THE EVOLUTIONARY FATE OF DUPLICATE TRNA GENES IN THE BACTERIUM *Pseudomonas fluorescens* SBW25

Dissertation

in fulfilment of the requirements for the degree of Dr. rer. nat.

of the Faculty of Mathematics and Natural Sciences

at Christian-Albrechts-Universität zu Kiel

Submitted by

Wing Yui Ngan

Kiel, 2024

First examiner: Dr Jenna Gallie

Second examiner: Prof. Dr Tal Dagan

Date of the oral examination: 7th June 2024

I, **Wing Yui Ngan**, declare that apart from my supervisor's guidance, the content and design of the thesis is my research work in its entirety. The thesis has been prepared following the Rules of Good Scientific Practice of the German Research Foundation. No academic degree has ever been withdrawn.

Part of this thesis was used to write the following research articles:

Khomarbaghi, Z., Ngan, W.Y., et al., *Large-scale duplication events underpin population-level flexibility in tRNA gene copy number in Pseudomonas fluorescens SBW25*. Nucleic Acids Res, 2024. **52**(5): p. 2446-2462.

A full version of the manuscript can be found here:
https://academic.oup.com/nar/article/52/5/2446/7595401

Authors' contributions:

Wing Yui Ngan: Carrying out population growth curve measuring, tRNA pool sequencing, characterising of mutants such as stability test, population sequencing and the data analysis and visualising of the data from the above experiment.

Zahra Khomarbaghi: Constructed the mutant (Δ*EGEG* and rWT) used in this study. Performed the daily serial transfer evolution experiment, sampled the isolates, isolate genome sequencing.

Gökçe B Ayan: Construction of mutant (Δ*glyGCC*) tRNA pool sequencing, analysis of mutants.

Sungbin Lim: Quantifying the mutants in population using PCRs and quantifying the results from gel electrophoresis.

Pabitra Nandy: Fitness assays.

Jenna Gallie: Conceptualization, performing stability assays and tRNA sequencing, writing of manuscript.

All authors contributed to data analysis, visualising of the data and reviewers' rebuttal process.

Signature:

Abstract

The transfer RNA (tRNA) content of cells affects the efficiency of protein synthesis. To study how organisms can adapt to novel translational demands, our laboratory uses two engineered strains of *Pseudomonas fluorescens* SBW25, each lacking one or more tRNA genes. Previously, we have shown that these slow-growing strains rapidly recover fitness in serial transfer evolution experiments by duplicating large chromosome segments (up to 1 Mb), each containing a small (~100 bp) compensatory tRNA gene. However, while adaptive, these large duplications are mechanistically unstable and hence are unlikely to persist over longer evolutionary time scales.

Here, this project investigates the evolutionary fate of the duplications and the new tRNA gene copies that they contain. It restarted and extended the evolution experiment to 100 transfers (~750 generations) and characterised the evolving lines. There are general adaptations observed in all lines, where part of the cells in the population lost the motile ability through various mutation pathways. It suggested the reserved energy from flagellate provide fitness advantage when being motile is not necessary in the growth condition.

Except for one SNP promoter mutant observed in one line, all the identified mutations compensating the tRNA gene(s) loss were duplication mutations. Within each evolving population, various duplication fragments rapidly arise and compete. Over time, progressively smaller – and hence, mechanistically more stable – duplication fragments arise and dominate in all lines. The smallest is a duplication fragment of only 236 bp, encompassing the compensatory tRNA gene and promoter. Thus, we have observed an evolutionary mechanism for bacteria to acquire an extra copy of a tRNA gene.

Population sequencing revealed the evolution dynamics of these identified mutations. The duplication mutants with small duplication fragments were increasing in relative abundance in the population. This suggested they will dominate in the population and may outcompete other less fit mutation.

Other than the duplication mutation, a SNP promoter mutant arose in one of the lines. This mutant is stable but did not displace the duplication fragments in the line in question. To

investigate this outcome further, we use mathematical modelling and simulations to identify key factors influencing the fate of the two mutational classes.

Together, the results provide a detailed, real-time example of a bacterial tRNA gene set evolving in response to translational challenges. This study investigated the enormously growing interest in how recombination-related mutations shape living organisms in evolution and provided information on the evolutionary success of bacteria via changing their gene copy numbers. The results further contributed to our understanding of evolution across the tree of life.

Zusammenfassung

(DeepL translated)
Der Gehalt an Transfer-RNA (tRNA) in Zellen beeinflusst die Effizienz der Proteinsynthese. Um zu untersuchen, wie sich Organismen an neue Translationsanforderungen anpassen können, verwendet unser Labor zwei manipulierte Stämme von Pseudomonas fluorescens SBW25, denen jeweils ein oder mehrere tRNA-Gene fehlen. Wir haben bereits gezeigt, dass diese langsam wachsenden Stämme in seriellen Transferevolutionsexperimenten ihre Fitness schnell wiedererlangen, indem sie große Chromosomenabschnitte (bis zu 1 Mb) duplizieren, die jeweils ein kleines (~100 bp) kompensatorisches tRNA-Gen enthalten. Diese großen Duplikationen sind zwar anpassungsfähig, aber mechanistisch instabil, so dass es unwahrscheinlich ist, dass sie über längere evolutionäre Zeiträume bestehen bleiben.

In diesem Projekt wird das evolutionäre Schicksal der Duplikationen und der neuen tRNA-Gen-Kopien, die sie enthalten, untersucht. Es hat das Evolutionsexperiment neu gestartet und auf 100 Transfers (~750 Generationen) erweitert und die sich entwickelnden Linien charakterisiert. In allen Linien wurden allgemeine Anpassungen beobachtet, wobei ein Teil der Zellen in der Population durch verschiedene Mutationswege die Bewegungsfähigkeit verlor. Dies deutet darauf hin, dass die reservierte Energie der Flagellaten einen Fitnessvorteil bietet, wenn die Motilität unter den Wachstumsbedingungen nicht erforderlich ist.

Mit Ausnahme einer SNP-Promotor-Mutante, die in einer Linie beobachtet wurde, waren alle identifizierten Mutationen, die den Verlust der tRNA-Gene kompensierten, Duplikationsmutationen. Innerhalb jeder sich entwickelnden Population entstehen schnell verschiedene Duplikationsfragmente und konkurrieren miteinander. Mit der Zeit entstehen immer kleinere - und damit mechanistisch stabilere - Duplikationsfragmente, die in allen Linien dominieren. Das kleinste ist ein Duplikationsfragment von nur 236 bp, das das kompensatorische tRNA-Gen und den Promotor umfasst. Wir haben also einen evolutionären Mechanismus beobachtet, der es Bakterien ermöglicht, eine zusätzliche Kopie eines tRNA-Gens zu erwerben.

Die Populationssequenzierung zeigte die Evolutionsdynamik dieser identifizierten Mutationen. Die Duplikationsmutanten mit kleinen Duplikationsfragmenten nahmen an relativer Häufigkeit

in der Population zu. Dies deutet darauf hin, dass sie in der Population dominieren und andere, weniger geeignete Mutationen verdrängen könnten.

Neben der Duplikationsmutation trat in einer der Linien auch eine SNP-Promotor-Mutante auf. Diese Mutante ist stabil, hat aber die Duplikationsfragmente in der fraglichen Linie nicht verdrängt. Um dieses Ergebnis weiter zu untersuchen, verwenden wir mathematische Modellierung und Simulationen, um die Schlüsselfaktoren zu identifizieren, die das Schicksal der beiden Mutationsklassen beeinflussen.

Zusammengenommen liefern die Ergebnisse ein detailliertes Echtzeit-Beispiel für einen bakteriellen tRNA-Gensatz, der sich als Reaktion auf translatorische Herausforderungen entwickelt. Diese Studie untersuchte das enorm wachsende Interesse daran, wie rekombinationsbedingte Mutationen lebende Organismen in der Evolution formen, und lieferte Informationen über den evolutionären Erfolg von Bakterien durch Veränderung ihrer Genkopienzahl. Die Ergebnisse trugen weiter zu unserem Verständnis der Evolution im Lebensbaum bei.

Acknowledgements

I am truly thankful and grateful for the opportunities I got from my PhD supervisor, Dr Jenna Gallie, to work with her for almost four years. Over the last 3.75 years, I have had some of the best time doing science research. Thanks for providing a well-organised team and friendly environment. Working in evolutionary biology excites me. I have not regretted the decision to move to this research area. I am only thankful for having such a great opportunity. I also would like to thank everyone from Dr Jenna Gallie's Microbial Evolutionary Dynamics Research Group: Gunda Dechow-Seligmann, Dr. Gökçe Ayan, Dr. Pabitra Nandy and Sungbin Lim. Without Dr Jenna Gallie's help, support and supervision, this thesis would not have existed. Also, Gunda helped with the lab work. Our technician has magician hands doing experiments. Moreover, our postdocs (Dr Gökçe Ayan and Dr Pabitra Nandy) and my PhD fellow Sungbin Lim, thank you for all the advice and discussion.

I started my PhD in 2020 with the pandemic. Also, I never had the experience of living in a town with only ~8000 people before moving to Plön. I got help and support from lots of people from or outside the institute. From the institute, I thank everyone from the Department of Theoretical Biology led by Prof. Dr Arne Traulsen, for all the good scientific discussion and good times. Outside the institute, I thank my teammates from the athletics club Kieler TB and handball team TSV Plön.

Particularly, I would like to thank my lifelong best friend, Dr Corby YC Cheung, for all the support. Also, my friend Anastasiia Strekneva for pushing me to finish my PhD.

I wish to present this PhD work to two people: my best friend's brother Quinn YW Cheung, who passed away on the day when I decided to do science research 9 years ago, and my grandma, who passed away in the first winter when I came to Germany. My grandma gave me lots of support in doing what I liked. Finally, huge thanks to the mega support I got from my son, my Mum, and my brother. Even if I am not sure if they know what I am doing.

16 years ago, I was an athletic student. I had never thought of attending university at that time. 16 years later, I am submitting this thesis. So far, I have enjoyed it a lot.

Doing science research is fun.

Table of Contents

Chapter I - Introduction

1.1 DNA to RNA and protein synthesis

DNA, short for Deoxyribonucleic acid, is a crucial molecule within living cells. It carries genetic information for the development and functioning of an organism. In 1871, a Swiss biochemist named Friedrich Miescher discovered DNA [1, 2]. He was studying the chemical makeup of white blood cells, using pus from fresh surgical bandages as his source [2]. While initially interested in all cell components, Miescher focused on the nucleus after noticing that when treated with acid, it formed a substance he called 'nuclein.' Miescher, along with Richard Altmann and Albrecht Kossel, further studied 'nuclein,' which was later renamed nucleic acid by Altmann. Kossel further found that nucleic acid contained purine and pyrimidine bases, a sugar, and phosphate [1]. In the 1930s, additional research identified the four bases and the presence of deoxyribose, leading to the name deoxyribonucleic acid (DNA). Erwin Chargaff discovered that DNA molecules from a species always contained the same amounts of adenine (A), thymine (T) cytosine (C), and guanine (G). In 1944, Oswald Avery, Colin MacLeod, and Maclyn McCarty carried out studies that confirmed DNA is a genetic material (Avery–MacLeod–McCarty experiment) [3]. Next, in 1953, James Watson and Francis Crick determined the 3-dimensional double helix structure of DNA [4]. Each strand has a sugar-phosphate backbone. Each sugar is attached to one of the four: A, T, C, and G. The two strands are connected by hydrogen bonds, with A paired with T and C paired with G. The DNA chain has a 5’ phosphoryl end and another end with a 3’ hydroxyl end. The two DNA strands run in opposite directions from the 5’ end to the 3’ end.

RNA is an intermediate between DNA and protein [5]. RNA, or ribonucleic acid, is essential for the functioning of living organisms. RNA typically exists as a single-stranded molecule. Some special RNA viruses are double-stranded [6]. RNA is made up of nucleotides, which are ribose sugars attached to nitrogenous bases and phosphate groups. The nitrogenous bases include adenine (A) (Noted that adenosine is the larger nucleotide molecule made up of adenine), guanine (G), uracil (U), and cytosine (C) [6]. There are mainly three types of RNA: messenger RNA (mRNA), ribosomal RNA (rRNA), and transfer RNA (tRNA). mRNA is transcribed from DNA and contains the genetic blueprint to make proteins. rRNA forms ribosomes, which are essential in protein synthesis. A ribosome contains a large and small ribosomal subunit. In

prokaryotes, a small 30S and large 50S ribosomal subunit make up a 70S ribosome. tRNAs translate mRNA into proteins. They consist of a 3' acceptor site, 5' terminal phosphate, D arm, T arm, and anticodon arm. The primary function of a tRNA is to carry amino acids on its 3' acceptor site to a ribosome complex with the help of aminoacyl-tRNA synthetase [6]. tRNA is important in protein synthesis and affects the efficiency of the translation process [7].

1.2 Translation by variable tRNA gene sets in bacteria

Genetic information from a cell is passed from DNA to mRNA (messenger RNA) through transcription. The process of the genetic information from mRNA to amino sequences and proteins is called translation. Transfer RNAs (tRNAs) serve as adaptor molecules that align amino acids along with the mRNA template, translating 61 types of codons into 20 kinds of amino acids. Each tRNA carries a particular type of amino acid and contains a 3-bp anticodon that recognises specific 3-bp codon(s) from mRNA (Fig 1.1). These two regions determine the translation specificity (thus, the amino acid carried 3' acceptor end and three base anticodon). If the tRNAs are classified by their anticodon, there are 64 of them, with three decoding stop codons (UAA, UAG or UGA) to terminate the protein synthesis in the translation process [8].

Translation is commonly acknowledged to proceed through three primary phases: initiation, elongation, and termination, with some perspectives also considering a recycling phase. Initiation is the beginning of the translation process. The small ribosomal subunit binds to the mRNA and the initiator tRNA carrying the amino acid methionine attaches to the start codon. During elongation, the ribosome moves along the mRNA, and tRNA molecules bring amino acids to the ribosome in the order specified by the mRNA codons. Peptide bonds form between adjacent amino acids, extending the growing polypeptide chain. Termination marks the end of the translation process. When the ribosome reaches a stop codon (i.e., UAA, UAG and UGA) on the mRNA, it releases the completed polypeptide chain, and the ribosomal subunits dissociate.

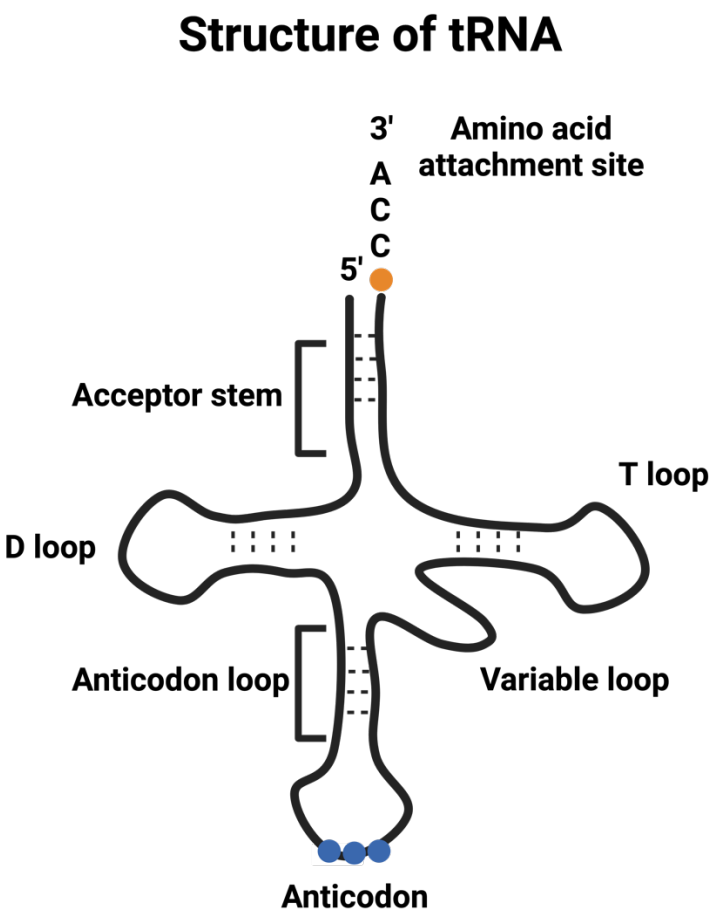


Fig 1.1. Cloverleaf representation of a tRNA structure. A tRNA consists of an acceptor stem, D loop, anticodon loop containing the anticodon (3bp denoted by blue dots here), variable loop and T loop. The discriminator base (orange dot) is the unpaired base at the 3′ end before the terminal CCA residues. Amino acids ligate onto the 3′ terminal. The figure was drawn using Biorender.com.

Accurate codon-anticodon matching is crucial to secure efficiency during translation [9]. However, codon-anticodon matching is not straightforward. Firstly, Watson and Crick base pairing rules allow C-G and A-U pairings, and many additional ("wobble") base pairings are also tolerated (*e.g.*, G-U, I-U, I-A, and I-C pairings) (reviewed in [10]). During the elongation of the protein chain in the translation process, each codon is sequentially matched to a tRNA by randomly sampling tRNA types from the mature tRNA pool until a correct match is found. This random sampling and matching process is the rate-limiting step of elongation [11]. Hence, the average time needed to translate any given codon heavily depends on the proportion of matching tRNAs in the pool; codons matched by a higher proportion of the pool will, on average, be translated more quickly than those with rarer matches. Therefore, how many and which tRNAs

are available to translate each codon plays a key role in determining translation efficiency and protein expression. The composition of the available pool of tRNAs (the "mature tRNA pool") is affected by numerous factors (reviewed in [12]), an important one of which is the tRNA genes encoded in the genome (the "tRNA gene set"). Since tRNA gene set composition affects the types and abundances of mature tRNAs present in the cell, it ultimately affects translational efficiency [11, 13, 14].

The composition of tRNA gene sets across the tree of life is not the same [15-17]. Different organisms may consist of varying tRNA genes in tRNA types and copy numbers. There are several possible ways tRNA sets may change. For example, tRNA genes may be gained by horizontal gene transfer, lost by deletion, and altered by mutation (such as codon changes and duplication) [18-20]. Given the critical roles of tRNAs in many biological processes, there were plenty of evolutionary studies on tRNAs from computational or bioinformatic approaches [18-24]. However, very few direct experimental works observe the adaptation of tRNA gene sets [13, 25, 26].

1.3 The evolutionary emergence of new bacterial tRNA gene sets by large-scale, tandem duplication events

Our laboratory has recently investigated the effect of deleting two different tRNA loci from the model bacterium *Pseudomonas fluorescens* SBW25. These allowed us to understand how tRNA gene sets evolve in adaption. More details are provided on the model organism and both tRNA studies below.

1.3.1 Using P. fluorescens *SBW25 as the model system*

P. fluorescens SBW25 is an environmental isolate from the leaf of a sugar beet plant in the United Kingdom, and similar strains are commonly found in soil and plant roots [27]. *P. fluorescens* SBW25 plays a crucial role in interacting with plants and promoting plant growth [28, 29]. In addition, it is also used as an evolutionary model system in the laboratory. For instance, it readily diversifies into ecotypes that form colonies with distinct morphologies [30]. These morphotypes have helped track genotypic and phenotypic changes in laboratory evolution experiments [30-

32]. The different colony morphologies include opaque, translucent, and wrinkly spreaders [27, 30, 33]. The adaptive radiation of SBW25 has been well studied for decades, and studies are still being carried out in this area.

As a result of the extensive use of *P. fluorescens* SBW25 in lab experiments, many tools are available. Examples include (i) a reliable, accurate reference genome sequence for the wild-type [34], (ii) a genetic engineering process that can be used to generate scar-free mutants (e.g., [30]), (iii) a neutrally marked wild-type strain for fitness assays [35], and (iv) an established mature tRNA pool sequencing protocol and pipeline [14]. Our laboratory is proficient in using all of the listed tools and techniques.

1.3.1.1 Mechanisms of the appearing of wrinkly spreaders phenotypes

P. fluorescens SBW25, as a strain widely used in bacterial evolutionary experiment, often observed the rising of wrinkly spreaders (WS) phenotypes in both static and shaking conditions [36]. *P. fluorescens* SBW25 contains the wss (Wrinkly Spreader Structural)operon, which is essential for the WS phenotype [9]. The wss operon in *P. fluorescens* SBW25 consists of ten genes, *wssA-J* [10]. Among these, WssBCDE consist cellulose synthase subunits, while WssFGHI are anticipated to participate in cellulose acetylation [10]. Additionally, WssA and WssJ are predicted to contribute to the proper cellular localisation of the synthase/acetylation complex. These mutants are called wrinkly spreaders as they form wrinkled colony morphology on agar. They also form mats that colonise the air-liquid interfaces to dominate the oxygen access [37].

Production of multiple structural components is involved in the mat formation process and is thus subject to sophisticated control mechanisms. The biosynthesis of many structural elements is regulated by the secondary messenger bis(3'-5')-cyclic dimeric guanosine monophosphate (c-di-GMP) [11, 12]. The intracellular c-di-GMP levels are determined by the expression levels of the two primary enzymes: diguanylate cyclases (DGCs) and phosphodiesterases (PDEs) (Fig 1.2) [13, 14]. DGCs synthesise c-di-GMP from two molecules of GTP (guanosine triphosphate). DGCs contain conserved GGDEF domains responsible for catalysing GTP's cyclisation into c-di-GMP [12]. On the other hand, PDEs degrade c-di-GMP into linear GMP. PDEs play a key role in maintaining the balance of c-di-GMP levels within the cell.

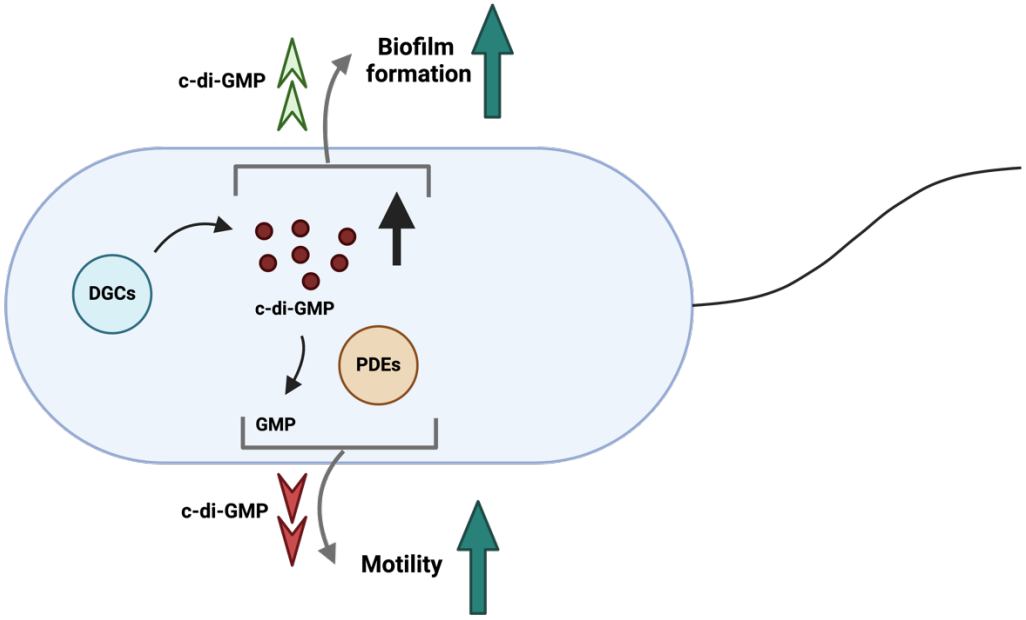


Fig 1.2. The regulation of c-di-GMP signalling by DGCs and PDEs. DGCs and PDEs control the global concentration of c-di-GMP and regulate biofilm formation, motility or other bacterial phenotypes. The figure was drawn using Biorender.com.

These enzymes typically contain EAL or HD-GYP domains, which are responsible for hydrolysing the phosphodiester bond in c-di-GMP [12, 13]. While DGCs and PDEs are primary enzymes involved in regulating c-di-GMP levels, other proteins and regulators can influence this signalling pathway. For example, PilZ domain proteins and c-di-GMP riboswitches. PilZ domain proteins are a widely recognised class of c-di-GMP receptors that binds c-di-GMP and mediates downstream effects [15]. c-di-GMP riboswitches are structured RNAs that regulate the expression of downstream genes in response to changing concentrations of the second messenger c-di-GMP [16]. The overall regulation of c-di-GMP levels is a dynamic process involving interactions between DGCs, PDEs, c-di-GMP-binding effectors, and riboswitches [12, 15, 16]. Through binding to effector molecules, c-di-GMP regulates diverse cellular processes, including motility, adherence, biofilm formation, virulence, development and cell cycle progression [11]. The mat formation involves the production of c-di-GMP through physiological signalling or mutation by one of three enzymes: WspR (DGC), AwsR (DGC), or MwsR (combined DGC/PDE) [12]. Increased c-di-GMP levels lead to more cellulose production. The production of cellulose contributes to the formation of air-liquid interface mats.

1.3.2 tRNA deletion experiment 1: the serCGA *locus*

Some codons can be translated by more than one tRNA type. For example, 5'-UCG-3' codons can be decoded by tRNA-Ser(CGA) (through Watson-Crick base pairing) or tRNA-Ser(UGA) (through G-U wobble base pairing) (see section 1.1 and Fig 1.3). In the first of the studies mentioned above, the effect of deleting the single-copy tRNA gene *serCGA* was investigated [14]. The deletion of *serCGA* led to the elimination of tRNA-Ser(CGA) from the mature tRNA pool, placing a higher demand on tRNA-Ser(UGA) when translating UCG codons. This increased demand resulted in lower translation efficiency and a reduced maximum growth rate. To compensate for the growth defect, the *serCGA* deletion mutant (Δ*serCGA*) was found to evolve by duplicating the tRNA gene *serTGA* (encoding tRNA-Ser(UGA)), leading to an increase in tRNA-Ser(UGA) in the mature tRNA pool. Duplication of *serTGA* was achieved by duplicating large sections of the bacterial chromosome; the duplication fragments range between 45 kb and 780 kb in size - more than 8,000 times the length of the 91 bp compensatory tRNA gene (Fig 1.3). These duplication fragments also carry many other genes (>700 in some cases) and thus change the length and symmetry of the chromosome and the dosage of many genes.

1.3.3 tRNA deletion experiment 2: the gluTTC-glyGCC-gluTTC-glyGCC *(EGEG) locus*

The second experiment mentioned above involved reducing the levels of essential tRNAs (rather than eliminating a non-essential tRNA type). This study deleted a locus containing four tRNA genes (*gluTTC-glyGCC-gluTTC-glyGCC*), giving strain *⊿EGEG* [38]. The deletion strain still contained two further copies of *gluTTC* and one further copy of *glyGCC*, elsewhere in the bacterial chromosome. The reduction in tRNA gene copy numbers was found to cause a growth defect, which was again compensated by the duplication of large segments of a second region of the bacterial chromosome (this time up to 1 Mb, and containing >900 genes), each containing the remaining *glyGCC* gene copy [38].

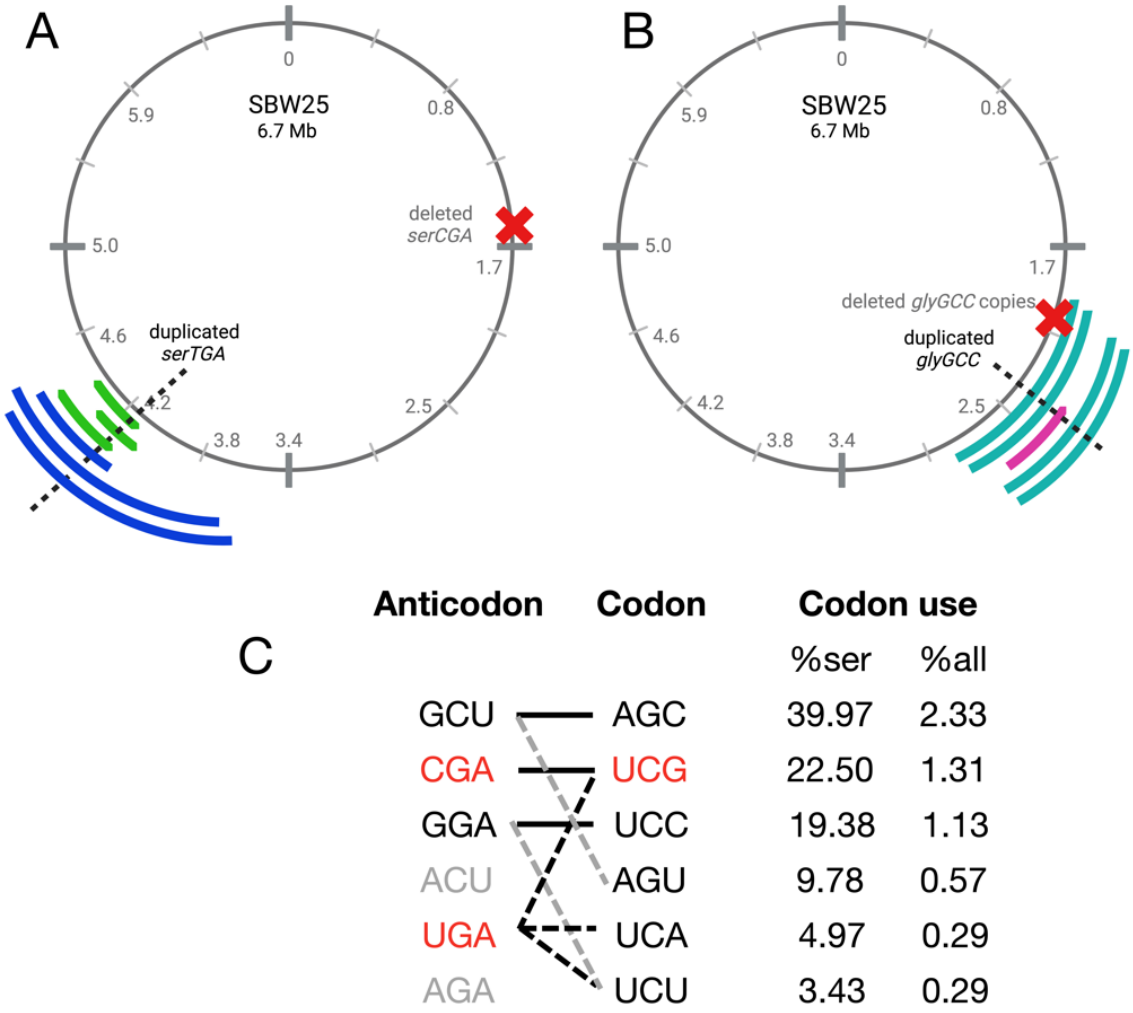


Fig 1.3. Large-scale duplications were found to compensate for the growth defect causing alterations in the tRNA gene set of *P. fluorescens* SBW25. The red crosses in **(A)** and **(B)** indicate the location of the tRNA genes targeted for deletion. **(A)** Duplications from seven Δ*serCGA* compensated strains were found to have different duplication fragments. Different colours denote different tRNA gene sets. The size of duplication fragments ranges between 45 and 780 kb. **(B)** Five Δ*EGEG* compensated strains containing duplications in two different tRNA gene sets (denoted by colors). **(C)** Theoretically possible seryl-tRNA anticodons to translate between seryl-tRNAs and serine codons. Anticodon CGA is deleted in the Δ*serCGA* strain, leading to higher usage of UGA to decode codon UCG. (red = present in SBW25, grey = absent, * = theoretically capable of translating codon UCG; solid black lines = Watson Crick pairing; black dotted lines = wobble pairing through post-transcriptional modification; grey dotted line = G:U wobble pairing). [14] (**(A)** and **(B)** were modified from Gallie, unpublished data; **(C)** is reproduced from [14]). Figures were drawn using Biorender.com and OmniGraffle.

Similar large-scale, tandem duplication events have been observed in many other bacteria (see sections 1.1 and 1.2) (reviewed in [39]). These duplications typically arise and are subsequently lost – without a trace – at extremely high rates [39, 40]. For example, despite the growth advantage afforded by our tRNA gene-containing duplication fragments [14, 38], they can be lost from overnight culture (Jenna Gallie, unpublished data). Such instability raises questions about the long-term evolutionary fate of the duplication segments and their new tRNA genes. That is:

If the serial transfer evolution experiments were to be continued, what would become of the duplication fragments and the tRNA genes they contain?

1.4 Large-scale, tandem duplication events are pervasive in microbial genomes

Changes in gene copy number via tandem duplication events are among the most frequent types of mutation [39]. Altering gene copy numbers by duplication and amplification can lead to the evolution of new functions [41, 42]. Duplications can be generated through homologous recombination (RecA-dependent) or illegitimate recombination (RecA-independent) [41]. Duplications are generally transient in nature due to the instability and associated fitness costs [43]. The importance of duplication in evolution is often underestimated [43]. Andersson and Hughes reviewed in 2009 and generalised gene duplication could occur in a fast response time within 10 generations [43]. While point mutations are slower and could take thousands of generations to occur [43]. Thus, up-regulation of a specific target gene through a promoter mutation or by inactivation of a regulatory repressor gene occurs at a frequency significantly lower, approximately ranging from three to six orders of magnitude, compared to a gene duplication event increasing the gene copy number [43].

Duplications generate an emergent join point between DNA regions that are normally not adjacent. Hence, this duplication junction is present in the duplication mutant but not the wild-type (Fig 1.4). One of the methods used to detect tandem duplications is high throughput Next-Generation Sequencing (NGS), where duplication events present as regions with unusual coverage, and reads spanning the new duplication junction can be identified. In cases where the approximate location of the duplication fragment is known, the duplication can be detected by designing a set of PCR primers spanning the unique duplication junction (C-B junction in Fig

1.4). The use of primers that gives no PCR product in strains would indicate that the specific duplication fragment is lacking (*e.g.*, wild-type has no C-B junction) [39, 41].

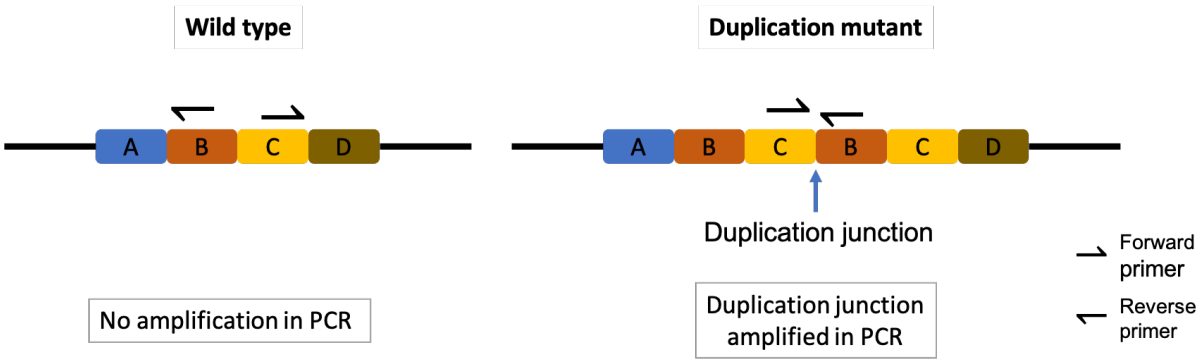


Fig 1.4. Depiction of the emergent duplication junction in strains carrying a tandem duplication fragment. Left is the wild-type sequence, and right is a duplication mutant. Each box represents a gene. The duplication mutant has a unique duplication junction (C-B), which can be amplified by PCR. Figures were drawn using PowerPoint.

1.5 Large-scale duplications are an unstable, short-term evolutionary solution

Unstable duplication fragments have a number of possible evolutionary fates [44]. The factors that could affect the fate of duplication are discussed and studied in the literature [40, 44, 45]. There are some main factors that would affect the fate of large duplications, such as the fitness costs of a duplication, the stability of a duplication, and how duplication can be maintained [44]. Possible fates of a duplication include loss of the extra copy resulting from (1) counterselection due to a fitness cost, or (2) instability of duplication, or stabilisation and maintenance of the duplication resulting from (1) accumulation of random mutations that inactivate one of two copies of essential genes on either fragment; (2) acquisition of new, beneficial functions by each fragment/gene ("neofunctionalization") [44].

To hypothesise the possible fate of duplicate genes, the innovation-amplification-divergence (IAD) model was proposed in 2007 [46]. The model suggests preexisting genes in bacteria may act as "innovation". A bacterial gene with a weak trace of a new function may be amplified under selection ("amplification"). Selection could result from environmental changes (such as the appearance of a new nutrient or a toxic compound) or internal changes (such as the fixation of a

deleterious mutation somewhere else in the genome) [44]. Followed by increasing copy numbers, the target size in which beneficial mutation(s) could occur increases too, which facilitates the accumulation of rare beneficial mutations ("divergence") [42]. This model is further supported in several empirical experiments using *Salmonella enterica* [42], yeast *Saccharomyces cerevisiae* in the evolution of the MALS family of a-glucosidases [47], and even fish species *Lycodichthys dearborni* in the evolution of antifreeze proteins evolved [44, 48].

In addition to the IAD model and the possible fates suggested by Andersson et al. in 2015 [44], there is also the Selection during Niche Adaptation (SNAP) hypothesis suggesting that parts of duplications may stabilise, leading to large-scale chromosomal rearrangements. The SNAP hypothesis suggests that, if the duplication fragment contains multiple essential (or highly beneficial) genes, deleterious mutations could occur in both fragments, leading to the retention and stabilisation of – parts of – both copies. After fixation, the mutated, non-functional parts of each copy may be lost, leading to rearrangements in gene order (Fig 1.5) [49]. The hypothesis was further observed in the laboratory evolutionary experiment *S. enterica* serovar Typhimurium [50]. The experimental results show after 2000 generations, only 31% of duplicated genes remained from 1.66 Mb large duplication in the chromosome. The gene order within the chromosome has been significantly altered. These results suggest that the rearrangement of chromosomes could be one fate of a large-scale duplication.

