only present in Pf0-1. The tRNA-Sec^{UCA} is an unusual tRNA – it carries the amino acid selenocysteine (the 21st proteinogenic amino acid; Stadtman, 1980). Selenocysteine is an unusual amino acid: only a small number of proteins are known to incorporate it, and it is – to date – totally absent from plant and fungi kingdoms (Copeland, 2005). The selenocysteine tRNA is a suppressor tRNA (Leinfelder *et al.*, 1988); in presence of specific sequences in mRNA, it reads through TGA stop codons, providing instead the selenocysteine amino acid (Copeland, 2005).

Though the selenocysteine tRNA is observed in many bacteria (Zhang *et al.*, 2005) including many *E. coli* strains and the two outgroups of this study, but among the *P. fluorescens* it is only seen in Pf0-1 in this study.

The observations about the subset of tRNAs in *P. fluorescens* strains aligns well with what was mentioned in the introduction; due to a combination of wobble base pairing and post transcriptional modifications only a subset of the 61 possible tRNAs (usually less than 50; Rak *et al.* 2018) is enough to decode all the universal 61 codons (Crick, 1966).

As the plots in Figure 3.3 show, there exist some commonalities among tRNA species with multiple copies in all eight *P. fluorescens* strains. Notably, there are five tRNA species with four to six copies in all the selected genomes. Across all strains, there is five or six copies of tRNA-Ala^{UGC} and tRNA-Ile^{GAU} in all eight *P. fluorescens* genomes. Looking at the genomic region of these genes shows that they are present as doublet tRNA genes in middle of ribosomal RNA (*rrn*) operons (see Figure 1.5-d in the introduction). The tRNA-Glu^{UUC} gene is present in four or five copies in each strain, while the tRNA-fMet^{CAU} and the tRNA-Gly^{GCC} gene has three to four copies in the genomes. From the finding above it can be concluded that the tRNA gene complements of *P. fluorescens* strains are highly similar in the tRNA species category and they differ principally in copy number. Typically, this difference in number does not exceed one additional copy.

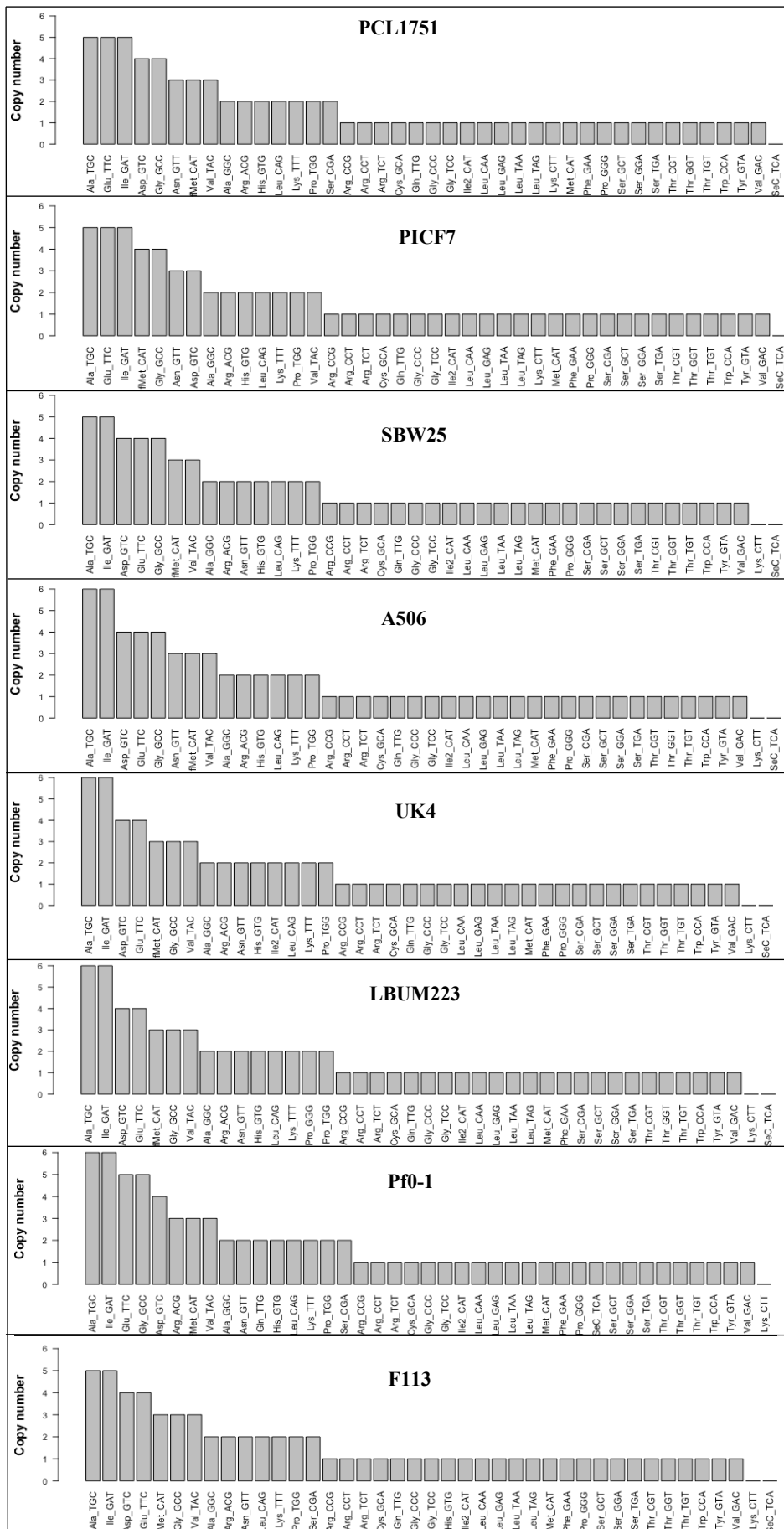


Figure 3.3. There is a similar pattern for tRNA gene copy number among eight selected *P. fluorescens* strains. Strains are ordered from up to down based on their phylogenetic order.

The primary sequences of these tRNAs 39 tRNA isotypes that are in common in the eight strains show that many of them have 100% identity; 17 out of the 39 have 100% homology between and within the strains, and difference between the other copies of tRNA mostly is not exceeding 10% - Supplementary Table S1). Thus, it can be hypothesized that multiple copies of a specific tRNA may be the product of gene duplication of a common ancestral gene. However, these identical copies are not always syntenic; for example, the four identical copies of tRNA-Glu^{UUC} are in similar genomic regions in A506, Pf0-1, and SBW25, but in the other five strains they are found in completely different regions of the genome.

The most diverse primary sequence belongs to tRNA-Asn^{GUU}. There are two or three copies of this tRNA gene in each of the eight *P. fluorescens* genomes. However, in every case, there appears to be at least two unique primary sequences of this tRNA *per* genome. The first gene copy is 100% identical across all the genomes, but the second copy is quite different both within and between the strains with 70 to 97% homology to the first one. This characteristic makes this tRNA a good model for comparing the expression level of each copy of tRNA genes of the same isotype (because they can be differentiated when deep sequencing the mature tRNA pool).

After systematic dissection of the tRNA complements of the *P. fluorescens* strains, three were selected for further studies with genetic manipulation and experimental evolution approaches (see section 3.2).

3.2 Expanding tRNA gene sets in *P. fluorescens* can be reduced by gene deletion

Three strains of *P. fluorescens* were selected for genetic engineering (SBW25, A506, Pf0-1). Looking at the tRNA gene complement of these strains shows that there is a set of 66 “core” tRNA gene all of them. That is, there are 66 tRNA genes that are present in all three strains. On top of this core set of 66, A506 and Pf0-1 have three and seven “accessory” tRNA genes respectively (Figure 3.4). These are tRNA genes that are absent from the SBW25 genome.

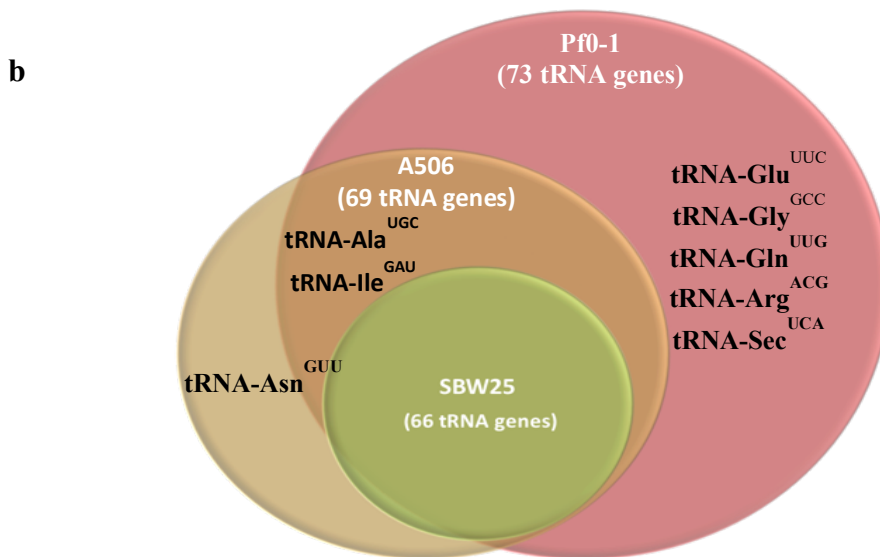
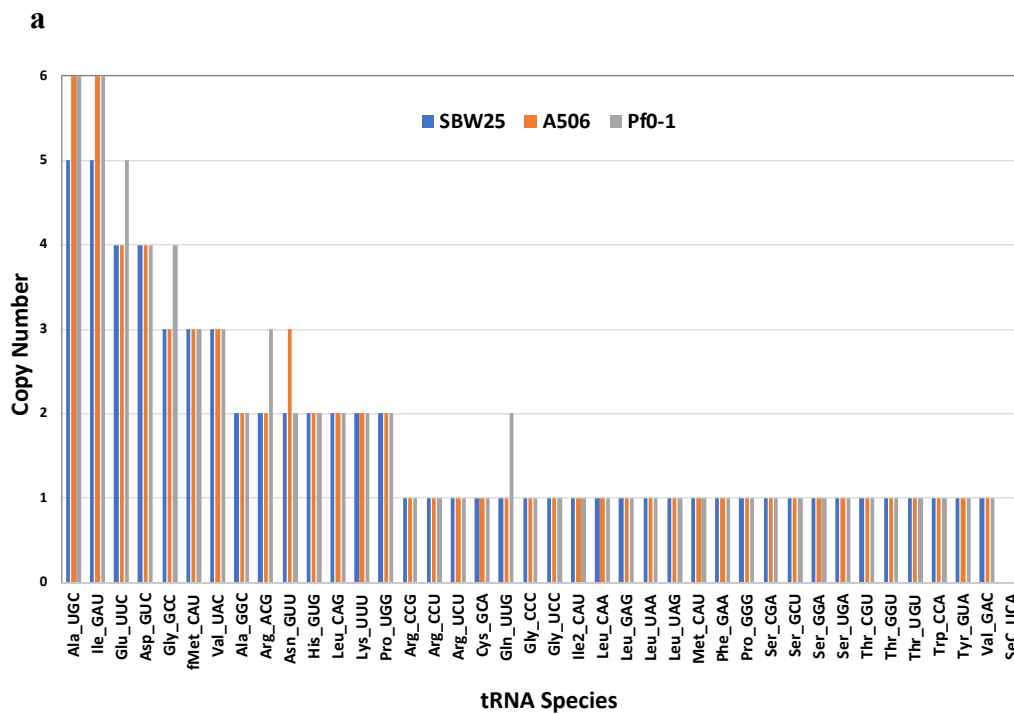


Figure 3.4. a) Expanding tRNA gene sets in *P. fluorescens* SBW25, A506 and Pf0-1. b) 66 tRNA genes are considered core genes (smallest circle) because they exist in all three genomes. A506 (second largest circle) has three accessory genes and Pf0-1 (largest circle) has seven accessory genes.

From the comparison of the copy number of the accessory tRNAs - which have one to six identical copies in the A506 and Pf0-1 genome - it was hypothesized that these genes may be functionally redundant. To test this hypothesis, an experiment was designed which involved sequential deletion of all accessory tRNAs from A506 and Pf0-1, so that the tRNA

complement of these two strains become similar to that of SBW25. To do this, SOE-PCR was used for constructing tRNA deletion fragments (Section 2.2.2.2). The pUIC3-ligated SOE-PCR fragments were used in two-step allelic replacement, which was done via biparental conjugation (Section 2.2.6.1). The colonies that have passed through the second rare recombination event (2XOs) were sequenced using primers that anneal the outside of the manipulation region in the genome (see section 2.2.7.1). Figures 3.5 and 3.6 show the alignment of the 2XO colonies with the corresponding wild type sequence and the deleted tRNA genes.

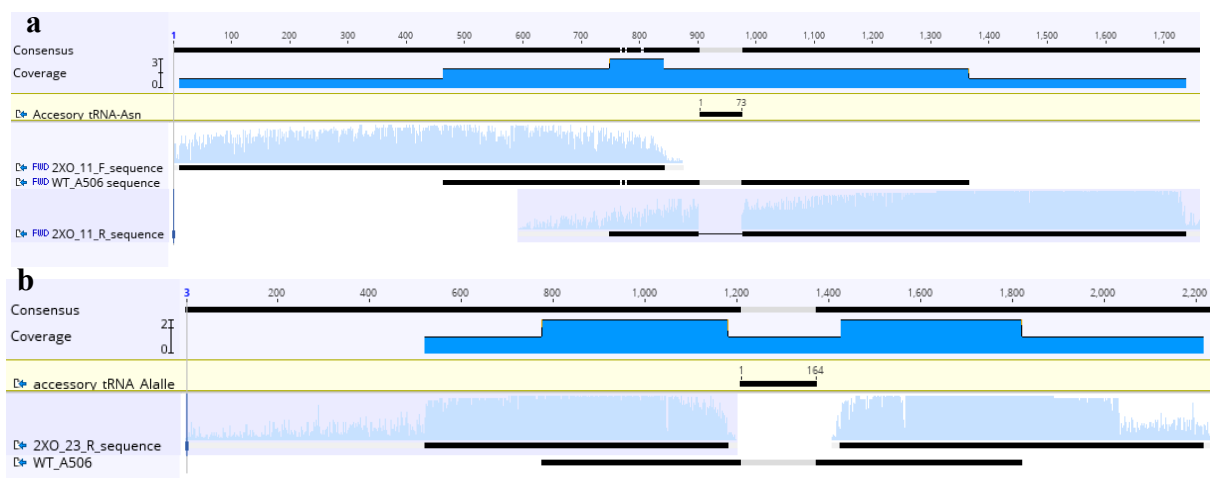


Figure 3.5. Alignment of the 2XO colonies sequencing results with wild type *P. fluorescens* A506 (the WT_A506 line in the figure) and the tRNA sequence indicates clean deletion of the tRNA from the mutant (the sudden drop in the coverage or the gaps right in start of the tRNA gene sequence is evidence for deletion of that gene; **a**) tRNA-Asn^{GUU} gene **b**) simultaneous deletion of tRNA-Ala^{UGC} and tRNA-Ile^{GAU} gene).

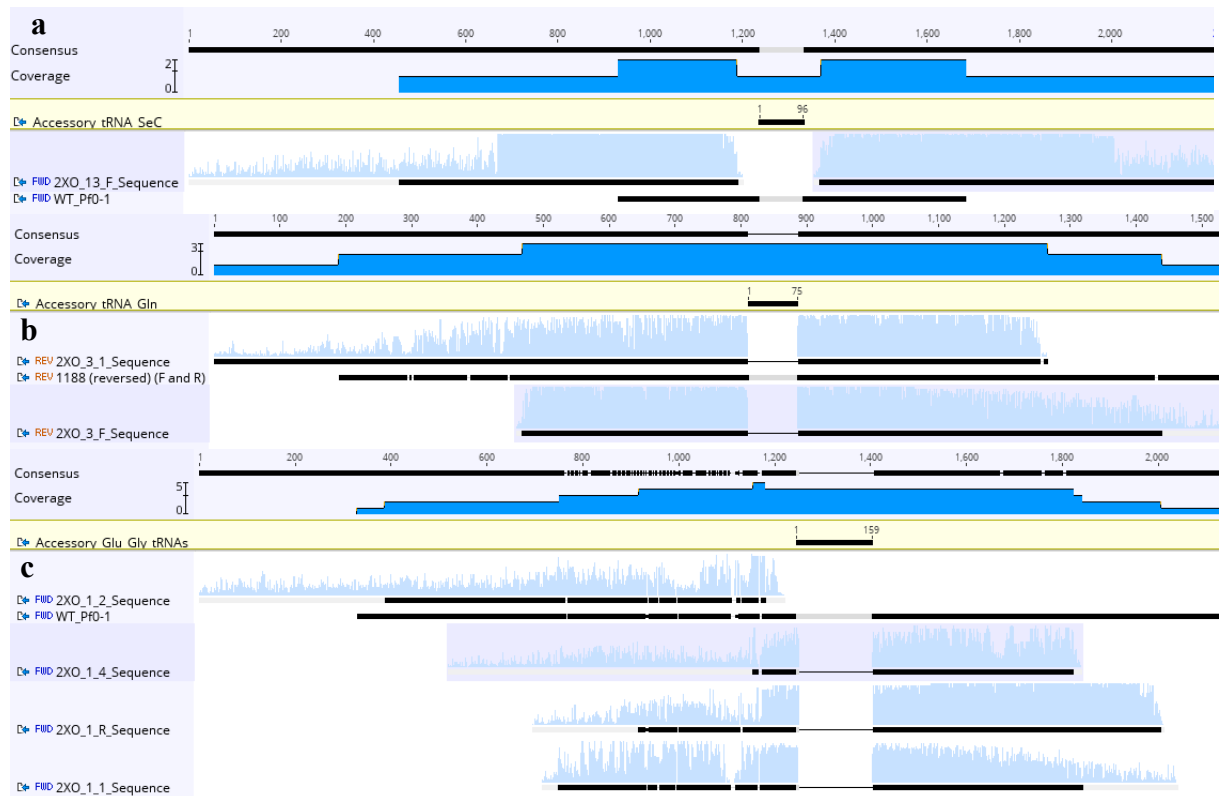


Figure 3.6. Alignment of the 2XO colonies sequencing results with the wild type *P. fluorescens* Pf0-1 (the WT_Pf0-1 line in the figure) and the tRNA sequence indicates clean deletion of the tRNA from the mutant (the sudden drop in the coverage or the gaps right in start of the tRNA gene sequence is evidence for deletion of that gene; **a**) tRNA-Sec^{UCA} gene **b**) tRNA-Gln^{UUG} gene **c**) simultaneous deletion of tRNA-Glu^{UGC} and tRNA-Gly^{GCC} gene). Note that deletion of tRNA-Ala^{UGC} and tRNA-Ile^{GAU} doublet and tRNA-Arg^{ACG} from Pf0-1 background was not successful.

After Sanger sequencing confirmation, two copies of each mutant were stored in glycerol saline stock 45% (v/v) in -80°C. The mutants were next subjected to systematic phenotypic analyses. The results of the analyses are explained with more details in next section.

3.3 Reduction of tRNA gene sets in *P. fluorescens* leads to limited phenotypic effects

For clarity, the phenotypic analyses are divided into two sections: those for the A506 background (section 3.3.1), and those for the Pf0-1 background (section 3.3.2):

3.3.1 Phenotypic analysis of the A506 tRNA deletion mutants

By deletion of all three “accessory” tRNA genes, the tRNA gene complement of SBW25 was successfully constructed in the A506 genome. Table 3.2 shows the list of all A506 tRNA deletion mutants constructed during this process. The mutant with the greatest number of tRNA deletion is called the “final mutant” and the rest are called “intermediate mutants”. The engineering control is an A506 that has passed through the engineering process for deletion of one of the tRNAs, but the sequencing of the manipulated area shows that it still has the wild type sequence. Some of the phenotypic assays including the colony size measurements and competition test were carried out only for the final strain, with the expectation that the strain with the greatest number of deleted tRNA genes is the most likely to show differences compared to the wild type A506. However, testing of growth in different media and temperatures was carried out for all the strains.

Table 3.2. Constructed tRNA deletions in A506 background

mutant	Number of deleted tRNA genes	explanation
A506 Δ Asn	1	Intermediate mutant
A506 Δ AlaIle	2	Intermediate mutant
A506 Δ AsnAlaIle	3	Final mutant
A506_EC	0	Engineering control

Initially, as a rudimentary measure of fitness, mutant and wild type colony size and shape of the final mutant was analyzed, and compared with that of A506. An assay was performed where the results show that there is no significant difference in the colony size of the mutant compared to the wild type (Figure 3.7).

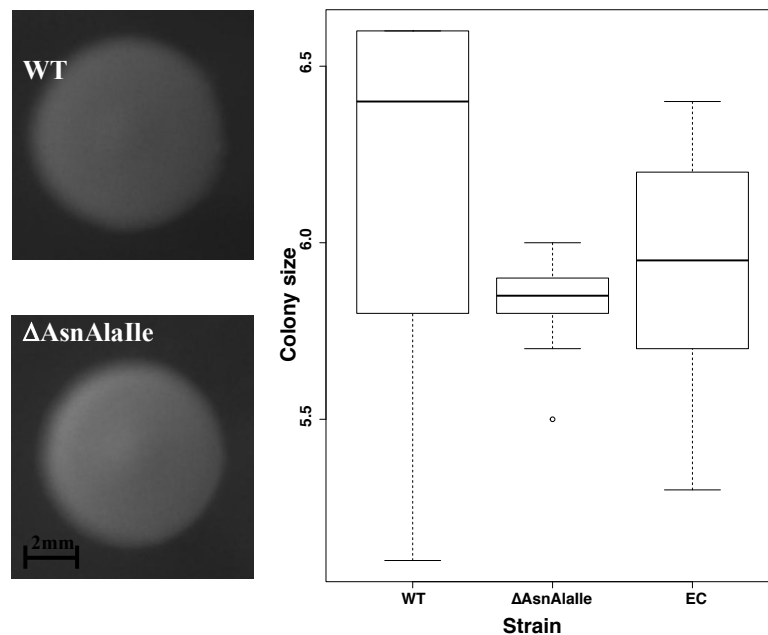


Figure 3.7. Deletion of three tRNA genes does not changes colony morphology in A506. Left: colony size comparison in wild type A506 (up) and A506 Δ AsnAlaIle (down). Colonies grown on KB agar at 28°C for 48 hours and the photos were taken under 12.5X magnification. Right: measurement of colony diameter from plates with equal numbers of colonies showed no significant difference between the mutant, wild type, and the engineering control (Kruskal-Wallis rank sum test p -value = 0.1384; WT= wild type A506, Δ AsnAlaIle= final mutant of A506, EC= engineering control of A506).

Next, the growth curve of the mutants was compared in full and minimum media. Terrific broth (TB) and King’s medium B (KB) were used as full media; KB is a complete medium for isolation of bacteria (King *et al.*, 1954). Certain pseudomonads including *P. fluorescens* produce the pyoverdine pigment that fluoresces under longwave UV light (Meyer and Abdallah, 1978). This media contains tryptone and glycerol, which provides the opportunity for fast growth of bacteria. TB contains an increased concentration of tryptone, glycerol, and yeast extract as a carbon source, which selects for faster growth of bacteria (Sambrook *et al.*, 1954). As the graphs in Figure 3.8 show, there is no significant difference in growth rate of the mutant versus the wild type in either of these rich media. Thus, the growth rate of the wild type and final mutant was assayed in a more challenging medium: M9. This medium contains minimal salts, and glucose as the sole carbon source (Sambrook *et al.*, 1954). The statistical analysis of the growth parameters (including the OD600 over time and the maximum growth rate) indicates that there is no significant difference between the final and intermediate mutants and the wild type *P. fluorescens* A506 (Figure 3.8).

a

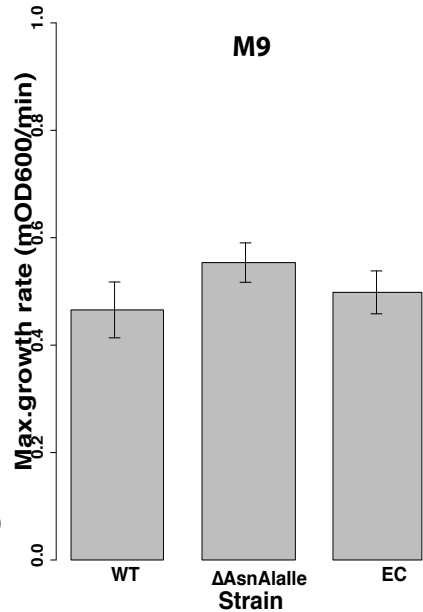
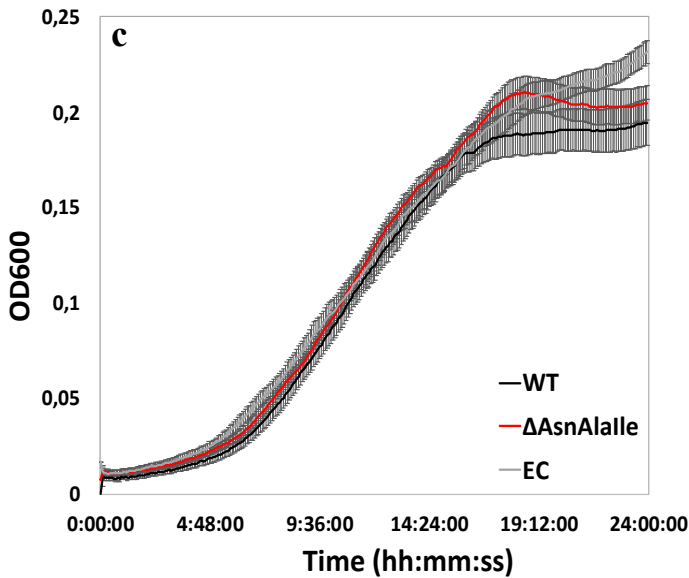
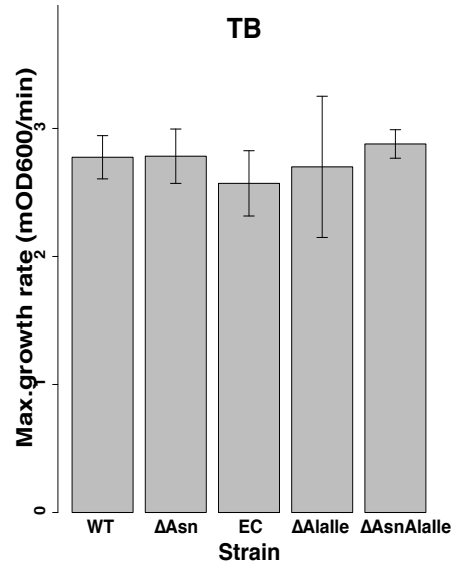
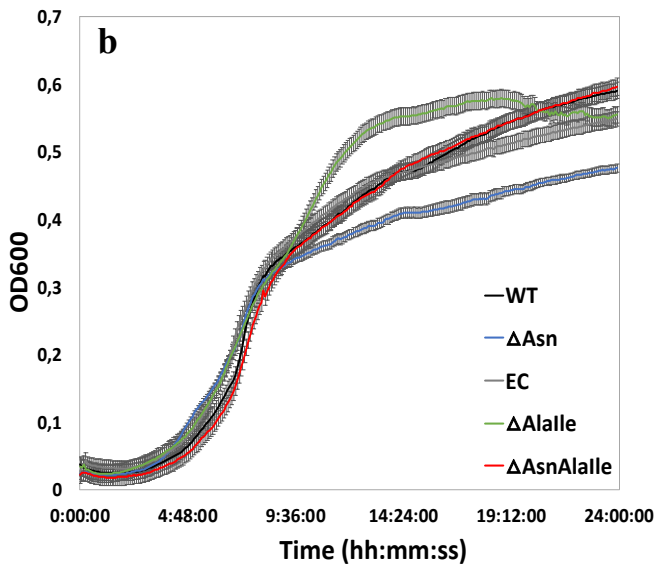
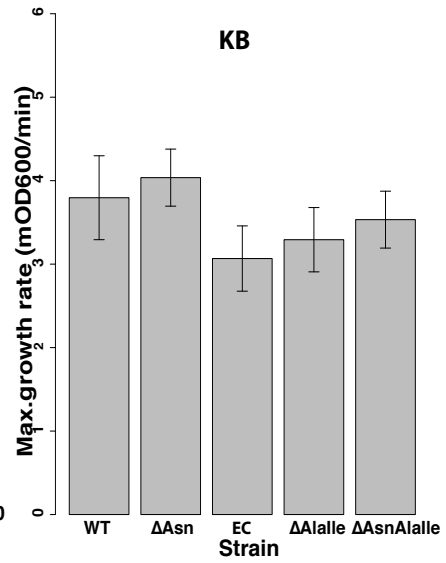
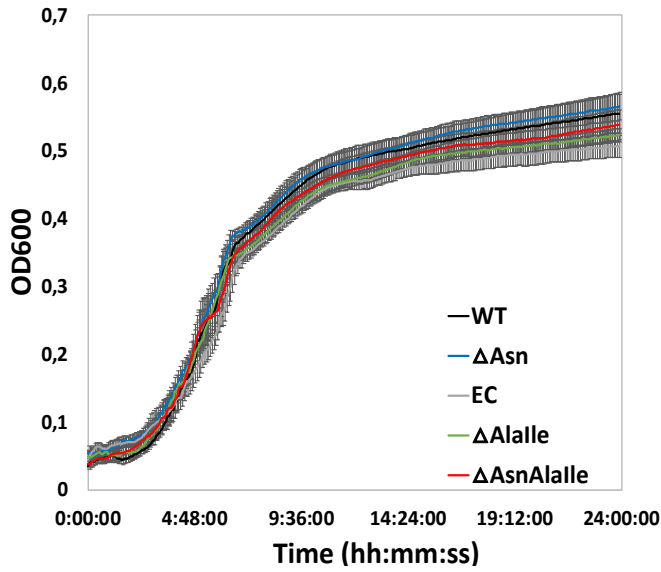


Figure 3.8. Three tRNA deletions have no significant effects on growth of *P. fluorescens* A506 in the conditions tested. 24 h growth curves of A506 versus the mutants in shaken medium - a) KB b) TB c) M9 - at 28°C. Measurements were taken every 5 min, with eight replicates for each strain (using the uninoculated medium as blank). The mean maximum growth rates were calculated using a sliding window of nine data points. (one-way ANOVA test was done for comparison of the variances of maximum growth rate in each strain for all media; p-value>0.05 for all comparisons, Error bars in all graphs are one standard error).

After not observing any significant growth difference in TB, KB or M9, the growth rates of the wild type and the mutants was assayed in in KB (as the most commonly used medium of the three for *P. fluorescens* SBW25 in the lab) in a variety of temperatures (Figure 3.9). From the growth analyses in different media and temperatures, it can be concluded that the deletion of three accessory tRNAs did not change the maximum growth rate of the mutants. As it is obvious from figure 3.9-b, the final OD600 of the A506 Δ AlaIle mutant in growth at 32 °C is visibly lower than the other mutants (and the wild type). Also, the Vmax of this mutant is smaller than others but it is not statistically significant. Regarding the fact that the A506 Δ AsnAlaIle did not show any significant Vmax or final OD600 compared to the wild type, it is unprobeable that the observed difference between the A506 Δ AlaIle and the wild type A506 is reliable (It is noteworthy that the Δ Asn deletion has been introduced to the A506 Δ AlaIle to construct the final A506 Δ AsnAlaIle mutant, thus it is expected that any probable fitness defect in the A506 Δ AlaIle should also be detected in the A506 Δ AsnAlaIle, but as the graphs show this is not the case). However, for solving this discrepancy, more repeats of growth curve assays and the competition of the A506 Δ AlaIle with the wild type A506 is suggested (In this study only the final A506 Δ AsnAlaIle mutant have been competed to the wild type).

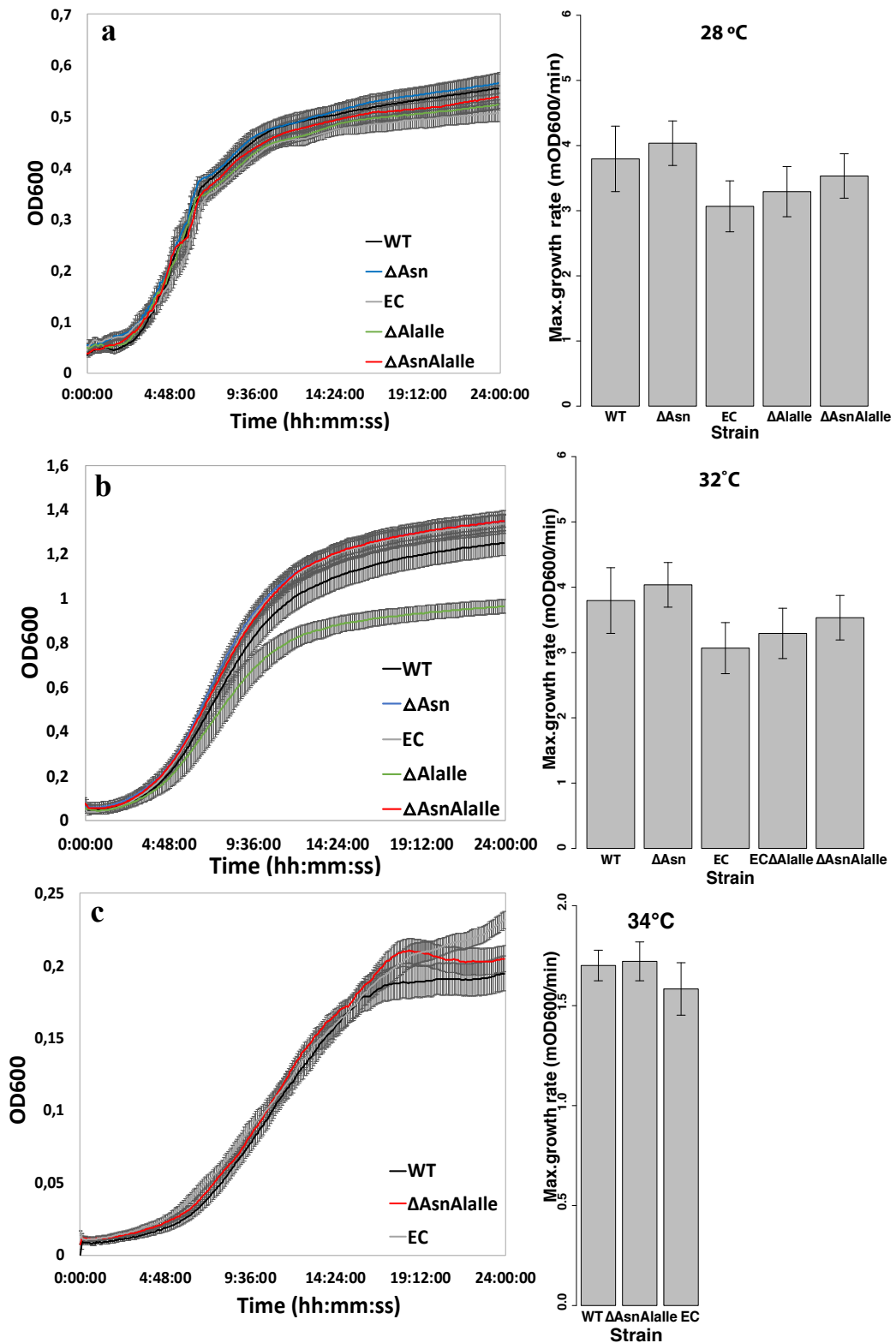


Figure 3.9. There is no difference in the maximum growth rate of the mutants with one to three tRNA deletions and the wild type *P. fluorescens* A506 in the tested range of temperature - from 28 °C to 34 °C; a) 28 °C, b) 32 °C, c) 34 °C. One-way ANOVA test was done for comparison of the variances of maximum

growth rate in each strain for all temperatures; p -value >0.05 for all comparisons, Error bars in all graphs are one standard error.

Next, an indirect competition assay between the mutants and the wild type was carried out; each strain was competed against the neutrally marked wild type strain, *P. fluorescens* SBW25-*lacZ*. The competition was done in eight replicates. Each replicate competition was inoculated with a low density of SBW25-*lacZ* and the strain of interest in a 1:1 ratio, and allowed to grow at 28°C, 200 rpm for 24 hours. After 24 hours, the colony forming units (CFU) and the Malthusian parameter formula (Lenski *et al.*, 1991) were applied to calculate the relative fitness of each strain to the SBW25-*lacZ* based on the equations explained in section 2.2.9.2. Based on the Malthusian parameter value of the mutant compared to the wild type, it can be concluded that the deletion of three tRNA genes from A506 genome has not affect the fitness and there is no significant difference between the relative fitness of the mutant and the wild type to the *P. fluorescens* SBW25 *lacZ* (Figure 3.10).

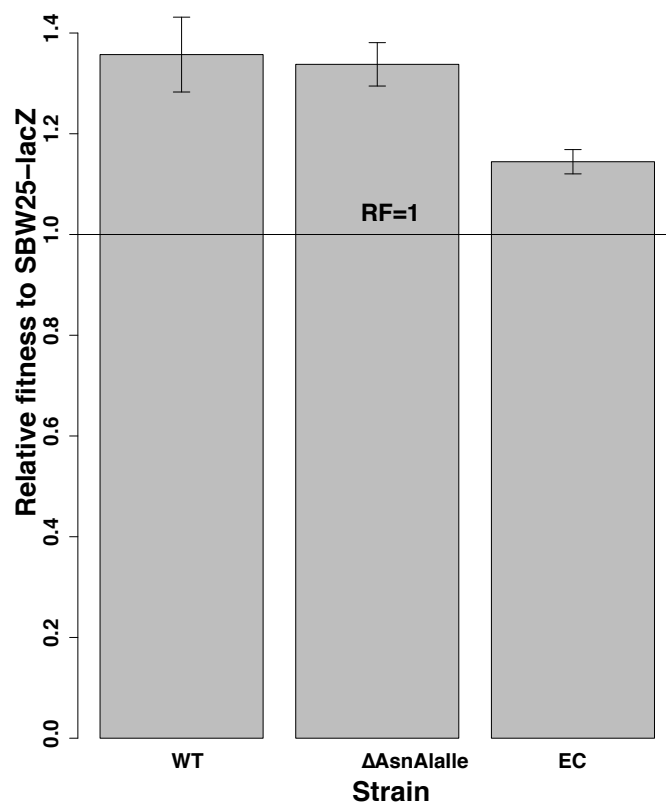


Figure 3.10. There is no difference in the relative fitness of the final mutant (A506 Δ AsnAlalle) compared to the wild type A506 (pairwise comparisons for WT and Δ AsnAlalle using Wilcoxon rank sum test, p -value =

0.46639). The relative fitness above 1 shows higher fitness than the SBW25_ *lacZ* strain. (Competition assays were performed with 1:1 starting ratio for 24h, see supplementary materials for the colony numbers of each competitor).

Generally, all the *P. fluorescens* A506 strains are growing faster than the *P. fluorescens* SBW25_ *lacZ*. Thus, in the competition experiment the relative fitness of all A506 strains is more than one (RF>1). The expectation from the competition of A506 was that the wild type and engineering control have lower fitness than the final mutant (A506ΔAsnAlaIle). However, from figure 3.10 it can be concluded that there is no significant difference between the relative fitness of the wild type and the mutants (see Appendixes tables S3 to S6 and figure S1 for more details of the statistics). Hence, the final conclusion here is that deletion of three accessory tRNAs (one out of three tRNA-Asn^{GUU}, one out of six tRNA-Ala^{UGC}, and one out of six tRNA-Ile^{GAU}) from *P. fluorescens* A506 genome does not result in a discernible phenotype.

3.3.2 Phenotypic analysis of the Pf0-1 tRNA gene deletion mutants

Several intermediate strains with an increasing number of deleted tRNA genes were constructed in *P. fluorescens* Pf0-1 (Table 3.3).

Table 3.3. Constructed tRNA gene deletions in the Pf0-1 background

Mutant	Number of deleted tRNAs	explanation
Pf0-1ΔASec	1	Intermediate mutant
Pf0-1ΔSecGln	2	Intermediate mutant
Pf0-1ΔSecGlnGluGly	4	Final mutant
Pf0-1_EC	0	Engineering control

The primary aim for this section was deletion of all the seven accessory tRNA genes in Pf0-1, to make the tRNA complement similar to that of SBW25. However, due to technical complexities, the construction of this strain was not successful. The explanation for this complexity could be the repetitive region of the tRNA genes to be deleted; The tRNA-Ala^{UGC} and tRNA-Ile^{GAU} doublet is located among the highly repetitive rRNA operons which make the allelic replacement very error prone. It is notable that deletion of this doublet from A506 background was only successful after many attempts and due to time constraints

the attempts for construction of this deletion in Pf0-1 was stopped. The tRNA-Arg^{ACG} is also located in close proximity to three other tRNA genes (see figure 1.5 in the introduction) and due to highly conserved sequences of tRNA genes, the homologues recombinations are very error prone. The strain with the greatest number of deletions is Pf01 Δ SecGlnGluGly (*i.e.*, with four of the seven accessory tRNAs deleted), and this strain is considered the “final mutant” in this section. This means that the phenotypic assays focused on this mutant, with the expectation that this mutant would be most likely to show phenotypic differences from the Pf0-1 wild type. As for A506 in the section above, fitness assays started with checking colony morphology in KB plates. As shown in figure 3.11, there is no statistically significant difference between the colony size of the final mutant and the wild type (and the engineering control), under the conditions tested.

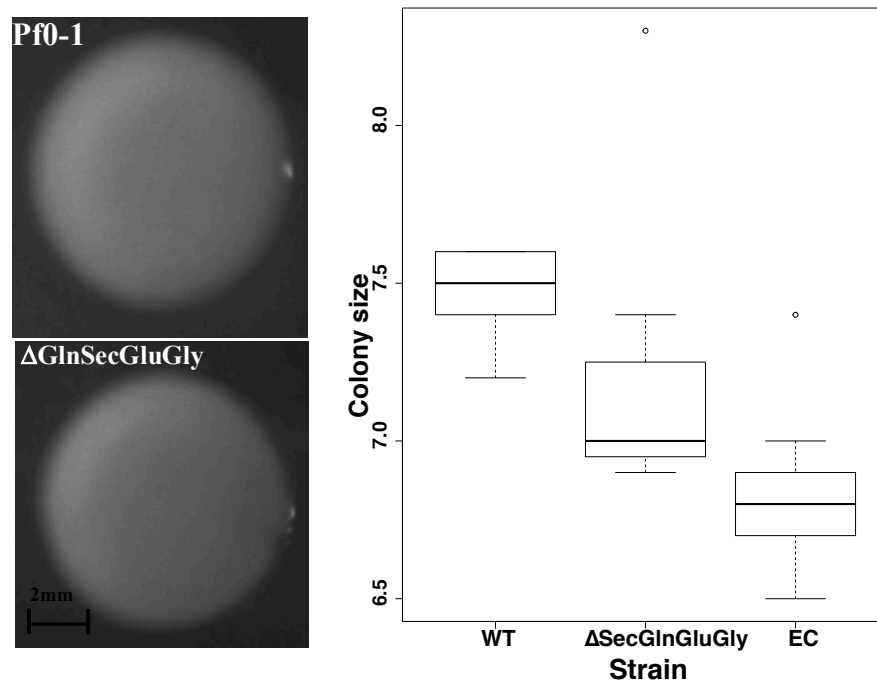


Figure 3.11. Deletion of four accessory tRNA genes does not change colony morphology in *P. fluorescens* Pf0-1 under the conditions tested. Left: colony size comparison in wild type Pf0-1 (up) and Pf0-1 Δ GlnSecGluGly (down). Colonies grown on KB agar at 28°C for 48 hours and the photos were taken under 12.5X magnification. Right: measurement of colony diameter from plates with equal numbers of colonies showed no significant difference between the mutant and the wild type (Pairwise comparisons using Wilcoxon rank sum test between the wild type and the mutant; p -value = 0.08978); WT= wild type Pf0-1, Δ SecGlnGluGly= final mutant of Pf0-1, EC= engineering control of Pf0-1).