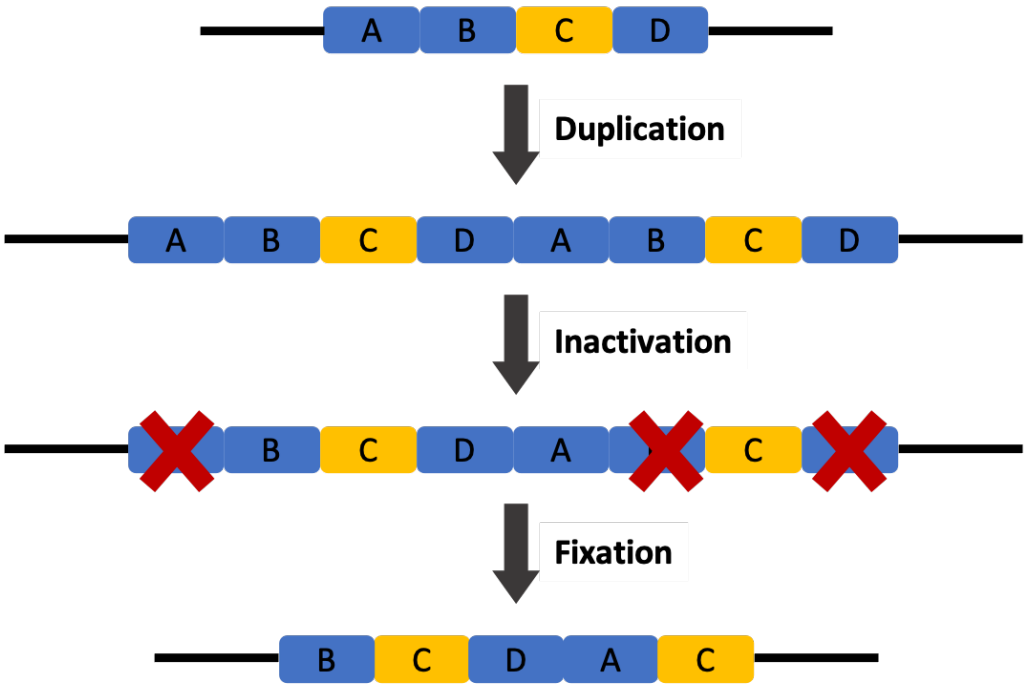


Fig 1.5. Schematic diagram demonstrating the Selection during Niche Adaptation (SNAP) hypothesis. ABCD genes were first duplicated with function beneficial gene C (yellow). The extra copies of the non-beneficial gene ABD (blue) were inactivated. Further fixation led to chromosomal rearrangement. The figure was drawn using PowerPoint.

1.6 Use of mathematical modelling to shed light on biological issues

Evolved strains are typically obtained from laboratory evolution experiments under particular environmental conditions [51]. Investigating how strains evolve under different conditions would require repeating the evolution experiment with new conditions. These variations in setup could, for example, include different bottleneck sizes or using strains with different growth rates or mutation rates. If the investigation is extended to a larger scale, the required time and resources may increase dramatically. To help with this issue, mathematical modelling can simulate evolution under various conditions. Modelling methods have been applied in various evolutionary simulations, such as simulating the formation and loss of mutations [52]. Many ready-to-use software are available online such as PopG, which was first released in 1993 and simulates multiple populations and the effects of natural selection, migration, and genetic drift (Website: https://evolution.gs.washington.edu/popg/). There are also many newly established tools to simulate mutation [53-56]. However, it is important to understand the mathematical methods behind each software and make sure the tools fit the investigation. In this study, the simulation can be focused on two ways: 1) how compensated mutants will take over in the population differently with different growth rates, mutation rates, and loss rates; 2) if there is more than one type of mutation, how these different types of mutation may change in abundance over time. Further, ask how the evolutionary dynamics may change if these mutations have different growth rates and formation/loss rates in mutation. The simulation will further answer the question of what is the critical factor in the evolutionary dynamics of the population.

1.7 Aims and objectives

Bacterial tRNA gene sets vary across the tree of life. Not only do organisms in different domains contain different tRNA types, but organisms in the same genus or between strains could have different tRNA gene types. The differences in tRNA gene sets allow bacteria to adapt to different niches and fit the translation efficiency.

This study has a few broad aims:

A) Investigating how bacteria could evolve to improve their translation efficiency.

Do they change their tRNA gene sets, or if they have other evolutionary paths to fulfil the need to improve their translation efficiency?

B) What mechanisms would allow the bacteria to change the translation efficiency or their tRNA gene sets? How could bacteria gain this mutation?

C) Would the evolving bacterial population contain multiple mutations or be dominated by the fittest mutation?

One of the direct approaches is constructing mutants with sub-optimal tRNA gene sets. Then, allow the bacteria to evolve. Previous studies from Ayan et al. 2020 [14], a PhD study [38] and Khomarbaghi, Ngan et al. 2024 [57], constructed two mutants from *P. fluorescens* SBW25. Both studies suggested evolving bacteria gained large-scale duplication mutations. The duplications were large and contained the tRNA gene to change the tRNA gene sets. However, these mutations are also unstable. The mutation can be lost from an overnight culture. The main focus of this PhD project is to investigate the fate of the large-scale duplications and the new tRNA gene copies that they contain. This will allow a better understanding of whether bacteria will finally gain a stable mutation to change their tRNA gene sets or remain unstable with a high reversion rate when they change their tRNA pool. Then, we will have a better understanding of how changes in tRNA gene sets in bacteria may play a role in the longer evolutionary term.

The above aims were achieved using the following three objectives:

(1) Quantification of the evolutionary dynamics of compensatory mutations in independently evolving tRNA deletion mutant populations. This will be done in the following three stages:

(a) The serial transfer evolution experiments started with the two constructed mutants (Δ*serCGA* and Δ*EGEG*; described in section 1.2 above) were restarted and extended to 100 transfers (~700 generations). In addition to daily transfers, this involved daily storage of samples, regular plating of populations on agar (2-3 days intervals), and regular PCR-based contamination checks.

(b) When the transfers were completed, subsequent definition and characterisation of emergent morphotype classes were undertaken.

(c) Isolates of interest were selected from each line and subjected to whole genome re-sequencing.

(2) Characterization of the biological effects of selected mutations identified in objective (1). Putative mutations were confirmed, and those of interest were tracked through lines (using PCR-based techniques). The confirmed mutations were checked through the corresponding assays (mat formation, biofilm assay, Congo Red assay, motility test and stability test).

(3) Investigation of the dynamics of these mutations (from objective 2) in the population. The factors influencing the fate of the unstable duplication were simulated using mathematical modelling.

(a) Identified mutations were checked in the population sequences. The dynamics of these mutations were estimated bioinformatically through the sequence analysis.

(b) Using mathematical models to simulate the evolutionary dynamics of two types of mutants in a transferring evolutionary experiment. The model allows changing parameters of the mutant, such as mutation rate, reversion rate and replication rate. Different "what if" scenarios were simulated using this model. For example, if the mutant has a higher mutation or replication rate.

At the end of the project, the evolutionary fate of the unstable large duplication was determined. Alternative mutants related or unrelated to the compensation of tRNA genes were predicted and confirmed. The population sequencing revealed the dynamics of these mutations. Mathematical modelling also successfully simulated different factors needed for a mutation to dominate the population.

Chapter II – Methods and Materials

2.1 Strains and growth conditions

The model bacterium *Pseudomonas fluorescens* SBW25 was used in this study with two tRNA gene deletion mutants engineered from this model bacterium. The first engineered strain with one tRNA gene (*serCGA*) deleted (Δ*serCGA*) [14]. Another engineered strain has four tRNA genes (*gluTTC-glyGCC-gluTTC-glyGCC*) deleted (Δ*EGEG*) [57]. Both strains were constructed and tested in previous studies. Previous studies observed that the engineered founder strains both had growth defects and rose as small colonies on King's medium B (KB) agar. Unless specified, in this study, strains were grown at 28°C with 200 rpm shaking condition in liquid KB medium for 24 hours. This is an optimal condition for *P. fluorescens* SBW25 to grow. Most KB plates were grown at room temperature with static conditions for 36-48 hours. The plates were mainly not grown at 28°C due to the time windows to give more significant observations between large (wild-type or compensated mutants) and small (non-compensated engineered strains) colonies. Colonies on plates grown at 28°C for 24 hours are too small to obverse. They were all too large for 48 hours of incubation at 28°C, and it was harder to distinguish the size differences.

2.2 Evolution experiment

After purification on an agar plate, four single colonies from the three founder strains (wild-type, Δ*serCGA* and Δ*EGEG*) were picked by sterile pipette tips and transferred to 9.9 mL sterile KB medium in a 50 mL falcon tube (SARSTEDT, Germany) to restart and extend the evolution experiment. Every 24 hours, 100 µL of grown cultures were transferred into fresh sterile 9.9 mL KB medium. Throughout the experiment, the population were regularly plated to assess the proportion of colonies with different phenotypes (especially the small types). Every day, glycerol stocks from each line were stored and kept at -80°C freezer. There were several pauses in this evolutionary experiment. 80 µL of 50% glycerol saline was mixed with 100 µL population samples and stored at -80°C freezer until the restart of the experiment. The 180 µL mixture of glycerol saline and bacterial samples was added to a 9.9 mL fresh and sterile KB medium to restart the experiment. Non-filter pipette tips were used in the first half of the experiment. It is suspected that during the restarting, when a 200 µL pipette tip was used to load 180 µL samples,

the risk of contamination increased. After the third restart of the experiment, contamination across lines was observed. After that, filter tips were used during the transfer, and no cross-population contamination was observed.

2.3 Population growth curve measuring

The engineered founder strains have growth defect and grow slower in liquid and solid media. Population growth curves were performed using a 96-well plate to assess the general growth compensation. A big chunk from each glycerol stock was taken and re-suspended in 200 µL sterile Ringers. 2 µL from each were transferred into 198 µL KB medium in each 96-well plate. The plate was wrapped by a Breathe-Easy sealing membrane (Sigma Aldrich, USA) and incubated at 28°C shaking condition. This process serves as a pre-culture of the samples. After 24 hours, the plate was taken out and Breathe-Easy sealing membrane was removed carefully to avoid contamination between wells due to shaky or shocks. 2 µL from each well was transferred into a new plate with 198 µL fresh KB medium in each well. The plate was wrapped with a new Breathe-Easy sealing membrane and placed into a plate reader (Bio Tek Epoch 2, USA). The plate reader was set up to read OD600 every 5 mins with 5s orbital shaking before each read. The readings were collected after at least 24 hours. The plate was checked for no contamination before discarding.

2.4 Contamination checking

Contamination checking is crucial in an evolutionary experiment. Any contaminant would easily ruin the experiment with a long transferring time. Regular checking is of paramount importance. In this study, some contamination controls were done. Population samples were regularly diluted and plated during the evolutionary experiment to observe for any potential external contamination. One of the advantages of using *P. fluorescens* as a model bacterium is that they turn fluorescence on an agar plate. It is relatively easy to distinguish from the appearance of the colonies. Mutants of *P. fluorescens* may have different morphotypes. Those suspicious colonies were isolated and further checked by PCR. No such contamination was observed in the experiment. Population PCR was also regularly performed to check for contamination between population. Primers were designed to target deleted regions in the engineered strains. This setup could distinguish between wild-type SBW25, Δ*serCGA* and Δ*EGEG*.

2.5 Sequencing and analysis of selected isolates

72 isolates were selected based on targeted lines and their diverse morphotypes. Bacteria were first purified on solid KB agar and grown overnight on fresh KB liquid medium. Genomic DNA (gDNA) was extracted using Blood and Tissue Kits (Qiagen, Germany). 2 mL of bacterial culture were collected and centrifuged at 10,000g for 20 mins (Micro Star 17, VWR, Germany). After removing the supernatant, the cells were lysis using proteinase K and the lysis buffer provided by the kit. The detailed protocol is attached in the appendix. Before sequencing, the qualities of the extracted and purified DNA samples were examined using a Nanodrop 2000 (Thermo Scientific, USA) to determine the optical density values and concentrations. Degradation of the DNA samples was checked by gel electrophoresis. All gDNA samples were sequenced in-house at the Max Planck Institute for Evolutionary Biology (Plön, Germany) using Illumina NextSeq platform for whole genome sequencing (150 bp, pair-ended).

After sequencing, the raw data were filtered to eliminate adapter pollution and low-quality data and ensure clean reads. Aligning the raw sequencing reads for each re-sequenced genome to the *P. fluorescens* SBW25 reference genome sequence NC_012660 [34] using *breseq* (v 0.37.1) [58, 59] gave an average of 80-fold coverage per genomic base. *breseq* is a pipeline to detect mutations from short-read sequences to a reference genome. The suggested duplication mutations were confirmed bioinformatically through junction identification and PCR techniques.

2.6 Isolates characterising and phenotype checking

Some mutations were often detected in the isolates' sequences. These included duplication mutations (in which the genome of the mutants was partially doubled), *fleQ* and *fli* mutations (where the mutants lose the swimming abilities), and *wss* and *wsp* mutations (forming wrinkly spreaders). Some of these mutations can be identified or characterised.

2.6.1 Duplication junction checking

The duplication junctions were checked both bioinformatically and by PCR. Firstly, the raw reads and SBW25 sequence (wild-type) were aligned to the emerged junction sequences using Geneious Prime (v2023.2.1). Most of the junctions are unique in the mutants. Therefore, the SBW25 sequence should give no hit in the alignment, and the mutants' sequence should provide

a convincing coverage (at least 20-30-fold). Next, the primers for junction PCR were designed. The pair of primers targeted the emerged junctions. Hence, they amplified the duplication mutants using PCR, and there should be no amplification in other mutants. In this study, selected mutants were tested and confirmed by both methods.

2.6.2 Stability test on selected mutants

Some of the identified mutations were known to be unstable. Hence, these mutations were lost quickly, even after an overnight culturing. To test the stability of these mutations, selected isolates were retrieved from the glycerol stock on KB solid media. A large single colony from each mutant were picked to grow in 4 mL fresh KB medium in a 13 mL culturing tube overnight at 28°C shaking condition. After 24 hours, the grown cultures were diluted and plated on KB agar and incubated at room temperature for around 40 to 48 hours until both small and large colonies could be identified visually. The large and small colonies were counted to measure the loss rate of the mutation.

2.6.3 Congo red assay

Congo Red assessments were performed to examine the ability of each morphotype to produce cellulose. KB plates with 0.5 g/L Congo Red powder (Sigma-Aldrich, USA) were prepared. Plates were hand-poured and dried under a dark box. Each plate was spotted 4x strains. 5 µL of fully grown cultures were spotted on the relevant ¼ of the plate. After spotting the four spots on the plates, they were dried for 20 mins. The plates were incubated at 28°C static conditions. Every 24 hours, the plates were taken for assessment

2.6.4 Bacteria motility assay

Standard mobility assays were modified from the standard methods [60]. 0.3 % semi-solid LB agar containing 3 g/L NaCl were prepared. After autoclaving, the media were left at 50°C for two hours. 3 mL of the agar was poured into each well of a 12-well plate by a 10 mL pipette. The plates were put in laminar flow for 30 mins. After that, the plates were dried at room temperature (23-25°C) overnight. After 20-24 hours, a sterile needle modified from an inoculation loop was first dipped into the grown bacterial culture and then dipped carefully into

the centre of the semi-solid agar well. After dipping all the strains into the semi-solid agar well, the 12-well plates were incubated at 28°C for 24 hours to examine.

2.6.4 Biofilm assay

Biofilm formation abilities of the 72 sequenced isolates were tested [61]. Firstly, the strains were grown in KB fresh medium for 24 hours. On the next day, 1.5 μL of the grown culture was transferred to a 96-well plate, with each well containing 148.5 μL fresh KB medium (1:100 dilution). The plates were then transferred into a 28°C static incubator for 30 hours. The next day, the whole plate was washed with deionised water, and all the liquids were shaken out. The washing process was repeated for the second time. After that, 170 μL of 0.1% crystal violet was added into each well and incubated at room temperature for 10-15 mins. The plate was then submerged in a tub of deionised water. All liquids were shaken out vigorously. This water process was repeated 3 times. The plate was then placed upside down and dried overnight. On the next day, 170 μL of 30% acetic acid was added to each well to solubilise the crystal violet. After incubation at room temperature for 10-15 mins, all liquids were transferred into a new 96-well plate. The plate was read in a microplate reader (Bio Tek Epoch 2, USA) at OD580. The readings were used to indicate the relative biofilm-forming abilities of each strain.

2.7 Population sequencing and bioinformatics analysis

Two time points during the rise of compensated mutants were selected for early-stage population sequencing (days 5 and 10 for line 11; days 10 and 16 for lines 4, 12 and 20). Four time points during the evolutionary experiment were selected for further population sequencing (days 25, 50, 75 and 100). To get these samples, days 4 and 9 (or 9 and 15) targeting the early-stage and days 24, 49, 74 and 99 targeting the whole experiment were retrieved from the glycerol stock. After fully thawing samples on ice, the samples were inverted several times for mixing. 100 μL of bacterial cultures were transferred into 9.9 fresh KB medium to grow under the same conditions as the evolutionary experiment. After 24 hours, the samples were collected and processed for gDNA extraction using the same method as the isolates' gDNA extraction and examination mentioned in section 2.5. All gDNA samples were sent to Novogene (UK) for sequencing. During the transfer, the samples were stored with ice packs in an isolation box to prevent thawing and degradation. Illumina Novaseq (150 bp, pair-ended) was used to give at least 1000-fold coverage for each sample. The raw sequences were first filtered to eliminate adapter

pollution and low-quality data and ensure clean reads using *fastp* (with settings: --detect_adapter_for_pe, --cut_right, minimum read length 50 bp). The filtered sequences were then analyzed using breseq for potential mutation detection. Junctions of interests were aligned using Geneious Prime (v2023.2.1) using the raw reads. The ratio of the number of matched reads over the total coverage of the whole samples was used to estimate the dynamics of the duplication mutation. The overview of coverage was exported from Geneious Prime and generated using RStudio (v2023.06.0+421).

2.8 YAMAT-seq and analysis

YAMAT-seq was used to sequence the mature tRNA pool. The method was developed by the Kirino lab [62], and has been adapted in previous studies for prokaryotes [14, 57, 63]. After measuring the growth curve of each strain, the mid-exponential phase was determined. A single colony was picked and grown in KB medium. Samples were collected one hour before the mid-exponential phase, at the mid-exponential phase and one hour after the exponential phase. These samples were diluted and plated on KB agar plates. The colony forming unit (CFU) of samples at these time points was estimated to ensure the samples were collected and sequenced at the mid-exponential phase. For the RNA extraction, the cells collected at the mid-exponential phase immediately performed the extraction using a TRIzol Max Bacterial RNA isolation kit. After removing amino acid from the mature tRNA using deacylation treatment with 20 mM Tris-HCl, Y-shaped, DNA/RNA hybrid adapters were annealed to the tRNAs. The tRNAs were converted to cDNA using SuperScript™ III reverse transcriptase. cDNA products were amplified by PCR. The samples were mixed in equimolar amounts after quantification by a Bioanalyzer and fluorescent Nanodrop. The mixed sample was purified on a 5% polyacrylamide gel to obtain the tRNAs with adapters and barcodes. The extracted gel was soaked in ultrapure water overnight. The sample was then centrifuged through a microcentrifuge tube (cellulose acetate membrane with pore size 0.45; Merck CLS-8162). The sample was then examined and sequenced at the in-house sequencing service at the Max Planck Institute for Evolutionary Biology (Ploen, Germany) using NextSeq Mid Output Kit (150 bp, single-end reads).

The raw sequences were analysed using Geneious Prime (v2023.2.1). The sequences were aligned to the 42 tRNA reference sequences from SBW25 (with setting: Maximum gaps per read: 10%, Maximum gap size: 2, maximum mismatches per read: 10% and maximum ambiguity: 5). The

unused reads were de novo assembled to check for any tRNA reads. The relative abundance of each tRNA type was calculated and the three independent replicates were used to measure the standard deviation. A heatmap showing the differences in each tRNA type was created using the R package DESeq2 [64].

2.9 Primers and PCR conditions

Sets of PCR primers were designed in this study, including primers to detect the founder strains (these were also used in contamination tests) and primers to confirm the duplication junctions. The single-nucleotide polymorphism (SNP) promoter mutant of *serTGA* was detected using temperature switch PCR, where two sets of primers with different melting temperatures were used. The locus specific (LS_f/LS_r) primers have higher melting temperatures and the nested locus specific (NLS_f/NLS_r) primers have lower melting temperatures. The outer LS primers amplify ~800-1000 bp genome DNA. The inner NLS primers target a 100-200 bp region, where the NLS_f primer contains the mutant SNP allele at its 3' terminus. Thus, the NLS_f primer only binds to the mutant allele.

All PCR reactions were run using a Biometra machine (analytikjena, Germany) under the following conditions (except when otherwise specified):

For each reaction:

5x green buffer	5.0
5xCES [65]	5.0
$MgCl_2$	3.0
dH_2O	2.0
dNTPs	1.0
Primer_f (5 pmol/µl)	4.0
Primer_r (5 pmol/µl)	4.0
GoTaq	0.5
Template	0.5
	25 µl

PCRs to be run on the following program:

1. 94-95˚C, 1-2 min
2. 94-95˚C, 15-30 sec
3. 57-58˚C, 15-30 sec
4. 72˚C, 1-3 min (for each 1 kb of the product, it sets for 1 min. Hence, 1.5 kb would be 1.5 mins)
5. cycle 2-4 29 more times
6. 72˚C for 5 min
7. 4˚C for 5 min

Table 2.1. PCR primers used in this study

Name	Sequence (5'→3')	°C	Target	Usage	Reference
Ser_CGA_f	GAAGATCTCAAGCGCATCGAACAACTTGATC	57.0	Δ*serCGA*	Checking of strains	[14]
Ser_CGA_r	GAAGATCTGATGTGGCTCTGACTCATTAGGAAG	56.9		Checking of strains	[14]
delEGEG_f	GTGACCTGACCATGTAGGGCTTC	59.4	Δ*EGEG*	Checking of strains	[57]
delEGEG_r	GCAAGGAAGTCCAACTGACCGTTAC	59.7		Checking of strains	[57]
junct_f3	GATGTCGGTGGAAGGCTTCAAG	58.1	Δ*serCGA* junct 183 kb	Confirmation of the duplication junction	[14]
junct_r3	CACCATCAACTTTACTGTGCTCAAGG	58.0			[14]
junct_f4	CTAACGCTACTTGTGCCAAAGCTG	58.4	Δ*serCGA* junct 290 kb	Confirmation of the duplication junction	[14]
junct_r4	CACGATTCGACCGCTCCTTG	58.4			[14]
M4L_D7_SmL_f	GTGTGGATTGGCGGCGAATG	59.6	Δ*serCGA* junct 656 kb	Confirmation of the duplication junction	This study
M4L_D7_SmL_r	GATTCGATGAGCCAGAGCATGGAG	59.5			This study
ZK3_junct_f	GTGTATGACGATGACGAGCAGAACC	59.3	Δ*EGEG* junct 480 kb	Confirmation of the duplication junction	[57]
ZK3_junct_r	CAGACAGCGATGAGCGAAAGC	58.3			[57]
ZK5_junct_f	CAACACCAACTCATCGGGGACAG	60.0	Δ*EGEG* junct 1 Mb	Confirmation of the duplication junction	[57]
ZK5_junct_r	GTAGAGCACGCAAATGGCTGGAG	60.4			[57]
serCGA_probe_r	CACTCTGGCATCTCTCCAACG	57.7	Paired with Ser_CGA_f to target the Δ*serCGA* region	Contamination checking across lines	This study
delEGEG_probe_r	CACCCGAAGGCAATACCGCTTAA	59.6	Paired with delEGEG_f to target the Δ*EGEG* region	Contamination checking across lines	This study
LS_serTGA_prom_f	CGAGTCACGCAGCTTGTTGGTCAG	62.2	*serTGA* promoter mutant	Locus specific primers to amplify the outer region to detect the SNP	This study
LS_serTGAprom_r	GGTCGAGAAAGCGGCATACCCTGATG	63			
NLS_serTGAprom_f	cggatggcgagagc	52.9	*serTGA* promoter mutant	Nested locus specific primers within the region where LS primers amplified. It only amplifies the SNP	This study
NLS_serTGAprom_r	ggCACATCAGTAGGTAGC	52.5			
Sam27/28_du4025_f	CATCACCGGGATGCCATC	56.1	4025 bp duplication mutant	Confirmation of the duplication junction	This study
Sam27/28_du4025_r	CGTTTCGCGGGTGTTGAG	57	4025 bp duplication mutant	Confirmation of the duplication junction	This study
Samp33_dup830bp_f	ACCAACAAGCTGCGTGACTC	58.2	830 bp duplication mutant	Confirmation of the duplication junction	This study
Samp33_dup830bp_r	CAAGGGAAATTGGCAGAGTGG	56.4	830 bp duplication mutant	Confirmation of the duplication junction	This study
DupTn_seq_r1	CAGGCAGGATGGATCGATCAAG	57.6	Transposon inserted mutant	Confirming and detecting of inserted transposon after losing of the duplication	This study
DupTn_seq_f1	cttgatcgatccatcctgcctg	57.6	Transposon inserted mutant	Confirming and detecting of inserted transposon after losing of the duplication	This study
DupTn_seq_f2	gttgtcagagcgaatgacttcc	55.9	Transposon inserted mutant	Confirming and detecting of inserted transposon after losing of the duplication	This study
DupTn_seq_f3	GTACTGGGAGACGGCCAATC	57.7	Transposon inserted mutant	Confirming and detecting of inserted transposon after losing of the duplication	This study
DupTn_seq_f4	CAACCAGCTTGATTCGTTGGTTC	56.9	Transposon inserted mutant	Confirming and detecting of inserted transposon after losing of the duplication	This study

2.10 Statistical analysis and data visualising

The statistical analysis and data visualising were done using Biorender.com, OmniGraffle, PowerPoint, SigmaPlot 11, GraphPad Prism 10, OriginLab (v2023) and RStudio (v2023.06.0+421). Before running One-way ANOVA, the dataset was first tested for normality. Kruskal-Wallis One-way ANOVA was used to test the statistical difference in the stability test between different mutants. The data of the stability test was fitted using the function of "linear fitting" in OriginLab (v2024) to perform a correlation test. The correlation coefficient was used to determine the correlation between fragment size and stability.

Chapter III - Investigating tRNA Gene Copy Number Flexibility in *Pseudomonas fluorescens* SBW25 Through Large-Scale Duplication Mutations: A Continued Investigation

Part of the work presented in this Chapter contributed to the following publication: Khomarbaghi, Z., Ngan W.Y, et al., *Large-scale duplication events underpin population-level flexibility in tRNA gene copy number in* Pseudomonas fluorescens *SBW25*. Nucleic Acids Res, 2024.

In this chapter, I carried out population growth curve measuring, tRNA pool sequencing, characterising of mutants such as stability test, population sequencing and the data analysis and visualising of the data from the above experiment. Zahra Khomarbaghi constructed the mutant (Δ*EGEG* and rWT) used in this study and performed the daily serial transfer evolution experiment, sampled the isolates, isolate genome sequencing. Gökçe B Ayan constructed of mutant (Δ*glyGCC*), conducted tRNA pool sequencing, analysing of mutants. Jenna Gallie conducted some stability test.

3.1 A summary of the previous study

To investigate how tRNA gene sets evolve, a previous PhD study constructed an engineered strain (Δ*EGEG*) from *P. fluorescens* SBW25 [38]. Two copies of each of *gluTTC* and *glyGCC* were deleted in this strain, leaving two copies of *gluTTC* and only a single copy of *glyGCC*. The deleted locus is 415 bp long. After the tRNA gene deletion, Δ*EGEG* showed a growth defect. The result suggested that without an optimal tRNA gene set, the translation efficiency was altered, leading to the growth defect [38].

To investigate how bacteria can adapt to changing translational needs, the previous study included a 28-day serial transfer evolution experiment with five independent Δ*EGEG* lines (and five wild-type lines). One compensated mutant from each line was isolated from day 21 and sent for whole genome sequencing. Sequencing results found the compensatory mechanism in all the five mutants was within-genome duplication events, where four of the duplications were almost 1 Mb in length. The final strain carried a duplication of ~0.49 Mb [38].

3.2 First part of follow-up experiments

The finding of large-scale duplications as a compensatory mechanism is interesting. These mutations significantly changed the genome size and dosage of more than 1,000 genes. Also, theoretically, the more DNA that needs replicating, the higher the cost. In addition, our group also constructed a mutant with one *glyGCC* copy deleted (Δ*glyGCC*), leaving four copies of *gluTTC* and two copies of *glyGCC*. Δ*glyGCC* did not have growth defects. Finally, it was observed that these large-scale duplications are unstable. Thus, when a compensated mutant was picked to grow in a liquid medium overnight, part of the bacterial population had lost the mutation. When plating the overnight culture, some cells in this overnight culture would have a growth defect and the small colony phenotype.

These observations led to some further questions to be addressed. For example:

(1) How does the stability of the duplication vary among mutants with different fragment sizes?

(2) How does the mature tRNA pool change after gaining the extra tRNA copy?

(3) How did the population improve with the duplication mutation?

(4) How diverse are the duplication mutations (in terms of size)?

(5) Is the instability related to the high cost associated with replicating a larger part of their genome in these mutants?

3.3 Aims

This chapter aims to address the questions arising from the previous PhD study [38] with the following broad aims:

(1) Characterising the mutants identified. This included assessing their stability and how the tRNA pool of the bacteria changed.
(2) Assessing the evolving population. This included getting an overview of the evolving population to check their growth improvement and performing a population sequencing

on the final day of the experiment to assess the mutants in the population.

Aim (1) was achieved by performing a detailed stability test on the identified mutants. The changes in the mature tRNA pool of these strains were tested by YAMAT-seq [62], a sequencing technique previously adapted for bacterial tRNA pools [14].

For aim (2), population growth curve measurements were taken on days 7, 14, 21, and 28 from glycerol stock of all the evolving populations. Population sequencing was done at the end of the experiment (day 28) to identify other duplication mutants (and to check for contamination events).

These experiments were crucial in characterising the identified mutants, especially the changes in the mature tRNA pool through the change in its tRNA gene sets.

3.4 Improving growth in the population

Population growth curves were measured to understand the growth improvement in the population of each line. Each population sample was retrieved to grow in 96-well plates. The growth curve measurement was done using a plate reader. Wild-type lines showed no significant improvements in growth (Fig 3.1). Among the Δ*EGEG* mutant lines, M1-M3 showed improvement in growth from day 7 (Fig 3.2). There were no observed differences between day 14 to day 28. This implies that the compensatory mutants rose from day 7, and the lines consisted of compensatory mutants with similar fitness after that. Two of the lines (M4 and M5) showed no improvement in growth. Later, external contaminants were found in these two lines (from day 21) by dilution plating from the glycerol stocks. The external contamination raised general concerns about contamination between lines. In this experimental setup, cross-contamination cannot be distinguished by PCR or phenotype checking. Therefore, population whole genome re-sequencing (M1-M3 and W1-W5) was performed, which is discussed in section 3.7.

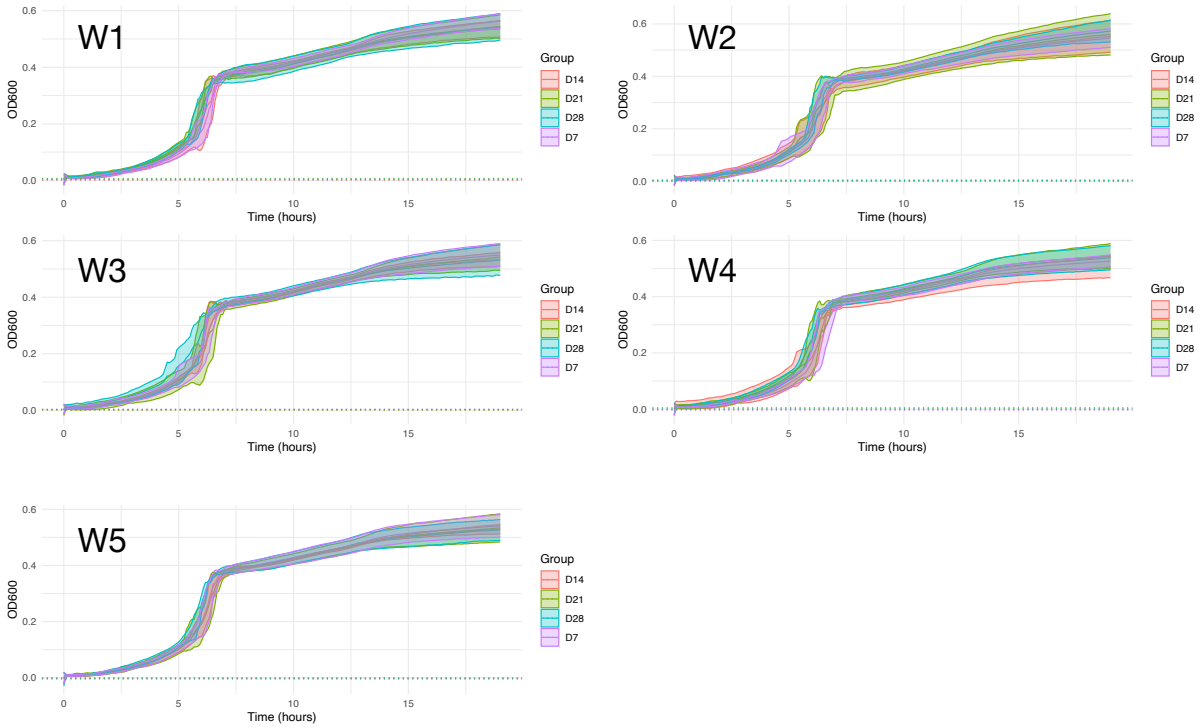


Fig 3.1. Population growth curve of the wild-type SBW25 lines W1-W5. Each figure is an independent replicate. Different colours denote the different sampling days in the evolutionary experiment (Purple from day 7, orange from day 14, green from day 21, and blue from day 28). No significant differences in the growth over the experiment indicate no observed growth improvements in the wild-type lines. Figures were drawn using RStudio.

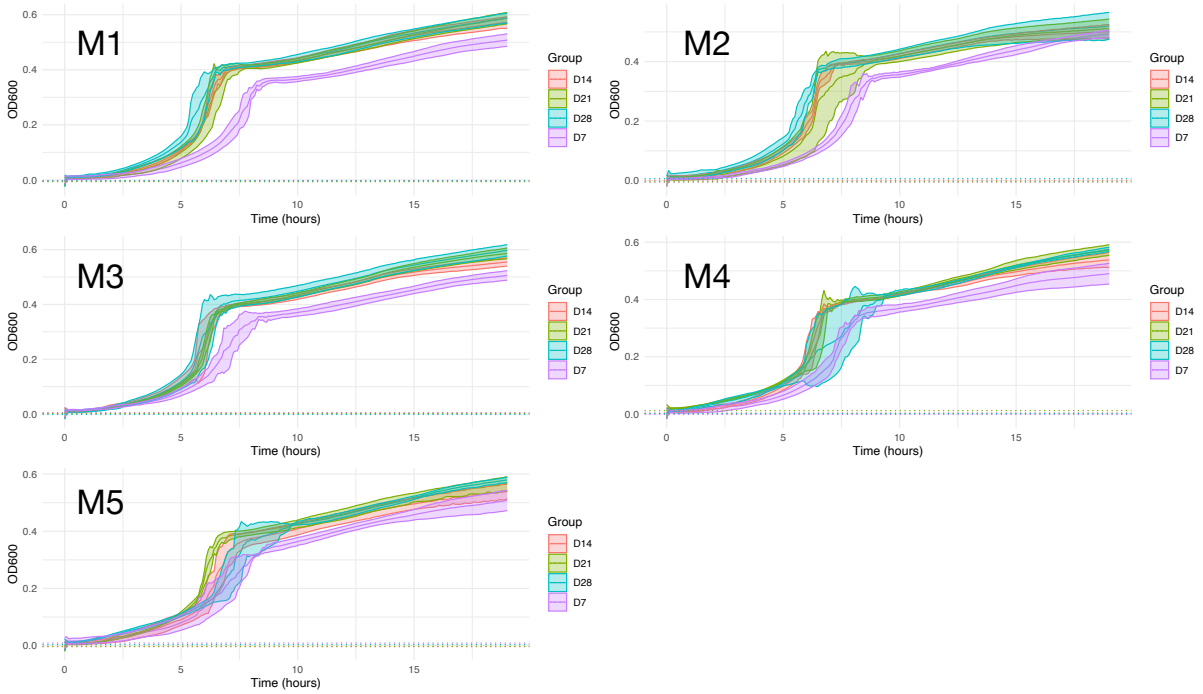


Fig 3.2. Population growth curve of the engineered founder (Δ*EGEG*) lines M1-M5. Figures' There were improvements in growth in M1-3 after day 7. M4 and M5 showed no improvements. Later, by

retrieving the samples, external contaminants were found in M4 and M5 lines. Different colours in each figure denote the different sampling days in the evolutionary experiment (Purple from day 7, orange from day 14, green from day 21, and blue from day 28). Figures were drawn using RStudio.

3.5 Stability test of the identified duplication mutants

Preliminary observations in our group found that some compensatory duplication mutants likely lost the mutation from overnight culturing, indicating the mutations were unstable. However, little was shown between the different mutants and how unstable they were. The extended investigation of the stability in this study used one compensated mutant from each line – four of which carried a 1 Mb duplication, the other a 0.49 Mb duplication (i.e., those in [38]).

The stability test measures the proportion of the cells that lost the mutation after an overnight culturing. The proportion of small colonies after overnight culturing would be directly related to the loss rate of the mutation. Results revealed that 1 Mb duplication fragments have mean loss rates from 12.9% to 21.8%. The 0.49 Mb fragment has a mean loss rate of 5.0%. Dunn's one-way ANOVA test showed a statistical difference between the 0.49 Mb and other 1 Mb fragments (M3-1 vs M1-1 with $p = 0.0212$ and other groups with $p < 0.0001$). This indicates that the duplication size highly affects the duplication mutation's stability. This is not surprising because the duplication loss occurs by exchanges between the two extensive identical copies of the duplicated region [39]. Thus, the longer repeated sequences generally have a higher loss rate. A non-parametric Kruskal-Wallis test using wild-type as a control group was performed on all tested strains (Fig 3.3). The single *glyGCC* deletion, wild-type engineering control and evolved wild-type strains do not significantly differ with actual wild-type. In general, the stability test indicates that the duplication mutations were unstable. The stability also differs among different fragment sizes with the shorter duplication fragment being more stable (Fig 3.3).