Figure 3.12 shows growth of the final mutant, Pf0-1 Δ GlnSecGluGly, in a variety of media. The mutant does not show any difference in growth rate (in terms of maximum OD₆₀₀ in logarithmic phase) than either the wild type Pf0-1 or the engineering process control (EC). The growth rate was also assayed in KB in the same variety of temperatures as for A506. As the figure 3.13 shows, no significant difference was detected in the maximum growth rate of the Pf0-1 and the mutants (with one to four tRNA gene deletions).

a

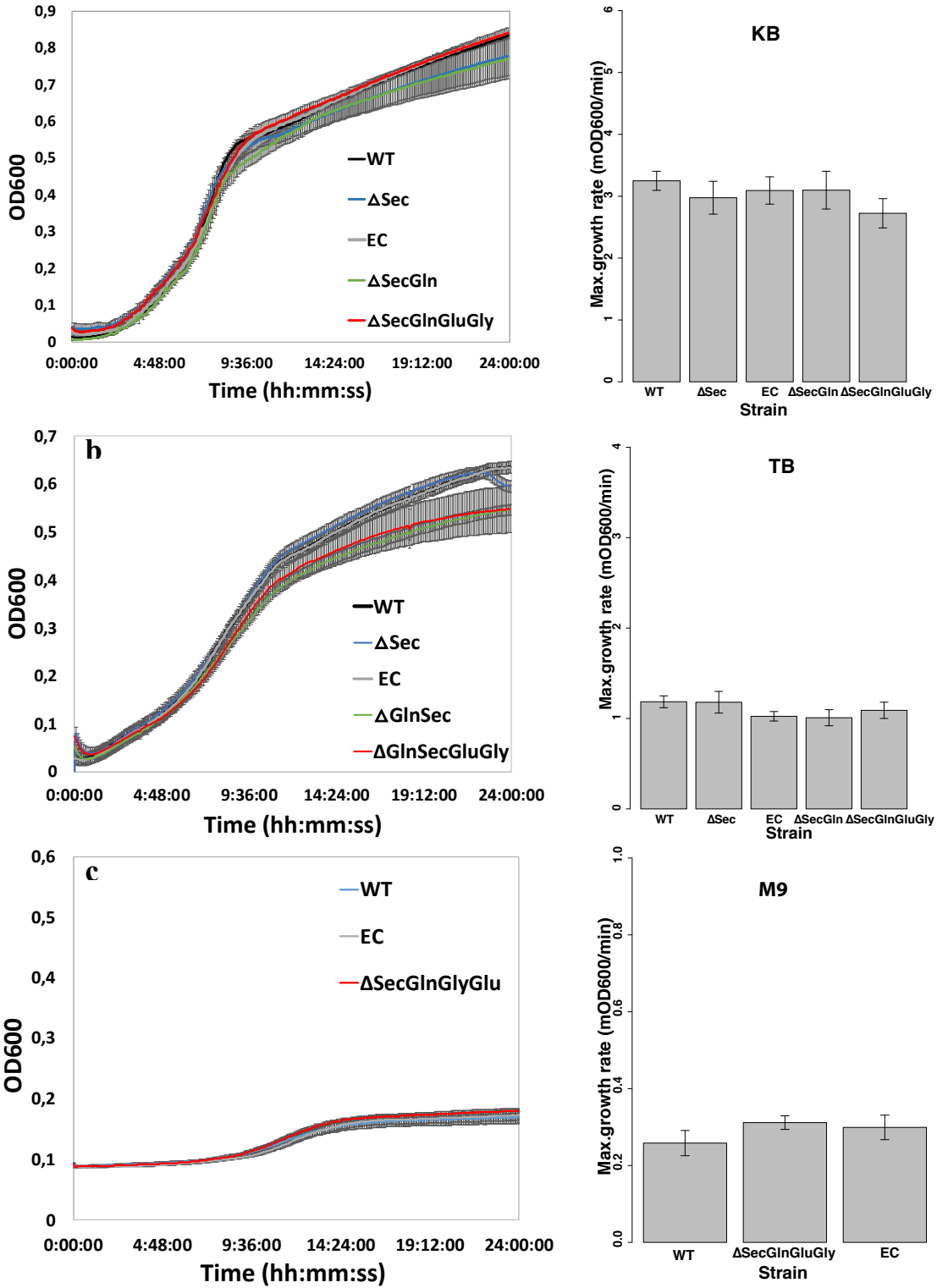


Figure 3.12. Four tRNA deletions have no significant effects on growth of *P. fluorescens* Pf0-1. 24h growth curves of Pf0-1 versus the mutants in shaken medium - a) KB b) TB c) M9 - at 28°C. (One-way ANOVA test was

done for comparison of the variances of Maximum growth rate in each strain for all media; p -value >0.05 for all the comparisons; WT= wild type Pf0-1, EC= engineering control).

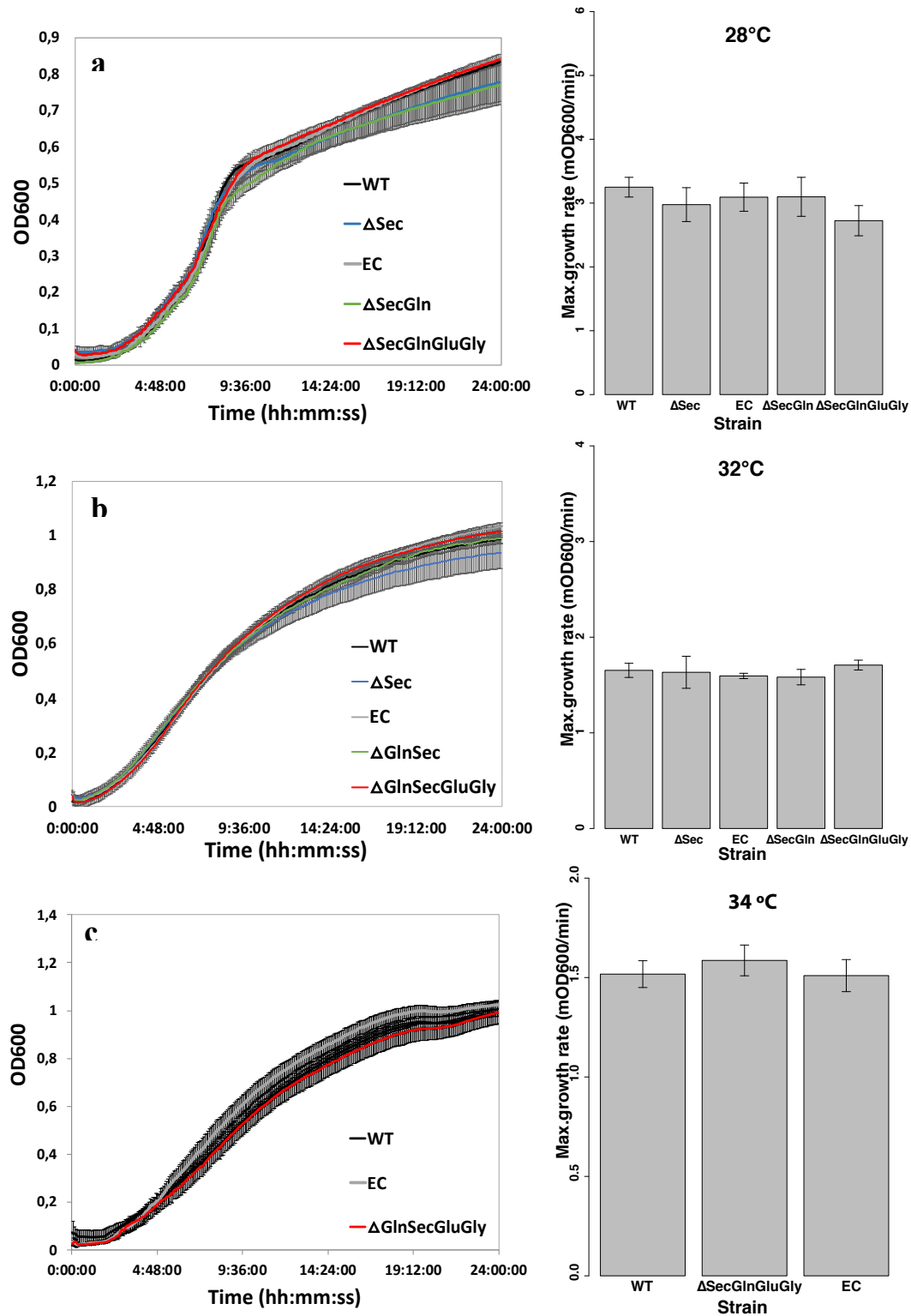


Figure 3.13. There is no difference in the growth rate of the mutants with one to four tRNA deletions and the wild type *P. fluorescens* Pf0-1 in the tested range of temperature -from 28 °C to 34 °C; a) 28 °C, b) 32 °C,

c)34 °C- (One-way ANOVA test was done for comparison of the variances of Maximum growth rate in each strain for all temperatures; p -value > 0.05 for all comparisons, Error bars in all graphs are one standard error).

After not seeing any phenotype in the growth analyses, an indirect competition assay between the mutants and the wild type was carried out as it was explained for *P. fluorescens* A506. From the comparison of the Malthusian parameters of the final mutant versus the wild type Pf0-1, it was concluded that deletion of the single copy of tRNA-Sec^{UCA}, one out of two tRNA-Gln^{UUG}, one out of four tRNA-Gly^{GCC}, and one out of five tRNA-Glu^{UUC} does not result in a discernible phenotype (Figure 3.14).

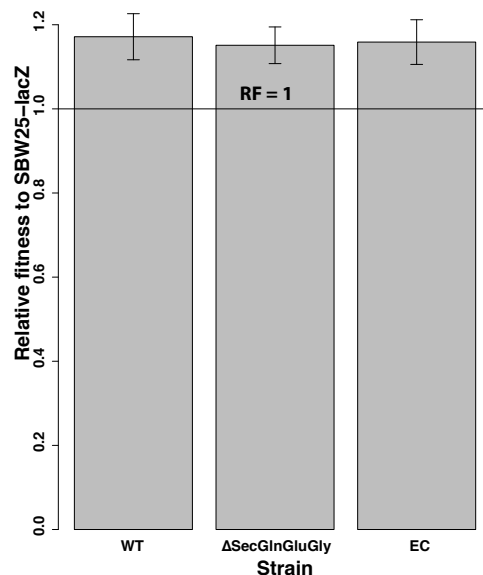


Figure 3.14. The relative fitness of the final mutant (Pf0-1ΔSecGlnGluGly) vs the wild type Pf0-1 (one-way ANOVA test, p -value = 0.96001). The relative fitness above 1 shows higher fitness than the SBW25_ *lacZ* strain. (Competition assays were performed with 1:1 starting ratio for 24h).

3.4 Passing the level of redundancy in tRNA gene families with multiple copies results in fitness defect in SBW25

Based on all the phenotypic assays of the A506 and the Pf0-1 mutants, it can be concluded that at least in the tRNA gene families with multiple copies that were manipulated in this study, deletion of one single tRNA gene has no observable phenotype and these genes are redundant copies and hence have no or little contribution to the fitness. The observed redundancy raises an interesting questions such as what is the level of this redundancy? *e.g.*

in the case of tRNA-Ala^{UGC} and tRNA-Ile^{GAU}, there are six copies of this doublet in A506 and Pf0-1 genome, thus after deletion of one copy, there are still five copies that can do the function of the missing gene. In case of deletion of two out of six copies, there are still four functional genes, *etc.* What is unknown here is that how many copies can be deleted before a discernible phenotype is observed? Focusing on tRNA genes that are represented more than twice in the genome would provide some clues to answer this question. The candidate strain for this section was *P. fluorescens* SBW25 genome. After scanning the genomic position of multiple-copies tRNA genes using the tRNAscan-SE tool and Geneious, the tRNA-Glu^{UUC} was selected as a candidate gene for manipulation because there are four identical copies in non-repetitive regions of the *P. fluorescens* SBW25 genome (Figure 3.15). Two of these genes are in close proximity of two tRNA-Gly^{UUC} genes. The other two are close to two tRNA-Ala^{GGC} genes.

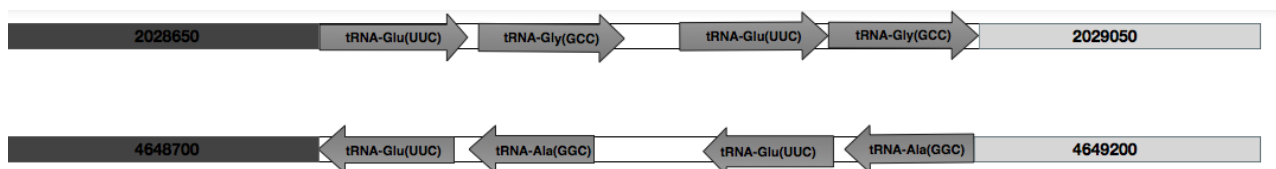


Figure 3.15- There are four identical copies of tRNA-Glu^{UUC} in *P. fluorescens* SBW25 genome. (Numbers shows the locus in the genome).

The experimental design of this section included the following deletions in *P. fluorescens* SBW25 genome:

- I. Deletion of any of the single tRNA-Glu^{UUC} genes to achieve the copy number transition from four to three.
- II. Deletion of either of the 4-tRNA set separately to achieve the decrease from four to two.
- III. A combination of I and II to achieve transition from four to one.

However, due to the complexity of the two-steps allelic exchange technique, only the construction of the SBW25 Δ GluGlyGluGly mutant was successful (experiment number II); This mutant is a derivative of *P. fluorescens* SBW25 in which a 500 kb fragment containing four tRNAs has been removed by genetic engineering. Additionally, the wild type form was reconstructed from the mutant and the accuracy of the deletions were approved by Sanger sequencing. The phenotypic characteristic of the mutants were then compared to the wild

type and as the figure 3.16 shows the deletion of the 4-tRNA gene set has changed the colony morphology. The mutant colonies are significantly smaller than the wild type (with 95% confidence interval of 2.28 to 2.94mm).

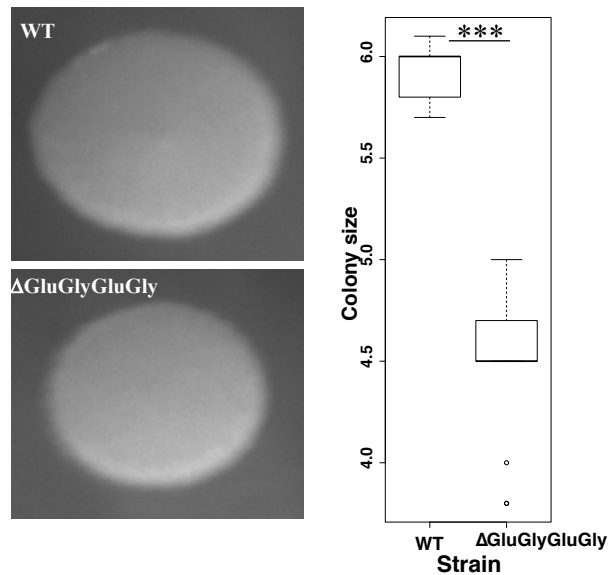


Figure 3.16 – Simultaneous deletion of four tRNA genes changes colony morphology in SBW25. Left: colony size comparison in SBW25 (up) and SBW25 Δ GluGlyGluGly (down). Colonies grown on KB agar at 28°C for 48 hours and the photos were taken under 12.5X magnification using dissecting microscope. Right: measurement of colony diameter from plates with equal numbers of colonies showed highly significant difference between the mutant and the SBW25 (two-sample *t*-test, *p*-value = 2.2e-16).

The growth phenotype of the mutant was analysed in KB as a representative for full medium for providing fast growth. The resulting growth curves show that the mutant has a slower growth in exponential phase compared to the wild type SBW25 and the reconstructed wild type (Figure 3.17).

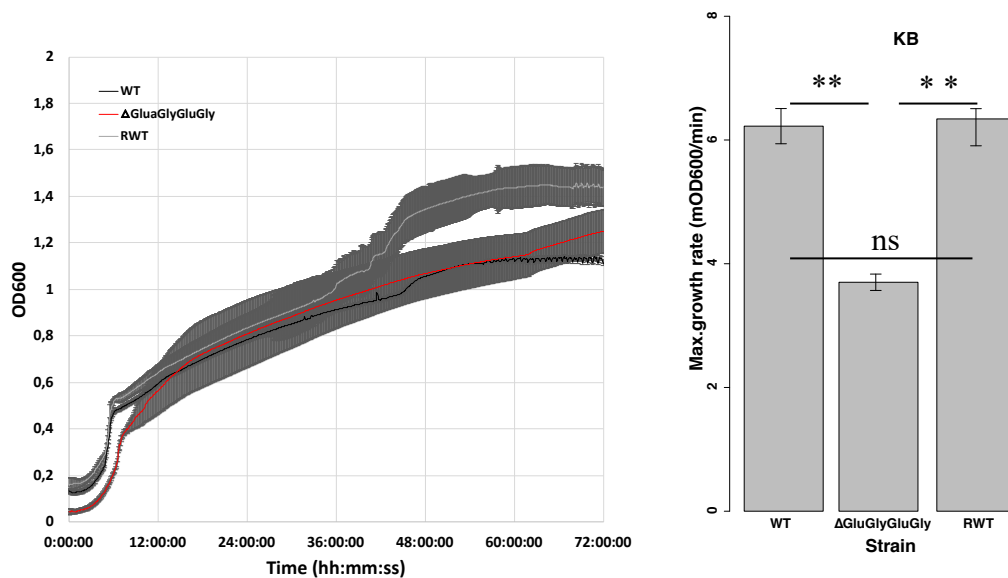


Figure 3.17. Deletion of four tRNA genes from the *P. fluorescens* SBW25 genome alters growth in KB. 72h growth curves of Pf0-1 versus the mutants in shaken KB medium at 28°C. (One-way ANOVA test was done for comparison of the variances of Maximum growth rate in each strain for all media; WT= wild type SBW25, RWT= reconstructed wild type).

From the phenotypic observation it can be concluded that the deletion of four tRNA genes has a deleterious effect on the *P. fluorescens* SBW25. The manipulation region involves two copies of tRNA-Glu^{UUC} and tRNA-Gly^{GCC} both of which have other backups in the genome: there are still two other copies of tRNA-Glu^{UUC} and one other copy of tRNA-Gly^{GCC} available in the mutant genome. The deletion of two of each tRNAs have deprived the mutant from 50% and 66% of its gene pool for these tRNA types respectively (supplementary Table S2). Each of these tRNA species also can decode other codons via wobble base-pairing (e.g., tRNA-Glu^{UUC} decode GAA via Watson-Crick base pairing and the GAG codon based on wobble base pairing). Furthermore, there are no tRNA species for decoding the GAG in the SBW25 genome, so shortage in the tRNA-Glu^{UUC} will affect both the decoding rate of GAA and GAG, which totally include 5.43% of the codon usage of SBW25 (Lowe and Eddy, 1997). In the tRNA-Gly^{GCC} case, the imposed change in the gene pool is even more dramatic. There is a 66% shortage in this tRNA gene pool for decoding 6.5% of the genome codon usage (4.95% for GCC via Watson-Crick base pairing and 1.55% for GGT via wobble base pairing; supplementary table S2). Either of these losses or a combination of both could be the reason for the deleterious fitness defect that observed in

SBW25 Δ GluGlyGluGly mutant. Construction of the SBW25 Δ GluAlaGluAla mutant (which involves deletion of the other 4-tRNA set) could provide some clues to understand the contribution of each tRNA type deletion to the observed defected phenotype. However, removal of the second 4-tRNA set is lethal to the bacteria, because there are no other backups for the two tRNA-Ala^{GCC} genes in the SBW25 genome. That is why in the homologous recombination experiments, no 2XO colonies were observed (see section 2.2.6).

In the previous section of this study, one copy of tRNA-Glu^{UUC}/tRNA-Gly^{GCC} doublet was deleted from Pf0-1 genome which provided the transition from 5 to 4 and 4 to 3 for tRNA-Glu^{UUC} and tRNA-Gly^{GCC}, respectively. The phenotypic assays showed that this transition in the copy number is not strong enough to result in a discernible phenotype. However, the deletion of two copies of the mentioned doublet from SBW25 genome, which provided the transition from 4 to 2 and 3 to 1, resulted in a fitness defect. These observations are consistent with the hypothesis that there is a threshold in the copy number of each tRNA gene, below which translation speed (and thus growth) drops. This threshold is likely to be different for each tRNA type, and could depend on many factors including the corresponding codon usage in the genome.

Thus far, the experiments were carried out with the assumption that each tRNA gene is contributing to the fitness equally. In other words, the expression levels of the identical tRNA copies are equal. However, the expression level is dependent on many factors including the genomic region (in particular the promoter and its regulation). In order to test the amount of contribution of each copy to the fitness, a tRNA gene with two copies in the genome can be used to design an experiment to make two mutants by deleting the alternate copy each time and then comparing the fitness of the mutants together (and to the wild type) with the expectation that if each copy is contributing equally to the cellular tRNA set, then deletion of either of the copies should be deleterious. The two copies of the tRNA-Asn^{GUU} gene in SBW25 were selected for doing such experiment. The first copy was deleted from the SBW25 genome successfully and interestingly it did not result in any difference in the growth rate (Figure 3.18).