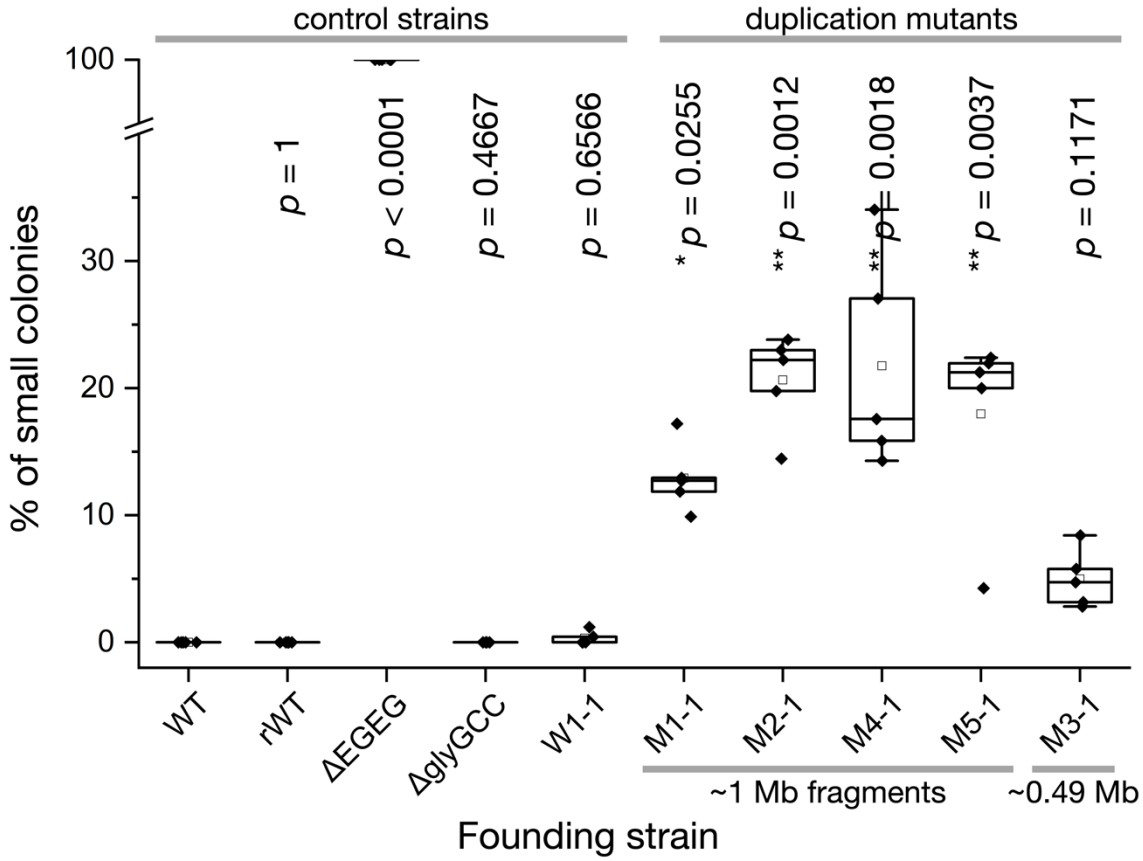


Fig 3.3. Stability test on the selected isolates. For the duplication strains, the percentage of small types after overnight culturing indicated the percentage of cells that lost the mutation. Each strain had five replicates, and at least 41 colony counts per replicate. The horizontal line in the boxplot showed the median and the small square box showed the mean value of each dataset. Statistical significance was tested by the non-parametric Kruskal-Wallis test using WT as a control group. Benjamini, Krieger and Yekutieli were used to correct multiple comparisons. Significance levels: *0.05 < P < 0.001, **0.01 < P < 0.001, ***P < 0.001. Figures were drawn using OriginLab and OmniGraffle.

3.6 Deep sequencing of mature tRNA pool in isolates

During DNA replication, bacteria have the flexibility to gain an extra copy of the tRNA gene and, in subsequent rounds of replication, to lose the duplication and revert to the original chromosome structure. This was also demonstrated in the stability test above. The next question is how the change in the tRNA gene *glyGCC* copy numbers affects the mature tRNA pool (and hence translation). The comparison of the duplications with different fragment sizes is of

particular interest. The mature tRNA pool (YAMAT-seq) deep sequencing tested three replicates of each selected strain. The wild-type (SBW25), evolved wild-type (W1-1; from day 21), deletion of four tRNA genes (*gluTTC-glyGCC-gluTTC-glyGCC*; Δ*EGEG*), deletion of single tRNA gene (Δ*glyGCC*) and three compensated Δ*EGEG* isolates (M1-1, M3-1 and M5-1). M1-1, M3-1 and M5-1 were isolated from different lines. Previous whole-genome sequencing showed M1-1 and M3-1 carried a tandem duplication fragment of ~1 Mb size, and M5-1 carried a fragment sized ~0.49 Mb, halving the size of the other two fragments. YAMAT-seq only sequences the tRNA in the mature tRNA pool [63, 66]. Hence, it shows the proportion of each tRNA in the strain but not the absolute concentration of it.

In the analysis, the engineering control and founder strains were first compared against the SBW25 (Fig 3.4A). Results showed that the engineering control does not have changes in the mature tRNA pool. tRNA-Gly^{GCC} has the highest proportion in both strains, which are 11.1% ± 1.76 % in SBW25 and 11.3% ± 1.00 % in the engineering control (reconstructed SBW25). The engineered strains with four tRNA gene deleted (Δ*EGEG*) has a lower proportion of tRNA-Gly^{GCC} in the mature tRNA pool with only 3.6% ± 0.34%. Deleting a single *glyGCC* gene copy does not greatly impact the overall tRNA-Gly^{GCC} proportion (Fig 3.4A).

Next, we compared the evolved strains against SBW25 (Fig 3.4B). The evolved wild-type (W1-1) showed no significant changes in its mature tRNA pool (Fig 3.4B). The three evolved mutants (M1-1, M3-1 and M5-1) showed decreased tRNA-Gly^{GCC} proportion compared with the wild-type even though all compensated for the growth defects. This is interesting because Δ*glyGCC*, M1, M3 and M5 contained two copies of tRNA gene *glyGCC*. The proportion of *glyGCC* in the mature tRNA of Δ*glyGCC* engineered strain is 6.9% ± 0.75%, higher than those in M1-1, M3-1 and M5-1 with 5.1% ± 0.34%, 6.3% ± 0.30% and 5.8% ± 0.42%, respectively. This could be because of the position of the gene in the genome. When the gene is closer to the origin of replication (*ori*), it will tend to be expressed at a high level (because replication starts from *ori*, leading to a higher gene dosage of genes closer to the *ori*). The duplication mutants (M1, M3 or M5) contain fragment sizes of ~0.49 Mb to ~1 Mb. This also leads to the lower level of the tRNA gene as it is farther from the origin of the genome. They need to replicate a larger part of the genome to replicate the tRNA genes.

A comparison between the compensated mutants (M1-1, M3-1 and M5-1) and their founder strain Δ*EGEG* demonstrated the level of *glyGCC* was increased after the evolutionary

experiment (Fig 3.4C). In the founder Δ*EGEG* strain, *glyGCC* contributed only 3.6% ± 0.34% to the mature tRNA pool, much lower than those in other strains in this study with range from (5.1% to 11.3%).

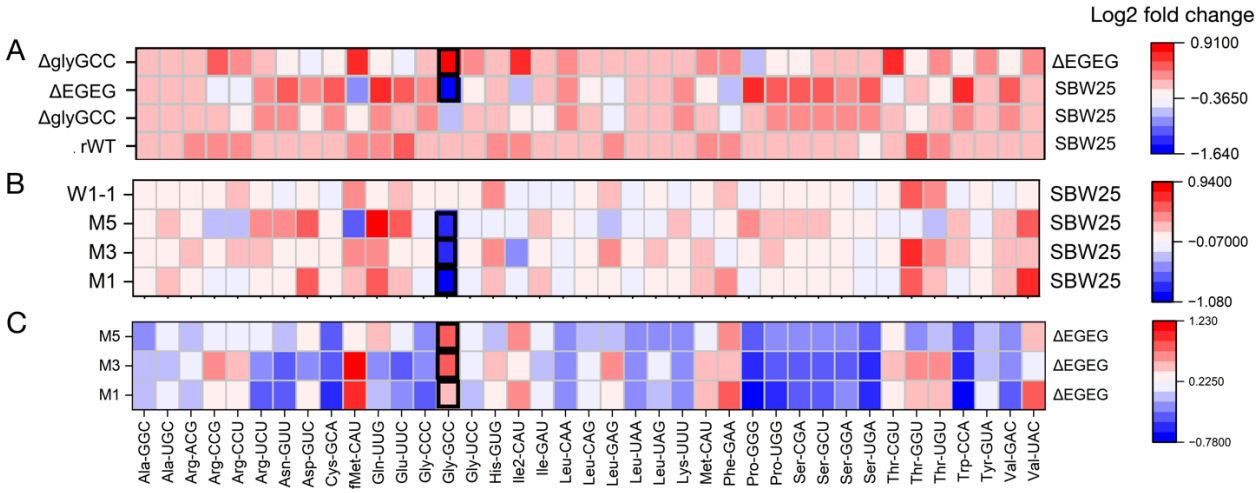


Fig 3.4. Heatmaps showing differences in the relative abundance of tRNAs. Differences were expressed as log2-fold change (strain on the left /strain on the right of the figure). (A) In comparing the founder strains, Δ*EGEG* with four tRNA gene deletions showed a decrease in *glyGCC* level compared with SBW25 wild-type and single gene deletion Δ*glyGCC*. (B) The evolved mutants were compared with the wild-type SBW25. W1-1 is isolated from the evolving wild-type line from day 21. M1-1, M3-1 and M5-1 were isolated from the evolving lines M1, M3 and M5 respectively. Even M1-1, M3-1 and M5-1 compensated the growth defect, the levels of tRNA *glyGCC* were still lower than the wild-type. (C) In comparing the evolved mutants with the Δ*EGEG* founder, M1-1, M3-1 and M5-1 have higher levels of *glyGCC*. Figures were drawn using OriginLab.

3.7 Population sequencing on day 28 samples

The contamination issue was discovered later in this study. By dilution plating the samples, lines M4 and M5 were confirmed to have contaminations. However, this could not ensure that the contamination did not happen in other lines. Contaminants may exist at a high level. Hence, it cannot be checked through dilution plating. There was also a higher concern about cross-contamination between lines. Therefore, population sequencings were performed in lines M1, M2, M3 and wild-type lines W1-W5 on day 28 samples. The population samples were retrieved and grown overnight. Next, the gDNA of the population was isolated and sent for Illumina short-read sequencing.

breseq was used to predict mutations computationally (Table 3.1). There are duplication hotspots in the genome, such as the highly repetitive region. *breseq* can predict point mutations or the actual position of a duplication mutation. We expect not to see the lines share too many point mutations or the exact size of duplication fragments. Also, the wild-type sequences should cover the whole reference genome evenly. While the Δ*EGEG* have the region of deletion covering four tRNA genes (*gluTTC-glyGCC-gluTTC-glyGCC*). If there is a lower coverage of the four tRNA genes in the wild-type sequences or the Δ*EGEG* sequences contain reads covering the deleted region, that could be explained by cross-contamination between lines. Also, if too many reads were unused in the mapping of sequences, this may indicate external contaminants in the sample.

Results showed that most reads were aligned to the references with an average of 9907599 out of 9917323 per sample, giving an average of 99.90% mapped reads and indicating a low chance of external contamination. *breseq* predicted nine duplication junctions from the evolving lines M1, M2 and M3 (Table 3.1). None were found in all wild-type lines.

Table 3.1. Duplication fragments were computationally predicted in day 28 populations from evolving mutant lines M1, M2 and M3.

Population	Evidence*	Start	End	Size (bp)	Junction locus (side 1; side 2)
M1 day 28	New junction	2375695	2379967	4273	*pflu2197; pgsA* (*pflu2191*)
	Marginal	2375280	3142292	767013	*pflu2872; pgsA* (*pflu2191*)
	Marginal	2267938	2712554	444617	*pflu2497;* IG (*pflu2087/8)*
M2 day 28	New junction	2375566	2379196	3631	*pflu2196; pgsA* (*pflu2191*)
	Marginal	2347702	2666045	318344	*pflu2447;* IG repeat (*pflu2165/7*)
	Marginal	2239941	2525738	285798	*soxB* (*pflu2319*); IG *rrn* (23S/5S)
M3 day 28	New junction	2296498	2786206	489709	*pflu2547;* IG repeat (*pflu2166/8*)
	New junction	2374401	2809371	434971	*pflu2553; uvrC* (*pflu2190*)
	New junction	2374589	2376758	2170	*ribD* (*pflu2193*); *uvrC* (*pflu2190*)

The predictions were made based on the reference genome (NCBI accession number: NC_012660.1). The genome positions also refer to the same reference genome. The position of the compensatory tRNA gene *glyGCC* is located from 2375805 to 2375880. *If there are

sufficient reads for breseq to predict the junction, it indicates "New junction". Junctions without sufficient reads will be "Marginal".

From the coverage of the sequenced reads, there were higher coverage in the tRNA gene *glyGCC* among population M1, M3 and M5 (Fig 3.5). No particular high coverage was observed in all wild-type lines.

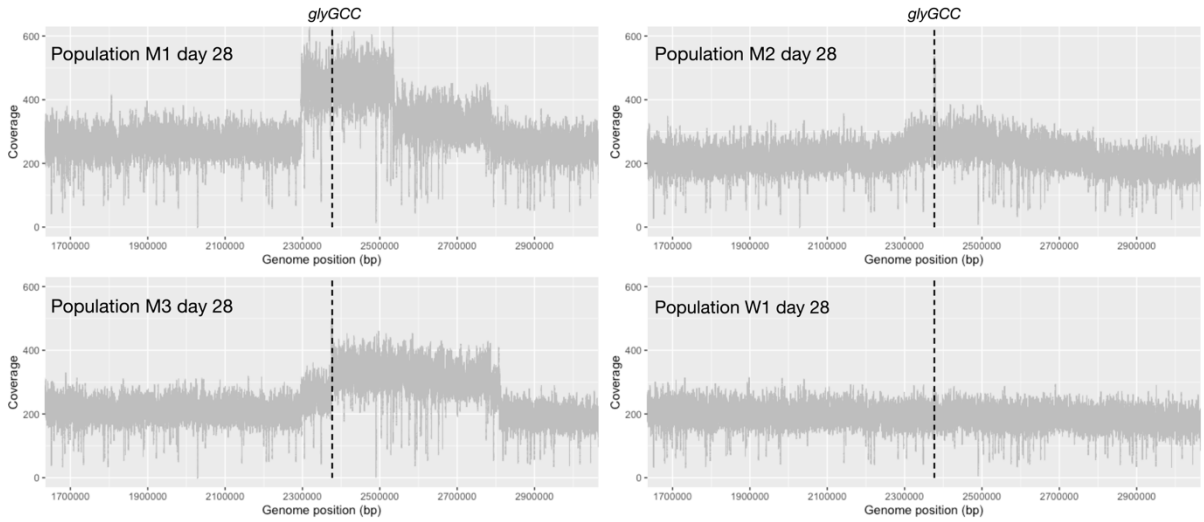


Fig 3.5. Coverage of population sequencing reads on evolving lines M1, M2, M3 and W1 from day 28. M1, M2 and M3 are the mutant lines and W1 is the evolving wild-type line. Figures covered 1.7 Mb to 3.0 Mb of the reference genome, where the compensatory gene *glyGCC* is located (dotted dash line). Figures were drawn using RStudio.

3.8 Conclusion

Previous work demonstrated a mechanism for a bacterial tRNA gene set to evolve: gaining extra tRNA gene copies through within-genome duplication events of 1 Mb or 0.49 Mb [38].

This follow-up study mainly focuses on characterising the identified mutants and assessing the population. This study first assessed the growth improvement through growth curve measurement. The population showed improvement in growth over the 28-day daily serial transfer evolutionary experiment. For the characterising of the mutants, the stability tests indicated these duplication mutants are very unstable. Up to 21% of the population lost the mutation through overnight culturing. The stability test also showed the 0.49 Mb duplication

mutant is more stable than those ~1 Mb in size, which suggested the stability of the duplication mutation can be related to the size of the duplication fragment carried.

Further deep tRNA pool sequencing revealed these duplication mutants increased the proportion of tRNA-Gly^{GCC}. This follow-up study was finalised by population sequencing. The population sequencing discovered duplication mutants as small as 2170 bp and 4273 bp (and demonstrated that there was no contamination between lines M1-M3, and W1-W5).

In the following chapters, I will focus on the evolutionary fate of these unstable duplication mutants. Thus, with a longer evolution time, what would become of these unstable duplications and the new copy of the tRNA gene?

Chapter IV – Restarted and Extended Evolution Experiment to Investigate the Evolutionary Fate of the Unstable Large-Scale Tandem Duplication Mutation Methods

In this chapter, I carried out most of the experiments, data analysis and visualising of the data from the above experiment except sequencing work. Sequencing work were done by Dr Sven Künzel and Gunda Dechow-Seligmann. Gunda Dechow-Seligmann also conducted the population PCR to detect the dynamics of duplication mutants.
Photo taking and post-processing were under the assist from Michael Schwarz.

4.1 Introduction

4.1.1 Brief introduction of previous work

Previous studies from Ayan et al. 2020 [14], PhD thesis [38] and Khomarbaghi, Ngan et al. 2024 [57] showed the two tRNA gene deletion mutants can quickly compensate for the growth defect by duplicating a large part of their genome (with the compensatory tRNA gene) [14, 38, 57]. Duplication mutation is a *de novo* mutation in the bacteria that changes the tRNA gene set by pre-existing genes (Fig 4.1). Large genome duplication in bacteria has been observed for decades. As early as 1970, many biologists were interested in studying tandem duplication [45, 67]. Tandem duplications appear at a high frequency and are lost at a high rate [43, 45]. Duplications are generally transient in nature due to the high intrinsic instability and associated fitness costs.

The previous studies showed that both engineered strains were quickly compensated by large-scale duplication. This demonstrated that gene duplication is a *de novo* mutation occurring at high frequency and can alter the growth advantage of bacteria by changing the gene copy number of pre-existing genes.

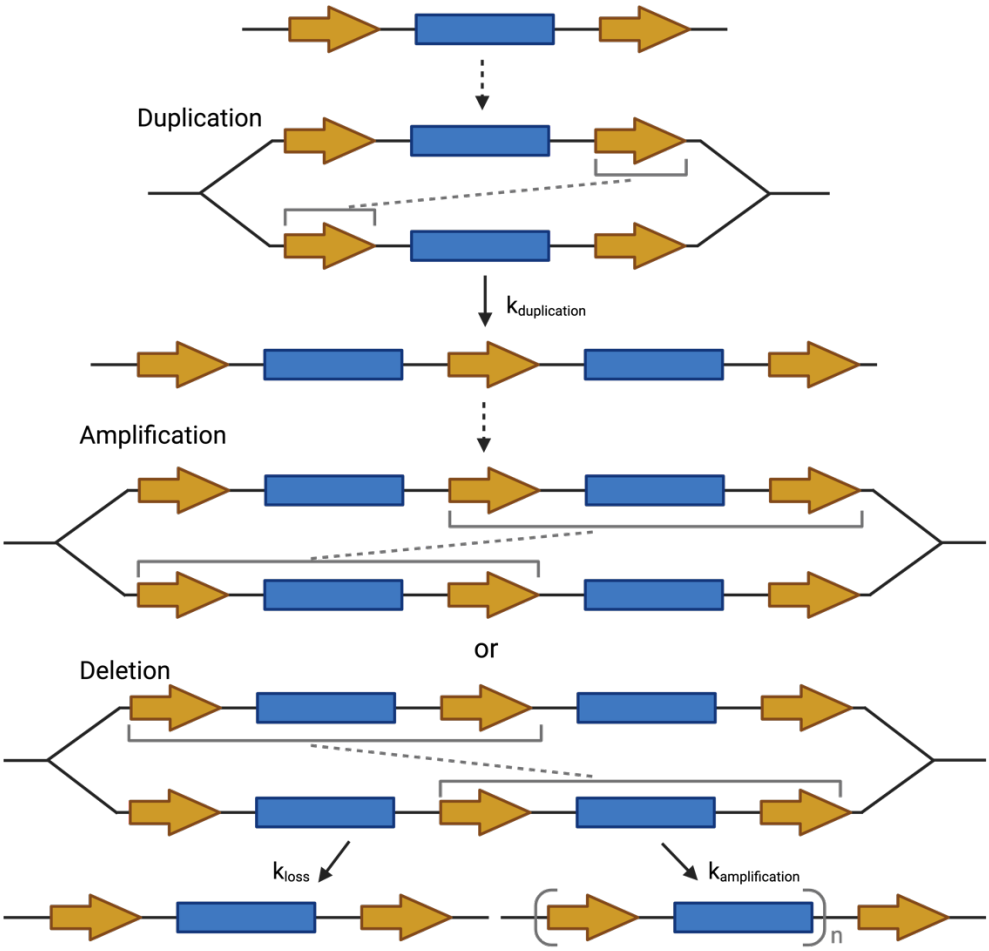


Fig 4.1. Schematic diagram of duplication, amplification and deletion. Duplication occurs at a high frequency through homologous recombination between repeat regions or through RecA-independent mechanisms between short- or no-homology regions. Yellow arrows denote homology regions (long or short) involved in recombination. Dashed lines denote recombination. Horizontal brackets indicate the region of homology involved in recombination. k denotes the rate of duplication, amplification or loss. Figure modified from [43]. Figures were drawn using Biorender.com.

In the previous chapter, some follow-up experiments were done for the investigation of compensatory duplication mutation in Δ*EGEG*. One of them is the stability test on the selected isolates due to the fact that gene duplication is unstable and could be lost due to the fitness cost. Results demonstrated that duplication mutations were unstable. The large-scale duplication (~1 Mb) can be lost at ~22% from an overnight culture. The smaller fragment (~0.49 Mb) is relatively more stable but still lost at 5% from an overnight culture. With such a high instability, it is hard to be convinced that these large-scale duplications are the final evolutionary "product". The results from the stability test in the previous chapter also suggested that having a smaller

fragment could be one of the mechanisms to gain a more stable duplication. In the introduction (Fig 4.2), the following hypotheses were mentioned about the evolutionary fate of the unstable duplication mutants:

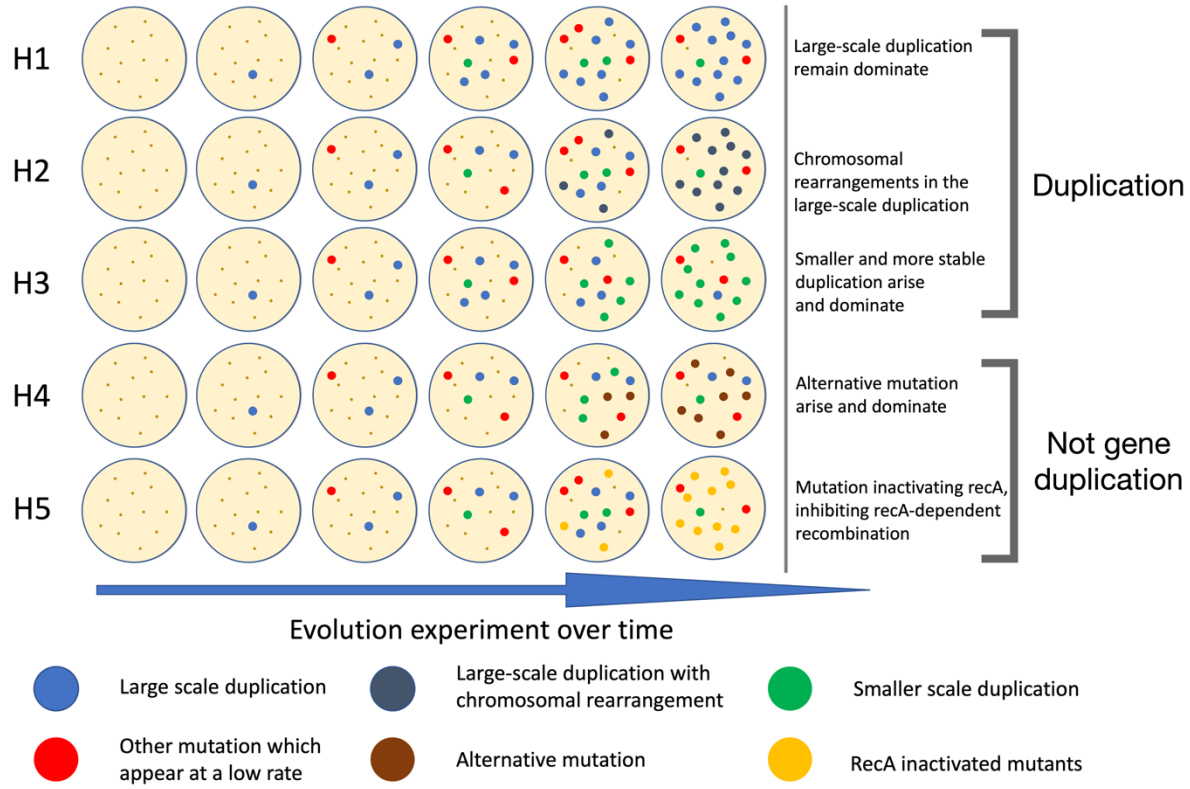


Fig 4.2. Hypotheses regarding the evolutionary fate of the duplications. Different colours show different types of mutations. Blue, grey, red, and green are duplication fragments. Brown and yellow (big) represent other potential mutations that may compensate for the deletion of tRNA gene(s). The figure was drawn using PowerPoint and OmniGraffle.

(H1) The large-scale duplication fragments remain dominant in the evolving populations;

(H2) Extra copies of genes that are not functional and beneficial may be inactivated [68, 69]. After gene fixation, the chromosome will be rearranged (Fig 4.3);

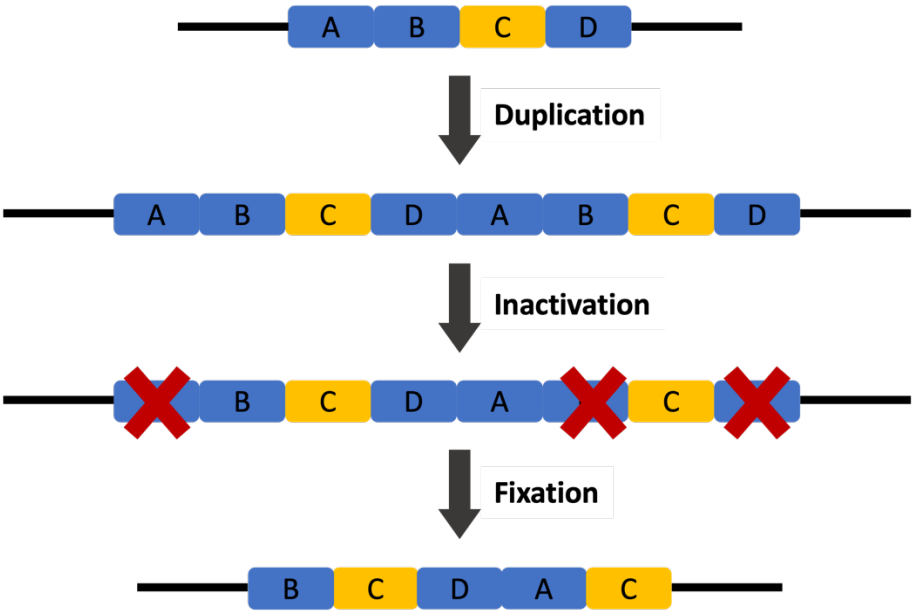


Fig 4.3. Schematic diagram demonstrating the Selection during Niche Adaptation (SNAP) hypothesis. ABCD genes were first duplicated with function beneficial gene C (yellow). The extra copies of the non-beneficial gene ABD (blue) were inactivated. Further fixation led to chromosomal rearrangement. Figure modified from [68]. The figure was drawn using PowerPoint.

(H3) Progressively smaller and, therefore more stable duplication fragments arise and dominate the population;

(H4) A stable alternative mutation arises and dominates the population;

(H5) Mutations inactivating *recA* led to inhibiting of recA-dependent recombination to stabilise the duplication mutations.

4.2 Aims

Previous studies carried out the evolution experiment for 14 - 28 days (~210 generations). Also, in general, only a few isolates were picked for whole-genome sequencing (one compensated isolate per evolving line). This limited the understanding of what would become of the unstable duplication, and the discovery of alternative mutations in the lines.

This chapter aims to carry out the evolution experiment for a longer time, allowing investigation of the evolutionary fate of the unstable large-scale duplications. This were achieved as follows:

1) The evolution experiments were started again with both founder strains (Δ*serCGA* and Δ*EGEG*) and the wild-type, and transferred for 100 days (~750 generations).

2) Phenotype changes were documented over the evolution experiment.

3) Compensated isolates were selected for whole genome sequencing. To broaden the understanding of the diversity of mutations contained in the evolving lines, a larger sampling of compensated mutants was performed with a much higher number of mutants selected (72 isolates in total). This allowed us to investigate, for example, if there are any mutations other than duplications.

4) These compensated mutants were characterised and further tested for the bioinformatically predicted mutation.

Experiments in this chapter investigated the evolution of bacteria to compensate for the growth defect due to sub-optimal tRNA gene sets for the translation need. Further experiments characterised these mutations in the compensated mutants.

4.3 Compensation of the bacterial strains over the evolution experiment

The evolution experiment was restarted and extended to investigate the evolutionary fate of the unstable large-scale duplication. Three founding strains (wild-type, Δ*serCGA*, and Δ*EGEG*), each comprising four independent replicates, underwent 1% daily serial transfers over 100 days, equivalent to approximately 750 generations (Fig 4.4).

Previous studies observed the two tRNA gene deletion strains quickly compensated for the growth defect, but neither had a general overview of how the populations were compensated. Therefore, in this study, the population samples were regularly plated at least every 3-4 days during the evolution experiment (Fig 4.4). The large and small colonies were counted to measure the proportion of compensated and non-compensated mutants (Fig 4.5). All lines for both engineered strains were compensated rapidly (by day 14). After day 14, most of the colonies were large. A few small colonies were observed, possibly due to the instability of large-scale duplication fragments, with the loss of a fragment leading to a reversal of phenotype. From day 14 until day 100, the populations mainly consisted of various subtypes of large colonies.

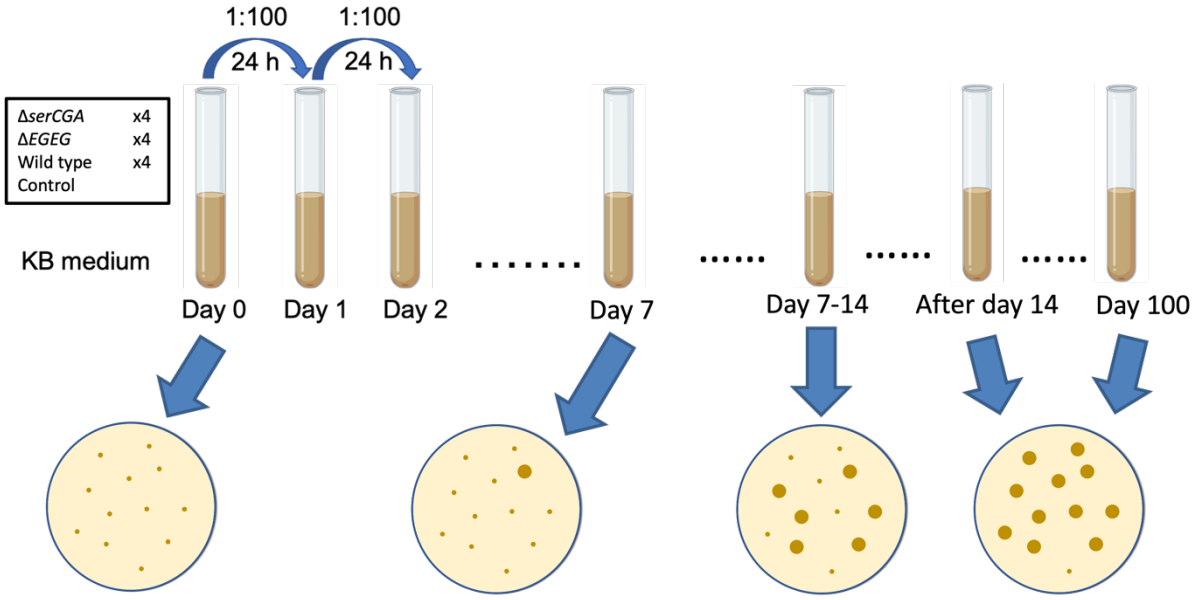


Fig 4.4. Schematic diagram of the serial transfer evolution experiment and the dilution plating. Each of 12 bacterial lines (and a medium control) is passaged daily through rich (KB) medium. The bottom figures are the schematic diagram of the dilution plated population. All lines started having compensated colonies before day 7. Most of the bacteria in the population are compensated after day 14. The 1% daily serial transfer was continued until day 100. Bacterial population samples were collected daily for further analysis. The glycerol stock was stored and used for re-examination. For contamination checking purposes, except for performing PCR on population samples, suspicious colonies in the plates used for colony counting were also examined in PCR. (Colony figures do not represent the phenotypes. Among all lines, not all colonies are smooth. See Fig 4.8 for detailed phenotype dynamics over the experiment.) The figure was drawn using PowerPoint.

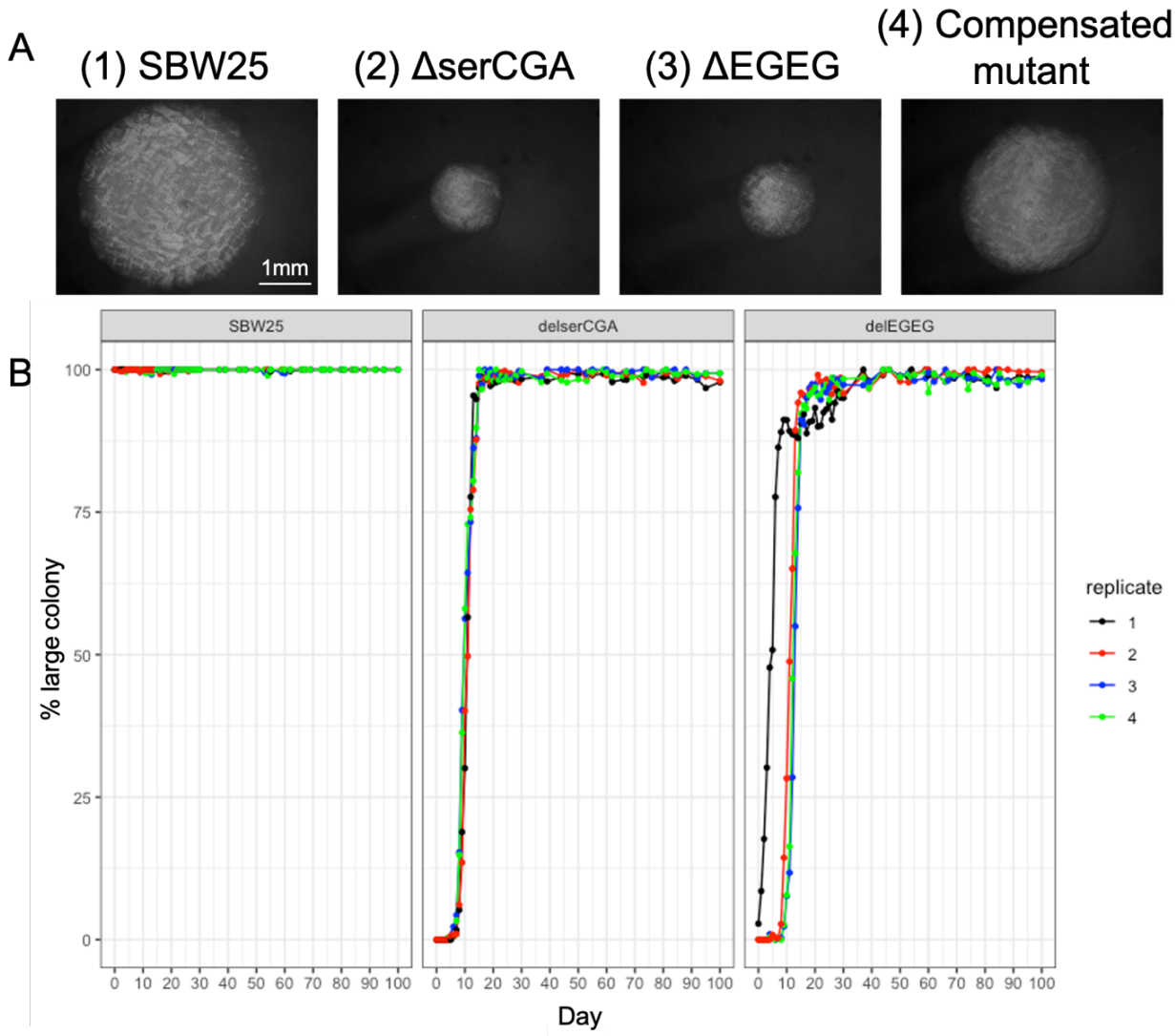


Fig 4.5. Colony phenotypes in evolving lines of the serial transfer evolution experiment. **(A)** Images of (1) SBW25; (2) Δ*serCGA*; (3) Δ*EGEG* founding colony and (4) compensated Δ*serCGA* colony. **(B)** Large colony-forming, compensated strains were first seen in Δ*serCGA* lines (middle) on day 4, with a little inter-population variation. Among the Δ*EGEG* lines, one replicate showed compensation on day 0 (*i.e.*, in the founding culture), while the remaining three lines showed compensation between days 6-9. After day 20 (Δ*serCGA*) and day 30 (Δ*EGEG*), the lines were dominated by large colonies. The extended evolution experiment revealed all lines were quickly compensated. Meanwhile, until day 100, small colonies were still observed in all Δ*serCGA* and Δ*EGEG* lines. This could be due to the populations containing unstable mutations that led to reversing in phenotypes or the uncompensated founder strains persisted in the population. Colony images were taken using VisiCam 1.3 (VWR, Germany). Figures were drawn using PowerPoint and RStduio.