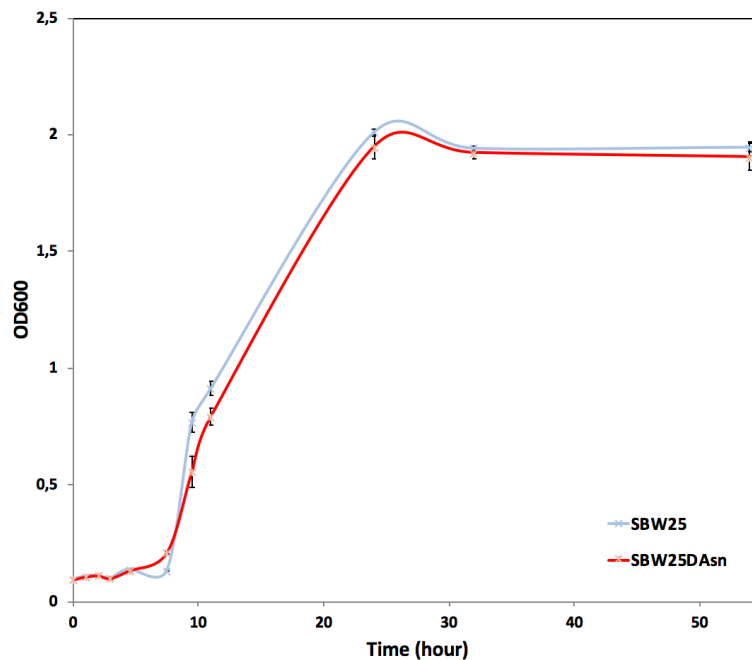


Figure 3.18. Deletion of one out of two tRNA- Asn^{GUU} genes from the *P. fluorescens* SBW25 genome does not alter growth in KB (Each time point is an average of three replicates- Error bars are one unit of standard error).

From this observation it can be concluded that the two copies of tRNA-Asn^{GUU} may not be expressed equally, and the second copy is making the majority of the cellular tRNA. Thus, maybe this explains why the construction of the second deletion was repeatedly unsuccessful, if the second tRNA-Asn^{GUU} gene is the major source of the cellular Asn-tRNA (with total codon usage of 3.13% for both codons of Asn), then the second crossing over should be lethal or highly deleterious. Combining this observation with the promoter prediction of these two tRNA genes (see Figures 1.4 and 1.5) brings us to this hypothesis that the structure of the promoter is a determining factor for the expression level of tRNA genes; *i.e.* the first copy of tRNA-Asn^{GUU} has its own promoter, while the second copy shares the promoter with two other tRNAs, thus, it is expected to be more active than the one in the first copy. Measuring the expression level of each of the mentioned tRNAs could provide enough data for testing such hypothesis. Interestingly analyzing the primary sequences of these two tRNA gene shows that there are 6 bp differences between them (Figure 3.19), thus, these two tRNAs can be distinguished in RNA sequencing methods.

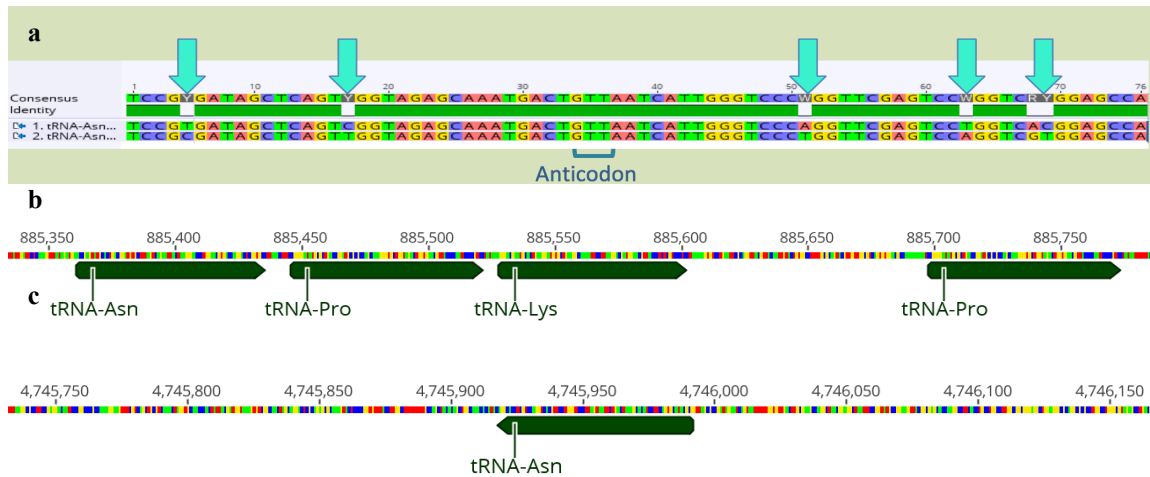


Figure 3.19. Two tRNA-Asn^{GUU} in the SBW25 genome are distinguishable due to 6x point mutations along their primary sequence (a) turquoise arrows show the point mutations in the tRNA sequence (b and c) show the location of two tRNA-Asn^{GUU} in the SBW25 genome.

The results so far showed that many tRNA genes have redundant copies in the genome. This redundant copies could have evolved in the genome via gene duplication (Nowak *et al.*, 1997) or horizontal gene transfer (Gogarten *et al.*, 2002). The next aim of this study was to study the evolution of tRNA genes in real time evolution. For studying the evolution of tRNA genes in the genome, a mutant with defective fitness caused by perturbation in its tRNA gene complement was evolved in the lab with the prediction that under the artificial imposed selective pressure, some genomes in the population will eventually evolve to compensate the fitness defect and will be taken over in the population. And further investigation of the evolved genomes would provide insights about how genomes evolve to keep up the required amounts of tRNAs during evolution. The SBW25ΔGluGlyGluGly mutant was selected for this purpose. More details of the results from experimental evolution are discussed in the next section.

3.5 SBW25ΔGluGlyGluGly fitness defect compensates fast in experimental evolution

Experimental evolution was performed for five independent lineages for both the mutant strain (SBW25 Δ GluGlyGluGly) and the wild type. During the evolution experiment, the colony morphology of the evolving populations was being observed regularly on KB agar plates. Interestingly, after one week, diverse morphotypes of SBW25 had emerged in both the mutant and the wild type lines (Figure 3.20). However, the majority of the colonies were smooth. Thus, it was assumed that although the other morphotypes exist in the population, they don't take over the population. Further, since they were seen in both wild type and tRNA deletion lineages, they can be considered a constant biotic factor across the experiment. A representative of each morphotype was purified, stored and their growth profiles were checked in KB.

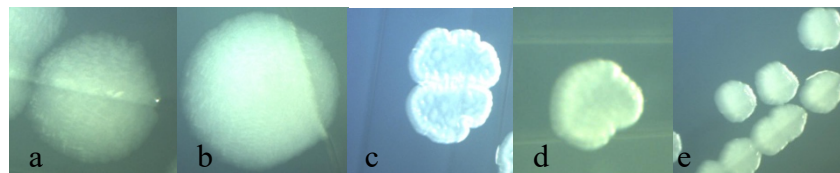


Figure 3.20. Evolution of different morphotypes from a common smooth ancestor suggests fast evolution in *P. fluorescens* SBW25. **a)** Ancestor mutant (smooth) with which the evolution started. **b)** Evolved smooth type **c)** large-sized evolved wrinkly spreader **d)** small evolved wrinkly spreader **e)** densely opaque evolved colonies.

The evolution experiment was carried out for 21 days. The generation time was calculated according to the method outlined in section 2.2.10. The results show that each day consisted of 10.50 and 10.23 generations of the wild type and the tRNA deletion mutant, respectively. Thus, each day of the evolution experiment was considered as an equivalent of 10 generations and hence it can be concluded that the experimental populations have passed through approximately 210 generations.

After 21 days (210 generations), the OD measurements showed that in the smooth colonies isolated from all mutant lineages, growth improvement has occurred (Figure 3.21).

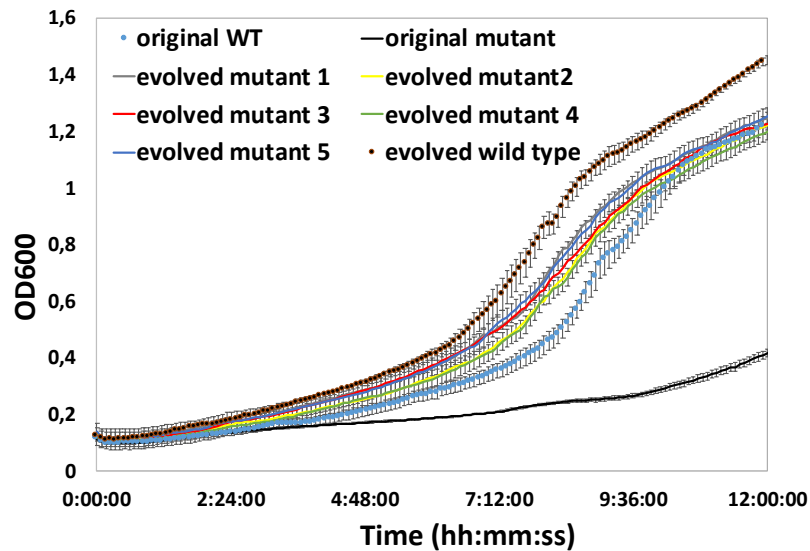


Figure 3.21. The growth profiles of the evolved mutant isolates are similar to the wild type after 21 days of evolution (12h growth curves in shaken KB at 28°C).

In experimental evolution, contamination between lineages is always a potential problem; if cells from a wild type lineage inadvertently passed into a tRNA deletion lineage, the wild type cells would likely spread quickly through the population. A quick way rule out that the compensated isolates are actually wild type contaminants is through PCR. Since a ~500 bp fragment was deleted from the wild type to construct the tRNA deletion mutant, the difference between the wild type and the mutant (and compensated mutants) can be easily distinguished by applying primers in the upstream and downstream of the manipulated region in the genome (Figure 3.22). This process showed that none of the compensated mutant isolates contained the deleted tRNA quad, ruling out contamination from the wild type lines.

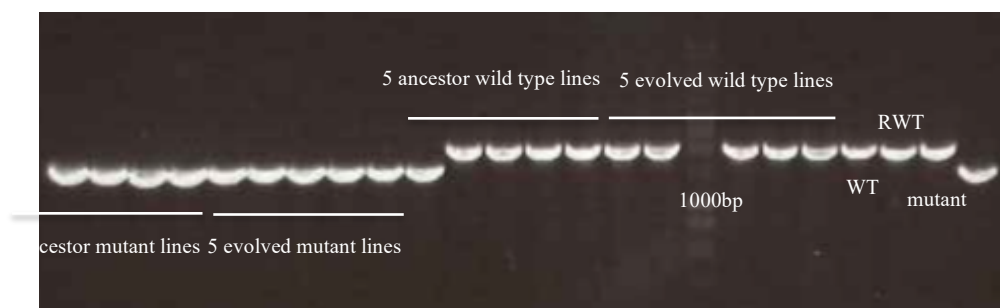


Figure 3.22. All the mutant lines (including the ancestor and the evolved ones) show smaller PCR products compared to the wild type, demonstrating that no contamination with wild type has occurred during experimental evolution.

Next, more precise phenotypic analyses were performed for one lineage of the mutant and the evolved wild type (An isolate of the evolved mutants from line 2 and one of the evolved wild type from line 3). Eight single colonies from each line were selected to perform a competition test with the marker *P. fluorescens* SBW25-*LacZ* strain. The expectations for the competition assays are no difference in fitness of the marker strain and the original wild type, evolved wild type, and the evolved mutant, but significant fitness difference in the original mutant. As it is indicated in figure 3.23, the relative fitness of the evolved mutant to the wild type is improved significantly (Figure 3.23).

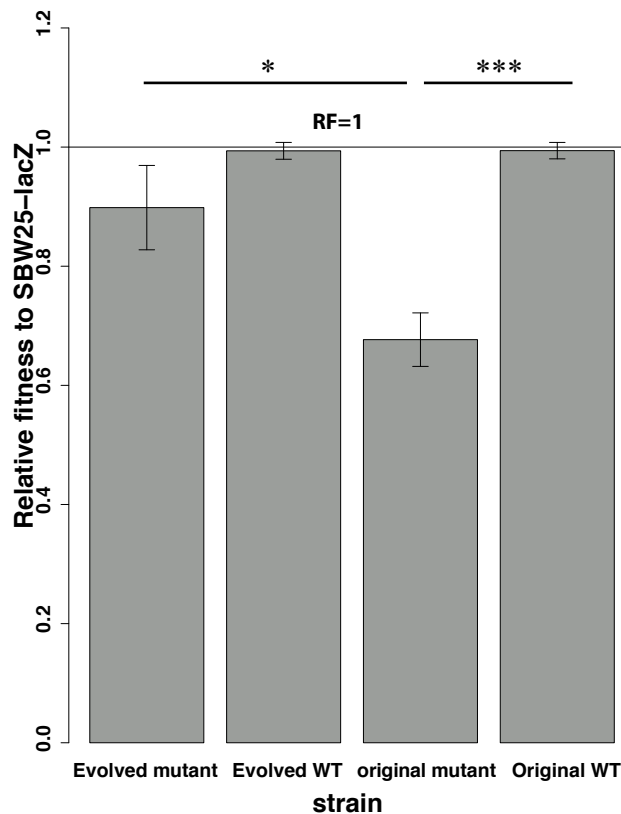


Figure 3.23. After 21 days (~210 generations) the relative fitness of the mutant was improved significantly (two-sample t-test, p -value for ancestor mutant and evolved mutant = 0.01929, p -value for the ancestor mutant and the ancestor wild type = 0.0001554, p -value for evolved mutant and the wild type = 0.2061). The relative fitness below one means lower fitness than the other competitor (wild type).

From the phenotypic analysis of the evolved mutant it can be concluded that the fitness defect of the mutant has improved significantly.

3.6 Genome sequencing of the evolved mutants shows large duplications in all lineages

Next, whole genome of one isolate from each evolved line was sequenced to study the genomic arrangements causing this fast adaptation. Single isolates from the Day 21 populations of each of the five independent lineages founded with SBW25 Δ GluGlyGluGly (Lines 1 to 5), plus one isolate from Day 21 of the first lineage founded by SBW25 wild type (Line 6), were sequenced using MiSeq v2, with 250 bp long, paired end reads. The sequencing results were aligned to the SBW25 genome and differences identified using a combination of breseq and Geneious. The breseq algorithm uses read alignment, missing coverage, and new junction evidence to predict the mutations in the studied genome. The mutations predicted by breseq in each of the six lineages are presented in Table 3.4.

Table 3.4. Mutations identified by NGS in the evolved lineages.

Mutation	Position	Annotation	Gene	Evidence	Description	Present in Lineage
Addition of G	45,881	intergenic	PFLU0045 → / ← PFLU0046	Read alignment	Putative lipoprotein/putative hydrolase	1,2,3,4,5,6
Addition of C	985,333	intergenic	PFLU0872 → / → PFLU0873	Read alignment	intergenic	1,2,3,4,5,6
Deletion of CC	3,447,984	intergenic	PFLU3154 → / ← PFLU3155	Read alignment	Putative helicase	1,2,3,4,5,6
G→A	6,048,401	T43I (ACC→ATC)	Pflu5527	Read alignment	50S ribosomal protein L3 (<i>rplC</i>)	5
G→T	3,694,377	intergenic	PFLU3336 → / → PFLU3338	Read alignment	Zinc-binding dehydrogenase	6
Δ 420	2,028,652	Two tRNA-Glu ^{UUC} /tRNA-Gly ^{GCC} doublets	PFLUt40– PFLUt43	Missing coverage/junction evidence	tRNA genes	1,2,3,4,5

Based on the table, the deletion of 420 bp is called in all the evolved mutant lineages, but not in the evolved wild type. This region contains the 4-tRNA set of interest, and its absence in isolates descended from the engineered deletion mutant demonstrates that engineering process worked, and that no contamination has occurred from the wild type

lineages. It also shows that during the experimental evolution the manipulated region of the genome has remained as it was in the founding genotype.

The first three mutations in the table (the addition of G in 45,881 position, addition of C in the 985,333 position, and deletion of CC in the 3,447,984 position) are common to all the lineages, including the evolved wild type. Thus, it can be concluded that they are pre-existing mutations; most likely they did not occur during the evolution experiment. There are only two unique point mutations: one transition in a ribosomal protein encoding gene (*rplC*) in Line 5, and one transversion in the zinc-binding dehydrogenase gene of the evolved wild type. Given that several evolved mutant lineages have no unique point mutations, none of the changes listed in Table 3.4 provides an obvious explanation for the observed adaptation.

The next step involved looking for deletions or insertions. For finding such changes, one can first look for significant increases or decreases in sequencing coverage across the genome (using a viewer such as Geneious). Figure 3.24 shows the genome-wide read alignments of each of the six evolved lines.

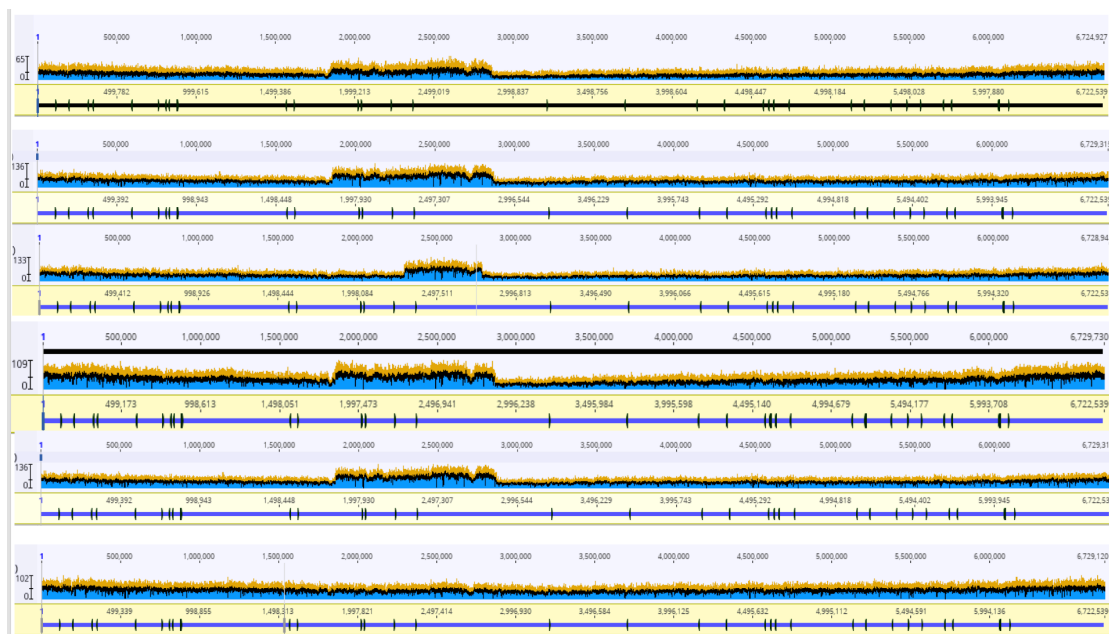


Figure 3.24. The aligned sequenced genome of the evolved strains to the *P. fluorescens* SBW25. From up to down: evolved mutant Lines 1,2,3,4,5, evolved wild type (Line 6). There is a common region in all the evolved mutant lines (around 0.48 Mb at approximately positions 2.3 to 2.78) with an immediately visible increase in coverage (y-axis).

To characterize the duplications in more detail, the upstream and downstream regions of the increased coverage area were examined carefully in Geneious. In four out of five evolved mutant lineages, the exact duplication junction was uncovered (Table 3.5). The fifth junction (that in Lineage 4), concerns a highly repetitive region of DNA. As such, definition of the junction in this lineage will require PCR amplification and Sanger sequencing (which can provide longer reads of up to 1 kb). The increased coverage region of the evolved mutants was mapped to the annotated SBW25 genome from NCBI (gbk format) and the genes located in the duplication regions were extracted by Geneious. The results show that there are ~900 genes in the duplicated regions in lineages 1,2,4, and 5, and ~424 genes in the lineage 3 (Table 3.5). This includes around 15% of the total identified 5956 genes in SBW25.

Table 3.5. The duplication junctions in the evolved mutants

Strain	Upstream	Downstream	Genes involved in the duplication area	Duplication size (bp)
ZK1	<i>pflu 2597</i> (2,864,810)	<i>pflu 1683/84</i> (1,848,645)	~916 genes (including 7 tRNA)	1,016,165
ZK2	<i>pflu 2597</i> (2,865,277)	<i>pflu 1686</i> (1,852,939)	~910 genes (including 7 tRNA)	1,012,338
ZK3	<i>pflu 2547</i> (2,786,206)	Intergenic region (R0) (2,296,498)	~424 genes (including 1 tRNA)	489,708
ZK4	~2,862,851	~1,854,450	~900 genes	The increased coverage area ~ 1,040,131
ZK5	<i>pflu 2597</i> (2,866,785)	<i>pflu 1686</i> (1,854,447)	~910 genes (including 7 tRNA)	1,012,338

The found junctions in each lineage were next aligned to the corresponding lineages as well as the evolved wild type. Interestingly, while there were 15 to 32 exact matches of each junction in the corresponding lineage, there were no exact match found in the evolved wild type (see supplementary table S3). These junctions can also be studied in the lab; that is by designing primers in the downstream and upstream of the junction, its presence in each lineage could be visualized by PCR and sequencing. Furthermore, the history of this junctions can be tracked back during the evolution using PCR techniques (*i.e.* by sequencing the genomes of the frozen populations in each time point from the start of the experimental

evolution). Figure 3.25 shows the schematic diagram of the genomic rearrangements that has occurred in the evolved wild types.