4.4 Morphotype identification over the evolution experiment across lines

Over the 100-day evolution experiment, different colony morphotypes arose from the founder strains. When plating the population out on agar plates, at least five morphotype classes were observed. The diversity in morphotypes also underpinned a higher diversity in genotypes in the population. For the follow-up analysis, these morphotype classes were systematically classified. Two of the morphotype classes were present in the founding genotypes: large smooth (LSm, wild-type), small smooth (SSm; Δ*serCGA* /Δ*EGEG*), and other emergent morphotypes: opaque (LOp), wrinkly spreader (WS) and others (Fig 4.6).

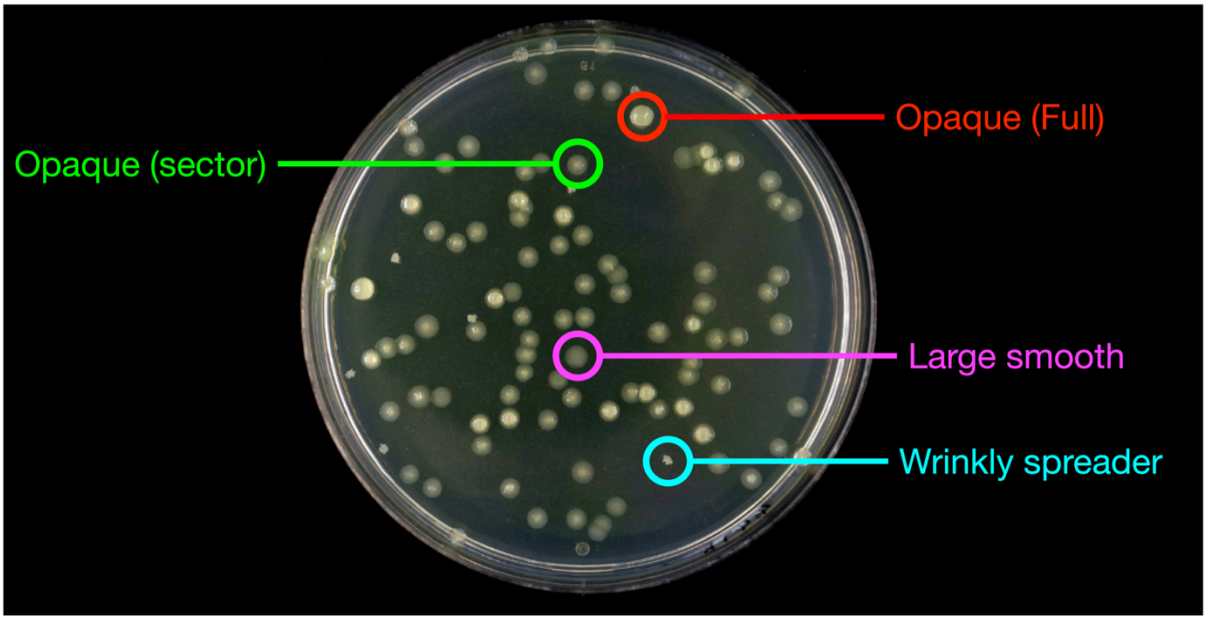


Fig 4.6. Phenotype identification from one Δ*serCGA* line on day 50. Opaque types overproduce the colanic acid-like polymer. Wrinkly spreaders are widely observed in *P. fluorescens* SBW25. The usual observed phenotypes also include large and small smooth types. Other rare phenotypes were also observed. Later, whole-genome sequencing results revealed these usually contain mutations in the BarA-UvrY two-component system. Labelling of the image was done using PowerPoint.

Notably, some individuals from Δ*serCGA* founded lines produce a colanic acid-like polymer (CAP) and opaque colonies because these duplication fragments contain the CAP biosynthetic genes located close to the compensatory tRNA gene (*serTGA*) in this founder strain [70]. Therefore, this phenotype is unique among the Δ*serCGA* lines.

The diversification in phenotypes suggests higher diversity in genotypes, which are awaiting revealing by genome sequencing.

4.5 Contamination checking in this study

There is always a chance that contamination occurs during any experiment, either from the external environment or from the inadvertent exchange of material between lines. This is important to mention what has been done to limit the chances of contamination. In this study, contamination checks are mainly done by phenotype checking or PCR.

The first contamination check can be done by examining the phenotypes on agar plates. Firstly, the grown *P. fluorescens* secrete a soluble fluorescent pigment [28]. Secondly, the opaque phenotypes should only appear among those Δ*serCGA* lines based on phenotype identification and whole-genome sequencing findings, where some duplication fragments contain the CAP biosynthetic genes. However, the duplication fragments compensating for Δ*EGEG* loss occur in a different region of the chromosome. Thus, the opaque phenotype is not expected to appear in lines founded by Δ*EGEG* (or wild-type). Otherwise, it suggests there may be a cross-line contamination. More phenotypes arose in the later phase of the evolutionary experiment, which increased the difficulty of identifying contamination during plating.

The contamination measures were also arranged without fully relying on phenotypes. There were also probe PCR primers designed for second contamination checking (Fig 4.7A). In short, one primer (forward) targets the deleted tRNA gene, and another primer (backward) targets outside the deleted gene. No PCR products should be amplified among lines evolved from the deletion strain, while strains/lines containing the tRNA gene will generate a PCR product (Fig 4.7B). Two types of PCR checking (population and single colony) were performed during the experiment, with two different sets of primers for each founder. The population PCR checking gave an overview of any potential contamination between lines. Meanwhile, PCR on colonies with unexpected phenotypes checked the colony at a higher sensitivity.

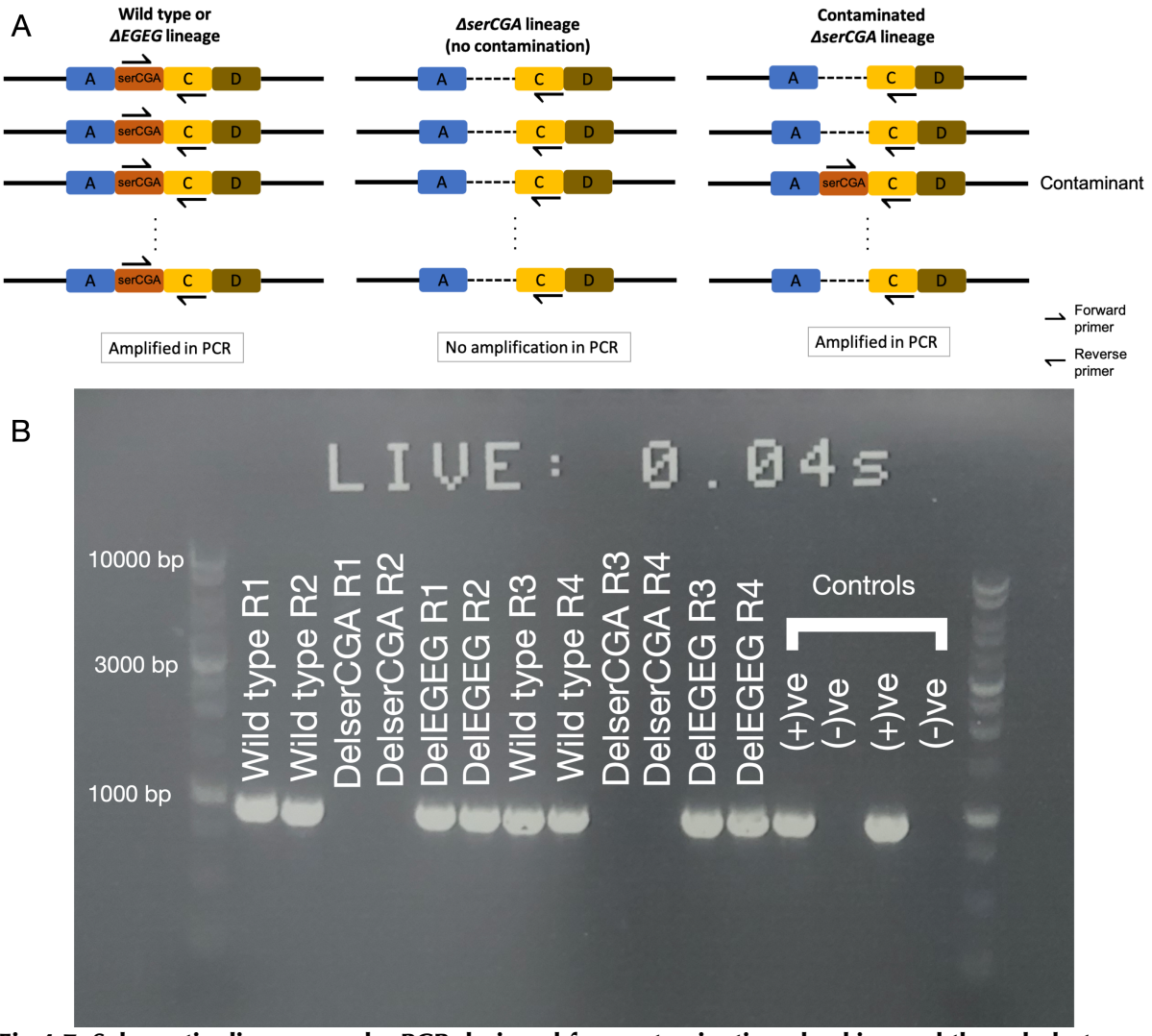


Fig 4.7. Schematic diagram probe PCR designed for contamination checking and the gel photo. **(A)** *serCGA* PCR is used as an example. *ΔserCGA* line/colonies will not be amplified if there is no cross-contamination between lines. **(B)** As an example, gel electrophoresis of PCR products of Day 100 samples using *serCGA* primers. All *ΔserCGA* lines showed no amplification. Day 0 of wild-type, *ΔserCGA* and *ΔEGEG* were used as positive controls. Templates for control lanes: (1) unevolved wild-type (+ve), (2) unevolved Δ*serCGA* (-ve), (3) unevolved Δ*EGEG* (+ve), and (4) sterile water (-ve). R1-4 are different isolates with unexpected phenotypes from day 100. Figures were drawn using PowerPoint.

Indeed, a contamination event was detected by observing opaque colonies in an ΔEGEG line. From the first round of the evolutionary experiment, day 75, samples were found to be contaminated across lines. Substantial PCR tests showed the contamination started before day 70.

When tracing back the transferring experiment record, it was paused on day 53 and day 61. The contamination was likely due to the restarting process. During the restarting of the experiment, 180 μl in total were taken from glycerol stocks. When the 200 μl pipette tips were close to full loads and careless repeating pipetting for mixing samples, there was a high risk for contamination by small residual leftover media on the pipette. Hence, it may be carried over when transferring the next line. To ensure the experiment can be restarted safely without carrying contaminants, day 53 samples (the second restarting point before contaminants were found) were chosen to retrieve the population. Filtered pipetting tips were used for all the transfers. Extra caution was put in before day 75 in the second round of the evolutionary experiment by more regular plating and with new PCR designed with higher sensitivities. No contaminants were found by the end of the experiment (day 100). These were confirmed with a contamination check by PCRs before every gDNA sample was sent for sequencing.

The contamination event indicated the importance of regular contamination checking using multiple methods in evolutionary experiments.

4.6 Temporal changes in morphotypes among lines

After finishing the evolution experiment, population samples were retrieved from glycerol stock. Samples were grown on KB agar again with dilution plating to test the morphotype dynamics. The phenotypes were quantified in all lines on samples from different days (Fig 4.8). Changes were observed among all lines over the evolutionary experiment. At least six different phenotypes were observed (Fig 4.9). Some phenotypes are only specific to the Δ*serCGA* founder due to overexpressing of polymer-producing related genes.

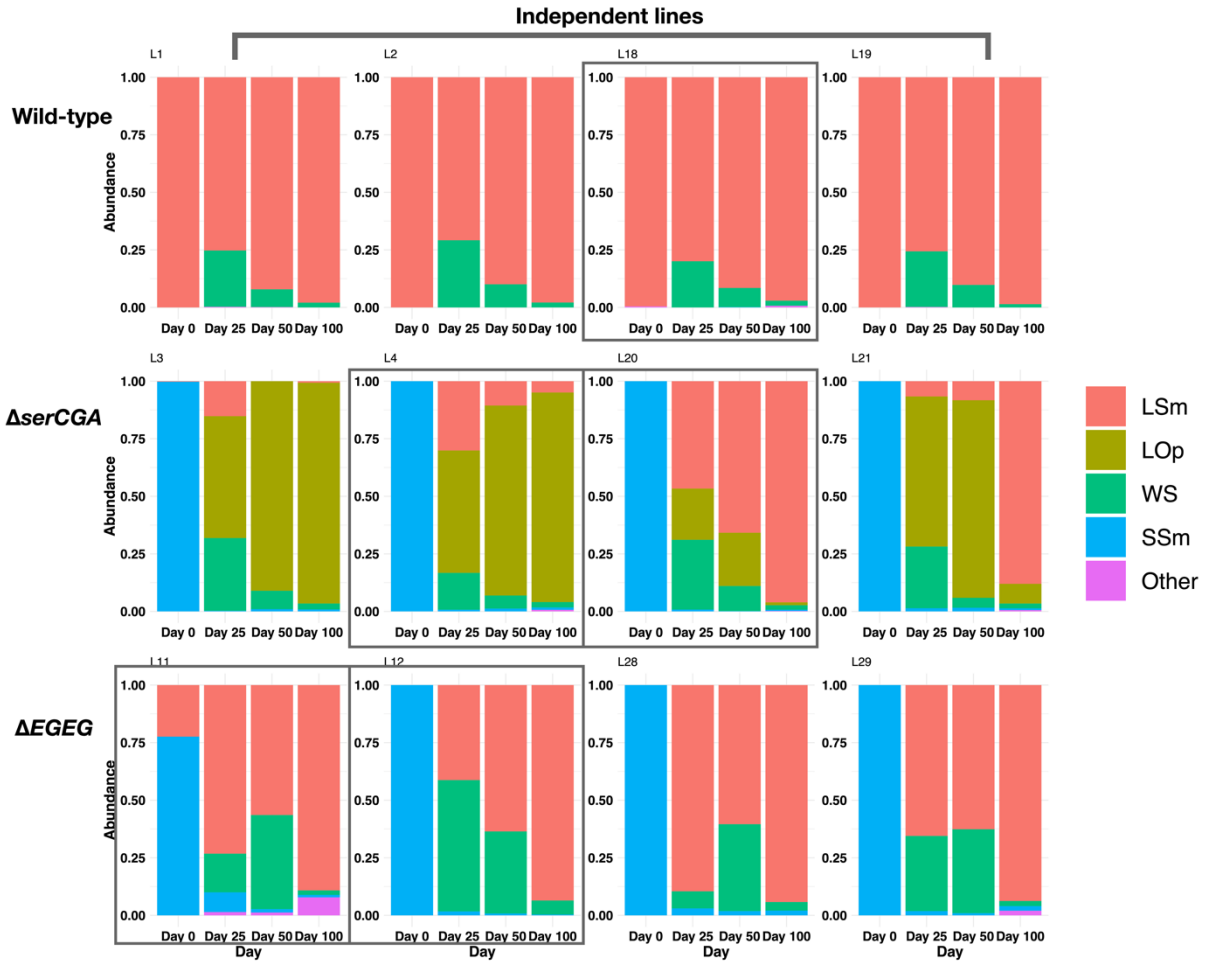


Fig 4.8. Relative abundance of morphotypes in each line. (Top) Among wild-type lines, large smooth (LSm) was the dominate morphotype. **(Middle)** Among *ΔserCGA* lines, opaque type were found in all lines since day 25. One line has large smooth type absent on day 100. **(Bottom)** One *ΔEGEG* line started compensated since the experiment started. Therefore, this line observed large smooth type since day 0. No opaque types were observed as this may indicate a potential contamination event from *ΔserCGA* lines. **(Top-bottom)** Both *ΔserCGA* and *ΔEGEG* lines observed small smooth colonies over 100-day passage and WS observed since day 25. The squared lines were later picked for in-depth sampling for whole-genome sequencing. Other lines were sampled at less depth. Figures were drawn using RStudio and OmniGraffle.

4.6.1 Phenotypic changes observed among all lines

Among all lines, colony sizes increased from day 1 to day 100, seemingly due to adaptation to the environment of the evolution experiment (Fig 4.10). This increased the difficulties when we compared the small types with unevolved founder strains over the evolution experiment.

However, at any particular time point, small(er) colonies can easily be distinguished from large(er) colonies. Also, the wrinkly spreaders (WS) were observed in all lines, mostly in day 25 and day 50 samples. WS types decreased in abundance by day 100 for all lines. This is unsurprising as WS types are favoured on air-liquid surfaces to form mats in static conditions [71]. In this experiment, the evolution experiment was performed in a shaking liquid media. During the experiment, no visible mat formation was observed as well. One of the possible reasons for WS to rise during the experiment is that they started between the media and the inner surface of the culture tube. Then, they suspended in the liquid media due to the shaking condition and co-existed with the non-WS type throughout the evolutionary experiment. Because these WS have lower fitness than the smooth type arose after day 50 in the shaking condition, their relative abundance in the population decreased after 100 days of the experiment (~750 generations).

Other than smooth and WS, other phenotypes were observed from day 100 samples. One of them is called Round Disc (RD). Round disc (RD) mutants were seen after day 50 in several lines. This morphotype is slightly smaller compared with typical large smooth colonies. The RD morphotype is whiter and slightly shiny on agar plates. This mutant was confirmed in the isolate sequencing related to *barA/gacS* mutation, which is covered later in this chapter.

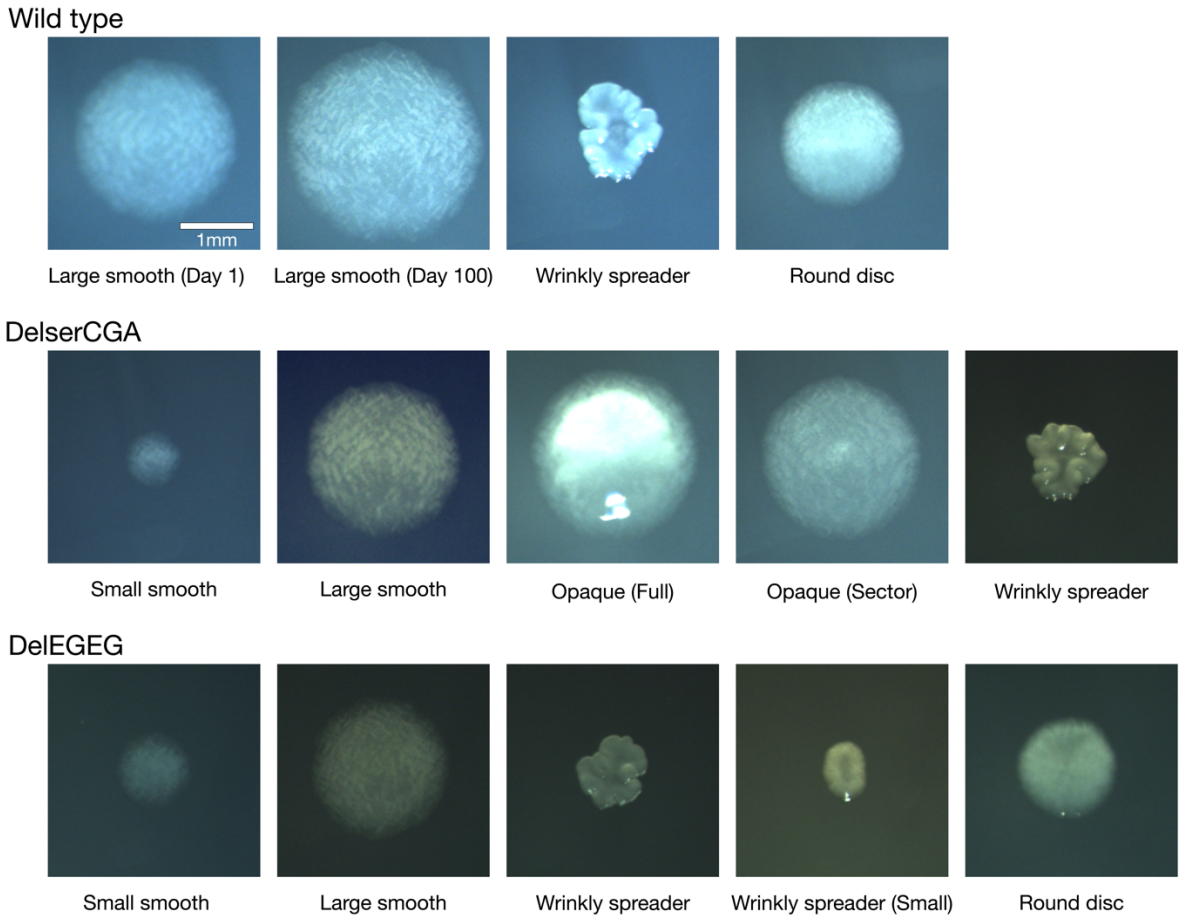


Fig 4.9. Different morphotypes were identified in each founder line. Large smooth colonies on day 100 were larger compared to day 1. These two small smooth colonies from Δ*serCGA* and *ΔEGEG* were isolated from day 1 samples. Large smooth colonies in both Δ*serCGA* and *ΔEGEG* were from day 25. Both full and sectored opaque types were raised from the large duplication, duplicating the CAP biosynthetic genes. Labelling of image was done using OmniGraffle.

4.6.2 Phenotypic changes observed among wild-types lines

All the phenotypes that arose in the wild-type lines were also observed in the other two founder strains (Fig 4.10). These are particularly the WS and RD types. Phenotypes that arose in wild-type lines suggested these are due to background mutations and have little to do with the compensation of tRNA genes. Results of isolate sequencing, later in this chapter, provided a further investigation of these mutations (see *general adaptation*).

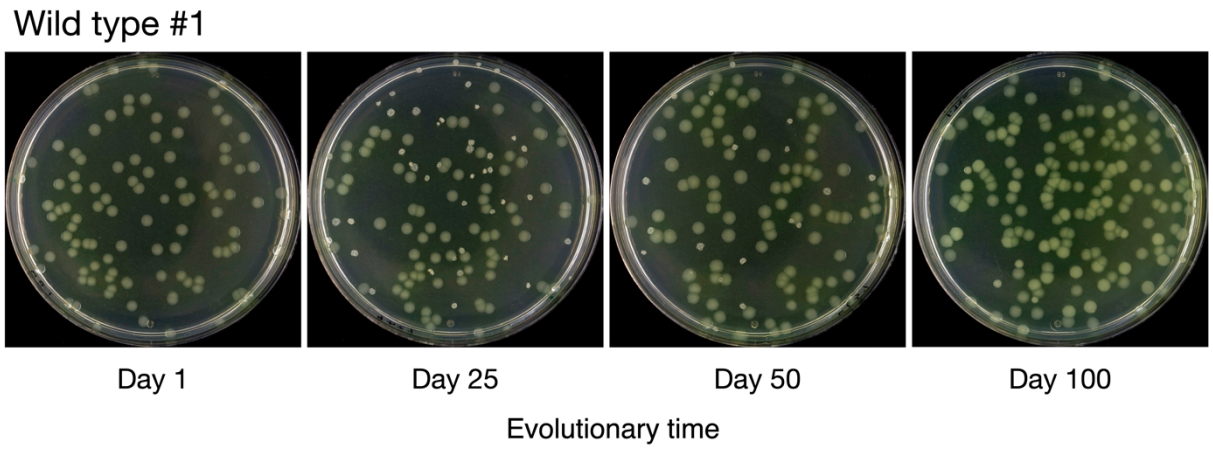


Fig 4.10. Changing colony morphotypes in a wild-type line over 100 days. Horizontal images were the population samples from different days over the evolution experiment after 72 hours of incubation. These images were taken from one of the wild-type lines. Wrinkly spreaders appeared before day 25. In the day 100 image, the proportion of wrinkly spreaders decreased. Instead, half of the smooth colonies changed in colour. The size of the colonies also increased over the experiment. Images from other lines can be found in Appendix (1-3). Labelling of images was done using OmniGraffle.

4.6.3 Changes observed among ΔserCGA lines

For Δ*serCGA* lines, day 1 samples were mainly uncompensated small colonies. From day 25, most colonies were compensated and only contained a few small colonies (Fig 4.8). The small colonies could be due to unstable duplication mutants who lost the mutation or the uncompensated strains persisted in the population. There were two opaque morphotypes: full opaque or sectored opaque. Full opaque phenotypes are those colonies producing full colanic acid (white colour and shiny with light). The sectored type has less colanic acid, leading to while sectors in the middle of the colonies (Fig 4.9; 4.11). Both phenotypes are expected to be those duplication mutants that also duplicate a locus containing genes (*pflu3655* to *pflu3678*) to synthesise the colanic acid-like polymer (CAP). The first gene, *pflu3655*, controls the expression of these genes in a self-perpetuating loop: *pflu3655* activates its own transcription, and hence polymer production [72-74]. Presumably, duplication of *pflu3655* (in some of our mutants) pushes the level of Pflu3655 high enough to activate the expression of the polymer genes and the self-perpetuating loop.

Another significant change was observed in Δ*serCGA* replicate#3: On day 100, the smooth, non-CAP morphotype is dominant (Fig 4.12). This morphotype may be underpinned by a duplication fragment(s) that does not contain the CAP-producing genes, or a different type of mutation altogether.

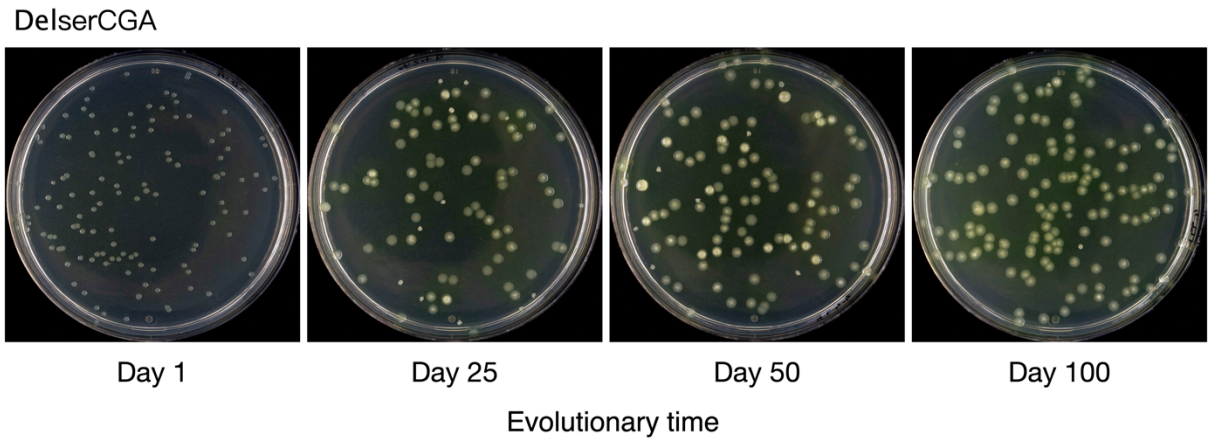


Fig 4.11. Changing colony morphotypes in a Δ*serCGA* replicate line #2 over 100 days. Horizontal images were the population samples from different days over the evolution experiment after 72 hours of incubation. These images were taken from one of the Δ*serCGA* lines. In this line, there are divergences in morphology in day 25 and day 50 samples. For example, full opaque, sectored opaque, wrinkly spreaders, and smooth morphotypes appeared in day 25 and day 50 samples without any as the obviously dominant type. Morphotype diversity reduced considerably by day 100, and sectored opaque types were the dominant type. Images from other lines can be found in Appendix. Labelling of images was done using OmniGraffle.

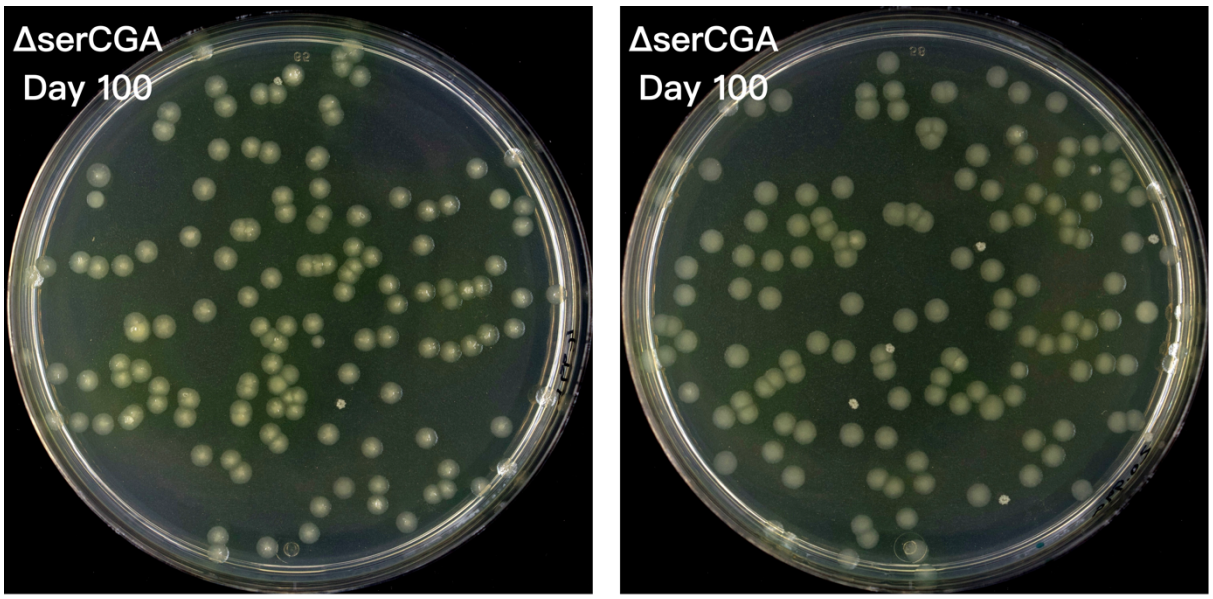


Fig 4.12. Different dominant morphotypes from the same founder strain after 100 days of the evolution experiment. There were four independent lines in each founder strain. The Δ*serCGA* lines have opaque morphotypes, which became dominant in some lines (Left) but not all (right, for example). The smooth type is the dominant for the other line (right). Knowing that the opaque type rises with the large duplication, line 20 is either dominant by a smaller duplication fragment or an alternative mutation without duplication. Later sections will reveal this with isolate sequencing and population sequencing. Labelling of image was done using OmniGraffle.

4.6.4 Changes observed among ⊿EGEG lines

For Δ*EGEG*, day 1 samples were small colonies. From day 25, most of the colonies were compensated large colonies. More wrinkly spreaders were in these lines compared with wild-type and Δ*serCGA*. For example, in one Δ*EGEG* line, WS is the most dominant type on day 25. When comparing day 50 by founder lines, WS decreased in relative abundance among wild-type and Δ*serCGA* lines. While all four Δ*EGEG* lines still have a significant proportion of WS (36% - 41%; Fig 4.8) on day 50. One possible reason could be the duplication Δ*EGEG* mutants are less fit than the duplication Δ*serCGA* mutants (or wild-type) in the suspension form in liquid media. The WS are mat-formers that start from colonising the well of the culturing tube and form mats. Hence, they may survive in niches different from the smooth phenotypes. The less fit Δ*EGEG* smooth mutants would take longer to outcompete the WS type once a fitter mutation appeared.

Unlike the Δ*serCGA* duplication strains, the duplication Δ*EGEG* strains do not produce the colanic acid morphotype. It is hard to predict the evolutionary fate of the large-scale duplication from the phenotypes. The evolutionary dynamic of the duplication fragments will be discussed in isolate and population sequencing.

Morphotype checking revealed the composition of each line based on morphotypes. Some rare and low relative abundance morphotypes were identified. From the morphotype checking, the lines with higher diversities were identified. This information could suggest which lines should have a wider sampling plan for whole-genome sequencing, which is covered in the next section.

4.7 Whole-genome sequencing of selected isolates

Different morphotypes in *P. fluorescens* SBW25 are expected to be underpinned by different mutations. Therefore, 72 isolates were selected based on the morphotypes identified, aiming to reveal as many mutations as possible. The isolates were picked from days 1, 25, 50 and 100, from each of the 12 evolutionary lines. Five lines were chosen for more in-depth sampling (wild-type replicate line #3; Δ*serCGA* replicate line #2, 3; and Δ*EGEG* replicate line #1, 2). The in-depth sampled lines were highlighted with squares in Fig 4.8.

Δ*serCGA* replicate line #2 was selected because of the discovery of a promoter mutant in early preliminary sampling. This line was dominated by opaque type on day 100. Therefore, Δ*serCGA* replicate line #3 was also chosen because it was dominated by the large smooth type on day 100. Δ*EGEG* replicate line #1 was chosen because compensated mutant arose since day 1 in this line. For Δ*EGEG* replicate line #2, it is because it has a high proportion of WS arose compared with other lines. Wild-type lines serve to provide crucial information about the background mutations in the population. Therefore, only one line was chosen for in-depth sampling for sequencing. The remaining seven lines were also sampled but at less depth.

After contamination checking, extracted gDNA was sent for Illumina NextSeq sequencing (150 bp pair-ended reads), with a mean coverage per bp of the reference genome of 75-fold per sample. There were four samples with coverage of less than 50-fold. The sample with the lowest coverage contained 36.4-fold coverage. Results from these four samples were examined with extra caution. Uncertain mutations with low read coverage were not included in the analysis (at least ~30-fold read coverage).

Mutations were predicted by *breseq* (v 0.37.1). **S**ingle **n**ucleotide **p**olymorphisms (SNPs), indels, inversions and emergent junctions were identified. The coverage was examined using Geneious Prime (v 2023.2.1). From isolates sampled from wild-type lines, a total of 32 mutations were identified (Fig 4.13). Among these, more than 65% were SNPs. 25% were insertions. Deletion and inversion contributed less than 10%. No duplication mutation was identified. In Δ*serCGA* lines, a total of 111 mutations were identified, where more than 41% were SNP. Duplication and insertion each contributed 28%. Less than 3% were deletions, and no inversion was identified. In Δ*EGEG* lines, SNP, insertion and duplication each contributed 28-32% of the 71 mutations identified in total. Around 8% were deletions, and less than 2% were inversion.

Among Δ*serCGA* and Δ*EGEG* lines, duplication mutations comprised more than 41% and 32%, respectively. The wild-type lines are dominated by SNPs or insertion mutants, which indicates a lot of the SNPs and insertions are background mutations in the population (Fig 4.13). This further revealed the evolving Δ*serCGA* and Δ*EGEG* lines likely contained substantial proportions of compensatory duplication mutants.

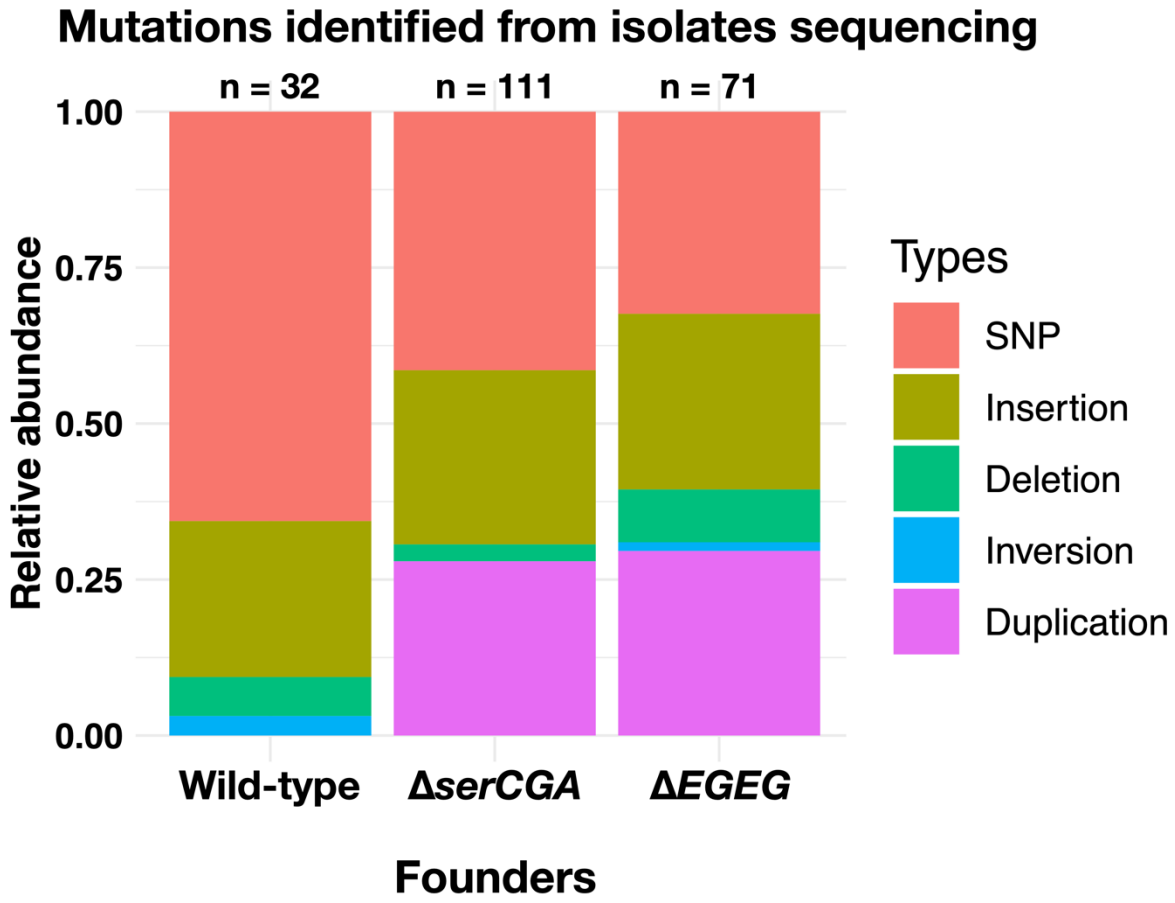


Fig 4.13. Relative abundance of types of mutations identified from isolate sequencing. In wild-type lines, most of the identified mutations are SNPs. In both Δ*serCGA* and Δ*EGEG* lines, SNPs, insertions and duplications have a similar relative abundance on the proportion of mutations identified. The figure was drawn using RStudio.