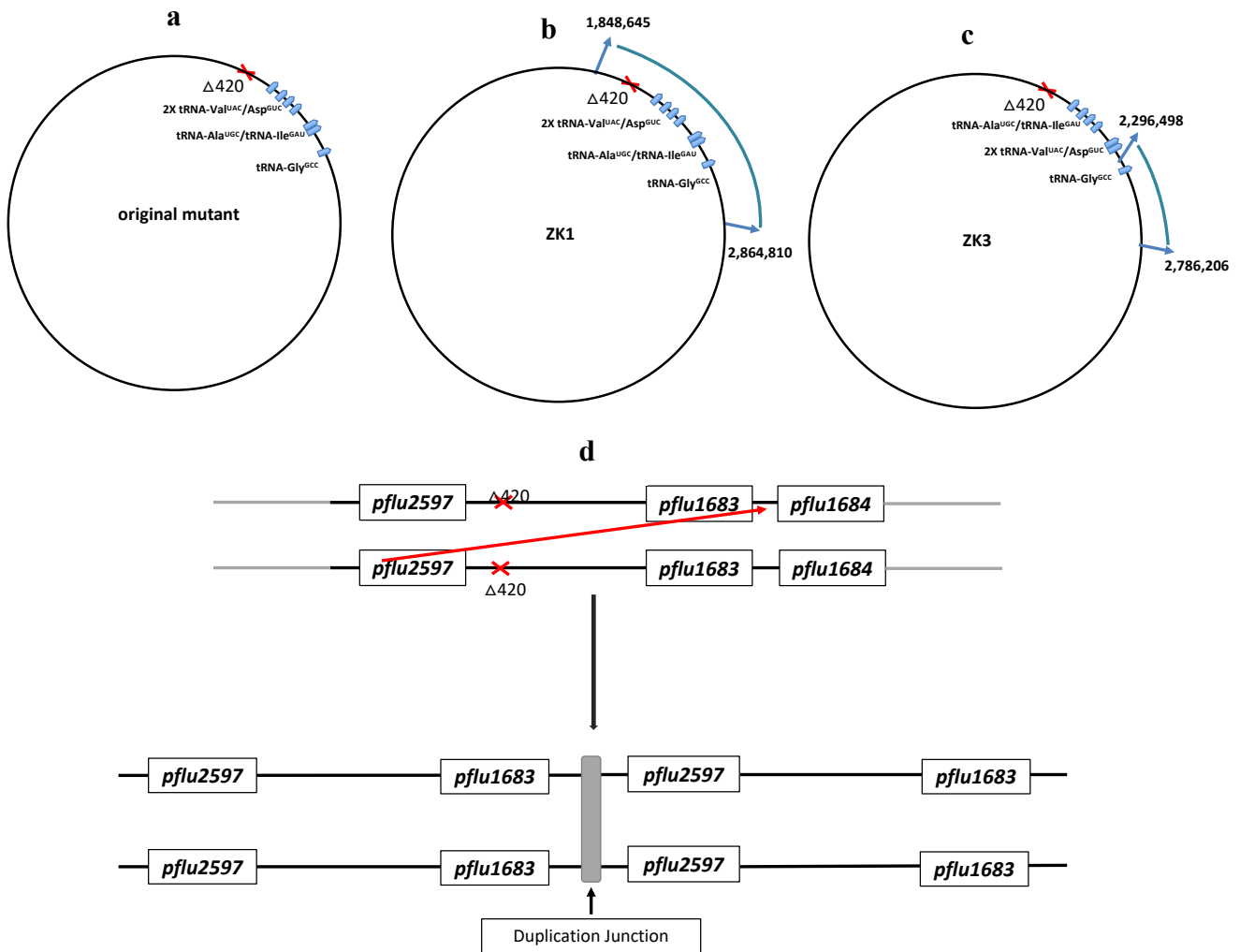


Figure 3.25. Schematic diagram of the genomic rearrangement in evolved mutants. a) The original mutant with the $\Delta 420$ deletion (only 7 out of the 62 tRNAs are shown). b) Evolved mutant ZK1 with duplication size of 1,016,165bp which has resulted in duplication of 7 tRNA genes including the tRNA-Gly^{GCC} c) Evolved mutant ZK3 with duplication size of 489,708bp which has resulted in duplication of the tRNA-Gly^{GCC} (The red cross shows the original manipulated region of the genome). d) A simplified scheme for the possible recombination which has resulted in attachment of the *pflu2597* to the intergenic region between *pflu1683* and *pflu1684* and duplication of this huge region of the genome. The grey rectangular is the duplication junction.

As it is shown in figure 3.25, the increased coverage region includes seven tRNA genes including the tRNA-Gly^{GCC}. Since two out of three copies of this tRNA gene had been deleted (by the engineering process), there was likely a shortage in the cellular tRNA-Gly^{GCC}. Such a shortage could conceivably result in slower translation and consequently the slower growth seen in Figure 3.17. During the course of DNA replication, many genomic rearrangement mechanisms including RecA-dependent and RecA-independent duplications occur continuously (Anderson *et al.*, 2009). If such duplications happen to include a beneficial gene (here the missing tRNAs), it results in improvement of fitness and the mutation may eventually spread through the population. The other six tRNAs in the duplicated regions includes two copies of tRNA-Val^{UAC}, two copies of tRNA-Asp^{GUC} and a copy of tRNA-Ala^{UGC}/tRNA-Ile^{GAU} double. One interesting question to be answered as future direction of this research is whether these extra “newborn” tRNA genes will increase the mature tRNA pool level for that specific tRNA in the evolved mutants. In four evolved strains the duplicated region includes the Δ 420 region and only the ZK3 duplication does not include the manipulated region (Figure 3.25).

3.7 Concluding statement

In this study by engineering the tRNA sets of A506 and Pf0-1 and a systematic phenotyping it was concluded that the multiple copies of tRNA genes are redundant copies and thus deletion of one copy results in no discernible phenotype. In next section, by deletion of more than one copy from two multiple copy tRNA gene families (tRNA-Glu^{UUC} and tRNA-Gly^{GCC}) from SBW25, it was experimentally shown that there is a threshold for such redundancy for each specific tRNA isotype and passing this threshold will result in significant phenotypic effect. The observed phenotypic defect could be due to shortfall of the deleted tRNA in the translation and thus slower growth in the logarithmic phase of cell growth. The fitness defect was quickly disappeared after evolving the mutant for ~210 generations. The evolved mutants showed some genomic rearrangements resulted in huge duplications which resulted in duplication of one to seven tRNA genes. The tRNA-Gly^{GCC} was the only common tRNA gene that had been duplicated in all the evolved lineages and it was concluded that it could be the main cause of such intriguingly fast adaptation.

4. Discussion

In the introduction, it was discussed that translation is the most energetically expensive process in transition of the genetic information from DNA to proteins (Alberts, 1983). Thus, there is strong selection for the efficiency and accuracy of translation. Translation is crucially important in the exponential growth phase of single-cellular organisms (Rocha, 2004) and the strong selection affects all the components involved in translation, including ribosomes, tRNAs, mRNA, and aminoacyl-tRNA synthetases. Previous studies have suggested that among all the components of the translation machinery, the cellular level of mature tRNA molecules is the rate-limiting factor during protein translation (Gingold *et al.*, 2012). This means, factors involved in shaping the mature tRNA pool are expected to be under strong selection throughout evolutionary history. These factors include the gene copy number of each tRNA species, the expression level of tRNA genes (and the components that affect the tRNA expression level, including promoter type and the genomic location of the tRNA gene), and the presence or absence of other types of tRNA (*e.g.* other cognate or near-cognate tRNAs compete with a specific types of tRNA for binding to the translation machinery components such as translation factors; Dong *et al.*, 1996).

This study focused on the copy number of tRNAs as a major component defining the cellular level of each mature tRNA type. Firstly, some perturbations were imposed on the tRNA gene sets of three phylogenetically distant strains of *P. fluorescens* (A506, Pf0-1, SBW25) to study the functional importance of multiple-copies of tRNA genes. A tRNA gene deletion library was constructed in the laboratory. This included deletion of one gene copy of each of seven multi-copy accessory tRNA species (tRNA-Sec^{UCA}, tRNA-Asn^{GUU}, tRNA-Ala^{UGC}, tRNA-Ile^{GAU}, tRNA-Gln^{UUG}, tRNA-Gly^{GCC}, and tRNA-Glu^{UUC}). The tRNA species to be manipulated were selected based on their presence in all the three strains - defined as “core”- or in a subset of the strains - defined as “accessory” tRNA genes. All three accessory tRNA genes were deleted from A506, and four out of seven accessory tRNA genes were deleted from Pf0-1.

From the phenotypic analyses it was concluded that the deletion of the chosen tRNA genes, did not affect growth or fitness in the environments tested (full and minimal media, at a range of temperatures). The tested tRNA types differ in several respects:

- 1- The manipulated tRNA genes carry a range of amino acids, belonging to different physicochemical groups. For example, isoleucine from the hydrophobic amino acids, aspartic acid from the acidic amino acids, and glutamine from the amidic amino acid group.
- 2- The deleted tRNA genes belong to tRNA gene families with copy numbers ranging from one (tRNA-Sec^{UCA} in Pf0-1) to six (tRNA-Ala^{UGC}/tRNA-Ile^{GAU} in A506 and Pf0-1).
- 3- The chosen tRNA genes encode anticodons whose cognate codons are used at varying frequencies in the genome. The lowest is 1.09% for tRNA-Gly^{GCC} in A506, and the highest is 4.77% for tRNA-Ala^{UGC} in Pf0-1 (Lowe and Eddy, 1997; see supplementary Table S2).

The members of the tRNA gene deletion library also have a common feature: they all are redundant tRNA gene copies, at least under some conditions, because their deletion has not resulted in any discernible phenotype. This result suggests that there is some redundancy in tRNA genes. Redundancy is a common feature in all domains of life, and refers to a situation where a component in a system has the same function of another component; thus, inactivation of one component can be compensated with the other component (Ohno, 1970; Wagner, 2005). Redundancy has been observed in genes involving in different biochemical pathways (*e.g.* the *GSL1* and *GSL5* genes involved in pollen formation in plants (Enns *et al.* 2005), the redundant role of *Vav* genes in platelet formation in humans and mice (Pearce *et al.* 2004), and many other isoenzymes that are the product of gene duplication and catalyse the same reaction (Zhang, 2012). Redundancy also exists beyond gene level. For instance, there are two distinct alternative enzymes for the pentose phosphate cycle in yeast: glucose-6-phosphate dehydrogenase and D-ribulose-5-phosphate 3-epimerase. Each of these can be deleted separately with no fitness effect (Harrison *et al.*, 2007). Redundancy in genomes may occur as a result of gene duplication during DNA replication (Nowak *et al.*, 1997), or horizontal gene transfer (Gogarten *et al.*, 2002). The redundant genes may prove useful, if any damage (*e.g.*, deleterious mutation) happens to one gene copy. Then, the intact redundant gene may take on the function of the (now) missing gene. This phenomenon provides a level of robustness of the genome to perturbations. The fluctuating environment imposes many challenges that may lead to organisms developing

buffering systems to survive in the long term. Buffering systems are seen at physiological (*i.e.* homeostasis) and genomic levels in time scales ranging from an individual life span to evolutionary scales (Furness *et al.*, 2015). The importance of having backups is especially important for genes involved in central cell processes such as translation, and all components involved in translation (Rocha, 2004). Among the main structural components of the translation machinery are the ribosomes. Experimental studies on rRNA operons (*rrn*), which are a crucial part of ribosomal structure and present in multiple copies in most genomes, have shown that deletion of some *rrn* copies leads to upregulated, compensatory expression of the remaining *rrn* operons (Condon *et al.*, 1993). Transfer RNAs are also among the central components of translation system. The concentration of mature tRNA molecules is the rate-limiting factor during elongation phase in translation (Varenne *et al.*, 1984); low concentrations of a tRNA species leads to translational pausing while waiting for the tRNA at the corresponding codon. This occupies the ribosomes for longer time, and can cause a global effect on the cellular translation efficiency (Sharp *et al.*, 1993). To avoid this, there is strong selection for an optimal cellular level of tRNA molecules. During evolutionary mechanism of controlling the cellular level of mature tRNAs is via altering tRNA gene copy numbers. Having similar backups for tRNA genes makes the cell robust to the deleterious mutations in one of the tRNA gene copies, because the spare can compensate this shortage.

Genetic redundancy is consistent with the observations in this study; except for the unique gene coding for tRNA-Sec^{UCA}, all the manipulated tRNA genes have at least one spare copy. Among single copy tRNA genes, the case of tRNA-Sec^{UCA} should be considered an exception; because there is only one copy of this tRNA gene, its deletion might initially be expected to cause some fitness defect. However, selenocysteine is an unusual amino acid. It is not universal; only a small number of proteins use it, and it is (so far) totally absent from the plant and fungi kingdoms (Copeland, 2005). The selenocysteine tRNA is observed in many bacteria (Zhang *et al.*, 2005). The selenocysteine tRNA is a suppressor tRNA (Leinfelder *et al.*, 1988); in presence of a specific mRNA sequence, known as selenocysteine insertion sequence (SECIS), tRNA-Sec^{UCA} reads through TGA stop codons, providing instead the selenocysteine amino acid (Copeland, 2005). The tRNA-Sec^{UCA} in Pf0-1 is likely to be functional, because Pf0-1 also encodes other SelC related machinery (including two enzymes for formation of selenocysteyl-tRNA from the seryl-tRNA precursor, and the specific elongation factor for recognizing the selenocysteyl-tRNA, both of which appear to

be under the control of a common promoter (Stadtman, 1980; Taboada and Verde, 2010). Finally, the Pf0-1 tRNA-Sec^{UCA} gene sequence scores highly in tRNAscan-SE, with the prediction of the cloverleaf secondary structure shared by traditionally functional tRNAs (Lowe and Eddy, 1997).

Although tRNA-Sec^{UCA} is present in many bacteria, including a number of *Pseudomonads*, among the studied *P. fluorescens* strains only Pf0-1 carries this tRNA. This raises an interesting question: why does the selenocysteine tRNA exist in Pf0-1 and not the other *P. fluorescens* strains? One possible explanation could be that Pf0-1 has acquired the whole SelC operon via horizontal gene transfer. The fact that other selenoproteins such as formate dehydrogenase - active in oxidoreductase activities of the cell - are clustered in the vicinity of the tRNA-Sec^{UCA} operon could be a sign of horizontal gene transfer of the whole system.

In the absence of tRNA-Sec^{UCA}, the UGA codon acts as a termination signal and thus the peptide chain is truncated. Nevertheless, the phenotypic assays in this work show that the eradication of tRNA-Sec^{UCA} gene from Pf0-1 genome does not have any effect on fitness under the conditions tested. A remaining question is: why does hindering the synthesis of the selenoproteome not have any fitness effects? One possible answer is that maybe there are parallel selenocysteine-independent oxidoreductive enzymes that can compensate the absence of selenoproteins. Thus, we can conclude that, similar to the case of redundant tRNA genes, the oxidoreductive enzymes also may have backups in the genome. Further studies are required to solve the puzzle of the evolution of selenoproteins in Pf0-1.

So far, I have discussed how redundancy can contribute to robustness in the tRNA complement of *P. fluorescens* species. One interesting question that raises is: how far does this robustness extend? In other words, what is the range of the buffered zone in each tRNA species in terms of gene copy number? In the case of tRNA species with two copies (tRNA-Asn^{GUU} and tRNA-Gln^{UUG}), deletion of a second gene copy would be lethal. However, in the case of tRNA species with more than two copies (the tRNA-Ala^{UGC}/tRNA-Ile^{GAU} doublet with five or six identical copies, and the tRNA-Glu^{UUC} /tRNA-Gly^{GCC} doublet with two or three identical copies, in the *P. fluorescens* genomes tested), deletion of more than one copy would begin to address this question. Due to having the highest copy number, the tRNA-Ala^{UGC}/tRNA-Ile^{GAU} doublet is an ideal model to design an experiment of a step-by-step

deletion of each redundant copy until a fitness defect is observed. Interestingly, in all the *P. fluorescens* species studied here, this tRNA gene doublet is in the middle of the rRNA operons (*rrn*) discussed above. This means that the DNA sequence surrounding each doublet is highly repetitive, making construction of deletions extremely difficult (see Figure 1.4 in the introduction). Hence, the next candidate is the tRNA-Glu^{UUC}/tRNA-Gly^{GCC} gene doublet, with three identical copies in the Pf0-1 genome, and two identical copies in the SBW25 genome. As previously mentioned, deletion of one copy of this tRNA gene doublet from Pf0-1 did not result in any discernable fitness defect (Figures 3.11 to 3.14); however, deletion of two copies of this doublet was deleterious in SBW25 (Figures 3.16 and 3.17). This observation leads me to hypothesize that there is a threshold where further loss of tRNA gene copies results in a large effect on translation speed and consequently on fitness. This threshold is likely different for every tRNA species, and gene copy number is only one determining factor for such threshold. This threshold could be estimated both experimentally and mathematically for each tRNA species. For instance, in one study the translation elongation rate has been calculated mathematically as a function of the cellular amount of each of 43 species of mature tRNA molecules in *E. coli*. The results show that, while the measurements differ quantitatively, each tRNA shows the same, three phase patterns (Rudorf, 2014):

- 1- At very small amounts of mature tRNA molecules, translation is inhibited very strongly.
- 2- After the concentration of mature tRNA molecules reaches a threshold, which is unique for every tRNA species, the elongation rate rapidly increases.
- 3- After the concentration of mature tRNA molecules reaches a maximum physiological level, the elongation rate decreases dramatically.

I conclude that the tRNA gene deletions made in the first section of this study did not change the corresponding concentration of mature tRNA molecules beyond phase 2, and that is why no significant change in translation rate and consequently fitness was observed. Thus, for passing the threshold, in the next section of this study, a more drastic change was imposed on the tRNA gene complement of SBW25 by the simultaneous deletion of two copies of tRNA-Glu^{UUC} and tRNA-Gly^{GCC} (resulting in construction of the SBW25 Δ GluGlyGluGly, see section 3.4). The phenotypic analysis of this deletion strain showed a statistically significant drop in both growth rate and fitness compared to the wild type (in KB medium; Figure 3.21). This mutant was used to found five independent evolutionary populations, to study the evolution of tRNA sets in real time. Experimental

evolution has been widely used to address questions about the fundamental features of evolution. A well-known example of long-term experimental evolution is Richard Lenski's ongoing long-term evolution experiment, on 12 populations of *E. coli*. Some of the findings of this famous experiment are related to the fitness dynamics of populations. For example, over the generations, fitness was shown to increase logarithmically, followed by a deceleration in subsequent generations (Lenski *et al.* 1991). In the study I have presented here, based on the growth analysis of individual evolved isolates from each population, all the lineages have shown significant fitness improvement compared to the initial founder, after a short time (~210 generations); two-sample *t*-tests showed no significant difference between the relative fitness of the evolved mutant compared to the wild type (Figures 3.21 and 3.23).

By whole-genome sequencing of isolates from all evolved mutants and one representative of evolved wild type, an intriguing example of genome evolution was provided. Firstly, there is a unanimous $\Delta 420$ bp in all the evolved mutants, showing that the four deleted tRNA genes are still absent from the genome (as expected). The point mutations summarized in the table 3.4 (except the unique transition in the ribosomal protein in line 5) are present both in the wild types and the evolved mutants, thus, it can be concluded that these mutations were either pre-existing or occurred in both the wild type and the mutant lines and they could not be the genetic mechanism of the observed improved fitness. Analysis of the raw read alignments in Geneious revealed a large area of increased coverage in all the evolved mutant lines (but not in the wild type). Further analyses led to the prediction of precise duplication regions in lines 1,2,3, and 5 (table 3.5), plus an inexactly defined duplication region in line 4. The region of increased coverage that is shared among all five evolved lines is around 0.48 Mb, at approximately position 2.3 to 2.78 Mb of the (original) SBW25 genome. Looking at the functional annotations of this region shows that the only remaining copy of tRNA-Gly^{GCC} in the genome of the tRNA quad deletion mutant is located in this region. Combing this observation with the facts about the codon usage of the manipulated tRNAs in this mutant leads to an interesting hypothesis about the duplication of tRNA genes during evolution: in *P. fluorescens* SBW25, two sets of tRNAs were deleted, leaving two other copies of tRNA-Glu^{UUC} and, one other copy of tRNA-Gly^{GCC} in the mutant genome. However, the transition from 4 to 2 (tRNA-Glu^{UUC}), and 3 to 1 (tRNA-Gly^{GCC}) genes causes a change in the mature tRNA pool, affecting translation system. The shortfall in the translation system is the probable cause of the observed fitness

defect and thus, during the course of the evolution in the laboratory, there are selective pressures against the mutant. There are spontaneous mutations including gene duplication-amplification processes that produce mutants in the population. This variation is the raw materials for adaptive evolution (Andersson and Hughes, 2009). Probably the duplication of the 0.48 Mb region - which includes many inverted repeats and the tRNA-Gly^{GCC} – compensates for the shortage of tRNA for decoding the GCC and/or GGT codons. The extra supplies of these tRNAs makes the mutant sub-population fitter than the others, allowing it to gradually spread through the population. This explains why after 210 generations we see such improvement in the fitness and growth rate of (i) the isolates, and (ii) the whole populations.