In functional terms, there are two types of mutations: a) general mutations identified in lines founded by all three founder strains, including SBW25 wild-type and; b) adaptions to the tRNA gene loss, which are mutations that compensate for the growth defect and improve translation efficiency. Each of these is discussed in more detail below.

4.7.1 General adaptions

The serial transfer evolution experiment included 100 days of transferring (around 750 generations). Evolution experiments typically involve gradual changes in the genetic makeup of a population over many generations. Often, the bacteria show adaption to the growth conditions of the experiment. Hence, these mutations are not expected to compensate specifically for the loss of the deleted tRNA gene(s). Among all the lines, three types of general adaptions are identified from isolate sequences: 1) mutations that affect flagella, 2) mutations that affect cellulose production, and 3) mutations in the master regulatory system BarA-UvrY.

4.7.1.1 Loss of motility during the evolutionary experiment

Mutations related to losing mobility in *fleQ, fliA* or *fliF* genes were identified among all the lines. A total of 24 mobility-related mutants are identified in all 12 lines. *fleQ* gene encodes the major flagellar regulator [75]. At least one *fleQ* mutation was identified in all of the 12 populations by day 100.

The *fli* genes are associated with flagellar formation [76]. Mutations in these genes affect flagellin expression resulting in non-motile, aflagellated bacteria [77, 78]. In our evolution experiment, the bacteria were cultured in shaking conditions. Under these conditions, forming and using flagella is costly but not crucial for survival or replicating. There could be a selective advantage in conserving energy by losing the ability to move. These bacterial motility adaptive mutations could be cost-benefit trade-off between motility and growth [79].

The motility of mutants was tested using 0.3% LB agar in 12-well plates (Fig 4.14). The 0.3% agar is a semi-solid agar. It allows bacteria with motility to swarm to reach more nutrients. Bacterial cultures of the isolates were dipped in the middle of the agar. Plates were examined the next day after incubating at 28°C in the static incubator. Bacterial cells with motility can swim in the agar and create a ring in the well. All the *fleQ*/*fli* mutants lost their motility, showing no swimming rings in the agar.

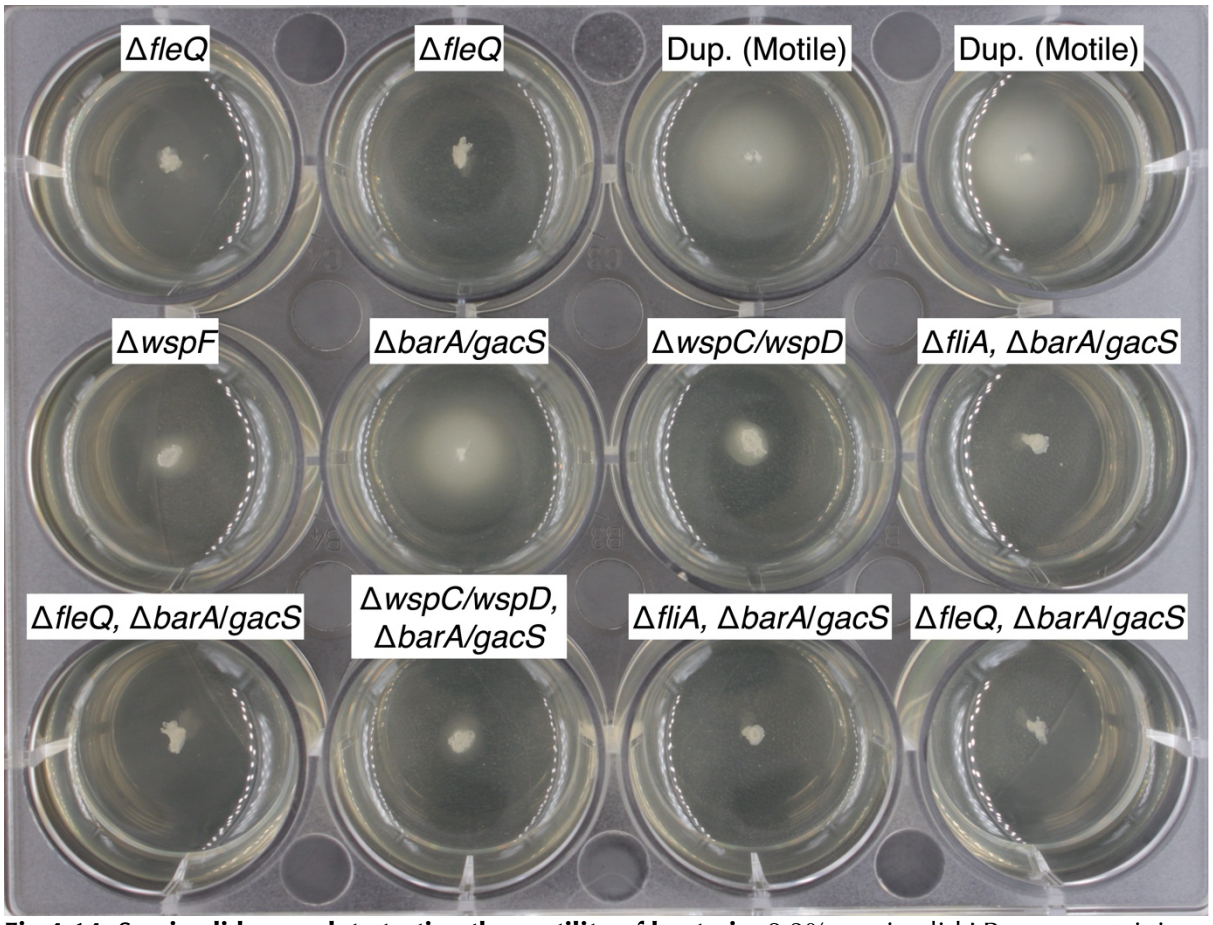


Fig 4.14. Semi-solid agar plate testing the motility of bacteria. 0.3% semi-solid LB agar containing 3g/L NaCl were prepared. The bacteria cultures were dipped in the middle of the well inside the agar. For example, the top left has a mutation in *fleQ,* leading to an aflagellate of bacteria. The motility assay showed that mutants with mutations in *wsp, fleQ* or *fli* led to aflagellate and motility defects. Mutants with only duplication or *barA/gacS* mutations were motile.

A study also found that *P. fluorescens* SBW25 could grow the flagella again even if *fleQ* is deleted via changes to gene regulatory network connections (also called rewiring) [80]. This shows a degree of adaptational flexibility in SBW25, with regard to changing between shaking or non-shaking conditions. In general, SBW25 demonstrated the phenotypic plasticity to change the trait of motile/aflagellate dynamically to match the prevailing conditions.

4.7.1.2 Rising of wrinkly spreaders

Wrinkly spreaders (WS) and mat formation are some of the adaptive radiations in static liquid microcosms observed in *P. fluorescens* SBW25. All the lines in this study had wrinkly spreaders appeared. From whole-genome sequencing and *breseq* analysis, 3 mutants were identified as having mutations in *wss* genes and 19 of them were identified as having mutations in *wsp* genes.

No mutation in *aws* or *mws* was identified from the isolate sequences in this study. These three pathways (Wsp, Mws or Aws) are subject to negative regulation of diguanylate cyclases (DGCs) [81]. Hence, these three pathways provided many mutational possibilities to activate their respective DGC and produce the WS phenotypes [82]. The *wss* operon is encodes the cellulose biosynthetic machinery related to the appearing of wrinkly spreader phenotype in *P. fluorescens* SBW25 [83]. Mutations in *wss* might interrupt the rising of WS phenotype. Studies also found after the deletion of *wssA-J* operon followed by evolution experiment in static microcosms, atypical wrinkly colonies [poly-beta-1,6-N-acetyl-D-glucosamine (PGA) wrinkly spreader] (PWS) were formed [82]. These involved *pgaABCD* encoding exopolysaccharide similar to cellulose or a *nlpD* amidase activator when defective causes cell chaining. These PWS types were not observed or identified in this study.

In this study, as mentioned in section 1.3.1.1, mutations that inhibit WspF function or activate WspE kinase activity lead to the activation of WspR and the synthesis of cyclic-di-GMP. Elevated cyclic-di-GMP levels subsequently induce the expression of the WS phenotype among the isolates sampled in this study.

From the wild-type WS isolates (they were from the same line), two mutants identified with *wspA* mutation and two with *wspF* mutation. One mutant contains a 15 bp insertion in the *wspF* and an SNP in *pflu0185*. PFLU0185 is a GGDEF/EAL domain-containing protein, which is related to the production of PDEs to degrade c-di-GMP. Both mutations are expected to lead to the appearing of the wrinkly spreader phenotype. In the two Δ*serCGA* sampled lines, all wrinkly spreaders were predicted to contain mutations in *wspA*. In the two Δ*EGEG*, mutants with *wspA* were predicted. There were two mutants in one line from different sampling timepoints were predicted to contain 11 bp deletion across *wspC* and *wspD*. All of these mutations were expected to lead to the wrinkly spreader phenotype.

Interestingly, some isolates not having a clear wrinkly spreader phenotype contains mutation(s) in *wsp* gene. They were identified as round disc (RD) in this study. Their looks denser and slightly smaller with a smooth surface. The reason is likely because they also have mutations in *wssB* and *barA/uvrY* two component system. When the function of *wss* operon is interrupted, they cannot form WS phenotype. While mutation in *wsp* system led to increase in expression of c-di-GMP. Hence increase in cellulose production. Therefore, the colonies look denser.

To compare the differences in cellulose production in these mutants, the biofilm assay was adapted from the microtiter dish biofilm formation assay [61]. In general, bacteria were grown in 96-well plates overnight at 28˚C. After washing with deionised water, the biofilm remained were stained by 0.1% crystal violet (CV) for 15 mins. After 3 washings with deionised water, the CV stains were dissolved by 30% acetic acid. Absorbances in the plate were read at 550 nm. Our strain, *P. fluorescens* SBW25, forms a biofilm at the air-liquid interface. The biofilm is sticky and some of them leave remains on the wall of the plates after washing. This method cannot accurately quantify the biofilm formation for our strain, but it gives comparable data for the biofilm formation between mutants. Biofilm assays on mutants showed that both *gacA*/*uvrY* and *wsp* mutants had increased biofilm production. The *wsp* mutants are wrinkly spreaders with good mat formation (Fig 4.17). They have the highest measures of mat formation in biofilm assay (Fig 4.15).

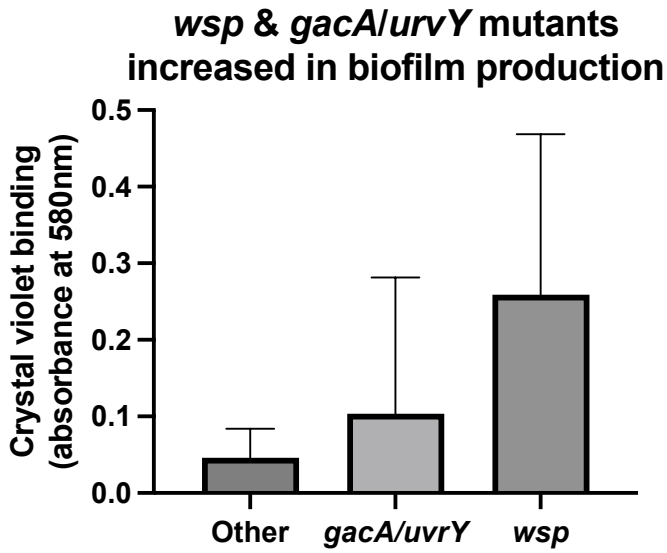


Fig 4.15. Crystal violet binding by isolates of interest. *gacA/uvrY* and *wsp* mutants showed an increase in biofilm production, while *wsp* mutants have the highest production.

Next, cellulose producing abilities of different morphotypes were tested by Congo Red assay and mat formation at air-liquid interfaces. The cellulose producing abilities of isolates obtained from this study were also qualified by Congo Red assay (Fig 4.16). Cellulose producing bacteria, which can form a good mat at air-liquid interfaces, will turn red on Congo Red agar.
WS types are mutants that overproduce cellulose-like polymers, leading to the distinctive morphotype [33]. In the Congo Red assay, all WS colonies turned red as expected.

RD and colanic acid-producing bacteria (LOp) were also of interest as they may still produce more cellulose. Results showed that RD phenotypes showed a slight increase in cellulose production, and LOp phenotypes did not (Fig 4.16; Table 4.1).

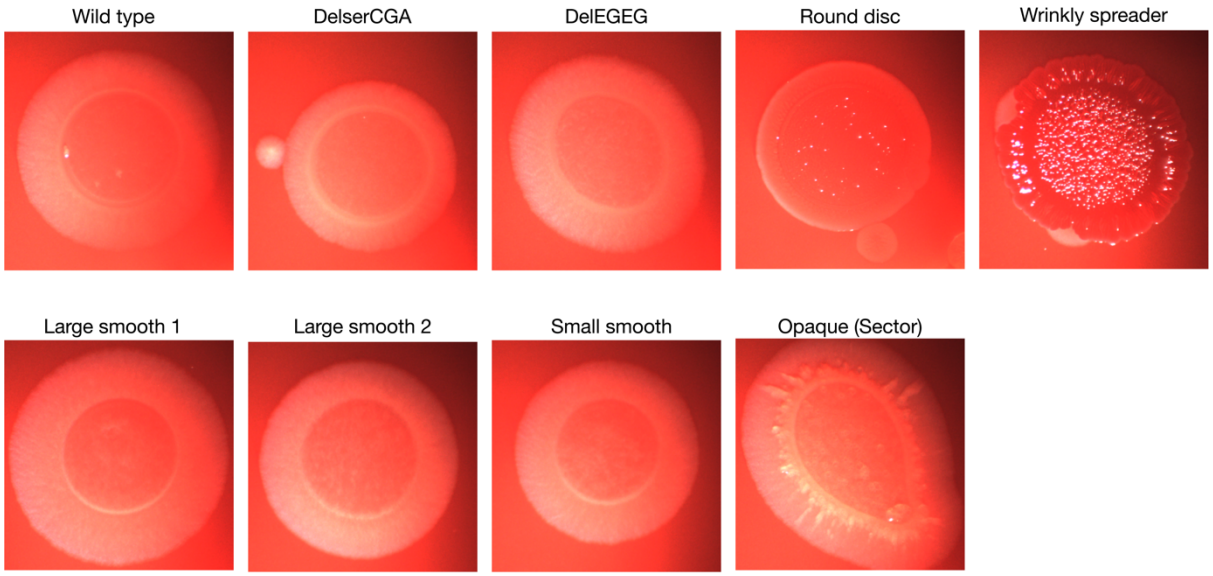


Fig 4.16. Congo red assessment in different morphotypes. Wrinkly spreader and round disc types were shown to produce cellulose. The wrinkly spreader turned red showing the production of cellulose in the colony. Round disc type also turned red but weaker than wrinkly spreader.

WS is also known to form a distinctive biofilm (or mat) at the air-liquid interface of static microcosms [84, 85]. In this study, the mat formation was further tested in static microcosms. All the WS mutants showed increased mat formation compared to SBW25 wild-type. All the evolved mutants do not form mat at air-liquid interfaces (Table 4.1; Appendix). Note that all unevolved strains (wild-type, Δ*serCGA and* Δ*EGEG*) form mats (Table 4.1; Appendix). This may be due to the strains losing mat-forming ability over the 100-day passage in a shaking environment.

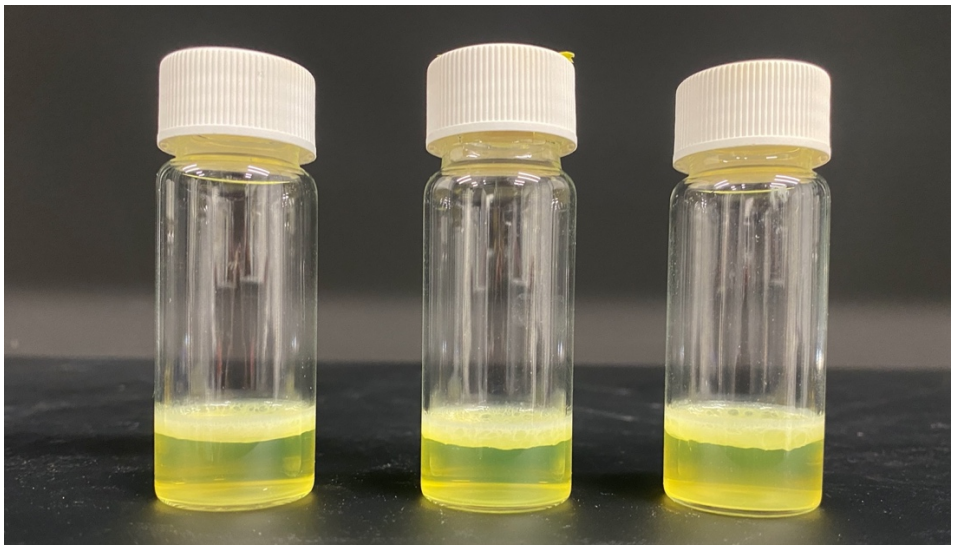


Fig 4.17. Formation of mats at the air-liquid interface of static microcosms of the wrinkly spreader (WS). A single colony of WS was picked to the liquid medium and left incubating under static conditions at 28°C for 48 hrs. Results show they are excellent mat formers at the air-liquid interface. Left to the right were different independent replicates. Photos of mat formation of other lines were prepared in Appendix.

Table 4.1. Qualitative phenotypical analysis of different morphotypes.

Morphotypes	Air-liquid mat formation	Congo red assay	Lines observed in
Wild-type	✓	✗	Unevolved (day 0)
Δ*serCGA*	✓✓	✗	Unevolved (day 0)
Δ*EGEG*	✓✓	✗	Unevolved (day 0)
Large smooth (LSm)	✗	✗	Wild-types, Δ*serCGA* and Δ*EGEG*
Small smooth	✗	✗	Δ*serCGA* and Δ*EGEG*
Wrinkly spreader (large)	✓✓✓	✓✓✓	Wild-types, Δ*serCGA* and Δ*EGEG*
Opaque	✗	✗	Δ*serCGA*
Rounded disc	✗	✓	Wild-types, Δ*serCGA* and Δ*EGEG*

4.7.1.3 Mutations in the BarA-UvrY two-component system

The BarA-UvrY two-component system regulates secondary metabolism, several phenotypic traits, and bacterial physiology [86, 87]. A BarA-UvrY two-component system was found in *Escherichia coli*, which belongs to the same orthologous family of BarA-SirA in *Salmonella enterica*, ExpS-ExpA in *Erwinia carotovora*, VarS-VarA in *Vibrio cholerae*, LetS-LetA in *Legionella pneumophila* and GacS-GacA in *Pseudomonas spp* [87-89]. This two-component system regulates the surface

structures of the bacteria such as the synthesis of capsules, exopolysaccharide (EPS), adhesion to abiotic and biotic surfaces, and regulation of surface adhesins [86]. Studies in *Pseudomonas* also found a role for the GacS-GacA system in virulence, quorum sensing, biofilm formation, and the production of biological control factors [87]. All GacA mutants identified in this study showed diverse morphotypes.

4.7.2 Adaptions to tRNA genes loss

In the *breseq* analysis, most of the mutants contain duplication fragments. In one of the Δ*serCGA* lines, there is one SNP promoter mutant, compensating for the growth defect without a duplication mutation. This section will cover the mutations related to adapting to loss of the tRNA gene(s).

4.7.2.1 Promoter mutation increasing the expression of a tRNA gene

The previous two studies only identified duplication mutations which compensate for the growth defect. Among the Δ*serCGA* founded mutants, some of the duplication mutants have the colonic-acid like phenotype (LOp), which is easy to identify. Some isolates were sampled for whole-genome sequencing in the early phase of the evolution experiment (day 14). Some large opaque (LOp) and large smooth (LSm) were sampled from Δ*serCGA* lines. Interestingly, one of the large smooth mutants does not contain any duplication. This can be easily seen by the coverage plot (Fig 4.18A). However, *breseq* identified some point mutation in the genome. One is a promoter mutation, with an a4163781g point mutation just outside the -35 box (Fig 4.18). This mutant is expected to increase the expression of the compensated tRNA-Ser^{UGA}. Hence, it also has a compensatory effect on the tRNA gene loss. The stability test showed that this mutation is stable. No small types were found after overnight culturing. This suggests there is a chance that this alternative mutation can overtake the population, considering the large-scale duplication mutations are unstable. However, after 100 days of transferring, this line was dominated by large opaque type, which suggests they are duplication mutants. There are still some large smooth types in the line. They were subjected to whole genome sequencing. Results from *breseq* suggest they are the same SNP promoter mutant. However, the large smooth types only comprised less than 5% of the population on day 100 (Fig 4.8). The duplication mutants with a fragment not covering the colanic acid-related genes (*pflu3655 - pflu3678*) may also have a smooth phenotype. This suggested the actual relative abundance of the SNP promoter mutant is actually lower than

the observed large smooth type. The dynamic of the SNP promoter mutant will be mentioned in the next chapter, with the results of population sequences.

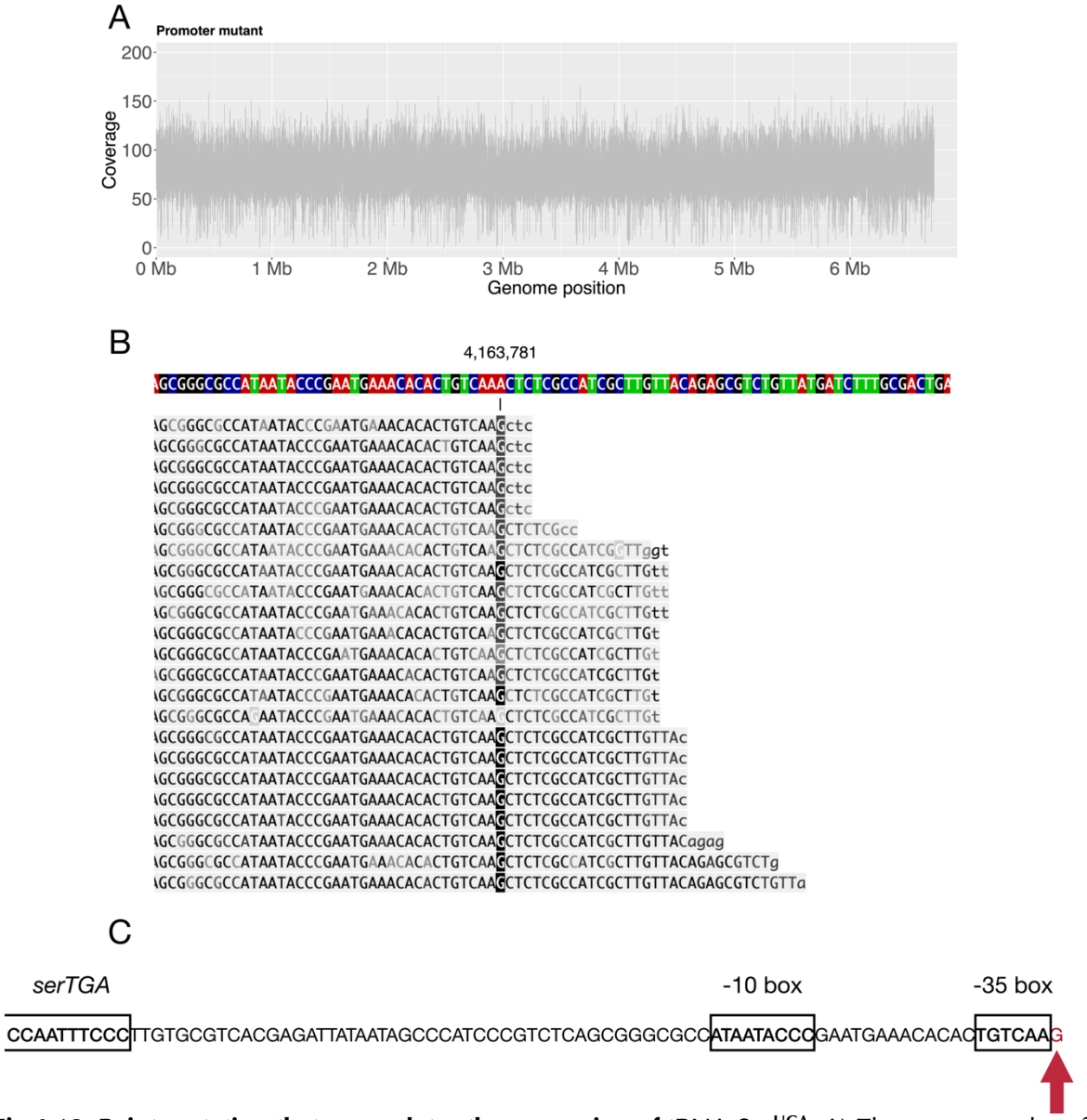


Fig 4.18. Point mutation that upregulates the expression of tRNA-Ser^{UGA}. A) The coverage plot of the mutant indicated there was no increase in coverage showing there was no duplicated region. **B)** The mutant had an A4163781G point mutation. **C)** The point mutation is just outside the -35 box of the tRNA-Ser^{UGA}.

4.7.2.2 Duplication containing the compensatory tRNA gene

The duplication fragments were predicted using *breseq* and bioinformatically confirmed by blasting the duplication junction reads into sequence reads. The key fragments were also

confirmed by amplifying the duplication junction by PCR (Fig 4.22). The junction sizes are summarised in Fig 4.20. The duplication fragments were generally getting smaller over the evolutionary time (Fig 4.19; 4.20). It is noticeable that mutants with large duplication fragments (hence, unstable) still presented in the 100-day population. This indicates the population still has a large flexibility to regain a new mutation by losing its large-scale duplication.

Among the 72 isolates sequenced, the smallest fragment identified from Δ*serCGA* lines is 830 bp and that from Δ*EGEG* lines is 236 bp (Fig 4.20; 4.21). Both fragments only fully duplicated the compensatory tRNA gene, indicating the mutants gained only one full copy of the tRNA gene (Fig 4.21). Considering that many bacterial tRNA genes exist in multiple copies within the bacterial genome, the within-genome duplication event is a mechanism by which bacteria can gain an extra, identical copy of a tRNA gene.

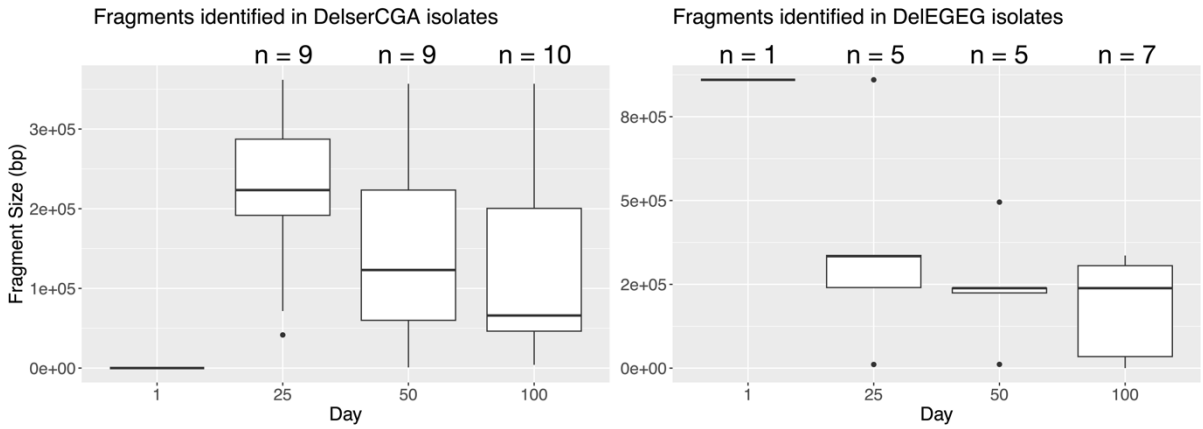


Fig 4.19. Decreasing in fragment sizes over the evolutionary experiment from isolate sequences. The left figure shows the identified fragment from isolates from Δ*serCGA* line and the right shows those from Δ*EGEG* lines. Both showed the mean value of the fragment sizes from each sampling days were decreasing in the evolving lines. Meanwhile, duplication mutants with large fragments were still present in both lines. This led to no statistical differences when comparing between days. Figures were drawn using RStudio.

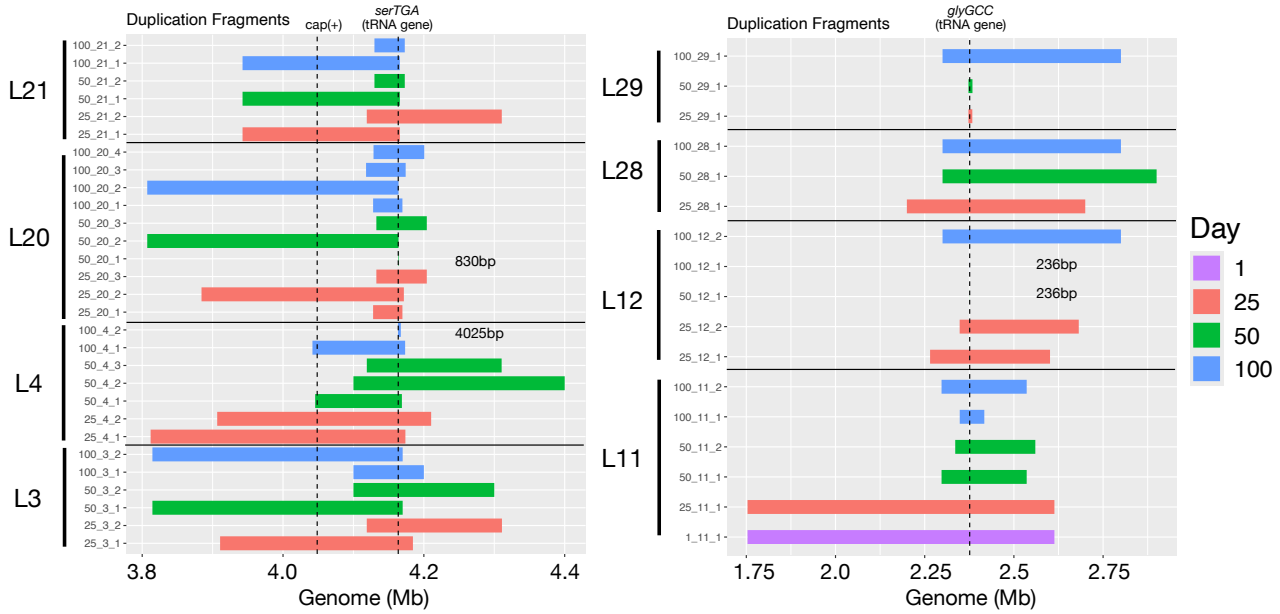


Fig 4.20. Duplication fragments identified from the isolate sequences. The left panel shows mutants from Δ*serCGA* lines and the right panel shows those from Δ*EGEG* lines. Different colours indicated sampling time points from the evolutionary experiment. Fragments in the chart were ordered by sampling lines. Small fragments were found starting from day 50 in the evolution experiment. The smallest identified was a 236 bp fragment from one of the Δ*EGEG* line.

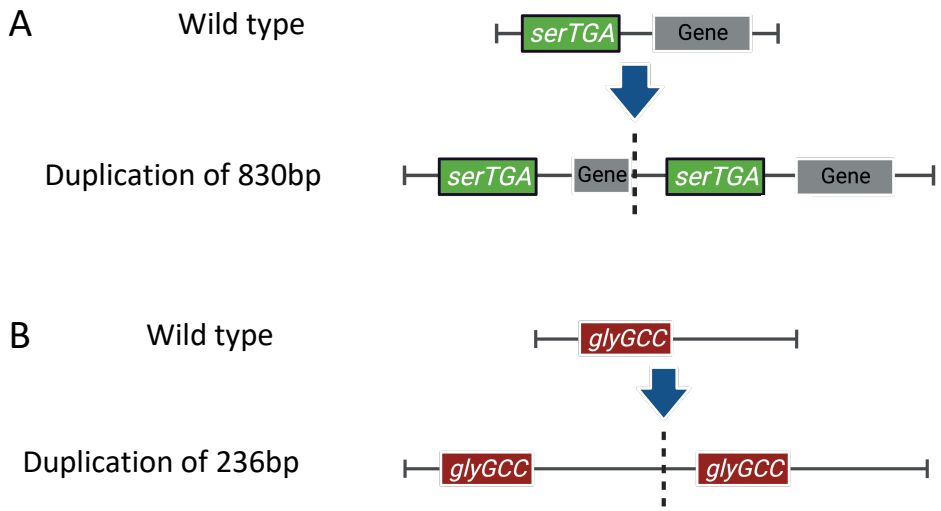


Fig 4.21. The smallest identified fragments from the isolate sequences. (A) 830 bp fragment identified from one of the Δ*serCGA* lines. It only duplicated *serTGA* tRNA gene fully and partially another gene (*pflu3767*). (B) 236 bp fragment identified from one of the Δ*EGEG* lines. It only duplicated the tRNA gene *glyGCC*, which gives an extra copy of the tRNA gene to the mutant.

The primers for the duplication junctions were designed to detect the duplication mutants. The primer was used for a population sample to analyse the frequency of each mutant in the

population. Two junction PCRs in Δ*serCGA* lines were performed in their sampled lines (Fig 4.22). In one of the lines, a 4025 bp duplication mutant rose before day 35 and increased in frequency (Fig 4.23). In another line, an 830 bp duplication mutant started as early as day 21 and increased in frequency until day 56-70 (Fig 4.23). Interestingly, the frequency of this mutant reduced in the population after day 70. The size of the duplication fragment of this mutant is very small (~0.012% of the genome). It only duplicates the compensatory tRNA gene fully (Fig 4.22). However, the population PCR indicated there are mutants with higher fitness presented in the population.

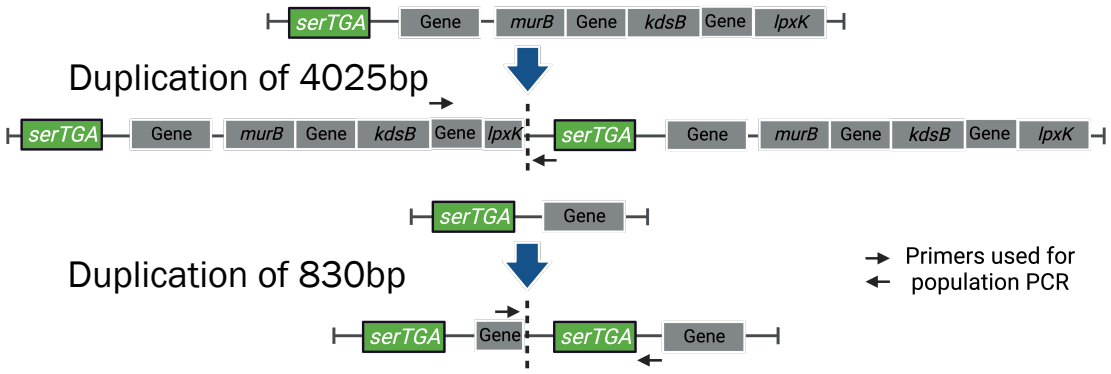


Fig 4.22. Junction PCR designed to detect the duplication mutants. The top figure shows the duplication of 4025 bp. The bottom figure shows the duplication of 830 bp. These two duplication mutants were identified in different lines.

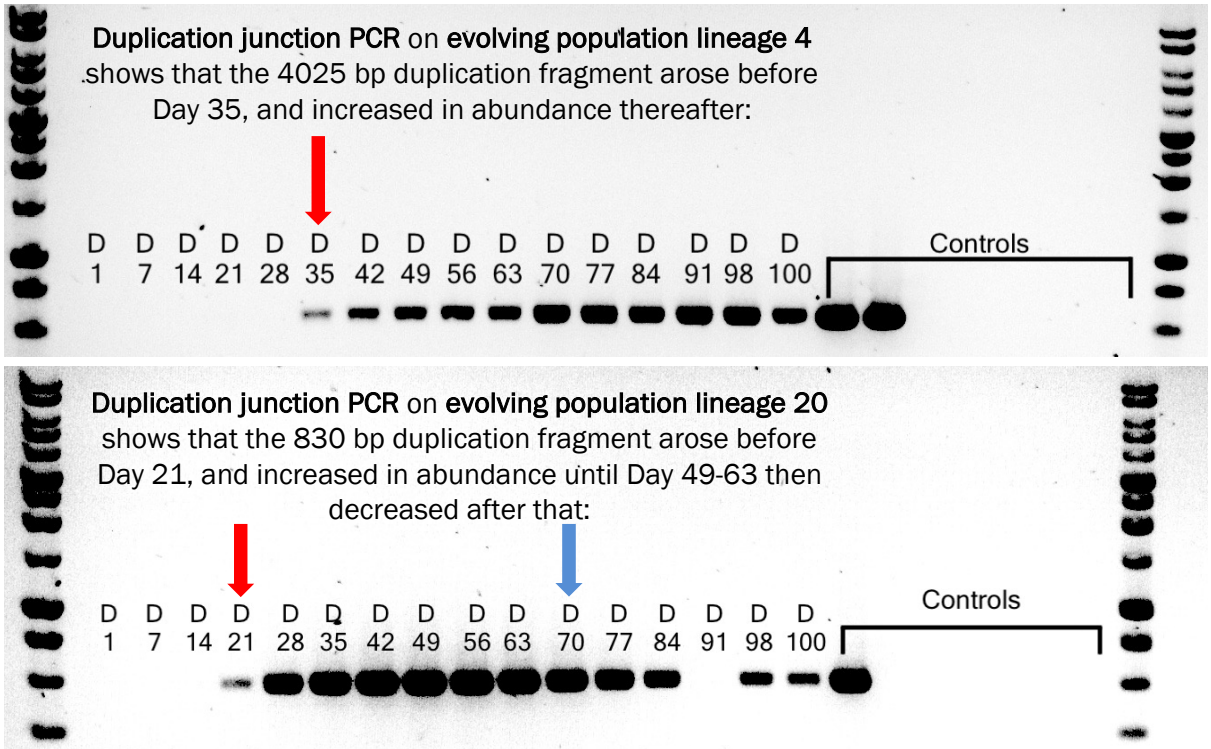


Fig 4.23. Population PCR detecting the change in relative frequency in the line. The top gel shows the 4025 bp duplication mutant started rising as early as day 35 and increased in the population afterwards. The bottom gel shows that the 830 bp duplication mutant arose as early as day 21 and decreased after day 70, indicating some mutants exist with a higher fitness than that carrying the 830 bp duplication mutant.

Other than duplication mutants, there is a SNP promoter mutant that persists in the evolution experiment until the end for ~750 generations (in one line only). However, from isolate and population sequences, this mutant did not overtake the duplication mutants in the population, even if it is stable. This was revealed by population sequencing, which is presented in the next chapter.

Also, a transposon cut-and-paste mutation was identified in two of the isolates. The transposon element is inserted between exactly the two duplication segments (Fig 4.24). The effects of the transposon on the stability of the duplication fragment were determined. It is interesting to check if the transposon can stabilise the duplication, and/or whether loss of the duplication fragment leads to stabilisation of a new transposon copy. The stability test showed mutation with a transposon cut-and-paste is not stable (Fig 4.25). The reversion rate is lower than other mutants very likely only because their fragment sizes are smaller.

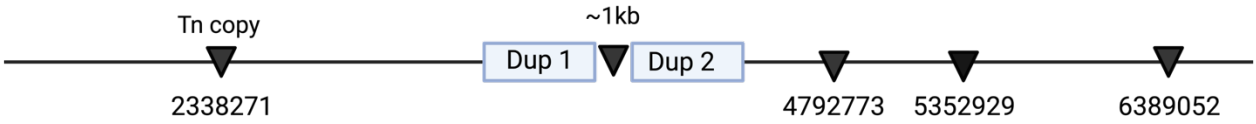


Fig 4.24. A mutant with a transposon cut-and-paste between the two duplication segments. This could be a new mechanism to stabilise the duplication mutation.

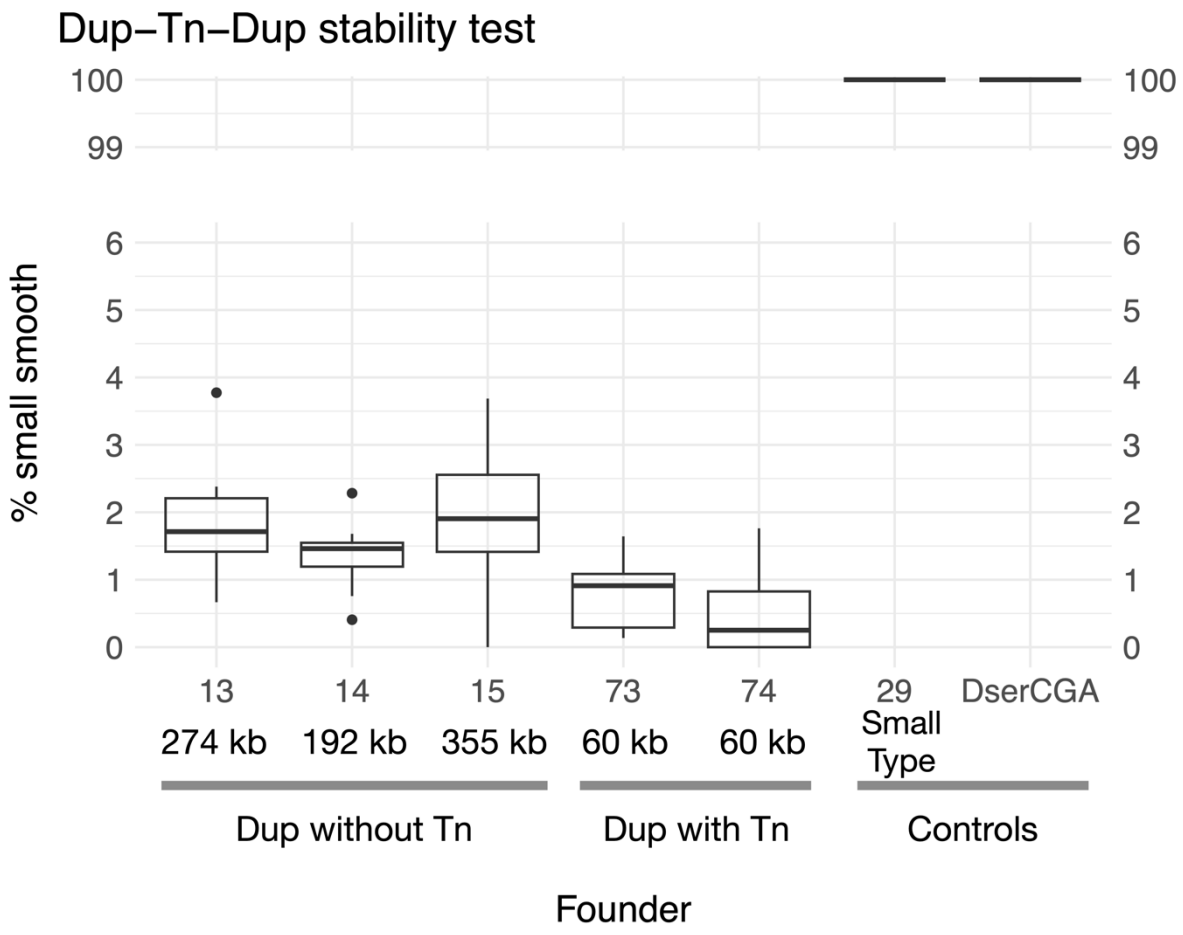


Fig 4.25. Stability test on the transposon mutant with other duplication mutants. The compensated colonies were picked to grow overnight. After dilution-plating, the small colonies were counted. Small indicated a growth defect without compensation. Thus, it lost its mutation. The duplication mutants with transposon inserted between two fragments is not stabilized.

This transposon cut-and-paste event is a very rare event in SBW25 evolution. To further investigate this, a new project has recently been established in collaboration with Dr Frederic Bertels and Dr Lavisha Parab (MPI Plön). They have isolated an *E. coli* C-derived mutant with a transposon inserted at the site of a large (~60 kb) deletion event [90]. We hypothesise that the transposon repaired the broken DNA strands during unequal cross over recombination events.

4.8 Stability of duplication mutants

The instability of duplication mutations was noticed in previous studies in Δ*EGEG* compensated mutants [57]. This is also a known nature of gene duplication or amplification [43]. 52 of the 72 sequenced isolates contained duplication mutations, whereas 34 of these duplications were unique (some mutants contained the same duplication fragment). In the stability test, 17 mutants were selected across four lines for the stability test (Fig 4.26A). When testing the dataset using linear regression to check the correlation between fragment size and stability. It shows a strong positive correlation (r >0.88) between fragment sizes and the instability (Fig 4.26B). There was one outliner as it is a large fragment. Likely rose from homology DNA regions. There was another linear regression test excluding this mutant (Fig 4.26C). A strong correlation (r > 0.81) is also observed between fragment sizes and the instability. The stability test suggested the larger the duplication fragment, the more unstable it is. Vice versa.

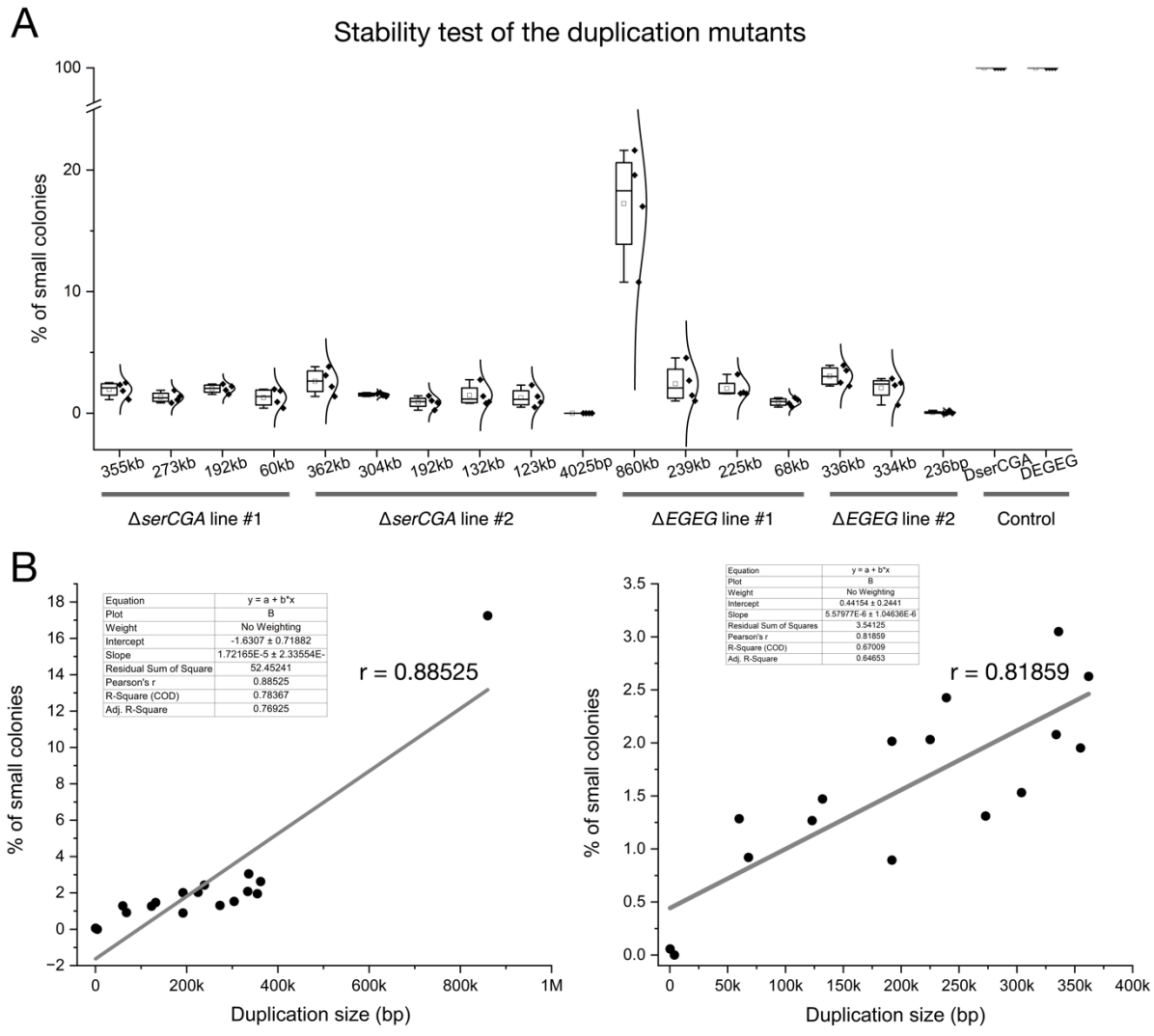


Fig 4.26. Stability test of the selected duplication mutants. A single colony of the mutant was cultured in liquid media overnight. After dilution plated the overnight culture on agar, the small phenotypes indicated the loss of their duplication mutations, showing the instability of the mutation. **A)** Isolates were selected from two lines from each of the founders. The 860 kb fragment showed a high instability. **B)** Linear regression between duplication size and the loss rate. (Left) All data were fit. A positive correlation was observed (r = 0.88525) between the duplication size and loss rate. (Right) Duplication can occur or be lost at a high rate in regions with DNA homology. Those in regions with little or no DNA homology have lower occurrence and reserving rates. The 860 kb fragment has a higher loss rate indicating other duplications were likely duplicated in the later mechanisms. To further compare these duplications, the 860 kb fragment was not included in the bottom right figure. A positive correlation was observed (r = 0.81859) between the duplication size and loss rate. Figures were tested and generated using OriginLab with labelling using OmniGraffle.

4.9 Discussion

To investigate the evolutionary fate of the unstable large-scale duplications, this study repeated and extended the evolution experiment of the wild-type and two engineered SBW25 strains (Δ*serCGA* and Δ*EGEG*). A previous study on Δ*serCGA* showed how large-scale duplication compensated the translational demands followed by the deep tRNA pool sequencing [14]. For the Δ*EGEG* study, large duplication was also found to be a major compensatory mechanism for the translational demands. The follow-up study discussed in Chapter III conducted the first stability test in both founder lines. It suggested the large-scale duplication are unstable and can reverse to their parent genotype.

However, in the previous two studies, it is remained unknown that how these duplication fragments and extra copies of tRNA gene would become of after a longer evolutionary time. Therefore, this PhD project restarted and extended the evolution experiment to investigate the evolutionary fate of these adaptive but unstable duplication mutations.

4.9.1 The advantages of sequencing isolates before performing population sequencing

In this study, after 100 days of transferring, phenotypes were first qualified and quantified before sampling for isolate whole-genome sequencing. Then, isolates were chosen based on the phenotypes. There are, in general, three advantages to using this approach: 1) diversifying the

nature of SBW25; 2) providing information on mutations contained in the population sequencing data; 3) identifying mutants containing multiple mutations.

1) The sampling of isolates was based on the diversifying nature of SBW25. Several studies have found that SBW25 diversify rapidly into niche-specific genotypes, and many of them appear in different phenotypes [91, 92]. Sampling isolates based on phenotypes could extend the understanding of mutation existing in the population.

2) The analysis of isolate sequencing is helpful to assist in understanding population data. The data is particularly helpful in identifying duplication fragments, because *breseq* actually cannot always predict duplication as realiably as it can predict SNPs.

3) A single mutant could contain multiple mutations. These mutations can be a diversifying mutation (e.g. WS) or potentially two mutations with compensatory effects on the tRNA gene loss. For example, in this study, a duplication mutant produced the colanic acid-like phenotype while also having WS phenotype and mutations in its genome. Currently, there is no analytical method to identify mutations from the respective mutant in intraspecies population sequences. Also, the sampled isolates could lead to further investigation. For example, the stability, cellulose-producing ability, swarming ability, or mutations of interest could be tested depending on the need. In this study, 72 isolates were sampled across all the lines. The duplication mutants were successfully predicted with *breseq*. The key fragments were further confirmed with PCR. Mutations in *fli/fleQ*, *wsp* and *barA/uvrY* were found in all lines. There were follow-up experiments characterising these mutants.

4.9.2 The progressively smaller duplication sizes in the population

The 72 isolate sequences revealed duplication mutants presented in the population from the time when compensatory mutants appeared until the end of the serial transfer evolution experiment. This may indicate duplication is a generally high fitness mutation for the strain to compensate the translational inefficiency. There was a SNP promoter mutant identified, which increased the expression of the compensatory tRNA gene. The SNP mutant is much more stable compared with the large-scale duplication. The preliminary results showed they increased relative abundance in the population in the early stage of the experiment (before a fitter duplication mutant appeared). However, this SNP mutant was overtaken by the duplication mutant with

small fragment sizes. This suggested that even if a promoter mutant increased the expression of the tRNA gene, it is less effective than containing an extra copy of the gene. The reason they can present in the population at a frequency of up to 40% could be only because of their stability and the fact that duplication mutants appeared at a high frequency. The dynamics of high/low frequency can be modelled mathematically. The results will be shown in the next chapter. Furthermore, analysis of the duplication mutants revealed the evolving populations contain progressively smaller duplication fragments throughout the experiment. The observed duplication fragments are as small as 236 bp, which only duplicates the compensatory gene (and its promoter). This compensated mutant gained an identical copy of the tRNA gene. Across the tree of life, organisms do not contain the same composition or copy numbers of tRNA genes. These tRNA genes can carry the same anticodon and encode the same amino acid but their nucleotide sequences can differ. This experiment provides a compelling example of the strain to gain an extra identical copy of the tRNA gene through evolution.

4.9.3 Is a mobile genetic element repairing the breaking DNA strands?

Transposons are one type of transposable elements (another is the retrotransposons) [93]. Transposable elements (TEs) are DNA sequences that have the ability to change their position within a genome [94]. Transposons and transposon sequences can contribute to genome plasticity [93]. TEs at distant genomic positions could provide homologous regions resulting in large-scale deletions, duplications, and inversions through recombination [94]. Also, transposons can create gene insertions, gene fusions, and DNA repair machinery [94, 95]. All of these could further result in genome rearrangements [94].

Most of the compensated mutants sampled in this study contained tandem duplications. One of those has a transposon cut-and-paste event exactly between the two fragments (hence, they are no longer strictly a tandem duplication). There was no hint to suggest that the duplication was initiated by the transposon. However, it is possible that the transposon repairs the damaged DNA strands. This could also explain the mutant found in the study from Dr Lavisha Parab and Dr Frederic Bertels [90], where the transposon was pasted into a deletion event (Dr Frederic Bertels, personal communication). A study reported in 2008 using *E. coli* found transposition events occur broadly around the point of the DNA break, suggesting DNA double-strand breaks in the chromosome stimulate transposition in transposon Tn7 [96]. As far as I am aware, a

transposition event fixing a DNA break – or indeed, any transposon movement event - has not been directly observed in SBW25 before.

4.9.4 Do P. fluorescens *SBW25 mutants overproducing cellulose have a further application?*

There were several phenotypes identified. They were further characterised by Congo Red assay, biofilm assay or mat formation testing. The appearance of wrinkly spreaders has been reported in many independent studies. They overproduce cellulose to form mats. In the microbial ecology, they colonise this niche to dominate oxygen access. This was discussed in section 4.7.1.2.

Not many bacterial species can synthesise cellulose. *Pseudomonas* is a bacterial genus frequently reported as a cellulose producer [97]. These bacterial celluloses have a wide range of applications from food-producing, personal care, household chemicals, water purification, and composite resin to medical purposes [97-101]. Bacterial cellulose has a high potential as an alternative food source [102, 103]. For example, kombucha uses the Scoby, a symbiotic community of bacteria and yeast, to brew the drink [104]. During the fermentation, layers of cellulose would be synthesised on the Scoby as a by-product. The Scoby is widely used as an alternative food source. In other industrial aspects, bacterial cellulose could be an alternative to petroleum-based polymers [105]. Bacterial synthesised polymers have the potential to overcome the difficulties of polymer-based nanocomposites that have incompatibility of structural strength and functionality and unsustainability of the preparation processes [105].

The mechanism in SBW25 to have wrinkly spreader appearing could raise the research interests in microbial engineering of bacterial communities. Future research could focus on improving the production rate of cellulose and reducing the cost by engineering them to brew using cheaper culture media.

4.9.5 Appearance of Wrinkly Spreaders in P. fluorescens *SBW25*

The appearance of wrinkly spreaders (WS) in this evolution experiment (in a shaken condition) is not surprising. WS appear through random mutation and have a significant fitness advantage over the founder strain. WS can present in the population for up to 30-50% in static conditions after 5 days of incubation. In shaking conditions, they can also represent ~10% of the population [36].

WS increased cellulose production, which facilitates the mat formation at the air-liquid interface. Bacteria in the niche have better access to oxygen to improve in growth. These mat formers could outcompete the non-mat former below the mat in the anoxic region. The bacterium SBW25 was isolated from the leaf of a sugar beet [106]. This adaptive radiation of the bacteria enables a successful colonisation of the soil, crop plants or weeds. The classic Red Queen hypothesis is one of the explanations in previous studies to suggest the diversification and adaptation of SBW25, where species must continually evolve new adaptations in response to evolutionary changes to avoid extinction [36].

It remains unknown why wrinkly spreaders form in a shaking laboratory condition. Forming mats need to overproduce cellulose and extracellular polymeric substances (EPS) to maintain the structure. One possible reason is that these WS cells were in a niche different from the smooth type. They initiated by colonising the contact surface between the well of the culturing tube and the air-liquid interface. Unless there is a mutation in the *wss* system, led to the strain lose the ability to mutate to wrinkly spreader phenotypes, wrinkly spreader could appear in the shaking condition. Another reason could be related to nutrient and resource availability. Previous studies reported in SBW25 that diversification and final population sizes depend on nutrient levels [107]. WS had a lower appearing frequency with reduced levels and single nutrient sources (with less peptone and glycerol) [107].

The overproduction of cellulose and EPS in *P. fluorescens* SBW25 is costly. In bacterial interactions, the availability of rich nutrients supports the formation of bacterial mats. Hence, these mutants can dominate the access to oxygen. Future studies may explore whether the frequency of wrinkly spreader (WS) appearing is influenced by the spatial dynamics at the air-liquid interface. To elucidate this, one approach could be assessing the impact of the edge area of the air-liquid interface on WS occurrence. Experimental designs may include comparing a larger culture vessel, such as a large glass flash or similar platform, with traditional culture tubes. The larger surface area in the former provides a higher interface for bacterial colonisation relative to the volume of the growth medium. It could shed light on whether the WS phenotypes appearing in *P. fluorescens* SBW25 is a phenomenon of niche occupying.

4.9.6 The importance of regular checking in an evolution experiment

Before the start of my PhD, we did not know that the two Δ*EGEG* lines from the previous study were contaminated. However, it was a good decision to restart the evolution experiment from the founder strains (rather than continuing from the previous transfer experiments). All the founder strains were also confirmed by PCR before starting the evolution experiment. This experience showed that it is vital to know how each sample was handled in previous studies and to perform a precise confirmation before any follow-up experiments. It is particularly important to check for contamination events in serial transfer evolution experiments when a long transfer time is needed. Also, contamination is always a concern in any microbial experiment. Establishing a proper contamination test for external contaminants and cross-contamination is important. This study discovered one cross-contamination on the first transfer to day 75. Multiple checks from glycerol stock and further backwards restarting are highly recommended in future studies. Because the contaminants might exist in the line at low amounts, which take a longer time to be observed in the population.

4.9.7 Follow-up experiments

The evolution experiment suggested that small duplication is the evolutionary fate of the unstable large-scale duplication. The stability test also showed the 236bp duplication is stable, with more than 2,300 colony counts but no loss of the duplication fragment. Hence, the loss rate is less than 0.04%. Some questions are left unknown. For example, how did this fragment appear throughout the 100 days of evolutionary time? Are there alternative mutants that contributed largely to the population but were overlooked during the isolate selections?

Also, the mutant with such a small duplication fragment appeared in a relatively short time. The hypothesis is that the system allows the founder to gain and lose the mutation. Hence, they have another chance to gain a new mutation. This is not because the large duplication fragments shrank in size directly.

To answer these questions, the next chapter covers mainly the results of two experiments: 1) population sequencing, and 2) sequencing of small colonies from overnight culturing of duplication mutant isolates.

Chapter V – Evolutionary Dynamics of the Evolving Populations

In this chapter, I carried out all the experiment, data analysis and visualising of the data from the above experiment except sequencing work. Sequencing work were done by Novogene (UK). Gunda Dechow-Seligmann also conducted genome DNA extraction for sequencing of small types.
The mathematical model was built by Florence Bansept and Arne Trauslen.

5.1 Introduction

5.1.1 From isolates to population sequencing, what was expected to be observed

In the previous chapter, the daily serial transfer evolution experiment was performed for 100 days (~750 generations). The phenotypes were identified and further characterised. Based on the phenotypes that appeared, isolates were selected for whole genome sequencing, where some lines were of particular interest. The computational predictions suggested there were mainly two categories of mutation: 1) general adaption, where mutations were observed across all the lines; and 2) adaptation compensating translation efficiency. Some key mutations were checked by PCR or corresponding assays. The sampled isolates suggested that large-scale unstable duplication still existed in the evolving population, and the population started to be dominated by duplication mutants with much smaller duplication fragment sizes (830 bp in Δ*serCGA* and 236 bp in Δ*EGEG* line).

There were some questions that remained in this project. For example, what are the dynamics of these mutants? Are there any mutants mostly represented in the population but were missed during the isolate sampling? Mechanistically, duplication fragments with smaller sizes are more stable. However, to what extent are these small fragments actually stable?

When tracking a mutation through a population, there are four hypothetical patterns of evolutionary dynamics (Fig. 5.1):

1) The mutation rose and fell afterwards;
2) The mutation rose and remained in the population;
3) The mutation rose and still increased in relative abundance in the population until the end of the experiment;

4) The mutation rose in a very low relative abundance, remained at a very low level.

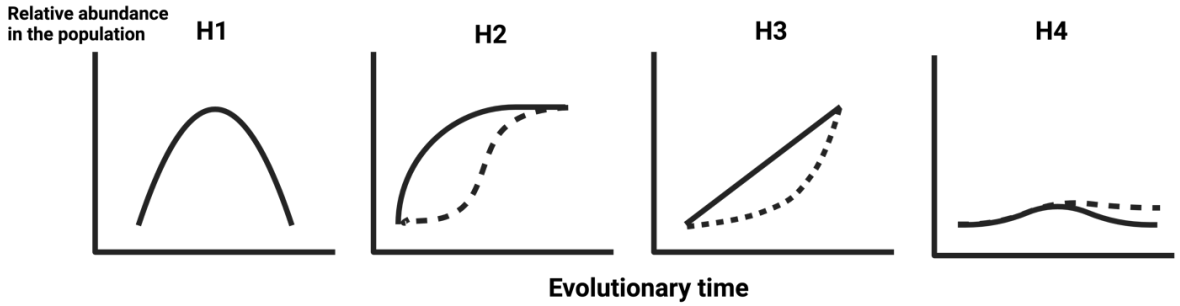


Fig 5.1. Hypothesis of the dynamics of the mutations in the population. H1: The mutation rose and fell. H2: The mutation rose and remained. H3: The mutation still increases in relative abundance in the population. H4: The mutation rose at a very low relative abundance and remained very low or fell below detection limits. The dotted lines indicated a different scenario under the same hypothesis. Figures were drawn using Biorender.com.

We know that large-scale duplications appear at a high rate, and are lost at a high rate. The SNP promoter mutant sampled is stable but did not fix in the population. Could we use mathematical modelling to simulate the conditions in which a mutation with low mutation frequency could take over the population? These questions were answered in this chapter with population data, stability test of the duplication mutations and mathematical modeling. The mathematical modelling was developed in collaboration with Dr Florence Bansept (CENTURI, France and MPI Plön) and Prof Arne Trauslen (MPI Plön).

Previous stability tests of the duplication mutations suggested the mutants lost their duplication fragment and reversed to the original type. This is the main mechanism by which the bacteria can gain and lose the mutation (and then gain a new mutation). In this study, after 100 days of serial transfer, the duplication mutant with 236 bp rose in the population, indicating a change in tRNA gene copy without changing the copy number of other genes. However, it is possible that the smaller fragment arise by losing part of the large duplication fragments. In order to distinguish these hypotheses, small types observed in the stability test will be isolated and their genome sequenced.

5.2 Aims

Some mutants were identified and characterised in the previous chapter. This chapter focuses on how these mutants appeared (and lost) over evolutionary time. The dynamics of mutations will be presented in this chapter to investigate each mutation's rise, maintaining or fall.
The sequencing of small types in stability tests is particularly important to confirm the mechanism in the system, where the unstable duplication lost the extra fragment and reversed to the parent genotype without a trace.

After analysing the evolutionary dynamics, the interaction of high and low mutation frequencies in a population will be simulated using mathematical modelling. This will be useful to simulate how rare mutations could overtake the population and whether the high frequency but unstable mutation will be able to remain in the population over a longer evolutionary time.

5.3 Sampling criteria of the evolving lines at selected time points for population sequencing

From the isolate sequencing results, five lines (one wild-type, two Δ*serCGA*, and two Δ*EGEG* lines) were selected for population sequencing. The selection criteria were based on the diversity of mutations identified. Replicate line #2 from Δ*serCGA* is the line containing the SNP promoter mutation and duplication mutants. Most of the day 100 compensated mutants from replicate line #3 from Δ*serCGA* had different phenotypes and they were mostly smooth large colonies. This may suggest the line contains different mutations of different fragment sizes or covers different regions. Δ*EGEG* replicate line #2 contains the smallest duplication fragment sampled in this study, with only 236 bp. Replicate line #1 from Δ*EGEG* founder compensated from the first day of the evolution experiment. Therefore, these four lines were selected for population sequencing.

The four selected lines were sampled at mainly two phases: two early-stage and four later-stage of the evolution experiment. The early stage covered the time when compensated mutants rose in the population based on the colony counting data, including the phases where compensation was still spreading through the population, and the initial point where almost all of the population was compensated (see previous chapter). For most of the line, these stages were days 10 and 16. One of the lines from Δ*EGEG* founder showed earlier compensation (from day 1). Days 5 and 10 were sampled in this line. For the sampling at later-stages of the evolution experiment, days 25, 50, 75 and 100 were picked to give an even distribution of sampling time. Naturally, the wild-type line did not have such a compensatory phase. Therefore, only the four

later-stage timepoints were sampled for the wild-type line. The extracted gDNA was sequenced with at least 1,000-fold coverage using an Illumina Nova-seq platform (pair-ended, 150bp; Novogene UK).

5.4 Evolutionary dynamics of the evolving population

In the previous chapter, whole-genome sequencing of compensated isolates – and a limited amount of population sequencing – showed that most compensated strains contain duplication mutations. One direct way to examine these would be to align the sequenced reads to the reference genome (NC_012660) to check the coverage. This is particularly potent in population sequences as it provides the overview of duplication fragments with different sizes. The sequenced reads were aligned in Geneious Prime (v2023.2.1) using the Geneious mapper. The following sub-sections shows the overviews of each the population(s) from each founder (wild-type, Δ*serCGA*, Δ*EGEG*).

5.4.1 Line evolved from wild-type P. fluorescens *SBW25*

The wild-type lines provide important information on general mutations in the evolution experiment (i.e., those mutations unrelated to the compensation of the lost tRNA gene). The overview of the coverage plot showed no increase in coverage in the population, confirming no observed duplication mutation (Fig 5.2). This also suggests no contamination from compensated duplication mutants from Δ*serCGA* or Δ*EGEG*.

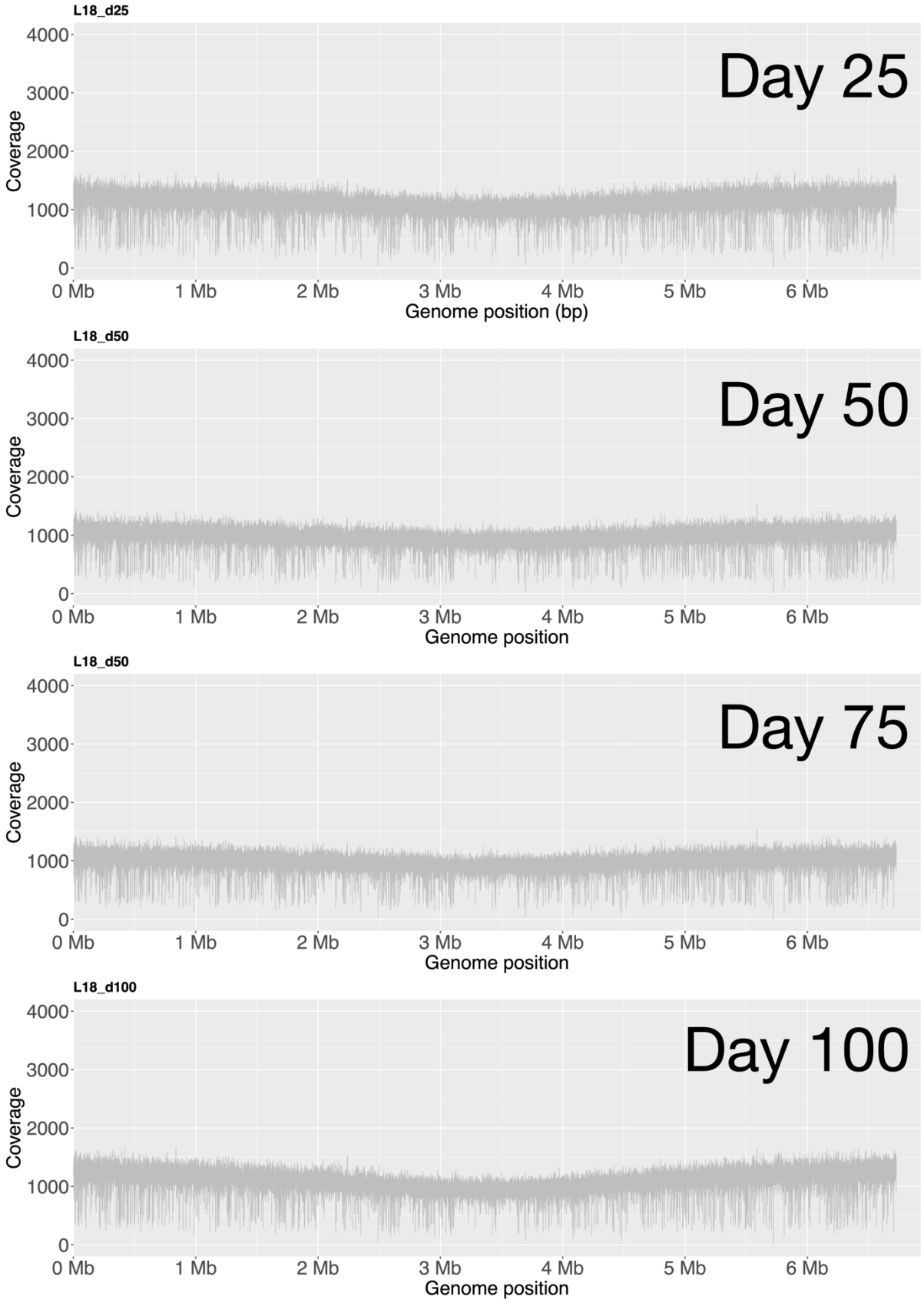


Fig 5.2. Coverage plots of the evolving wild-type line replicate #3. The sequenced reads were mapped to the reference genome (NC_012660). There are no obvious regions with high coverage observed, suggesting there a lack of duplication mutants in this line (as expected).

5.4.2 Lines evolved from ⊿serCGA founder

The coverage plots of the two selected Δ*serCGA* lines (replicate #2 and #3) are shown in Fig 5.3.

In Δ*serCGA* replicate line #2, during the compensation phase (days 10 and 16), the identified duplication fragments ranged from 104,468 bp to 380,791 bp. On day 100, the largest and smallest duplication fragments identified were 337,954 bp and 4,025 bp, respectively. The dynamics of all of the fragments were quantified computationally, as follows. Each set of population raw reads was independently mapped to each proposed duplication junction, and read numbers were normalised for mean genome coverage. Thus, the dynamics of particular fragments could be visualised (and, to some degree, quantified) through time/lines. (Some estimations were not reliable because they were repeated in the genome. These fragments' dynamics were not shown in the dynamics panel. Also, some fragments were identified through isolate sequencing. Among these, some were at a very low relative abundance in the population. They failed to be detected from population sequences.)

In Δ*serCGA* replicate line #2, on day 100, duplications with (relatively) large fragment sizes were still present in the population. One fragment of size 131,601 bp remained at around 25% on day 100. While the dynamics analysis showed the smallest fragment (4,025 bp) increased in the population. In Δ*serCGA* replicate line #2, from isolate sequencing, a promoter mutation was identified. From the phenotype analysis, this type gives rise to a smooth phenotype that was not the most dominant on day 100. A similar result was observed from the mutation prediction in the population sequencing (Fig 5.3).

For the promoter mutant identified in this line, it increased in the population and decreased at around day 25. Then dropped below detection limits (Fig 5.5). This suggested the stable promoter mutation is less fit and outcompeted by the duplication mutants

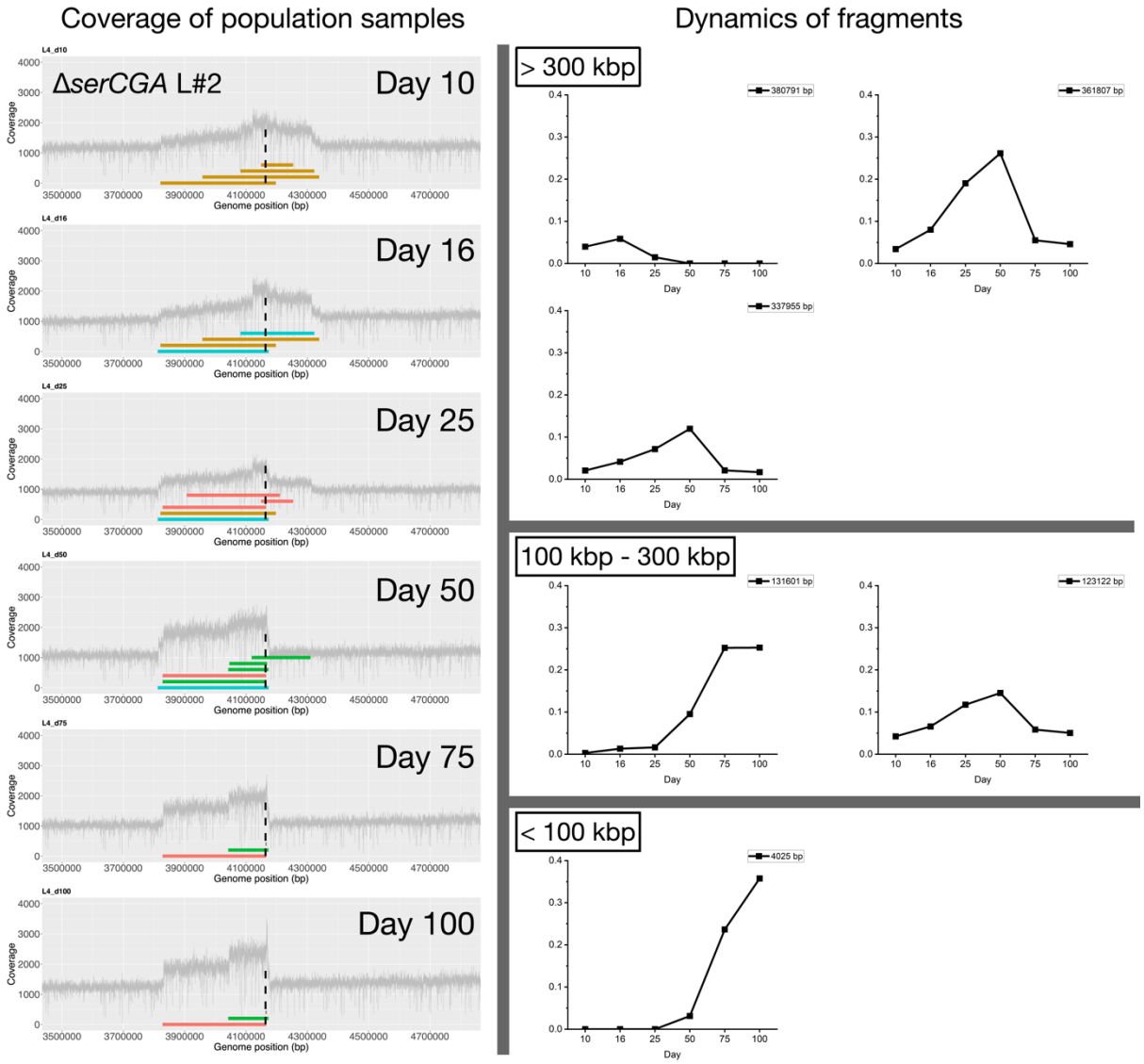


Fig 5.3. Population analysis of Δ*serCGA* replicate line #2. The left panel shows mean genome coverage using population sequences. The colour bars show the sizes of duplication fragments identified from isolates and population sequences. The colour coding indicates various discovery dates: deep yellow for day 10, cyan for day 16, pink for day 25, green for day 50, light brown for day 75, and blue for day 100. The dotted line shows the genome position of the compensatory tRNA gene. On the right panel, the computationally-estimated dynamics of the fragments are shown. The fragments were classified by size.