The shared 0.48 Mb duplication region that is present in all the evolved lineages, expands into an area as big as 1 MB in four lineages (all except line number 3). The annotation of this region in the SBW25 reference genome shows that it contains a total of seven tRNA genes. These tRNA genes are: one quadruplet including tRNA-Val^{UAC}-Asp^{GUC}-Val^{UAC}-Asp^{GUC}, one tRNA-Ala^{UGC}-Ile^{GAU} doublet, and the lone tRNA-Gly^{GCC} gene. Among all these duplicated tRNA genes, it appears that the duplication of tRNA-Gly^{GCC} gene is responsible for the observed improved fitness in this experimental evolution, because it is the only tRNA gene that shows evidence of duplication in all the evolved lineages (including line number 3 which has the shortest duplication region). In the future, further evidence could be provided for this hypothesis, *e.g.* plasmid-mediated expression of extra tRNA-Gly^{GCC} copies in the original mutant genome followed by growth assays.

Another interesting observation was prediction of junctions in similar (but not identical) regions of lines 1, 2, and 5. This is likely due to similar mechanisms of homologous recombination in approximately similar regions of genome. This similarity in DNA sequences across the genome may also explain the repeatability of the duplication events. Further evidence for the accuracy of the predicted junctions could be provided by amplifying and sequencing the junction regions in the evolved mutant genomes. Also, the exact time of the duplication occurrence can be estimated by performing the junction-amplifying PCR across each of the frozen (daily) intermediate populations.

This study has provided experimental evidence for tRNA gene evolution by genome duplication. As mentioned before, gene duplication-amplification (GDA) process is a

common mechanism of adaptations in bacteria including the members of *Pseudomonas* genus. For example, Toussaint *et al.* (2017) observed a GDA event covering several genes in experimental populations of *Pseudomonas aeruginosa* PA0148 that enabled the bacteria to improve growth on adenosine. Normally, *P. aeruginosa* strains are not capable of growing on nucleosides, but this GDA event led to this emergent ability (Toussaint *et al.*, 2017). Recently, Nicoloff *et al.* have shown that heteroresistant sub-populations of *Salmonella enterica*, *Klebsiella pneumoniae*, *E. coli* and *Acinetobacter baumannii* emerged in presence of different antibiotics. The genome sequencing of such populations demonstrated that most of the heteroresistance cases had resulted from spontaneous tandem amplifications in known resistance genes (Nicoloff *et al.*, 2019). Both instances showed GDA during rapid adaptation in response to an environmental stressor (nutrient limitations and antibiotic presence). In my work, GDA happened as a response to an internal stressor (the drastic change in the genome by deletion of four tRNA genes).

Previous experimental studies on tRNA genes have shown anticodon switching as a mechanism of evolution of tRNA genes, e.g., by deletion of tRNA-Arg^{CCU} gene in yeast, Yona *et al.* found an anticodon mutation in one of the 11 spare copies of tRNA-Arg^{UCU} gene, that changed the anticodon to that of the missing tRNA gene (Yona *et al.*, 2013). In the SBW25ΔGluGlyGluGly genome, there is only one copy of each of the essential tRNA-Gly^{CCC} and tRNA-Gly^{UCC} genes in the genome, thus any anticodon switching in these tRNAs is expected to be lethal. Studies about GDA often emphasize the transitory nature of gene duplications: they are usually costly and/or unstable (Anderson and Hughes, 2009). During evolution GDAs can either be tandemly repeated (leading to increase of the duplication area) or deleted (leading to gradual loss of the duplications from genome). The duplications in fact help the organism to overcome the growth restriction and after the population grow and dominate, further adaptation mechanism (e.g. point mutations) can act on the duplicated gene. The observation from different evolved lines aligns with this; line 3 has a short duplication area that is inside the longer duplication area. It is possible that this line also originally had the large 1 Mb duplication region as in other lines, and subsequently reduced it during evolution. Whether this fate is possible (or probable) could be tested by repeating the evolution experiment, this time using a strain carrying the 1 Mb duplication as a founder.

In this study, redundancy was experimentally observed in several tRNA genes: loss of one copy of a tRNA gene – either by targeted deletion, or by accumulation of mutations

in a natural environment, the phenotype of the organism does not change. This led to the hypothesis that tRNA gene redundancy renders the genome robust to perturbations. A number of phylogenetic studies have proposed sources of such redundancies including horizontal gene transfer and duplications (Iben and Maraia 2012; Fujishima and Kanai, 2014; Wald and Margalit, 2014). Interestingly, the experimental evolution section of this study provided real time evidence for large-scale genome duplication as a source of redundancy in genomes; up to 14% of the gene content of three evolved lines was duplicated. This observation brings our attention to the opening sentence of this thesis, which was written by Ohno many years before advances in whole genome sequencing: “*natural selection merely modified, while redundancy created*” (Ohno, 1970). This quote draws attention to the fact that gene/genome duplications allow opportunities for the function(s) of duplicated genes to diversify, take on novel functions, and bring about evolutionary innovations.

References

- Alberts B. 1983. Molecular biology of the cell. New York: Garland Pub, 1983.
- Altman S. 1971. Isolation of Tyrosine tRNA Precursor Molecules. *Nature New Biology* 229: 19 – 21.
- Altman S. 1980. Transfer RNA: Structure, properties, and recognition. Monograph 9A. *Cell* 19: 1064 – 1065.
- Andersson DI and Hughes D. 2009. Gene amplification and adaptive evolution in bacteria. *Annual Review of Genetic* 43: 167 – 195.
- Barrick JE, Yu DS, Yoon SH, Jeong H, Oh TK, *et al.* 2009. Genome evolution and adaptation in a long-term experiment with *Escherichia coli*. *Nature* 461: 1243 – 1247.
- Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. 2009. Experimental evolution of bet hedging. *Nature* 462: 90 – 93.
- Bermudez-Santana C, Attolini CS-O, Kirsten T, Engelhardt J, Prohaska SJ, Steigele S, Stadler PF. 2010. Genomic organization of eukaryotic tRNAs. *BMC Genomics* 2010, 11:270.
- Bertani G. 1951. studies on lysogenesis I. The Mode of Phage Liberation by Lysogenic *Escherichia coli*. *Journal of Bacteriology* 62:293.
- Bertels F and Rainey P. 2011. Within-Genome Evolution of REPINs: a New Family of Miniature Mobile DNA in Bacteria. *PLOS Genetics* 7(6): e1002132.
- Bloom-Ackermann Z, Navon S, Gingold H, Towers R, Pilpel Y, Dahan O. 2014. A Comprehensive tRNA Deletion Library Unravels the Genetic Architecture of the tRNA Pool. Copenhaver GP, editor. *PLoS Genet* 10: e1004084.
- Bruger EL and Marx CJ. 2018. A decade of genome sequencing has revolutionized studies of experimental evolution. *Current Opinion in Microbiology* 45: 149 – 155.
- Bulmer M. 1987. Coevolution of codon usage and transfer RNA abundance. *Nature* 325: 728 –730.
- Cabanás CGL, Schiliro`E, Valverde-Corredor A, Mercado-Blanco J. 2014. The biocontrol endophytic bacterium *Pseudomonas fluorescens* PICF7 induces systemic defense responses in aerial tissues upon colonization of olive roots. *Frontiers in Microbiology* 5 (Article 427).
- Condon C, French S, Squires C, Squires CL. 1993. Depletion of functional ribosomal RNA operons in *Escherichia coli* causes increased expression of the remaining intact copies. *The EMBO Journal* 12: 4305 – 4315.

- Copeland PR. 2005. Making sense of nonsense: the evolution of selenocysteine usage in proteins. *Genome Biology* (2005) 6:221.
- Crick FHC. 1966. Codon—anticodon pairing: The wobble hypothesis. *Journal of Molecular Biology* 19: 548 – 555.
- Crick FHC. 1968. The origin of the genetic code. *Journal of Molecular Biology* 38: 367 – 379.
- Deatherage DE and Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. *Methods in Molecular Biology*. 1151:165-88.
- Dong H, Nilsson L, Kurland CG. 1996. Co-variation of tRNA Abundance and Codon Usage in *Escherichia coli* at Different Growth Rates. *Journal of Molecular Biology* 260: 649 – 663.
- Enns LC, Kanaoka MM, Torii KU, Comai L, Okada K, Cleland RE. 2005. Two callose synthases, GSL1 and GSL5, play an essential and redundant role in plant and pollen development and in fertility. *Plant Mol. Biol.* 58: 333 – 349.
- Eriani G, Delarue M, Poch O, Gangloff J, Moras D. 1990. Partition of tRNA synthetases into two classes based on mutually exclusive sets of sequence motifs. *Nature* 347: 203 – 206.
- Furness AI, Lee K, Reznick DN. 2015. Adaptation in a variable environment: Phenotypic plasticity and bet-hedging during egg diapause and hatching in an annual killifish. *Evolution* 69: 1461 – 1475.
- Gallie J, Bertels F, Remigi P, Ferguson GC, Nestmann S, Rainey PB. 2019. Repeated Phenotypic Evolution by Different Genetic Routes in *Pseudomonas fluorescens* SBW25. *Molecular Biology and Evolution* 36(5): 1071 – 1085.
- Giegé R, Sissler M, Florentz C. 1998. Universal rules and idiosyncratic features in tRNA identity. *Nucleic Acids Research* 26:5017 – 5035.
- Gingold H, Dahan O, Pilpel Y. 2012. Dynamic changes in translational efficiency are deduced from codon usage of the transcriptome. *Nucleic Acids Research* 40(20):10053 – 63.
- Glick BR, 2012. Plant Growth-Promoting Bacteria: Mechanisms and Applications. *Scientifica* Volume 2012, Article ID 963401,15 pages.
- Gogarten JP, Doolittle WF, Lawrence JG. 2002. Prokaryotic Evolution in Light of Gene Transfer. *Molecular Biology and Evolution* 19(12): 2226 – 38.
- Goodenbour JM, Pan T. 2006. Diversity of tRNA genes in eukaryotes. *Nucleic Acids Research* 34: 6137 – 6146.
- Grosjean H, de Crécy-Lagard V, Marck C. 2010. Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Letters* 584: 252 – 264.
- Haubold B, Klötzl F, Pfaffelhuber. 2015. andi: Fast and accurate estimation of evolutionary distances between closely related genomes. *Bioinformatics* 31: 1169 – 1175.

- Hagervall TG and Björk GR. 2005. Transfer RNA Modification. EcoSal Plus 4.6.2.
- Harrison R, Papp B, Pal C, Oliver SG, Delneri D. 2007. Plasticity of genetic interactions in metabolic networks of yeast. *Proc Natl Acad Sci USA* 104: 2307 – 2312.
- Heineman RH, Molineux IJ, Bull JJ. 2005. Evolutionary robustness of an optimal phenotype: re-evolution of lysis in a bacteriophage deleted for its lysin gene. *Journal of Molecular Evolution* 61: 181 –191.
- Helmann JD and Chamberlin MJ. 2003. Structure and function of a bacterial sigma factor. *Annual Review of Biochemistry* 57: 839 – 872.
- Ho SN, Hunt HD, Horton RM, Pullen JK, Pease LR. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene* 77: 51 – 59.
- Hoagland MB, Stephenson ML, Scott JF, Hecht Li, Zamecnik PC. 1958. A soluble ribonucleic acid intermediate in protein synthesis. *Journal of Biological Chemistry*. 231: 241 – 257.
- Holley RW, Apgar J, Everett GA, Madison JT, Marquisee M, Merrill SH, Penswick JR, Zamir A. 1965. Structure of a ribonucleic acid. *Science* 19: 147(3664): 1462 – 1465.
- Hou YM. 2010. CCA addition to tRNA: Implications for tRNA quality control. *IUBMB Life* 62: 251 – 260.
- Ibba M and Söll D. 2003. Aminoacyl-tRNA Synthesis. *Annual reviews of biochemistry* 69: 617 – 650.
- Iben JR, Maraia RJ. 2012. tRNAomics: tRNA gene copy number variation and codon use provide bioinformatic evidence of a new anticodon:codon wobble pair in a eukaryote. *RNA* 18: 1358 – 1372.
- Iben JR, Maraia RJ. 2014. tRNA gene copy number variation in humans. *Gene* 536:376 – 384.
- Fujishima K and Kanai A. 2014. tRNA gene diversity in the three domains of life. *Frontiers in Genetics* 26(5) :1 – 11 (Article 142).
- Kanaya S, Yamada Y, Kudo Y, 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: 143 – 155.
- Kim W, Silby MW, Purvine SO, Nicoll JS, Hixson KK, Monroe M, Nicora CD, Lipton MS, Levy SB. 2009. Proteomic detection of non-annotated protein-coding genes in *Pseudomonas fluorescens* Pf0-1. Kelso J, editor. *PLOS ONE* 4: e8455.
- King EO, Ward MK, Raney DE. 1954. Two simple media for the demonstration of pyocyanin and fluorescein. *The Journal of Laboratory and Clinical Medicine* 44: 301 – 307.

- Kitten T., Kinscherf T.G., McEvoy J.L., Willis D.K. 1998. A newly identified regulator is required for virulence in *Pseudomonas syringae*. *Molecular biology* 28: 917 – 929.
- Koza A, Kusmierska A, McLaughlin K, Moshynets O, Spiers AJ. 2017. Adaptive radiation of *Pseudomonas fluorescens* SBW25 in experimental microcosms provides an understanding of the evolutionary ecology and molecular biology of A-L interface biofilm formation. *FEMS Microbiology letters*. 3: 364 (12).
- Lalucat J, Bennasar A, Bosch R, Garcia-Valdes E & Palleroni NJ. 2006. Biology of *Pseudomonas stutzeri*. *Microbial Molecular Biology* 70: 510 – 547.
- Lamrabet O, Martin M, Lenski RE, Schneider D. 2019. Changes in Intrinsic Antibiotic Susceptibility during a Long-Term Evolution Experiment with *Escherichia coli*. *Molecular Biology* 10(2): 189 – 219.
- Lang GI, Desai MM. 2014. The spectrum of adaptive mutations in experimental evolution. *Genomics* 104(0): 412 – 416.
- Lang GI, Rice DP, Hickman MJ, Sodergren E, Weinstock GM, *et al.* 2013. Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations. *Nature* 500: 571 – 574.
- Leinfelder W, Zehelein E, Mandrand-Berthelot MA, Böck A. 1988. Gene for a novel tRNA species that accepts L-serine and cotranslationally inserts selenocysteine. *Nature* 331:723 – 725.
- Lenski RE, Barrick JE, Ofria C. 2006. Balancing Robustness and Evolvability. *PLoS Biol.* 4: e428.
- Lenski RE, Rose MR, Simpson SC, Tadler SC. 1991. Long-Term Experimental Evolution in *Escherichia coli*. I. Adaptation and Divergence During 2,000 Generations. *The American Naturalist* 138:1315 – 1341.
- Lenski RE, Travisano M. 1994. Dynamics of adaptation and diversification: a 10,000-generation experiment with bacterial populations. *PNAS* 91(15): 6808 – 6814.
- Li Z and Deutscher MP. 1996. Maturation Pathways for *E. coli* tRNA Precursors: A Random Multienzyme Process In Vivo. *Cell* 86:503 – 512.
- Lindow SE, Desurmont C, Elkins R, McGourty G, Clark E, Brandl MT. 1998. Occurrence of Indole-3-Acetic Acid-Producing Bacteria on Pear Trees and Their Association with Fruit Russet. *Phytopathology* 88: 1149 – 1157.
- Lindsay RJ, Pawlowska BJ, Gudelj I. 2019. Privatization of public goods can cause population decline. *Nature Ecology & Evolution* 3(2019)1206 – 1216.
- Loftfield RB. 1963. The frequency of errors in protein biosynthesis. *Journal of Biochemistry*. October 1963: 82 – 89.
- Loftfield RB, Vanderjagt D. 1972. The frequency of errors in protein biosynthesis. *Journal of Biochemistry* 128(5): 1353 – 1356.

- Long A, Liti G, Luptak A, Tenailon O. 2015. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nature Publishing Group* 16:567 – 582.
- Lorenz C, Lünse C, Mörl M. 2017. tRNA Modifications: Impact on Structure and Thermal Adaptation. *Biomolecules* 7(2): 1 – 35.
- Lowe TM and Eddy SR. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Research* 25:955 – 964.
- Mallick B, Chakrabarti J, Sahoo S, Ghosh Z, Das S. 2005. Identity Elements of Archaeal tRNA. *DNA Research* 12: 235 – 246.
- Manrique-Poyato MI, Cabrero J, López-León MD, Perfectti F, Gómez R, Camacho JPM. 2019. Interpopulation spread of a parasitic B chromosome is unlikely through males in the grasshopper *Eyprepocnemis plorans*. *Heredity* 106: 332 – 10.
- Meyer JM, Abdallah MA. 1978. The Fluorescent Pigment of *Pseudomonas fluorescens*: Biosynthesis, Purification and Physicochemical Properties. *General Microbiology* 107: 319 – 328.
- Meyer JR, Dobias DT, Weitz JS, Barrick JE, Quick RT, Lenski R. 2012. Repeatability and contingency in the evolution of a key innovation in phage Lambda. *Science* 335(6067): 428–432.
- Murphy FV and Ramakrishnan V. 2004. Structure of a purine-purine wobble base pair in the decoding center of the ribosome. *Nature Structural and Molecular Biology* 11(12):1251 – 1252.
- Nelson DL, Lehninger AI, COX MM. 2008. *Lehninger principles of biochemistry*. New York : W.H. Freeman, 2008.
- Nicoloff H, Hjort K, Levin BR, Andersson DI. 2019. The high prevalence of antibiotic heteroresistance in pathogenic bacteria is mainly caused by gene amplification. *Nature Microbiology* 4:504 – 514.
- Novoa EM and de Pouplana LR. 2012. Speeding with control: codon usage, tRNAs, and ribosomes. *Trends in Genetics* 28(11):574 – 581.
- Nowak MA, Boerlijst MC, Cooke J, Smith JM. 1997. Evolution of genetic redundancy. *Nature* 388:167 – 171.
- Ohno S. 1970. *Evolution by gene duplication*. Springer-Verlag New York Inc.
- Palleroni NJ, 1992. Introduction to the Pseudomonadaceae. *The Prokaryotes, A Handbook on the Biology of Bacteria, Ecophysiology, Isolation, Identification and Applications*, 3: 3071 – 3085. Springer, New York.
- Paule MR and White RJ. 2000. Survey and summary transcription by RNA polymerases I and III. *Nucleic Acids Research* 28(6): 1283 – 98.

- Pearce AC, Senis YA, Billadeau DD, Turner M, Watson SP, Vigorito E. 2004. Vav1 and vav3 have critical but redundant roles in mediating platelet activation by collagen. *Biological Chemistry*. 279: 53955 – 53962.
- Percudani R, Pavesi A, Ottonello S. 1997. Transfer RNA gene redundancy and translational selection in *Saccharomyces cerevisiae*. *Molecular Biology* 268: 322 – 330.
- Perry JC, Joag R, Hosken DJ, Wedell N, Radwan J, Wigby S. 2016. Experimental evolution under hyper-promiscuity in *Drosophila melanogaster*. *BMC Evolutionary Biology* (2016) 16: 131.
- Pettersson BMF. 2009. tRNA Gene Structures in Bacteria. Dissertation submitted to Faculty of Science and Technology 628:1 – 68.
- Phizicky EM, Hopper AK. 2010. tRNA biology charges to the front. *Genes & Development* 24: 1832 – 1860.
- Raina M and Ibba M. 2014. tRNAs as regulators of biological processes. *Frontiers in Genetics*. 2014(5): 171 – 185.
- Rainey PB, 1999. Adaptation of *Pseudomonas fluorescens* to the plant rhizosphere. *Molecular Biology*:1 – 15.
- Rainey PB, Bailey MJ. 1996. Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Molecular Microbiology* 19: 521 – 533.
- Rainey PB, Travisano M. 1998. Adaptive radiation in a heterogeneous environment. *Nature* 394: 69 – 72.
- Rak R, Dahan O, Pilpel Y. 2018. Repertoires of tRNAs: The Couplers of Genomics and Proteomics. *Annual Review of Cell and Developmental Biology* 34: 239 – 264.
- Ralser M, Querfurth R, Warnatz H.J, Lehrach H, Yapso M.L, Krobitsch S. 2006. An efficient and economic enhancer mix for PCR. *Biochemistry Biophysics Research Community* 347: 747 – 751.
- R Core Team. 2017. R: A A Language and Environment for Statistical Computing. <https://www.R-project.org/>.
- Rich A, RajBhandary UL. 1976. Transfer RNA: molecular structure, sequence, and properties. *Annual Reviews of Biochemistry* 45: 805 – 860.
- Riehle MM, Bennett AF, Long AD. 2001. Genetic architecture of thermal adaptation in *Escherichia coli*. *PNAS* 98(2): 525 – 530.
- Rocha EPC. 2004. Codon usage bias from tRNA's point of view: redundancy, specialization, and efficient decoding for translation optimization. *Genome Research*. 14: 2279 – 2286.
- Rudorf S. 2015. Protein Synthesis by Ribosomes. PhD Thesis, Max Planck Institute of Colloids and Interfaces.