In Δ*serCGA* replicate line #3, the largest fragment identified arose before day 10, with a size of 1,046,974 bp. Among all the repeated experiments of Δ*serCGA* and Δ*EGEG* evolving lines, this is the largest duplication fragment that has been identified. The second largest duplication fragment identified in this experiment was 712,163 bp. Both the population coverage and dynamics analyses suggest that these two large fragments were out-competed by fitter mutants

before day 50. Indeed, the coverage plots from day 25 had a shorter region of high coverage. Hence, we hypothesise that smaller fragments arise and overtake in the population as the experiment progresses. On day 100, the coverage plot shows the fragments in the population were the shortest throughout the evolution experiment.

From the dynamics analysis, the two smallest identified fragments (56,225 bp and 830 bp) show, somewhat unexpectedly, decreases in relative abundance before day 100. Isolate sequencing from the previous chapter predicted the 830 bp duplication mutant also contains a mutation in *fliA*. The *fliA* mutation not only affects flagellar gene expression but also gives rise to changes in global transcription (including c-di-GMP expression). It seems probable that the effects of the *fliA* mutation lead the 830 bp duplication to be out-competed by other, larger duplication mutants (without other mutations).

Overall, the results indicate that duplication mutants with small fragment sizes are generally fitter than those with larger fragments. However, background mutations can also drastically change the trajectory of a duplication fragment.

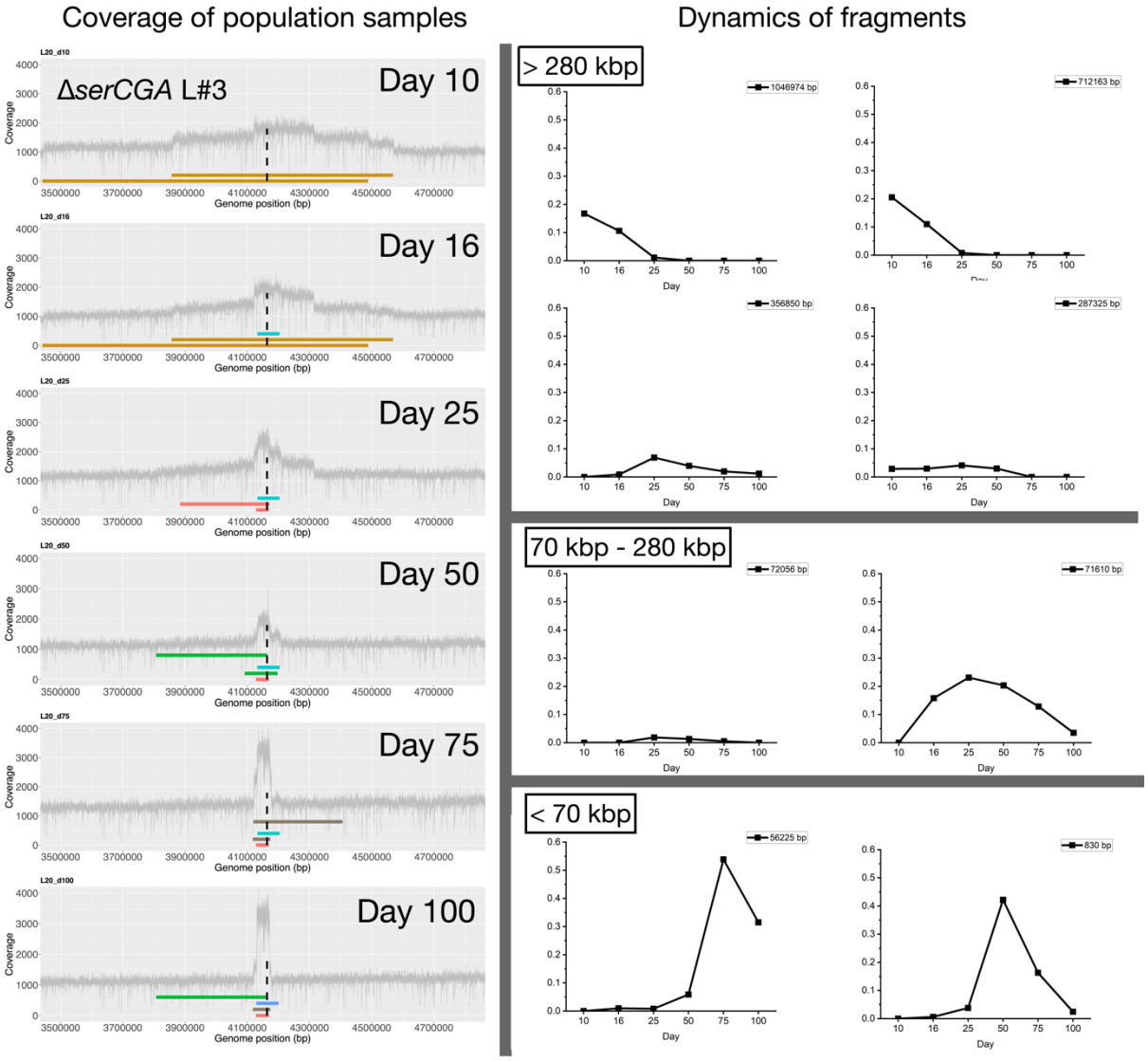


Fig 5.4. Population analysis of Δ*serCGA* replicate line #3. The left panel shows the coverage of the population sequences. Colour coding is the same as Fig 5.3. The right panel shows the dynamics of the fragments. All the identified fragments show a (final) decrease in abundance in the population, which suggests some fitter mutations or fragments are yet to be identified.

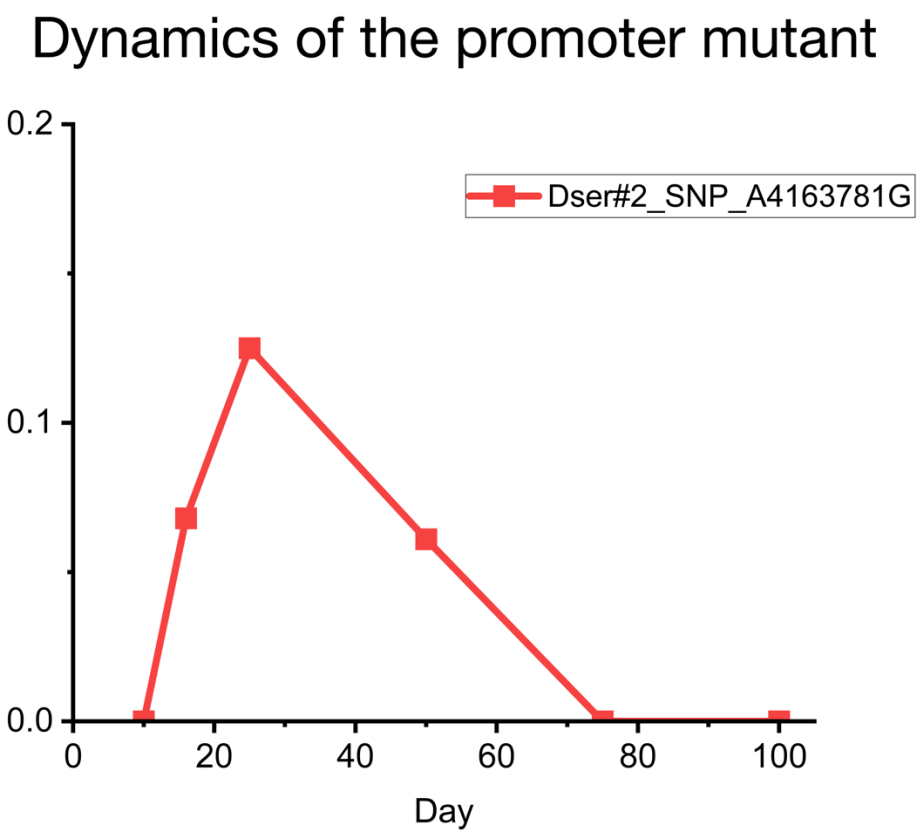


Fig 5.5. Dynamics of the tRNA-Ser^{UGA} promoter mutation. The mutant has a point mutation from A to G at 4163781 (NCBI ref: NC_012660). This mutant is expected to increase the expression of the compensated tRNA-Ser^{UGA}. The relative abundance of this mutation increased in the early phase of the evolution experiment. It decreased in relative abundance in the population before day 25.

5.4.3 Lines evolved from ΔEGEG founder

A similar general pattern to that seen in the evolving Δ*serCGA* lines was observed in the two evolving Δ*EGEG* lines: the identified fragment sizes decreased throughout the evolution experiment (Fig 5.6).

In Δ*EGEG* replicate line #1, compensated colonies were observed since day 1. Therefore, day 5 and day 10 were sampled for population sequencing for the early compensatory stage (instead of day 10 and day 16 in other lines). The largest identified fragment is 860,074 bp, which arose before day 5. In the early phase (day 5 and day 10), the smallest identified fragment was 238,328 bp. The largest 860,074 bp fragment was not included in the dynamics analysis because the duplication junction is highly repetitive in the genome (making the alignment process error-prone and difficult to interpret relative to other duplication fragments). However, the coverage

plot suggests that the duplication was very likely present in the population until at least day 25, as the fragment matches with the regions having higher coverage.

From day 50 onwards, the coverage plots suggest smaller fragments started overtaking the largest fragment in the population. The largest fragment from day 50 was 786,367 bp. However, the coverage showed that they had a low relative abundance in the population. The smallest identified fragment after day 50 was 68,115 bp. This fragment was able to be detected in the dynamics analysis. It increased to its highest point on day 50, and decreased thereafter. Interestingly, the coverage plot suggested the same results. From the isolate sequencing, this mutant contained mutations in *wspC*, *wspD* and *barA*/*gacS*. They also had a wrinkly spreader phenotype (as expected for *wspC*/*D* mutants). This suggested the wrinkly spreader phenotype has a lower fitness advantage than at least one other morphotype under the conditions of the evolution experiment.

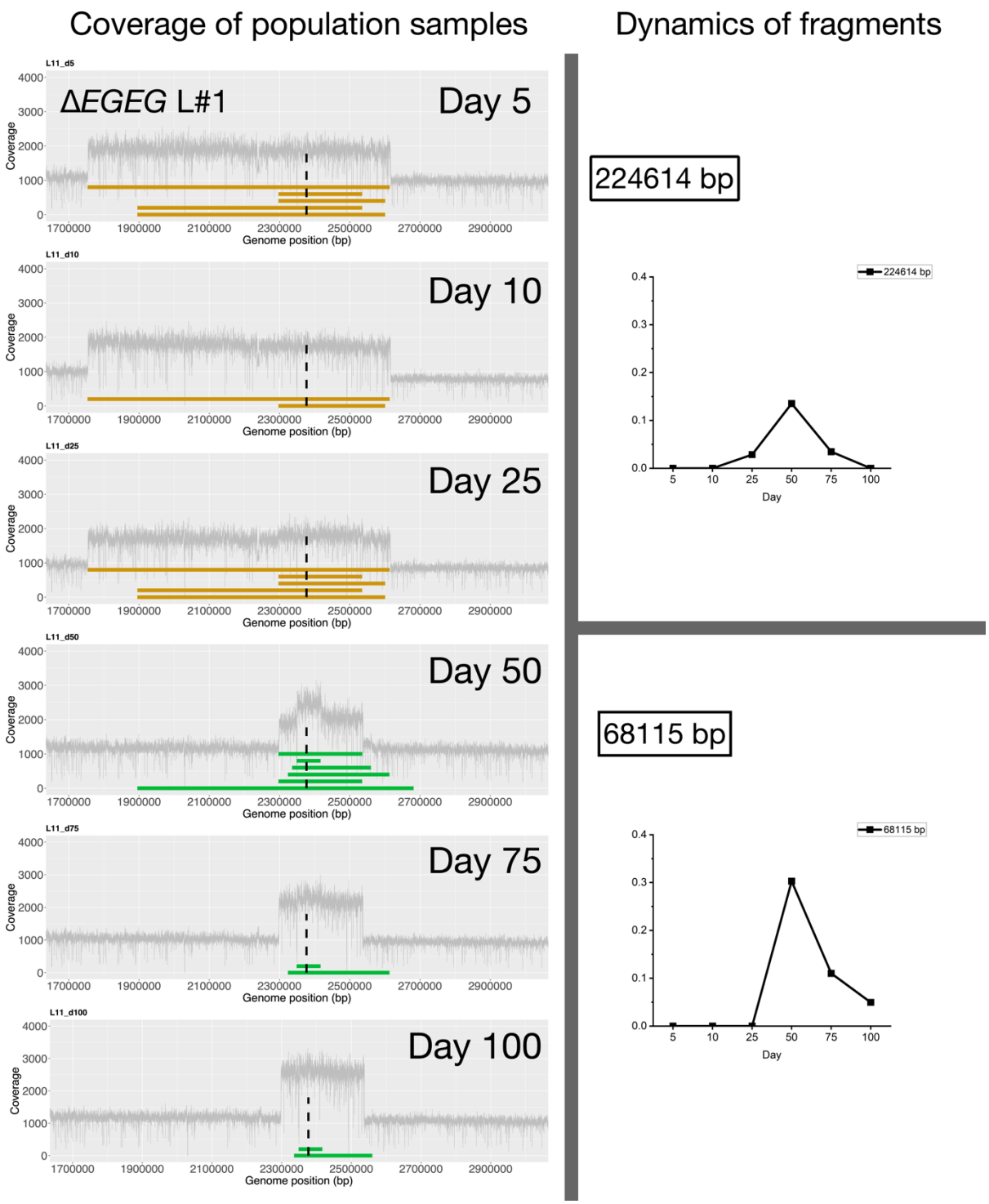


Fig 5.6. Population analysis of Δ*EGEG* replicate line #1. The left panel showed the coverage of the population sequences. Colour coding is the same as Fig 5.3. The right panel showed the dynamics of the fragments. In this line, due to the highly repetitive region, only two duplication fragments (224,614 bp and 68,115 bp) were able to run the dynamics analysis. Both fragments showed a decrease in abundance in the population after day 50, which suggests some fitter mutations or fragments were yet to be identified.

In Δ*EGEG* replicate line #2, the largest fragment was 879,978 bp in the early compensatory stage (Fig 5.7). This was identified on day 16, but does not exclude the possibility that they were already present on day 10. Indeed, the coverage plot suggests some duplications failed to be identified bioinformatically. One reasonably large fragment with a size of 336,320 bp was able to be tested in the dynamics analysis. It shows that they increased in the population until day 10 and decreased afterwards. The coverage plots suggest the fragments from day 25 were generally smaller compared with those on day 10 and day 16. On day 100, large duplication fragments did not present in a large amount in the population, presumably due to the rise in frequency of a 236 bp duplication. This fragment was also identified from the isolate whole-genome sequencing. From population sequencing, it appears that this 236-bp duplication first arose as early as day 10, and gradually increased in the population.

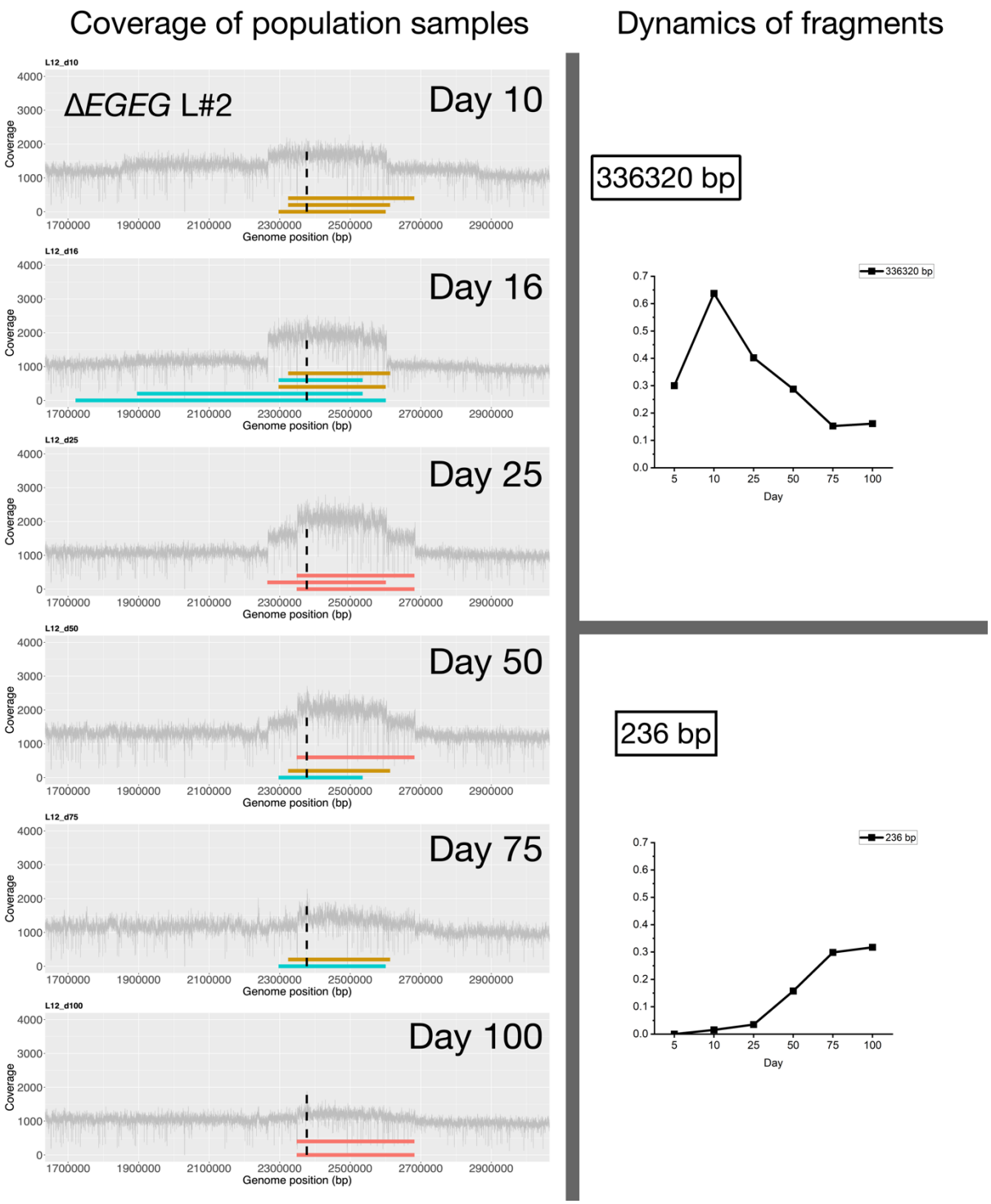


Fig 5.7. Population analysis of Δ*EGEG* replicate line #2. The left panel shows the coverage of the population sequences. Colour coding is the same as Fig 5.3. The right panel shows the dynamics of the fragments. In this line, due to the highly repetitive region, only two duplication fragments (336,320 bp and 236 bp) were able to be used for the dynamics analysis. The 336,320 bp fragment arose early. It peaked on day 10 and decreased afterwards. The 236 bp is the smallest duplication

fragment identified in this study. It arose as early as day 10 and increased until the end of the evolution experiment.

5.5 Sequencing of reversed genotypes

From the stability test in the last chapter, the small colonies were sampled for whole-genome sequencing. This is to confirm the mechanism of the duplication mutant when they lose the mutation. Hypothetically, from homologous recombination, these mutants are expected to have cleanly lost their duplication fragments and reverted to the parent genotype. The smaller duplication fragments identified are expected to occur via relatively rare mutational events, from these "new" founder genotypes, rather than through direct shrinking of larger fragments.

In the whole genome sequencing of the small types, all 17 showed scar-free loss of their duplication fragments. Results demonstrated that scar-free fragment loss occurs in the system. By this process, the scar-free founder strain is available for new mutations to independently arise.

5.6 General adaptions in the evolving populationd

In isolate sequencing from the previous chapter, there were some general mutations identified. These can be mainly classified into the following categories: 1) aflagellate related mutation; 2) wrinkly spreader related; 3) BarA/UvrY two component system related. The phenotype quantification showed the mutants related to BarA/UvrY two component system contributing a low amount in the population. Further, they were not able to be identified in the population sequence analysis using *breseq*. Hence, in this section, only the first two categories of mutations will be covered.

*5.6.1 Aflagellated mutations leading to loss of motility (*fleQ *or* fli*)*

The *fleQ* gene encodes a major flagellar gene regulator. Previous studies showed it positively regulates many flagellar genes [108]. These include genes involved in flagellar export (*flhA* and *fliLMNOPQ* operon), flagellar apparatus localisation and regulation (*flhF* and *fleN*), and structural components of the flagellar basal body and motor switch complex (*fliEFG* and *fleSR*) [78, 108]. Mutations in *fleQ* result in non-motile and aflagellated bacteria [78]. In the laboratory experiments with shaken condition, it is not a surprise for the bacteria to adapt into immotile to reserve energy from developing their flagellar.

From the population sequences, the frequency of motility-related mutations were extracted (Fig 5.8). *fleQ* mutations were identified in all the sequenced lines. Most of them showed increasing relative abundance in the population. In some line, the estimated frequency is almost as high as 70%. While the *fliA* showed decease from around day 50 to 75. This could be due to *fliA* is a flagellar sigma factor [109]. Interrupting the function of sigma factor may affect the fitness of cells too. Nevertheless, results showed mutation in *fleQ* provide a higher fitness advantage compared with those in *fliA* or *fliM*.

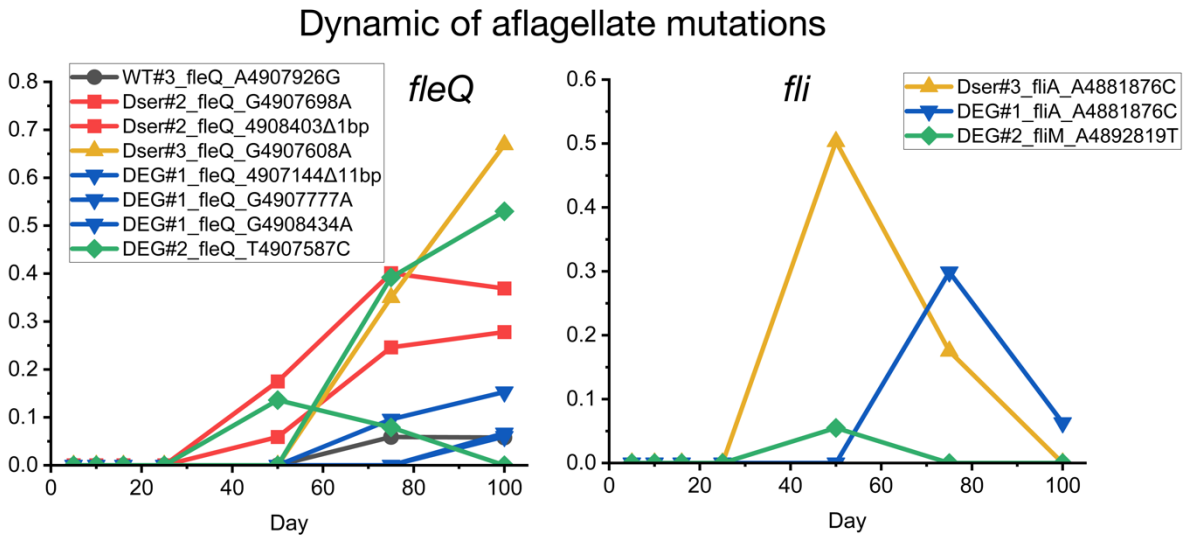


Fig 5.8. Dynamics of aflagellate mutations from population sequencing. Population sequences identified mutations in *fleQ*, *fliA* and *fliM*, which led to the aflagellate bacteria. The left figure shows mutations in *fleQ*. Most of the identified *fleQ* mutations increase in the population over the evolution experiment. In the right figure, all the identified *fli* mutations increase initially, but then decrease in relative abundance before the experiment ended.

*5.6.2 Wrinkly spreader related mutations (*wsp *and* wss*)*

Both *wsp* and *wss* mutations are wrinkly spreader related. In the previous chapter, these mutations were discussed. In short, a *wsp* mutation is likely to lead to the wrinkly spreader phenotype, and bacteria need the *wss* operon to form the wrinkly spreader phenotype. Mutations in *wss* are expected to cause loss of the wrinkly spreader phenotype. In this experiment, wrinkly spreader phenotypes were observed in all lines (wild-type, Δ*serCGA* and Δ*EGEG*).

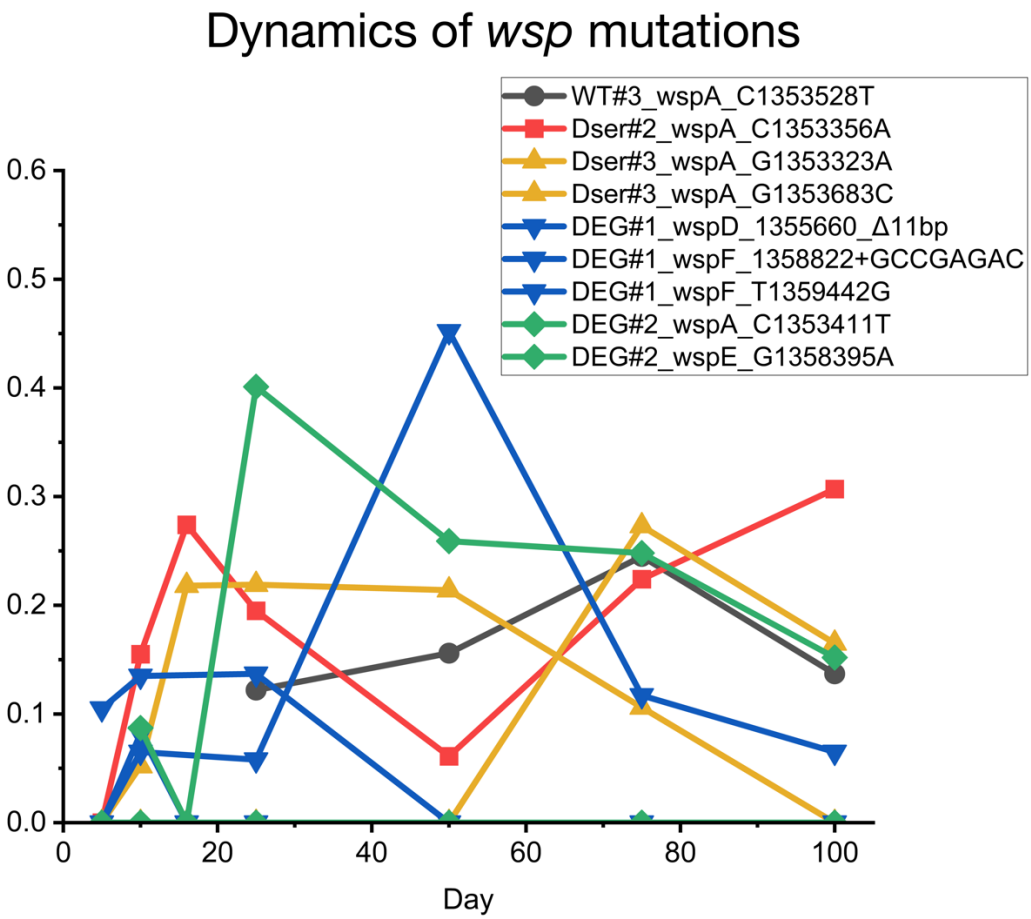


Fig 5.9. Dynamics of *wsp* mutations in the evolving lines. The mutants with *wsp* mutations are predicted to have wrinkly spreader phenotypes. Most of the identified mutations decreased before the end of the experiment.

5.7 Mathematical modelling of the dynamics of high and low frequency mutations

Mathematical modeling was done in collaboration with mathematicians Dr. Florence Bansept and Prof. Arne Traulsen (Theory Department, MPI Plön). They built the mathematical models, with my input on the biological components and data.

In this study, large-scale duplication mutants were formed in the early phase of the evolution experiment. They appeared, and were lost, at high rates. The instability of these mutants leads to unusual evolutionary dynamics in these evolving populations. Some rare and stable mutations were observed in the later phase of the experiment including a SNP and duplication mutants with very small fragments (as small as 236 bp). The dynamics of high-frequency but unstable mutants and rare but stable mutants in the evolving population are interesting. For example,

could the high-frequency but unstable mutants remain in the population when the rare but stable mutant is dominant? For the SNP promoter mutant, both quantitative phenotype analysis and population sequencing showed they could not overtake the population even if they were stable. The next question is *under what conditions will each type of mutation dominate the evolving population?*

Another situation is that if the mutation rate of stable mutants increases – thus increasing the chances of their appearance earlier in the experiment, when duplication mutants have not yet swept - would they be able to out-compete the duplication mutants and dominate the population more easily?

$$\frac{\partial R}{\partial t} = (r_R(1-\tilde{\mu}_{R\to D}-\tilde{\mu}_{R\to S})R + r_D\tilde{\mu}_{D\to R}D + r_S\tilde{\mu}_{S\to R}S)\left(1-\frac{D+R+S}{K}\right) \quad (1)$$

$$\frac{\partial D}{\partial t} = (r_D(1-\tilde{\mu}_{D\to R})D + r_R\tilde{\mu}_{R\to D}R)\left(1-\frac{D+R+S}{K}\right) \quad (2)$$

$$\frac{\partial S}{\partial t} = (r_S(1-\tilde{\mu}_{S\to R})S + r_R\tilde{\mu}_{R\to S}R)\left(1-\frac{D+R+S}{K}\right) \quad (3)$$

r = replication rate
D = population of duplication
R = population of original type
S = SNP mutant
μ = mutation rate

To answer these questions, it is extremely time-consuming to construct all strain combinations and perform the required assays. To this end, mathematical modeling can help to simulate different scenarios of evolutionary dynamics. For example, we could simulate how the abundance of founder strains and mutants may change with varying rates of growth or mutation rates. To achieve this, we first established the mathematical model. The first part of the equations simulates each type of mutation's replication and mutation rate. The second part is the simulation of the saturation of the bacteria due to limited growth in the culture. Equations (1)-(3) represent the dynamics of each strain [(1) R is the unevolved strain, (2) D is the high-frequency mutation and (3) S is the rare mutation.] The 1% carry-over condition was also simulated as the transferring bottleneck (Fig 5.10).

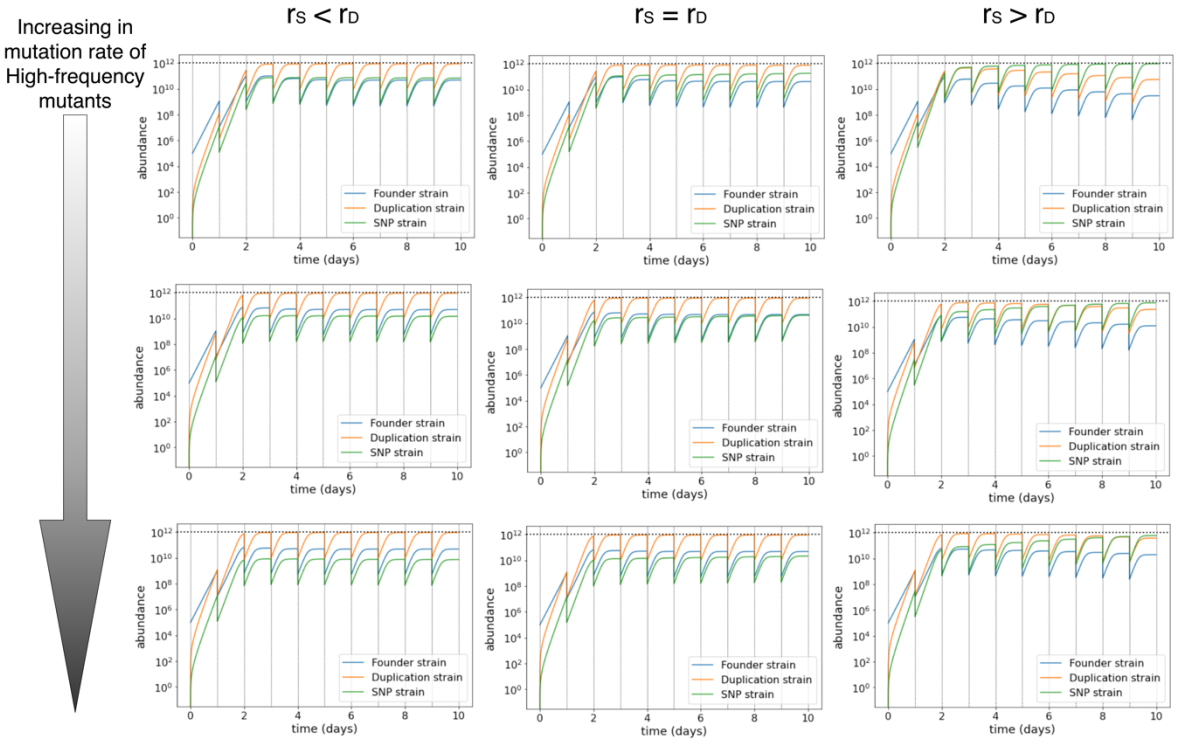


Fig 5.10. Simulation of the abundance of two types of mutants with different growth rates, mutation rates, and 1% bottleneck in transferring. D is the high-frequency mutation. S is the rare and stable mutation. In the figures, SNP is an example of a rare mutation. The mutation rate of duplicate mutants was chosen as a variable to demonstrate because the *serTGA* promoter mutant is a SNP mutant with a stable mutation rate. While mutation or reversion rate of duplication mutant could be varied. Horizontal graphs simulate different replication rates in each mutant. Vertical graphs simulate increasing in duplication mutation. The mutation rate of duplication mutants is expected to be higher than the SNP mutant. In the first row of the graph, the duplication mutation rate was 10 times higher than the SNP mutation rate. It was 50 times higher in the second row, and in the third row, it was 100 times higher. Blue, orange, and green lines indicated the founder strain, duplication strain, and SNP mutant (*serTGA* promoter mutant). Figures were drawn using the interactive interface built by Dr Florence Bansept and modified in OmniGraffle.

Simulations show that both the replication rates of the mutants and the mutation rate of the high-frequency mutants affect the dynamics of the two types of mutations. If the replication rate of a rare mutant is lower than or equal to the high-frequency mutant, the high-frequency mutant will be dominant (Fig 5.12). Rare mutations can dominate the population only if they have a higher replication rate. Hence, it implies that the rare mutation can only take over in the population when they have a higher fitness advantage. Interestingly, when the replication rate of the rare mutant is higher, the high-frequency mutant will be first rose in higher abundance but

will be overtaken by the rare mutation later (Fig 5.11). Even with increases in the mutation rate of the high-frequency mutant, the rare mutant will still dominate the population. Simulation suggested changes in replication rate determine which mutant will dominate over the evolution experiment.

Changing the high-frequency mutant's mutation rate will affect the rare mutant's abundance. In Fig 5.11, the vertical figures were the simulation when the high-frequency mutations have increasing mutation rates. Increasing the mutation rate of high-frequency mutants may lower the relative abundance of the founder strain and the rare mutant in the population. In the scenario when the rare mutation has a higher fitness advantage (Fig 5.11; right), increasing the mutation rate of the high-frequency mutant will increase the time for the rare mutation to overtake in the population.