- Sakano H and Shimura Y. 1978. Characterization and in vitro processing of transfer RNA precursors accumulated in a temperature-sensitive mutant of *Escherichia coli*. *Journal of Molecular Biology* 123: 287 – 326.
- Sambrook J, Fritsch EF, Maniatis T. 1989. *Molecular cloning: a laboratory manual*. 2nd edition, Cold Spring Harbour Laboratory Press.
- Schulman LH. 1991. Recognition of tRNAs by Aminoacyl-tRNA Synthetases. *Progress in Nucleic Acid Research and Molecular Biology* 41(1991): 23 – 87.
- Sharp PM and Matassi G. 1994. Codon usage and genome evolution. *Current Opinion in Genetics & Development* 4: 851 – 860.
- Sharp PM, Stenico M, Peden JF, Lloyd AT. 1993. Codon usage: mutational bias, translational selection, or both? *Biochemical Society Transactions* 21(4):835-41.
- Sharp SJ. 2008. Structure and transcription of Eukaryotic tRNA gene. *Critical Reviews in Biochemistry*, 19(2): 107 – 144.
- Shepherd J and Ibba M. 2015. Bacterial transfer RNAs. *FEMS Microbiology Reviews* 39: 280 – 300.
- Shrivastava R, Basu A, Phale PS. 2011. Purification and characterization of benzyl alcohol- and benzaldehyde- dehydrogenase from *Pseudomonas putida* CSV86. *Archives of Microbiology* 193: 553 – 563.
- Silby MW, Cerdeño-Tárraga AM, Thomson NR. 2009. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome biology* 10 R51.
- Silby MW, Winstanley C, Godfrey SAC, Levy SB, Jackson RW. 2011. *Pseudomonas* genomes: diverse and adaptable. *FEMS Microbiology Reviews* 35: 652 – 680.
- Smith D, Yarus M. 1989. Transfer RNA structure and coding specificity. II. A D-arm tertiary interaction that restricts coding range. *Journal of Molecular Biology*. 206(3): 503 – 511.
- Söll D. 1990. The accuracy of aminoacylation - ensuring the fidelity of the genetic code. *Cellular and Molecular life sciences* 46 (11-12): 1089 – 1096.
- Sonti RV, Roth JR. 1989. Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. *Genetics* 123:19 – 28.
- Stadtman TC. 1980. Selenium-dependent enzymes. *Annual Review of Biochemistry* 49: 93 – 110.
- Taboada B and Verde C, research EMNA, 2010. High accuracy operon prediction method based on STRING database scores. academic.oup.com.
- Todar K. Online textbook of bacteriology.
https://www.slideshare.net/sarah_jumali/calculating-bacterial-growth-7571318.
- Toh Y, Hori H, Tomita K, Ueda T, Watanabe K. 2001. *Transfer RNA Synthesis and Regulation*. Chichester, UK: John Wiley & Sons, Ltd.

- Toussaint J-P, Farrell-Sherman A, Feldman TP, Smalley NE, Schaefer AL, Greenberg EP, Dandekar AA. 2017. Gene Duplication in *Pseudomonas aeruginosa* Improves Growth on Adenosine. O'Toole G, editor. *Journal of Bacteriology* 199: e00261–17.
- Travis M, Lenski RE. 1996. Long-term experimental evolution in *Escherichia coli*. IV. Targets of selection and the specificity of adaptation. *Genetics* 143(1): 15 – 26.
- Turner TL, Stewart AD, Fields AT, Rice WR, Tarone AM. 2011. Population-Based Resequencing of Experimentally Evolved Populations Reveals the Genetic Basis of Body Size Variation in *Drosophila melanogaster*. *PLoS Genetics* 7(3): 1001336.
- Varenne S, Buc J, Lloubes R, Lazdunski C. 1984. Translation is a non-uniform process: Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. *Journal of Molecular Biology* 180: 549 – 576.
- Wagner A. 2005. Distributed robustness versus redundancy as causes of mutational robustness. *BioEssays* 27: 176 – 188.
- Wald N, Margalit H. 2014. Auxiliary tRNAs: large-scale analysis of tRNA genes reveals patterns of tRNA repertoire dynamics. *Nucleic Acids Research* 42: 6552 – 6566.
- Wendenbaum S, Demange P, Dell A, Meyer JM, Abdallah MA. 1983. The structure of pyoverdine Pa, the siderophore of *Pseudomonas aeruginosa*. *Tetrahedron Letters*. 24 (44): 4877 – 4880.
- Westhof E, Dumas P, Moras D. 1985. Crystallographic refinement of yeast aspartic acid transfer RNA. *Journal of Molecular Biology* 184: 119 – 145.
- Woese C. 1970. Molecular Mechanics of Translation: A Reciprocating Ratchet Mechanism. *Nature* 226: 817 – 820.
- Yona AH, Bloom-Ackermann Z, Frumkin I, Hanson-Smith V, Charpak-Amikam Y, Feng Q, Boeke JD, Dahan O, Pilpel Y. 2013. tRNA genes rapidly change in evolution to meet novel translational demands. *2*: 927 – 17.
- Zhang J. 2012. Genetic Redundancies and Their Evolutionary Maintenance. In: *Evolutionary Systems Biology*. Vol. 751. *Advances in Experimental Medicine and Biology*. New York, NY: Springer New York. 279 – 300.
- Zhang X-X, Rainey PB. 2007. Construction and validation of a neutrally-marked strain of *Pseudomonas fluorescens* SBW25. *J. Microbiol. Methods* 71: 78 – 81.
- Zhang Y, Fomenko DE, Gladyshev VN. 2005. The microbial selenoproteome of the Sargasso Sea. *Genome Biology* (2005) 6:6 6: R37.

Appendixes

Supplementary table S1 – Level of homology in the primary structure of the tRNA complements in eight *p. fluorescens* strains

tRNA	Length (bp)	homology	Copy number in each strain
tRNA-Ala ^{GGC}	76	100% in all except F113 and Pf0-1 (with 92% homology to the rest)	2 in all strains
tRNA-Ala ^{TGC}	76	100% in all except F113 and Pf0-1 (with 97.3% homology to the rest)	5 in SBW25, PICF5, F113, PCL1751 6 in A506, LBUM223, Pf0-1, UK4
tRNA-Arg ^{ACG}	77	100%	2 in all strains, 3 in Pf0-1
tRNA-Arg ^{CCG}	77	100% in all except in F113 and Pf0-1 (with 96,104% and 94% homology respectively)	1 in all strains
tRNA-Arg ^{CCT}	77	100% in A506, LBUM223, PICF7, PCL1751, SBW25. 97.40% in Pf0-1 and UK4, and 96.10% in F113	1 in all strains
tRNA-Arg ^{TCT}	77	100%	1 in all strains
tRNA-Asn ^{GUU}	76	The most diverse tRNA with 76 to 100% homology both between and within the strains	2 in all strains
tRNA-Asp ^{GUU}	77	100%	4 in all strains
tRNA-Cys ^{GCA}	74	100% in all except F113 (with 94.5% homology to the rest)	1 in all strains
tRNA-Met ^{CAT}	77	100%	1 in all strains
tRNA-Gln ^{UUG}	75	100%	1 in all strains 2 in Pf0-1
tRNA-Glu ^{UUC}	76	100%	4 in A506, F113, LBUM223, UK4, SBW25 5 in PCL1751, PICF7, Pf0-1
tRNA-Gly ^{GCC}	76	100% in all except one copy in A506 (with 98% homology to others)	3 copies in A506, F113, LBUM223, UK4 4 copies in PCL1751, Pf0-1, PICF7
tRNA-Gly ^{CCC}	74	100% in A506, PCL1751, PICF7, SBW25, 100% in LBUM223 and UK4, 100% in F113 and Pf0-1, with 77,79, and 86% homology between the groups	1 in all strains
tRNA-Gly ^{TCC}	74	100%	1 in all strains
tRNA-His ^{GTG}	76	100% in all except Pf0-1 and F113 (with 97.36% homology to the rest)	2 in 7 strains 1 in F113
tRNA-Ile ^{GAT}	77	100%	5 in F113, SBW25, PCL1751, PICF7 6 in A506, Pf0-1, LBUM223, UK4
tRNA-Ile ^{2CAT}	77	100% homology in F113 and Pf0-1 100% homology in A506 and SBW25	1 in all strains

		Others have 68.63% to 98% homology to these two groups	
tRNA-Leu ^{CAA}	85	100% in all strains except the one in LBUM223 which has 97.64% homology to others	1 in all strains
tRNA-Leu ^{CAG}	87	100% in A506, LBUM223, SBW25, PCL1751, PICF7. 100% homology in F113, UK4, Pf0-1 (97.64% between the two groups)	2 in all strains
tRNA-Leu ^{GAG}	86	100%	1 in all strains
tRNA-Leu ^{TAA}	87	100% in all except the ones in F113 and Pf01 (with 97.7% homology)	1 in all strains
tRNA-Leu ^{TAG}	85	100% in all except the one in F113 (with 94.118% homology to others)	1 in all strains
tRNA-Lys ^{CUU}	75	100%	1 in PCL1751 and PICF7
tRNA-Lys ^{UUT}	76	100%	2 in all strains
tRNA-fMet ^{CAT}	77	The first two copies in each strain have 100% homology, the third copy in each strain has 97.40% with the first two copies	3 in in all strains, PICF7 has the fourth copy
tRNA-Phe ^{GAA}	76	100%	1 in all strains
tRNA-Pro ^{GGG}	77	100%	1 in all strains, LBUM223 has the second copy with 81% homology to the rest
tRNA-Pro ^{TGG}	77	100% in both copies of A506, F113, SBW25, UK4 and in 1 copy of the LBUM223, PCL1751, Pf0-1, PICF7, 97% with the second copy of the laUUer four strains	2 in all strains
tRNA-Sec ^{TCA}	96	-	1 in Pf0-1
tRNA-Ser ^{CGA}	The longer copy:90bp, the shorter copies: 70bp	Every strain has 1 identical copy (100%, except the A506 and F113 which have 89% homology to the other 6), furthermore, F113 ,Pf0-1 and PCL1751 have the second shorter copy with 49% homology to the longer one	2 in F113 and Pf0-1, and PCL751, 1 in each of the other 6 strains
tRNA-Ser ^{GCT}	91	100%	1 in all strains
tRNA-Ser ^{GGA}	90	100% in PICF7, PCL1751, SBW25, A506, and LBUM223 and 97.7% in F113, Pf0-1, UK4 (which are 100% identical to each other)	1 in all strains
tRNA-Ser ^{TGA}	F113 and Pf0-1(88) rest: 91	100% in A506, PCL1751, PICF7, SBW25. These 4 have 96% similarity to the ones in LBUM223 and UK4 (which are 100% identical to each other)	1 in all strains
tRNA-Thr ^{CGT}	73 in SBW25, UK4, PICF7, PCL1751 And 76 in other 4 strains	100% in A506, LMUM223, PCL1751, Pf0-1, PICF7. These 5 have 95% similarity to the ones in F113, SBW25, and UK4	1 in all strains
tRNA-Thr ^{GGT}	76	100%	1 in all strains
tRNA-Thr ^{TGT}	76	100% except in Pf0-1 (97.36%)	1 in all strains

tRNA-Trp ^{CCA}	76	100%	1 in all strains
tRNA-Tyr ^{GTA}	85	100% except in Pf0-1 (98.82%)	1 in all strains
tRNA-Val ^{GAC}	77	100%	1 in all strains
tRNA-Val ^{TAC}	76	100% between SBW25, A506, UK4, LBUM223 and one copy of Pf0-1 100% between PICF7, PCL1751, F113 and two copies of Pf0-1 96.05% between two groups	2 in PICF7 3 in other 7 strains

Table S2- List of manipulated tRNA genes and their corresponding codon usage in three strains of *P. fluorescens* (Lowe and Eddy,1997)

strain	Manipulated tRNA	Codons decoded (cognate bold, non-cognate normal)	# isotype in the genome	# all isoacceptors in genome	Corresponding codon usage (for cognate)	Corresponding codon usage (for non-cognate)	Total codon usage (cognate+non-cognate)	% loss after deletion
SBW25	tRNA-Asn ^{GUU}	(AAC-AAT)	2	2	2.37%	0.76%	3.13%	50%
	tRNA-Glu ^{UUC}	(GAA-GAG)	4	5	2.99%	2.44%	5.43%	50%
	tRNA-Gly ^{GCC}	(GGC-GGT)	3	5	4.95%	1.55%	6.5%	66%
Pf0-1	tRNA-Sec ^{TCA}	UGA	1	1	NA	0	NA	100%
	tRNA-Gln ^{TTG}	(CAA-CAG)	2	2	1.29	3.16%	4.45%	50%
	tRNA-Gly ^{GCC}	(GGC-GGT)	4	6	4.77	1.73%	6.5%	25%
	tRNA-Glu ^{TTC}	(GAA-GAG)	5	5	3.36	2.31%	5.67%	20%
A506	tRNA-Asn ^{GUU}	(AAC-AAT)	3	3	2.38	0.8%	3.18%	33%
	tRNA-Ala ^{UGC}	(GCA-GCG)	6	8	1.09	3.25%	4.34%	16%
	tRNA-Ile ^{GAU}	(ATC-ATT)	6	7	3.53%	1.13%	4.67%	16%

Table S3 – Duplication junctions sequences and the number of matches in each line

Lineage	Junction sequence	# exact matches in the corresponding lineage	# exact matches in the evolved wild type
ZK1	GCATAGCGGTCTCACCACCCGCAAAGGTGA	22	0
ZK2	GTTGTTGCCATCAGGCAGGGTTTTTCTTACGAGCCGACCGG CGTGGTCT	25	0
ZK3	CAAGCCCCCTCCCACGATAGGTGGACGGCA	32	0
ZK4	NA	NA	NA
ZK5	aatcgccgcTTCCGCCAAACTATTGGTAATCGCAGGCCGAGAAA CCGTAGGCAAGGGCAGGGCAGTCGACCGATTTTCATGATC TGTCCACACCACCGAATCACCGAAACTCAA	15	0

- **Malthusian Parameters and Relative fitness of final mutants**

To compare each tRNA deletion mutant to its wild type strain, three types of competition experiments were set up:

- SBW25-*lacZ* vs wild type
- SBW25-*lacZ* vs engineering control wild type
- SBW25-*lacZ* vs tRNA deletion mutant

The methods for competition assays have been described in section 2.2.8.3 and the results has been explained in section 3.3 and 3.4 (Figures 3.10, 3.14, and 3.23).

Competition results for A506 strains

Table S4. SBW25-*lacZ* vs Wild type A506

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25- <i>lacZ</i>	A506	D	V	SBW25- <i>lacZ</i>	A506	A506	SBW25- <i>lacZ</i>	
1	10 ⁻²	50	87	57	10 ⁻⁶	100	13	61	6.62	8.58501579	1.29756826
2	10 ⁻²	50	73	57	10 ⁻⁶	100	7	61	6.17	8.58501579	1.39081663
3	10 ⁻²	50	63	64	10 ⁻⁶	100	1	44	4.37	8.14249974	1.86154342
4	10 ⁻²	50	46	54	10 ⁻⁶	100	12	68	7.17	8.74771685	1.21945599
5	10 ⁻²	50	68	64	10 ⁻⁶	100	10	51	6.60	8.29013574	1.25602968
6	10 ⁻²	50	49	50	10 ⁻⁶	100	10	68	6.93	8.82467789	1.27377763
7	10 ⁻²	50	50	48	10 ⁻⁶	100	7	51	6.55	8.57781781	1.30937454
8	10 ⁻²	50	35	60	10 ⁻⁶	100	5	44	6.57	8.20703826	1.24892479
Mean RF	1.357186										

D= dilution

V= plating volume (µl)

Mean RF = mean of Relative fitness (last column) 0

Table S5. SBW25-*lacZ* vs A506 Engineering control (EC)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25_ <i>lacZ</i>	EC	D	V	SBW25_ <i>lacZ</i>	A506	EC	SBW25- <i>lacZ</i>	
1	10 ⁻²	50	52	50	10 ⁻⁶	100	36	118	8.15	9.37585481	1.15048667
2	10 ⁻²	50	29	22	10 ⁻⁶	100	32	51	8.62	9.35797637	1.08616234
3	10 ⁻²	50	35	27	10 ⁻⁶	100	17	74	7.80	9.52542142	1.22198204
4	10 ⁻²	50	20	33	10 ⁻⁶	100	9	39	7.72	8.68424728	1.12509407
5	10 ⁻²	50	57	45	10 ⁻⁶	100	10	37	6.78	8.32144861	1.22794508
6	10 ⁻²	50	16	30	10 ⁻⁶	100	17	52	8.58	9.06723953	1.05705667
7*	10 ⁻²	50	0	53	10 ⁻⁶	100	0	126	NA	NA	NA
8	10 ⁻²	50	25	38	10 ⁻⁶	100	18	88	8.15	9.37585481	1.15048667
Mean	1.144485										

* unequal starting ratio of the competitors, so the data for this replicate was discarded from further analyses

Table S6. SBW25-*lacZ* vs A506Δ*AsnAlalle* (M)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25_ <i>lacZ</i>	M	D	V	SBW25_ <i>lacZ</i>	M	M	SBW25- <i>lacZ</i>	
1	10 ⁻²	50	51	63	10 ⁻⁵	50	37	147	6.59	7.75505314	1.17735428
2	10 ⁻³	100	28	8	10 ⁻⁵	50	40	143	5.65	8.18172046	1.44681372
3	10 ⁻²	50	30	23	10 ⁻⁵	50	16	95	6.28	8.32613795	1.32599833
4	10 ⁻²	50	144	77	10 ⁻⁵	50	32	172	5.40	7.71144433	1.42707328
5	10 ⁻²	50	89	49	10 ⁻⁵	50	24	138	5.60	7.94318867	1.41914303
6	10 ⁻²	50	60	79	10 ⁻⁵	50	15	137	5.52	7.45828835	1.3507817
7	10 ⁻²	50	44	13	10 ⁻⁶	100	9	51	6.93	9.88406947	1.4262257
8	10 ⁻²	50	42	62	10 ⁻⁶	100	11	41	7.18	8.10363087	1.12904528
Mean	1.337804										

Based on the above three tables, the relative fitness values from each competition set were compared together. The Levene test was done for testing the equal variances of the sets. This value is 0.64523, which means the sets come from groups with approximately equal variances. However, the Shapiro test shows that the groups don't have normal distribution (p-value = 0.00112), thus, the non-parametric Kruskal-Wallis test was carried out to compare the mean ranks of the relative fitness of each group. The result showed that one or more of the group means are not identical (p-value = 0.005297). Looking at the comparative stripchart and boxplots (Figure S1) of these three group shows that probably the Engineering control (The SAE in the plots) is creating this difference:

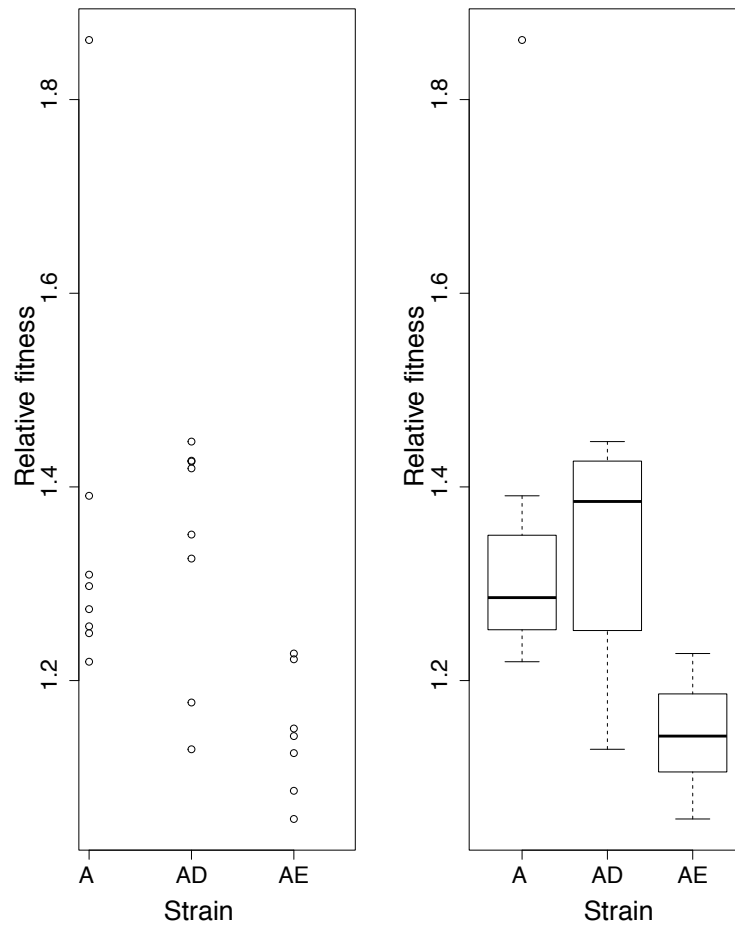


Figure S1. Comparative stripchart (left) and boxplot (right) of the data in tables S4 to S5 (Abbreviations: A= wild type A506, AD= A506 Δ SecGlnGluGly, AE= A506_Engineering control).

Next, pairwise mann whitney tests was done to compare each pair independently. (Bonferroni adjusted p -value for pairwise comparison of SA and SAD = 1.0), thus it was concluded that the mutant relative fitness is not significantly different with the one with the wild type.

Competition results for Pf0-1 strains

More details of three competition sets results explained in Figure 3.14 are explained in the following tables:

Table S7. SBW25-*lacZ* vs wild type Pf0-1(WT)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25_ <i>lacZ</i>	WT	D	V	SBW25_ <i>lacZ</i>	WT	WT	SBW25- <i>lacZ</i>	
1	10 ⁻²	50	48	12	10 ⁻⁵	40	22	23	5.90	7.33519929	1.24231645
2	10 ⁻²	50	25	9	10 ⁻⁵	40	5	2	5.08	5.18053433	1.02075998
3	10 ⁻²	50	18	5	10 ⁻⁵	40	12	7	6.28	7.02108396	1.11815895
4	10 ⁻²	50	33	6	10 ⁻⁵	40	6	5	4.98	6.50229017	1.30571651
5	10 ⁻²	50	45	8	10 ⁻⁵	40	7	7	4.82	6.55108034	1.3580579
6	10 ⁻²	50	15	7	10 ⁻⁵	40	13	4	6.54	6.12499594	0.93632741
7	10 ⁻²	50	17	7	10 ⁻⁵	40	28	20	7.18	7.73443385	1.07667893
8	10 ⁻²	50	31	8	10 ⁻⁵	40	8	11	5.33	7.00306546	1.31387967
Mean	1.171487										

Table S8. SBW25-*lacZ* vs Pf0-1 Engineering control (EC)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25_ <i>lacZ</i>	EC	D	V	SBW25_ <i>lacZ</i>	EC	EC	SBW25- <i>lacZ</i>	
1	10 ⁻²	50	8	5	10 ⁻⁵	50	26	26	8.09	8.5564139	1.05812265
2	10 ⁻²	50	8	2	10 ⁻⁵	50	25	21	8.05	9.25913054	1.15060425
3	10 ⁻²	50	17	1	10 ⁻⁵	50	22	20	7.17	9.90348755	1.3820907
4	10 ⁻²	50	7	1	10 ⁻⁵	50	58	32	9.02	10.3734912	1.14976279
5	10 ⁻²	50	15	1	10 ⁻⁵	50	23	18	7.34	9.79812704	1.33576835
6	10 ⁻²	50	1	1	10 ⁻⁵	50	22	14	10.00	9.54681261	0.95479605
7	10 ⁻²	50	2	1	10 ⁻⁵	50	31	18	9.65	9.79812704	1.01549777
8	10 ⁻²	50	5	1	10 ⁻⁵	50	19	24	8.24	10.0858091	1.22359666
Mean	1.158780										

Table S9. SBW25-*lacZ* vs Pf0-1ΔSecGlnGluGly (M)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25_ <i>lacZ</i>	M	D	V	SBW25_ <i>lacZ</i>	M	M	SBW25- <i>lacZ</i>	
1	10 ⁻²	50	247	36	10 ⁻⁵	50	31	16	4.83	6.09682506	1.26166768
2	10 ⁻²	50	16	1	10 ⁻⁵	50	16	4	6.91	8.29404964	1.20068666
3	10 ⁻²	50	21	5	10 ⁻⁵	50	80	11	8.25	7.69621264	0.93341061
4	10 ⁻²	50	30	4	10 ⁻⁵	50	17	5	6.34	7.13089883	1.12478803
5	10 ⁻²	50	14	2	10 ⁻⁵	50	14	14	6.91	8.85366543	1.28169935
6	10 ⁻²	50	10	1	10 ⁻⁵	50	11	6	7.00	8.69951475	1.24224381
7	10 ⁻²	50	3	1	10 ⁻⁵	50	26	10	9.07	9.21034037	1.01578218
8	10 ⁻²	50	7	1	10 ⁻⁵	50	23	11	8.10	9.30565055	1.14922324
Mean	1.151188										

Based on the above three tables, the relative fitness values from each competition set were compared together. Levene test was done for testing the equal variances of the sets. This value is 0.60401, which means the sets come from groups with approximately equal variances. Shapiro test shows that the groups have normal distribution (p -value = 0.3833), thus, one-way ANOVA test was applied to compare the equality of the of the relative fitness mean of each group. The result showed that there is no significant difference between the mean relative fitness of the three groups (p -value = 0.96001). The comparative stripchart and boxplots (Figure S2) also provides more evidence for this conclusion.

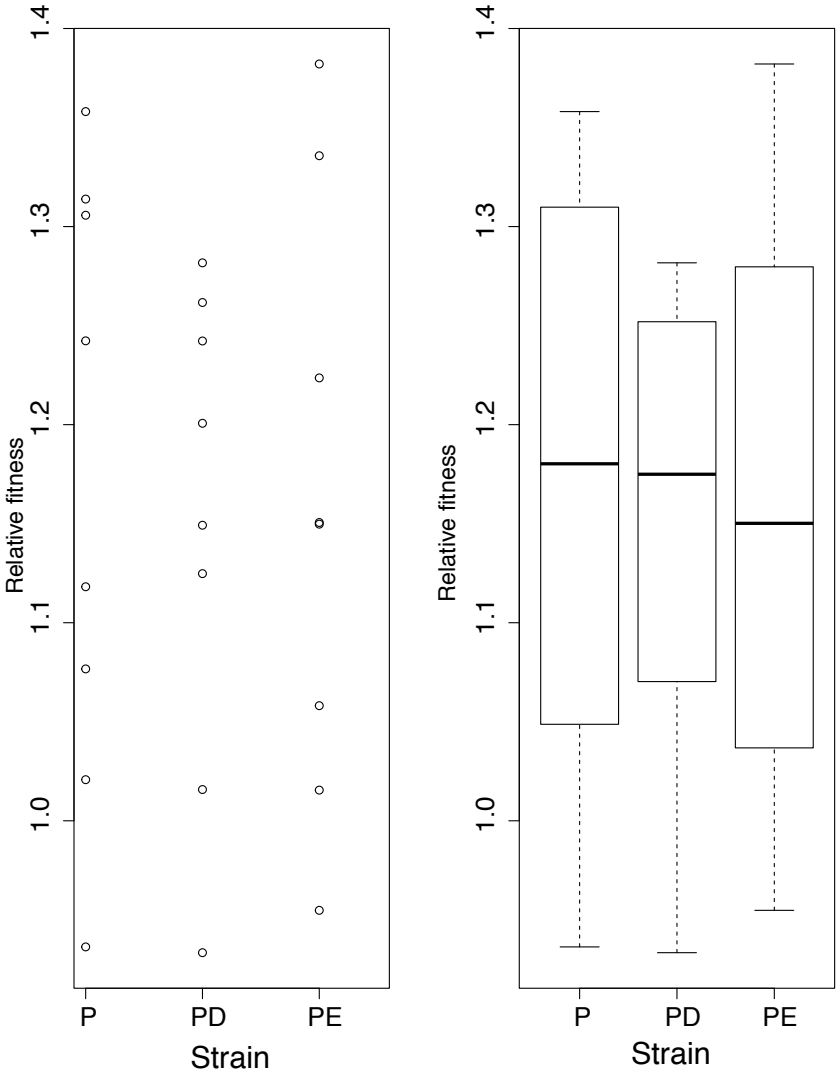


Figure S2. Comparative stripchart (left) and boxplot (right) of the data in tables S7 to S9 (Abbreviations: P= wild type Pf0-1, PD= Pf0-1 Δ SecGlnGluGly, PE= Pf0-1_Engineering control).

Competition results for SBW25 strains

There were four sets of competitions for this mutant including the original and the evolved wild types and the mutants (Figures 3.23). More details of the raw data and statistical analyses are explained below:

Table S10. SBW25-*lacZ* vs original wild type SBW25 (OWT)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25_ <i>lacZ</i>	OWT	D	V	SBW25_ <i>lacZ</i>	OWT	SBW25- <i>lacZ</i>	OWT	
1	10 ⁻²	50	68	94	10 ⁻⁶	100	66	80	8.4873402 3	8.3559250 4	0.9845163 3
2	10 ⁻²	50	80	86	10 ⁻⁶	100	118	117	8.9058511 8	8.8250198 3	0.9909237 9
3	10 ⁻³	100	16	30	10 ⁻⁶	100	104	105	8.7795574 6	8.1605182 5	0.9294908 4
4	10 ⁻²	50	65	57	10 ⁻⁶	100	78	86	8.6995147 5	8.9284892 2	1.0263203 7
5	10 ⁻³	100	42	60	10 ⁻⁶	100	300	300	8.8738681 4	8.5171931 9	0.9598061 5
6	10 ⁻²	50	90	70	10 ⁻⁶	100	38	44	7.6549696 8	8.0528875 8	1.0519816 4
7	10 ⁻³	50	19	33	10 ⁻⁵	50	50	81	5.5727542 1	5.5031117 8	0.9875030 5
8	10 ⁻²	50	66	58	10 ⁻⁵	50	100	103	7.3232707 2	7.4820412 6	1.0216802 8
Mean	0.9940										

Table S11. SBW25-*lacZ* vs evolved wild type SBW25 (EWT)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25_ <i>lacZ</i>	EWT	D	V	SBW25_ <i>lacZ</i>	EWT	SBW25- <i>lacZ</i>	EWT	
1	10 ⁻³	100	49	42	10 ⁻⁶	100	35	31	6.57128304	6.60407287	1.00498987
2	10 ⁻³	100	51	29	10 ⁻⁵	50	122	91	6.17051278	6.44188104	1.04397824
3	10 ⁻³	100	30	24	10 ⁻⁶	100	30	29	6.90775528	7.09699728	1.02739559
4	10 ⁻³	100	33	27	10 ⁻⁶	100	55	47	7.4185809	7.46206601	1.00586165
5	10 ⁻³	100	39	41	10 ⁻⁵	50	72	73	5.91142184	5.87520474	0.99387337
6	10 ⁻³	100	21	27	10 ⁻⁶	100	29	29	7.23052867	6.97921424	0.96524259
7	10 ⁻³	100	41	88	10 ⁻⁵	50	77	99	5.92855072	5.4161004	0.91356229
8	10 ⁻³	100	60	68	10 ⁻⁶	100	76	83	7.14414406	7.10708818	0.99481311
Mean	0.9937										

Table S12. SBW25-*lacZ* vs original SBW25ΔGluGlyGluGly (OM)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25- <i>lacZ</i>	OM	D	V	SBW25- <i>lacZ</i>	OM	SBW25- <i>lacZ</i>	OM	
1	10 ⁻³	100	36	22	10 ⁻⁵	50	134	6	8.91522332	3.99903438	0.44856245
2	10 ⁻³	100	57	30	10 ⁻⁶	100	60	5	6.95904857	5.11599581	0.73515736
3	10 ⁻²	50	117	103	10 ⁻⁵	50	205	15	7.46859132	4.98107649	0.66693654
4	10 ⁻³	100	37	21	10 ⁻⁶	100	145	10	8.27357111	6.16581793	0.74524264
5	10 ⁻³	50	69	36	10 ⁻⁵	50	136	8	5.28371857	3.10109279	0.58691483
6	10 ⁻²	50	117	18	10 ⁻⁶	100	60	4	7.84936382	7.01311579	0.89346296
7	10 ⁻³	100	90	41	10 ⁻⁵	50	58	2	4.85895071	2.27789248	0.46880337
8	10 ⁻³	100	40	18	10 ⁻⁵	50	90	7	6.10924758	4.35385576	0.71266645
Mean	0.6767										

Table S13. SBW25-*lacZ* vs evolved SBW25ΔGluGlyGluGly (EM)

Replicate	Count at 0h				Count at 24h				Malthusian		Relative fitness
	D	V	SBW25- <i>lacZ</i>	EM	D	V	SBW25- <i>lacZ</i>	EM	SBW25- <i>lacZ</i>	EM	
1	10 ⁻²	50	59	36	10 ⁻⁵	50	54	11	6.81920188	5.72213161	0.83912043
2	10 ⁻²	50	132	50	10 ⁻⁵	50	74	9	6.32901845	5.19295685	0.82049956
3	10 ⁻²	100	101	37	10 ⁻⁵	50	104	17	7.63017284	6.82319789	0.89423897
4	10 ⁻²	50	74	32	10 ⁻⁶	100	49	4	5.8023633	6.43775165	1.1095051
5	10 ⁻²	50	90	33	10 ⁻⁵	50	82	15	6.81466486	6.11929792	0.89796021
6	10 ⁻²	50	54	36	10 ⁻⁵	50	76	16	7.24950457	6.09682506	0.84099886
7	10 ⁻³	100	69	54	10 ⁻⁵	50	115	25	8.11172808	4.52820914	0.5582299
8	10 ⁻²	50	88	42	10 ⁻⁶	100	52	9	5.688515	6.97674815	1.22646212
Mean	0.8983										

For comparing the relative fitness of four groups, the raw data were first plotted in stripchart and boxplots (Figure S3). The significant difference between the relative fitness of original mutant with other groups is already obvious.

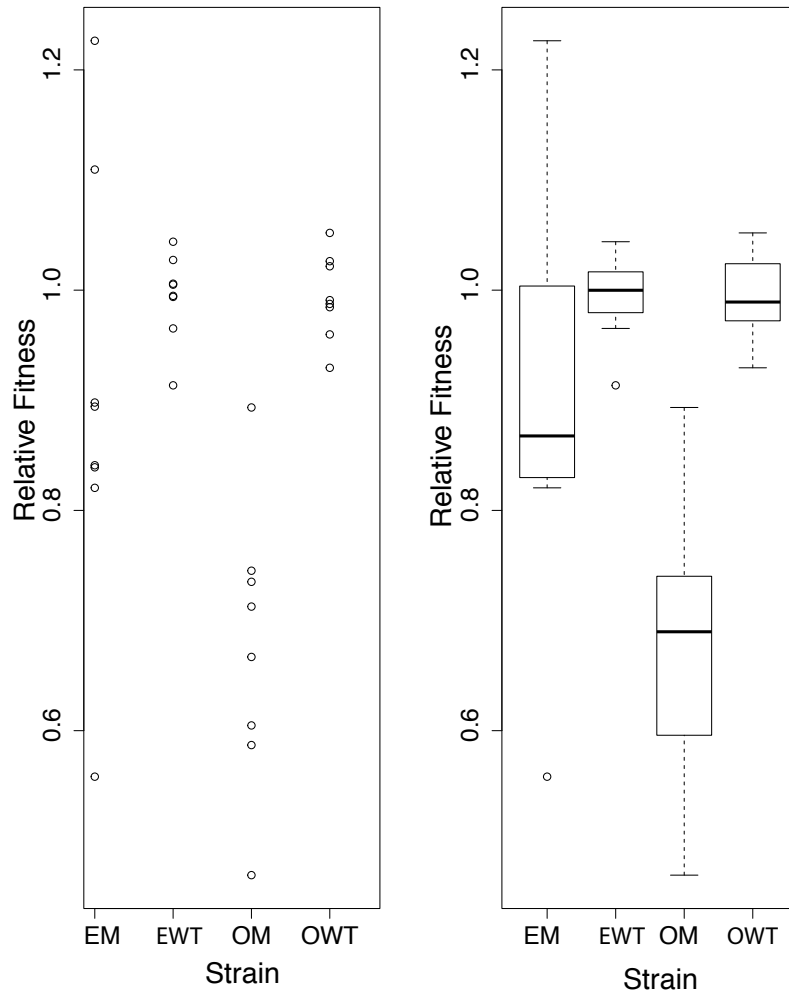


Figure S3. Comparative stripchart (left) and boxplot (right) of the data in tables S10 to S13 (Abbreviations: EM= evolved mutant (SBW25 Δ GluGlyGluGly), EWT= evolved wild type SBW25, OM= original mutant (SBW25 Δ GluGlyGluGly), OWT= original wild type SBW25).

The Shapiro test and Levene- test results indicated that the normal distribution and equal variance assumptions were violated (p -value <0.05). Thus, thus, the non-parametric Kruskal-Wallis test was carried out to compare the mean ranks of the relative fitness of each group. The result showed that one or more of the group means are not identical (p -value <0.05). Next, data set was analyzed separately for normal distribution and equal variances, where this assumptions were met, two-sample t -tests were applied to compare the means of each two groups indicated in table S14, and where any of the assumptions were not met, Welch two-sample t - test was applied (shaded in grey in Table S14).

Table S14. p-values for comparing the means of SBW25 strains

Strain	Compared to	<i>p</i>-value	Significance level
OWT	EWT	0.9876	ns
OWT	OM	0.0001554	***
OWT	EM	0.2061	^a ns
EWT	OM	0.0001198	***
EWT	EM	0.2070	ns
OM	EM	0.01929	*

^ans= not significant

grey shading = Welch two-sample *t*- test

No shading = two-sample *t*-test

Contributions to Zahra Khomarbaghi's thesis

Thesis title: The function and origin of multi-copy tRNA genes in *Pseudomonas fluorescens*

Student's name: Zahra Khomarbaghi

The thesis was written by the candidate Zahra Khomarbaghi. Below is a detailed listing of individual contributions to the topics covered in this thesis.

Topic I – Manipulation and characterization of accessory tRNA genes in *P. fluorescens* A506 and Pf0-1:

JG conceived the study. **ZK** performed the phylogenetic analyses of *Pseudomonas* strains with input from FB and JG. JG and **ZK** designed the protocol for making gene manipulations in A506 and Pf0-1. **ZK** performed tRNA gene deletions in A506/Pf0-1. **ZK** collected and analyzed growth and fitness data, and interpreted the results.

Topic II – Making and characterizing multi-copy tRNA gene deletions in *P. fluorescens* SBW25:

JG and **ZK** conceived the study, with some discussions with HJP. **ZK** performed the tRNA gene deletions in SBW25, and collected growth and fitness data. **ZK** analyzed the data.

Topic III – Compensating for a suboptimal tRNA gene set in *P. fluorescens* SBW25:

JG and **ZK** conceived the study. **ZK** designed and performed the evolution experiment. **ZK** performed and analyzed the growth and fitness assays. JG and **ZK** analyzed the next generation sequencing data and interpreted the results.

Publications: No part of the thesis has yet been published, or submitted to any journals yet.

Author Names in Alphabetical order

AM: Anuradha Mukherjee

FB: Frederic Bertels

HJP: Hye-jin Park

JG: Jenna Gallie

ZK: Zahra Khomarbaghi

Jenna Gallie (PhD supervisor)

Zahra Khomarbaghi (PhD candidate)

Acknowledgement

First and foremost, I thank my supervisor Dr. Jenna Gallie for funding this PhD and for all the fruitful discussions we had and for all the scientific lessons that I learnt from her. I am very grateful to her for helping me improve my scientific writing as a non-native English speaker. I express my deep gratitude to Jenna for her patience and for her honest feedbacks on my work and for helping me to try to be realistic about scientific achievements.

I also thank my thesis advisory committee members including Prof. Eva Stukenbrock and Prof. Bernhard Haubold for their valuable comments on the progress of my PhD.

I am grateful to the examiners committee including Prof. Dr. Eva Stukenbrock, Prof. Dr. Hinrich Schulenburg, and Dr. Jenna Gallie for taking the time for reading and examining this thesis. I also thank Prof. Dr. Stanislav Gorb for accepting being the chair person in my oral examination.

I would like to thank Dr. David Rogers for mentoring me in the first year of my PhD.

Thank you to my colleagues from the Gallie lab, for sharing and discussing the protocols and the lab experiences. Also, many thanks to Gunda for her technical help and for her patience while she would face a list of hundreds of plates to be prepared!

I am grateful to Derk for his support not only in the IT related problems but also in all issues that he could help. I am also very grateful to Britta and Petra for their administrative support and helping with understanding tons of German letters. Also, thanks to Sieggi, the night janitor for his kind and heart-warming smile and “GUTEN ABENDS” in the long working nights during my PhD.

Many, many thanks to people who helped me to improve the quality of this thesis by reading and proofreading it; especial thanks to my friends Alice and Gillian for their comments on my writing and thanks a lot to Micheal Sieber for translating the abstract of this thesis to German very fast! and thanks to Nico for reviewing it.

I would like to thank all my kind caring friends for sharing the good memories during my PhD.

Finally, I would like to wholeheartedly thank my nice big family for their unconditional supports from far away, Thank you to my lovely nice mum and dad for encouraging me to follow my goals, deep thanks to my sweet sister Narges for always being there for me; Narges your kindness always keep my heart warm even when we are thousands kilometers far away from each other. Many thanks to my sisters Somayeh and Samaneh and my lovely sweet niece Nila for traveling the long way to Plön to visit me and thank to my brother and his cutie Parnyan and my nice in-laws for their love and support.

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 a 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, December 2nd, 2019.

Zahra Khomarbaghi

Curriculum vitae



Zahra Khomarbaghi

Nationality: Iranian

E-mail khomarbaghi@evolbio.mpg.de

Address Rautenburg Strasse, 2a, Plön 24306, Germany

Education

2017- 2019 PhD

Max Planck institute for evolutionary biology

2010- 2013 Master of Science in Microbiology

University of Tehran, Tehran, Iran

Thesis: Investigation of bacterial community dynamics during bioremediation in hydrocarbon contaminated soil using genetic methods

2004-2008 Bachelor of Science (Cellular & Molecular Biology-Microbiology)

University of Tehran, Tehran, Iran

2001.2004 Diploma of Experimental Science

Alzahra state high school, Tehran, Iran

Professional Experiences

2015 –2016 Research associate in Chemistry and Chemical Engineering Research Center of Iran (CCERCI), Tehran, Iran

2013 - 2015 Biology teacher in Zoha school complex, Tehran, Iran

2010 – 2011 Research assistant in the project of **Biological Control of Cyanobacteria causing “Harmful Algal Blooms”** Research Institute of Biotechnology, University of Tehran

2009 – 2010 Research assistant in the project of **production of biological Nitrate Fertilizer** Research Institute of Biotechnology, University of Tehran

2006 - 2007 Biology teacher in summer schools organized by student charity group in Khozestan and Hormozgan state high schools, Iran

Conference events in the PhD

[1] Zahra Khomarbaghi, Fredric Bertels, Hye jin Park, Jenna Gallie **“Dissecting the evolution of tRNA gene pools in *Pseudomonas fluorescens*”** RAPID EVOLUTIONARY ADAPTATIONS POTENTIALS AND CONSTRIANTS DFG SPP1819 INTERNATIONAL MEETING 17TH-19TH OF OCTOBER 2018 WISSENSCHAFTS ZENTRUM KIEL, GERMANY

[2] Zahra Khomarbaghi, Fredric Bertels, Jenna Gallie **“evolution of tRNA gene pools”** 16th INTERNATIONAL CONFERENCE ON PSEUDOMONAS 05 - 09 SEPTEMBER 2017, LIVERPOOL, UK

Publications

[1] **Z. Khomarbaghi**, M.A. Amoozegar, M. Shavandi, S.M.M Dastgheyb **“bacterial community dynamics during bioremediation of alkane and PAHs contaminated soil in Siri island: Persian Gulf a microcosm study”** international journal of environmental science and technology, 2019.

[2] **Z. Khomarbaghi**, M.A. Amoozegar, M. Shavandi, S.M. Dastgheyb **“Investigation of the potential of indigenous bacteria of SIRI island soil in remediation of petroleum contaminants by a microcosm method”** Persian journal of petroleum research, 2014.