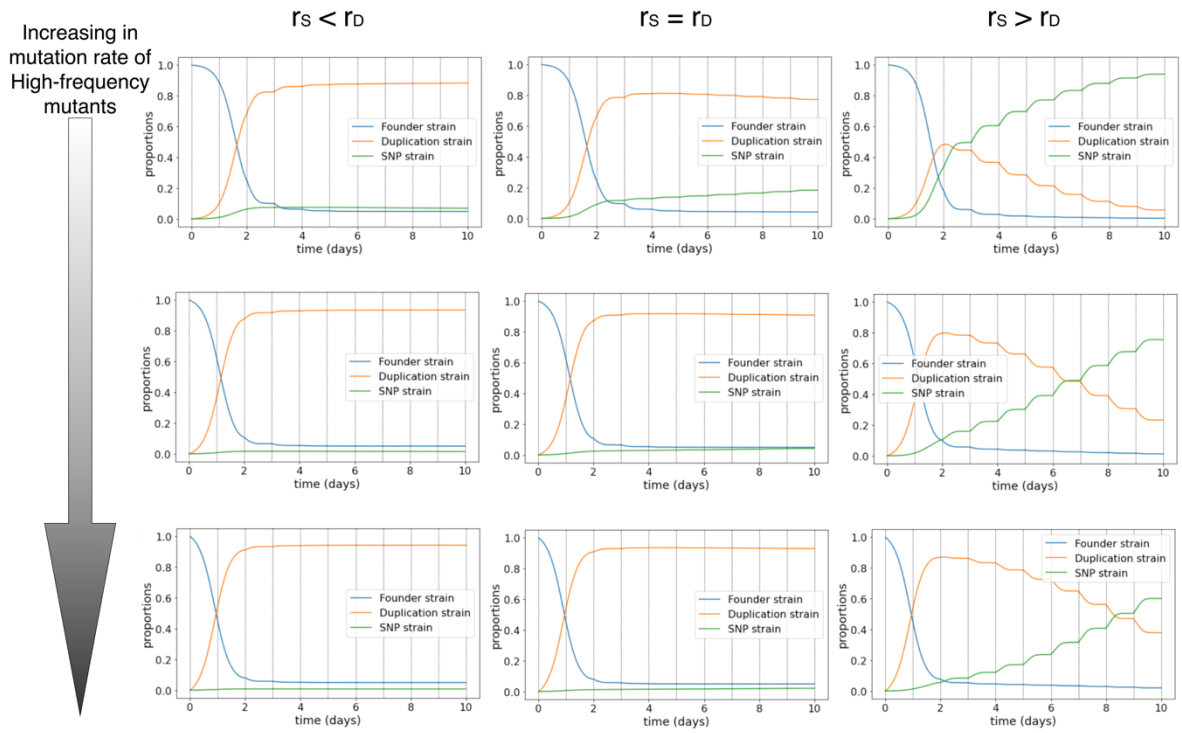


Fig 5.11. Simulation of evolutionary dynamics of two types of mutants with different growth rates and mutation rates. Simulation showed replication rate of mutants will determine which types would dominate over the evolution experiment. Mutation rate of duplication will affect the time needed for the overtaking. The setting of replication rate and mutation rate for the founder, duplication, and SNP strains and their colors are the same as Fig 5.10. Figures were drawn using the interactive interface built by Dr Florence Bansept and modified in OmniGraffle.

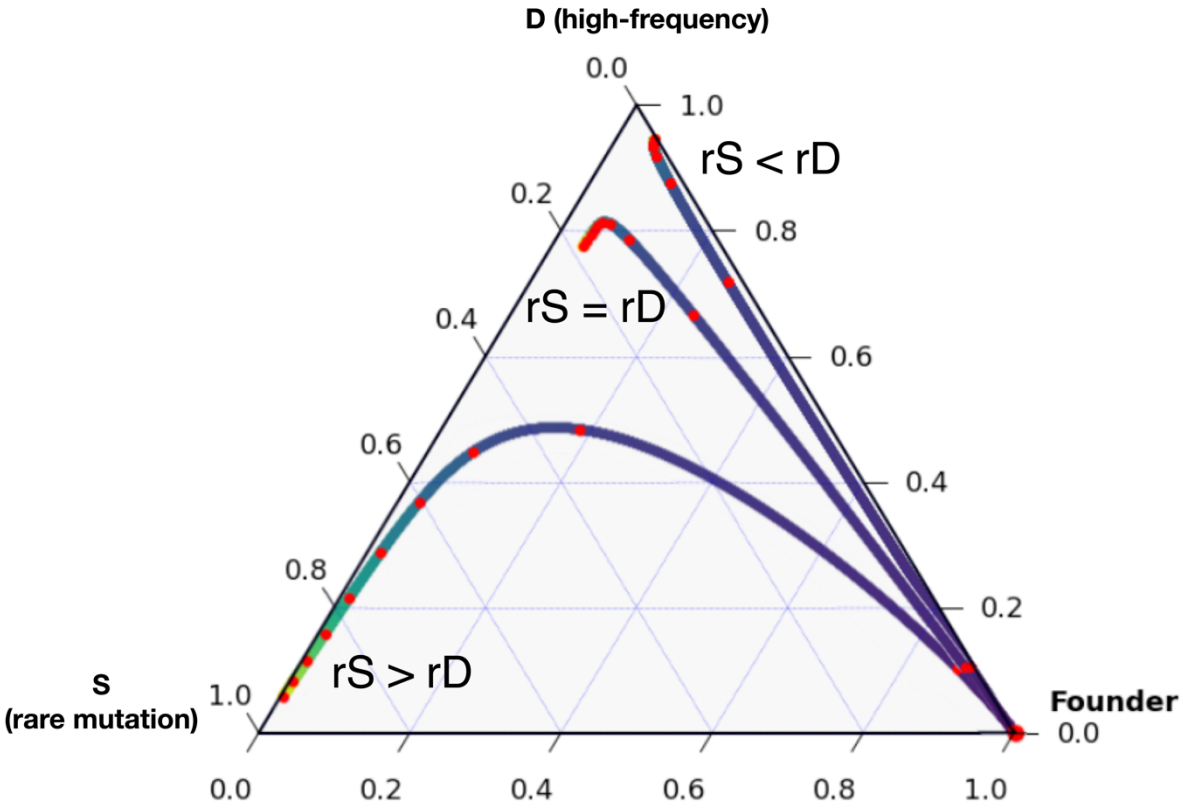


Fig 5.12. The phase plot showing replication rate can affect the evolutionary dynamics. The growth dynamics under different growth rate conditions are plotted on the simplex. It starts from the bottom right corner when the population began with all founder strains. Figures were drawn using the interactive interface built by Dr Florence Bansept and modified in OmniGraffle.

5.8 Conclusion

5.8.1 Rise and fall of duplications

The population sequencing demonstrated how the population evolved over the 100-day evolution experiment. In the early phase of the experiment, duplication mutants with large fragments (up to 1 Mb) arose. Fragments with 800 kb to 1 Mb fell quickly, with the rise of fragments of ~100-300 kb. In one of the lines, fragments as small as 236 bp rose since day 10 but it only gradually increased in the population until day 100 (~750 generations). This suggested other mutations (likely duplication mutants with larger fragments) existed in the population due to the high-frequency mutation. They could persist in the population for a relatively long time.

5.8.2 Further investigation of the evolutionary dynamics of the mutants through population sequencing

The compensated populations contain smaller fragments over the evolution experiment. The population consisted fragments with different sizes. It is interesting to track the dynamics of these mutants to see how mutations with different fragment sizes vary in their evolutionary dynamics. Therefore, the fragments identified by *breseq* were picked to further check for the dynamics. The current analysis in *breseq* does not provide the frequency of duplication mutations. The dynamics were estimated by the coverage of the duplication junctions over the average coverage in the whole genome. SNP can be easily estimated by the ratio of changed alleles, which provides higher accuracy.

Until the end of the experiment, the smallest identified fragments were found to increase in relative abundance in two lines (Δ*serCGA* replicate line #2 and Δ*EGEG* replicate line #2). In Δ*serCGA* replicate line #2, the smallest fragment is 4,025 bp. Δ*EGEG* replicate line #2, the smallest fragment is 236 bp. Both of these were also identified in isolate sequencing.

However, not all the lines were dominated by the smallest fragment called by *breseq*. The other two lines were not dominated by the smallest identified fragment. The possible reasons could be: 1) In the bioinformatic analysis, the smaller fragment was not called by the software. This could be due to the duplication rising in highly repetitive regions; 2) Some mutations decreased the fitness of the mutants with small fragments; 3) There are other mutants containing alternative mutations which increase the fitness.

Both fragments (830 bp from Δ*serCGA* replicate line #3 and 68,115 bp from Δ*EGEG* replicate line #1) were also isolated previously. The 830 bp duplication mutant occurred together with a *fliA* mutation (which causes aflagellate cells, and changes to downstream gene regulation). For this mutant, it is likely that, the mutant with *fliA* mutation gained this rare 830 bp duplication mutant. In the early phase, when the other co-existing duplication fragments in the population are relatively large, the 830 bp duplication with a *fliA* mutation is still fitter, leading to an increase in relative abundance. In this line, for example, there was a 56,225 bp duplication mutant. This fragment is much larger than 830 bp but if it did not contain other mutation that may alter its fitness, it will still be fitter than the one with 830 bp and mutation in *fliA*. Indeed, for these two mutants, on day 75 from the dynamics analysis, the 56,225 bp mutant was estimated to have more than 0.5 in the population but 830 bp mutant contributed less than 0.2 in the population.

This suggests that fragment sizes generally affect fitness. The smaller the fragment, the fitter the bacteria can be. However, other mutations that the mutants are carrying may alter the fitness. Regardless of the dynamics of these mutants, all the evolving lines from Δ*serCGA* and Δ*EGEG* observed decreases in duplication sizes over the evolutionary time.

5.8.3 Mechanisms of reversing to the founder genotypes

Over the 100-day evolutionary experiment, small phenotypes were often observed. The repeated stability tests confirmed the duplication mutations are unstable. These mutants could lose the extra fragment and reverse to the original genotypes. This still needs further confirmation because the small colonies observed from the population could be the non-compensated strains that remained in the population and carried over during transferring. In the stability test, all small colonies sampled were grown from mutants containing a single duplication mutation. The sequencing suggested it is very likely the small colonies observed in the population were from the unstable duplication mutants when they lost their duplication fragments and reversed to the founder genotypes. It also confirmed the mechanisms when losing the duplication fragment is scar-free. Hence, no evidence suggests that the small fragments arose from the shrinking of large fragments.

5.8.4 Occurrence of aflagellate cells and how they change the fitness of the bacteria

General adaptions were expected. It was helpful to include the wild-type strains in this experiment. It suggested that over the experiment, cells evolved to adapt to the growing conditions in the experiment. For example, *fleQ* mutation in all lines rose and increased over the evolutionary experiment, which caused the aflagellate cells and loss of motility. This is mainly due to the continuously shaking condition in the evolution experiment. The bacteria do not need to spend resources and energy to develop flagella to obtain nutrients. Those aflagellate cells may have higher fitness, as they did not invest energy on their flagella and swimming.

5.8.5 Wrinkly spreader rose and fell in the experiment

It is common to observe the wrinkly spreader (WS) in *P. fluorescens* SBW25. They are mat-formers in static conditions at the air-liquid interface [33] However, many studies also observed these mat formers in a shaking condition. In this study, some lines at a particular timepoint were observed

to have more than 50% of wrinkly spreaders in the population. It is hypothesised that these mutants colonised the interface of the air-liquid and wall of the culturing tubes. In this study, these wrinkly spreaders' dynamics varied throughout the evolutionary experiment. All evolving lines, including the wild-type, observed an increase in WS and a decrease afterwards. Wild-type line has the least WS in the population. Δ*EGEG* lines had the most among lines from three different founders. The previous cross-contamination event suggested Δ*EGEG* generally have lower fitness than Δ*serCGA*. This may suggest that the WS phenotype has fewer effects on reducing fitness due to the loss of the tRNA gene or duplicating too large parts of the genome. This indeed may suggest the WS phenotype is more reliant on the niche, where they initially colonising the interface of the air-liquid and wall of the culturing tubes. Meanwhile, those in suspension form without WS phenotype would have fewer advantages when competing for the limited nutrients from overnight growth.

5.8.6 Overview of the chapter

This chapter demonstrated how the evolving population changed. In the lines from Δ*serCGA* and Δ*EGEG* founders, large-scale duplications were observed in the population. In the population, they contained duplication mutants with smaller and smaller fragments. Some mutants gained a perfect extra copy of the tRNA gene (hence, without other extra genes) within a relatively short time (~700 - 770 generations). This is likely owing to the system where bacteria will not die after losing the extra fragment but will be reversed to the parent genotype. Hence, they provide continuous opportunities to gain new duplication fragments (or other mutations). The population sequences also provided insight into the dynamics of some identified indels, SNP and/or duplication mutations. Coupled with the mathematical modelling, the chapter further simulated how two types of mutation may vary in the population under different conditions.

Chapter VI – General Discussion: What we have learnt

6.1 Overview of the project and each chapter in this thesis

In this thesis, chapter 1 explains the function of transfer RNA (tRNA) and how it affects the translation and functioning of cells. It further discusses the background of using *Pseudomonas fluorescens* SBW25 as the model system, and the two mutants constructed from two previous studies [14, 38]. In chapter 2, provides the materials and methods used for the evolution experiments, further characterising tests, PCRs, sequencing, and bioinformatic analyses.

Chapters 3 to 5 are the results chapters. Chapter 3 describes the results from the follow-up study on the constructed strain Δ*EGEG*. The previous PhD study [38] constructed the strain Δ*EGEG* and performed a 28-day transfer evolution experiment. It provided evidence that the compensatory mechanism is through large-scale duplication. Some follow-up studies were required, and I performed these as part of my PhD. For example, I performed a stability test to examine the stability of these duplication mutants. The deep sequencing of tRNA in the mature tRNA pool of some selected isolates also revealed the change in tRNA proportion in the mutants due to the large-scale duplication mutation containing an extra copy of the tRNA gene. I further measured the changes in population growth through growth curve measurement and population sequencing. The population sequencing showed a diversity of duplication fragments in the bacterial population. The population sequencing could also check for any potential contamination between lines. These results were combined with previous data and published as a co-first author manuscript [57]. The application of mature tRNA pool sequencing also led to another co-authored publication, in collaboration with Dr Deepa Agashe and Parth Raval [63]. Results from that project were not included in this thesis.

Chapter 4 describes the results of the restarted and extended evolutionary experiment. The evolution experiment was restarted with Δ*serCGA*, Δ*EGEG* and wild-type strains. Each of these had four independent replicate lines. The experiment was continued for 100 days (~750 generations). The morphotype classes occurring across each line were identified and quantified. The diverse morphotype classes (at least 9) indicated high diversities in genotypes. Isolates were sampled from all lines with in-depth sampling from five lines, where isolates of different morphotype classes and from different days were selected. The background mutations and compensatory mutations for tRNA gene(s) losses were successfully identified. These mutations

were further confirmed with corresponding characterising assays or PCRs. Among the mutations compensating for the tRNA gene(s) loss, within-genome duplication is the major mechanism (except for one SNP in a tRNA gene promoter). A trend of decreasing in duplication fragment sizes among the sampled isolates over the evolution experiment was observed from both Δ*serCGA* and Δ*EGEG* founder lines. This suggested that smaller, and thus more stable, duplication occurred in the evolutionary experiment. The smallest duplication fragments were 830 bp from a Δ*serCGA* line and 236 bp from a Δ*EGEG* line. Further stability tests also revealed that duplication mutants with smaller fragments are more stable.

Chapter 5 further extended our understanding of the evolving population through (i) population sequencing, and (ii) mathematical modelling. Based on the isolate sequencing results, five lines were selected for population sequencing. The sampling was separated into two phases: in the early compensatory phase and over the evolution experiment. The early phase is when compensatory mutants start occurring in the population. The time points when half of the population was compensated and the when population was only just fully compensated were sampled. The other sampling phase was over the later stages of the evolution experiment from days 25, 50, 75 and 100. The two sampling phases provided information about how the population started compensating, and how compensation continued over the 100 days of the experiment. Specific mutations were tracked in the population sequences, revealing their evolutionary dynamics. Results indicated the duplication fragment sizes of the mutants in the population decreased over the evolution experiment. The sampling of small phenotypes from the overnight culture of mutant isolates revealed the mechanisms where unstable compensated mutants reversed back to the founder genotype. The continued generation of the founder genotype provided further chances to sample new mutations. No evidence supports that the large duplication fragments shrank to smaller fragments (i.e., lost internal parts of the fragment). The high flexibility of the bacteria in changing their tRNA gene sets was observed. By the end of the experiment, some mutants only duplicated a region as small as 236 bp, containing only the compensatory tRNA gene.

These small fragments, for example, the 236 bp, are mechanistically much more stable than the large duplications. This was also demonstrated by the stability test. The results of the stability test suggested bacterial populations can flexibly alter their tRNA gene copy numbers by duplicating parts of their genome during replication, and then losing the duplications in successive

replication rounds. Overall, this project has demonstrated one mechanism that may lead to the diversity of tRNA gene composition across the tree of life.

6.2 How evolution in tRNA gene sets change our understanding of the tree of life

Transfer RNAs are essential for protein synthesis. They translate the genetic information in messenger RNA (mRNA) into proteins. There are at least two ways to classify tRNAs. Based on the amino acid they carry, there are 20 different tRNA isoacceptors. Based on their anticodon, there are 61 hypothetically possible tRNA species, one matching each sense codon (64 codons – 3 stop codons = 61 sense codons).

The composition of tRNAs across the tree of life is different. The three domains (bacteria, archaea and eukaryota) have different general tRNA compositions (Fig 6.1). Some tRNAs are unique in one of the domains. When looking at organisms in the same domain, or even genus, they can also have a diverse composition of tRNA types [16, 17]. Not only are they different in the types of tRNAs they have, but also their tRNA gene copy numbers. Some organisms contain many multiple copies of identical tRNA genes.

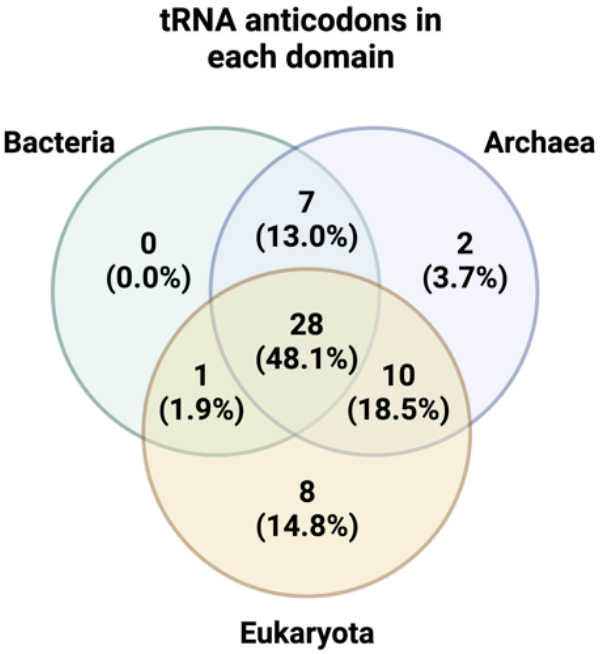


Fig 6.1. tRNA anticodon compositions in the three domains. Different domains have different tRNA anticodons. Figure was modified from [110] and drawn from Biorender.com.

When considering such a diversity in tRNA types in the tree of life, it brings up some interesting evolutionary questions. For example, how do such diverse compositions of tRNA sets arise? Why do organisms have different compositions of tRNA genes? Why do they have different numbers of identical copies? How could they change their tRNA gene sets if there are changing translational needs?

To investigate these questions, two previous studies from our lab constructed two engineered strains of *P. fluorescens* SBW25 with one or more tRNA gene(s) deleted (Δ*serCGA* and Δ*EGEG*) [14, 57]. There are also other studies on bacteria showing duplication or gene amplification is one *de novo* mechanism to change the bacteria tRNA gene sets to deal with new translational needs [25, 69, 111].

6.3 How large-scale duplication mutation shapes bacterial cell function

The last section discussed how tRNA gene sets of bacteria could change to meet translational need. Such tRNA set changes can alter bacterial physiology and function. For example, in antibiotic resistance [25, 111]. Some bacteria have gained resistance to a high dosage of antibiotics through duplicating/amplifying the genes responsible for resistance [111]. This underpinned how gene duplication could change the medical treatment strategies towards pathogenic bacteria.
Some duplication mutations occur at a higher rate in some regions of the chromosome, particularly with homologous DNA [39, 43]. For example, in *P. fluorescens* SBW25 used in this study, there are two *rhs* regions at positions ~1.86 Mb and ~2.86 Mb, with ~3 kb of homologous sequence. RecA-mediated ecombination between these generate the ~1 Mb duplications observed frequently in evolving Δ*EGEG* populations.

The repetitive regions through which RecA-mediated recombination occurs are often located considerable distanced away from each other in their genome position [57]. Hence, these high-frequency duplications often duplicate large parts of the genome. Many studies on duplication mutations had mutants with >1 Mb duplication fragments [14, 57, 112]. These extremely large duplications themselves generate long stretches of DNA homology, and hence can be lost by RecA-mediated recombination (larger duplication = more homology = higher loss rate) [57]. The

evolutionary impacts of these duplications in repetitive regions can be underestimated because of their transient nature [57, 112].

6.4 How this study further extends our current understanding of unstable large-duplication and evolution of tRNA gene sets

To address the questions on the composition and evolution of tRNA gene sets, the previous two studies using Δ*serCGA* or Δ*EGEG* were conducted [14, 57]. In both studies, the evolving population strains duplicated a large part of their genome to alter the translational needs. With the transient nature of these duplication fragments, there are further questions on how these transient mutations impact the evolution of bacteria. For example, do they affect the evolutionary path of the bacteria to change the composition of the genes they contain and further alter the bacteria's function? What roles do these transient duplications have in the evolution of bacteria? These questions were not answered in both previous studies, which instead provided evidence that these questions must be addressed. For example, my follow-up experiments on one of the studies indicated the high instabilities in Δ*EGEG* duplication fragments. My further studies also characterised these mutants and, in the case of Δ*EGEG*, showed how the composition in the mature tRNA pool altered in these duplication mutants, revealing the advantages provided by the duplications.

One direct way to investigate the evolutionary fate of the large-scale, unstable duplication mutations is to restart and extend the whole experiment. This study restarted using both Δ*serCGA* and Δ*EGEG* with an extended evolution time of ~750 generations. The restart of the experiment did not use the evolved population from both previous studies [14, 38, 57]. This generated four more independent evolving populations from each founder. Results of whole genome sequencing of 72 isolates suggested the robustness of large-scale duplication mutation to compensate for the translational needs as they were found in all lines. Except for one alternative mutation, duplication mutations are the only compensatory mutations observed from the sampled isolates. Duplications were identified in all the eight lines from Δ*serCGA* or Δ*EGEG* founders. Three identified fragments were smaller than 5 kb (236 bp, 830 bp or 4,025 bp). Presumably, these duplications occurred through RecA-independent mechanisms, which require little or no DNA homology on both ends of the duplication fragments [39, 43]. These types of duplications usually occur at much lower rates, indicating these are (relatively) rare mutations [39, 43, 57].

The small duplication mutants - for example, the identified 236 bp duplication - are relatively rare mutations with a very low mutation frequency. The number of mutational opportunities is one of the factors affecting the arising time of this (and other) mutation. For example, if the experiment were repeated millions of times, the smaller duplications may, very occasionally, be observed from day 1 of the evolution experiment. The discovery of this 236 bp duplication mutant demonstrated a mechanism for bacteria to evolve to change the copy number of tRNA genes and answered how bacteria can contain identical copies of tRNA genes.

6.5 Population sequencing revealed how bacterial populations evolve to change the tRNA gene composition

The isolate sequencing provided information on which mutants each population contained. The whole genome sequencing and characterising of these mutants suggested the unstable large-scale duplication mutants were outcompeted by small duplication mutants, where some of them only changed the copy number of their tRNA genes without affecting other gene copy numbers. The population sequencing further investigated the evolutionary dynamics of the mutants. Except for the mutations compensating for the tRNA gene(s) loss, all the evolving populations had general adaptations to the evolution conditions. For example, the mutation in *fleQ* causes non-motile bacterial cells, suggesting that energy could be conserved by removing flagella and swimming.

In the case of duplication mutations, all the evolving Δ*serCGA* and Δ*EGEG* populations contained smaller duplication fragments over the evolution experiment (Fig 5.3, 5.4, 5.6 and 5.7). The sequencing of small phenotypes suggested that the unstable duplication lost the mutation and that the reversed genotypes could gain a new mutation (section 5.5). There was no evidence to suggest the smaller duplication fragments in this study shrank directly from the larger duplications. Results of the stability test suggested a positive correlation between duplication size and instability ($r = 0.88525$). Both 236 bp and 4,025 bp duplication mutants were very stable. No reversed genotypes were isolated during the stability test with more than 2,200 and 1,200 cell counts respectively. The dynamics analysis using the population sequences revealed the dynamics of these mutations. Both still showed an increasing trend in the population after ~750 generations.

These results actually suggested that the unstable large-scale duplication could be a stepping

stone for the genomic evolution of bacteria. The unstable adaptive large-scale duplication first arose in the population at a high frequency. The duplication fragments contained the compensatory elements, the tRNA gene providing fitness advantages. The instabilities of the mutations allow them to regenerate the (scar-free) parent genotype, giving them new chances to gain a new mutation. The fitter and more stable, but lower-rate, mutations tend, on average to emerge later. However, when they do, they outcompete the less fit and less stable, but higher-rate, mutations to dominate the population.

6.6 Future experiments: how this project can be extended

This project focuses on the evolutionary fate of unstable but adaptive large-scale duplication mutations to address the question of how bacterial tRNA gene sets can evolve. Current results from experiments and the mathematical modelling provided sufficient support for the current finding:
1) The adaptive but unstable large-scale duplications are evolutionary stepping stones for the bacteria (based on the results from population sequencing, stability test and the sequencing of small types);
2) With large-scale duplication as a stepping stone, mutants with duplication only changed the tRNA gene copy numbers and began to dominate the population (based on isolates and population sequencing).

A few experiments can be performed to support these findings if the project can be further extended. These will focus on investigating these mutants' fitness advantages. This can be achieved by measuring the growth curve and conducting the competition assay. The growth curve measurement of each mutant can indicate how much improvement in growth between mutants. The competition assay will further provide direct evidence of the fitness advantages of different mutations. The duplications with smaller fragments are expected to be fitter as less DNA has to be replicated. Many of the isolates identified in this study contained other background mutations, such as wrinkly spreader or other aflagellate mutations (*fleQ* or *fli*). *fliA*, for example, is a sigma factor that regulates bacterial flagella gene expression [76]. However, it is also involved in modulating the intracellular c-di-GMP concentration bacterial physiology and other gene expression [113]. These background mutations, thus, would affect the measuring when determining the fitness differences between mutants with different fragment sizes.

Therefore, the comparison has to be separated into groups with or without the specific background mutation.

Also, from population sequences, large-scale duplications persisted in the population even when stable and small duplications arose. The mathematical model (developed by Dr Florence Bansept and Prof. Arne Traulsen) can simulate how long it will take for these unstable large duplications to be extinct in the population. The current model needs the input of some parameters from each mutant, including their replication rate, mutation rate and reversing rate. The growth assay can suggest the replication rate. The stability test has already provided information about the reversing rates. The mutation rate can also be estimated based on literature or simulated using different values.

These extended experiments measure the fitness advantages of each duplication mutation. Hence, we could also estimate the balance of the cost to replicate more DNA and the fitness advantages to increase the gene copy numbers.

6.7 Future perspective: future direction in investigating the diversity across the tree of life

This project demonstrated how bacteria evolve to change the tRNA gene composition, having large-scale duplication as a stepping stone. Then, a small duplication where only the extra copy of the tRNA gene arose. The transient nature of duplication increased the plasticity and flexibility in evolution. Across the tree of life, not only tRNA genes but also other genes could have different copy numbers. These genes are identical and close to each other in genome position. Future projects can investigate how duplication shapes gene composition. Actually, some studies showed that the copy number of resistance genes can be altered through duplication [111, 114]. Many studies investigating duplication mutation discovered this as a mechanism to change the gene copy number. They can gain extra copies through recombination. Future research direction should also focus on how deletion reduces the gene copy numbers. Another study on the evolution of tRNA genes in yeast, *Saccharomyces cerevisiae*, demonstrated that anticodon mutation is an adaptive mechanism that meets the new translational need. Of course, horizontal gene transfer is also considered by many evolutionists as a major evolutionary force that reshaped genomes throughout evolution [26, 115-117].

In this study, most of the mutants were shown to be fitter by gaining the extra duplication fragments as a main compensatory mechanism. This study was conducted in laboratory conditions with optimal growth conditions (temperature, rich nutrients and shaking). That was actually a selection of fast-growing bacteria. In nature, fast-growing may not always be beneficial. For example, some studies suggested that some phenotypic survival to antibiotics has been linked to slow growth, low metabolic activities or bacterial dormancy [118-120]. Also, a study has also suggested slower growth *E. coli* has a longer survival when deprived of nutrients due to the lower maintenance rate [121]. In the biofilm physiology context, when cells colonise the surface and are surrounded by the "house" of extracellular polymeric substances. These are food sources for protozoan grazers. The fast-growth bacteria will likely grow higher when there is insufficient space. They will be closer to the surface of the biofilm and have a higher chance of getting grazed by these predators.

Gaining extra or more gene copies can be deleterious and not necessarily always provide fitness advantages. For example, if the bacterial genome has too many repeating identical genes, these sites have a high frequency, causing recombination and genome rearrangement. This increased the genome instability. When too many gene copies may not always be beneficial, this raised the question of how bacteria delete their genes to increase their fitness advantages. Most studies investigating recombination-related mutation would involve how the bacteria evolve to increase the gene copy numbers, leading to adaptive evolution [43, 111]. The study on how bacteria reduce their genome to increase fitness is overlooked. Nevertheless, both duplication and deletion processes demonstrated how bacterial genomes could evolve rapidly, leading to the diversity in gene copy numbers across the tree of life.

This study investigated the enormously growing interest in how recombination-related mutations shape living organisms in evolution. For example, how could organisms have a high diversity of tRNA gene composition even if they are phylogenetically close? This study, by investigating the evolutionary fate of unstable large-scale duplication and its mechanisms, provided information on the resilience and evolutionary success of the bacteria through varying their gene copy numbers. These further contributed to our understanding of evolution across the tree of life.

References

1. Minchin, S. and J. Lodge, *Understanding biochemistry: structure and function of nucleic acids.* Essays Biochem, 2019. **63**(4): p. 433-456.
2. Hall, K. and N. Sankaran, *DNA translated: Friedrich Miescher's discovery of nuclein in its original context.* Br J Hist Sci, 2021. **54**(1): p. 99-107.
3. Avery, O.T., C.M. Macleod, and M. McCarty, *Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii.* J Exp Med, 1944. **79**(2): p. 137-58.
4. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
5. Spitale, R.C. and D. Incarnato, *Probing the dynamic RNA structurome and its functions.* Nat Rev Genet, 2023. **24**(3): p. 178-196.
6. Wang, D. and A. Farhana, *Biochemistry, RNA Structure*, in *StatPearls*. 2024: Treasure Island (FL).
7. Berg, M.D. and C.J. Brandl, *Transfer RNAs: diversity in form and function.* RNA Biol, 2021. **18**(3): p. 316-339.
8. Swart, E.C., et al., *Genetic Codes with No Dedicated Stop Codon: Context-Dependent Translation Termination.* Cell, 2016. **166**(3): p. 691-702.
9. Saint-Leger, A. and L. Ribas de Pouplana, *The importance of codon-anticodon interactions in translation elongation.* Biochimie, 2015. **114**: p. 72-9.
10. Agris, P.F., et al., *Celebrating wobble decoding: Half a century and still much is new.* Rna Biology, 2018. **15**(4-5): p. 537-553.
11. Varenne, S., et al., *Translation Is a Non-Uniform Process - Effect of Transfer-Rna Availability on the Rate of Elongation of Nascent Polypeptide-Chains.* Journal of Molecular Biology, 1984. **180**(3): p. 549-576.
12. Rak, R., O. Dahan, and Y. Pilpel, *Repertoires of tRNAs: The Couplers of Genomics and Proteomics.* Annual Review of Cell and Developmental Biology, Vol 34, 2018. **34**: p. 239-264.
13. Yona, A.H., et al., *tRNA genes rapidly change in evolution to meet novel translational demands.* Elife, 2013. **2**: p. e01339.
14. Ayan, G.B., H.J. Park, and J. Gallie, *The birth of a bacterial tRNA gene by large-scale, tandem duplication events.* Elife, 2020. **9**: p. e57947.

15. Wint, R., A. Salamov, and I.V. Grigoriev, *Kingdom-Wide Analysis of Fungal Protein-Coding and tRNA Genes Reveals Conserved Patterns of Adaptive Evolution.* Mol Biol Evol, 2022. **39**(2): p. msab372.
16. Fujishima, K. and A. Kanai, *tRNA gene diversity in the three domains of life.* Front Genet, 2014. **5**: p. 142.
17. Goodenbour, J.M. and T. Pan, *Diversity of tRNA genes in eukaryotes.* Nucleic Acids Res, 2006. **34**(21): p. 6137-46.
18. Rawlings, T.A., T.M. Collins, and R. Bieler, *Changing identities: tRNA duplication and remolding within animal mitochondrial genomes.* Proc Natl Acad Sci U S A, 2003. **100**(26): p. 15700-5.
19. Withers, M., L. Wernisch, and M. dos Reis, *Archaeology and evolution of transfer RNA genes in the Escherichia coli genome.* RNA, 2006. **12**(6): p. 933-42.
20. Higgs, P.G. and W. Ran, *Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage.* Mol Biol Evol, 2008. **25**(11): p. 2279-91.
21. Rogers, H.H., C.M. Bergman, and S. Griffiths-Jones, *The evolution of tRNA genes in Drosophila.* Genome Biol Evol, 2010. **2**: p. 467-77.
22. Bermudez-Santana, C., et al., *Genomic organization of eukaryotic tRNAs.* BMC Genomics, 2010. **11**: p. 270.
23. Sabi, R., R. Volvovitch Daniel, and T. Tuller, *stAIcalc: tRNA adaptation index calculator based on species-specific weights.* Bioinformatics, 2017. **33**(4): p. 589-591.
24. McDonald, M.J., et al., *The evolutionary dynamics of tRNA-gene copy number and codon-use in E. coli.* BMC Evol Biol, 2015. **15**: p. 163.
25. Nilsson, A.I., et al., *Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes.* Proc Natl Acad Sci U S A, 2006. **103**(18): p. 6976-81.
26. Bedhomme, S., et al., *Evolutionary Changes after Translational Challenges Imposed by Horizontal Gene Transfer.* Genome Biol Evol, 2019. **11**(3): p. 814-831.
27. Rainey, P.B. and M.J. Bailey, *Physical and genetic map of the Pseudomonas fluorescens SBW25 chromosome.* Molecular Microbiology, 1996. **19**(3): p. 521-533.
28. Trippe, K., et al., *Pseudomonas fluorescens SBW25 produces furanomycin, a non-proteinogenic amino acid with selective antimicrobial properties.* BMC Microbiol, 2013. **13**: p. 111.
29. Preston, G.M., N. Bertrand, and P.B. Rainey, *Type III secretion in plant growth-promoting Pseudomonas fluorescens SBW25.* Molecular Microbiology, 2001. **41**(5): p. 999-1014.
30. Gallie, J., et al., *Repeated Phenotypic Evolution by Different Genetic Routes in Pseudomonas fluorescens SBW25.* Molecular Biology and Evolution, 2019. **36**(5): p. 1071-1085.

31. Beaumont, M.A., et al., *Adaptive approximate Bayesian computation.* Biometrika, 2009. **96**(4): p. 983-990.
32. McDonald, M.J., et al., *Adaptive Divergence in Experimental Populations of Pseudomonas fluorescens. IV. Genetic Constraints Guide Evolutionary Trajectories in a Parallel Adaptive Radiation.* Genetics, 2009. **183**(3): p. 1041-1053.
33. Rainey, P.B. and M. Travisano, *Adaptive radiation in a heterogeneous environment.* Nature, 1998. **394**(6688): p. 69-72.
34. Silby, M.W., et al., *Genomic and genetic analyses of diversity and plant interactions of Pseudomonas fluorescens.* Genome Biol, 2009. **10**(5): p. R51.
35. Zhang, X.X. and P.B. Rainey, *Construction and validation of a neutrally-marked strain of Pseudomonas fluorescens SBW25.* Journal of Microbiological Methods, 2007. **71**(1): p. 78-81.
36. Spiers, A.J., *A mechanistic explanation linking adaptive mutation, niche change, and fitness advantage for the wrinkly spreader.* Int J Evol Biol, 2014. **2014**: p. 675432.
37. Koza, A., et al., *Environmental modification and niche construction: developing O2 gradients drive the evolution of the Wrinkly Spreader.* ISME J, 2011. **5**(4): p. 665-73.
38. Khomarbaghi, Z., *The function and origin of multi-copy tRNA genes in Pseudomonas fluorescens.* 2020, Christian-Albrechts-Universität: Kiel.
39. Reams, A.B. and J.R. Roth, *Mechanisms of gene duplication and amplification.* Cold Spring Harb Perspect Biol, 2015. **7**(2): p. a016592.
40. Anderson, P. and J. Roth, *Spontaneous Tandem Genetic Duplications in Salmonella-Typhimurium Arise by Unequal Recombination between Ribosomal-Rna (Rrn) Cistrons.* Proceedings of the National Academy of Sciences of the United States of America-Biological Sciences, 1981. **78**(5): p. 3113-3117.
41. Elliott, K.T., L.E. Cuff, and E.L. Neidle, *Copy number change: evolving views on gene amplification.* Future Microbiology, 2013. **8**(7): p. 887-899.
42. Nasvall, J., et al., *Real-time evolution of new genes by innovation, amplification, and divergence.* Science, 2012. **338**(6105): p. 384-7.
43. Andersson, D.I. and D. Hughes, *Gene amplification and adaptive evolution in bacteria.* Annu Rev Genet, 2009. **43**: p. 167-95.
44. Andersson, D.I., J. Jerlstrom-Hultqvist, and J. Nasvall, *Evolution of new functions de novo and from preexisting genes.* Cold Spring Harb Perspect Biol, 2015. **7**(6): p. a017996.
45. Anderson, R.P. and J.R. Roth, *Tandem genetic duplications in phage and bacteria.* Annu Rev Microbiol, 1977. **31**: p. 473-505.

46. Bergthorsson, U., D.I. Andersson, and J.R. Roth, *Ohno's dilemma: evolution of new genes under continuous selection.* Proc Natl Acad Sci U S A, 2007. **104**(43): p. 17004-9.
47. Voordeckers, K., et al., *Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms underlying evolutionary innovation through gene duplication.* PLoS Biol, 2012. **10**(12): p. e1001446.
48. Kelley, J.L., et al., *Functional diversification and evolution of antifreeze proteins in the antarctic fish Lycodichthys dearborni.* J Mol Evol, 2010. **71**(2): p. 111-8.
49. Brandis, G. and D. Hughes, *The SNAP hypothesis: Chromosomal rearrangements could emerge from positive Selection during Niche Adaptation.* PLoS Genet, 2020. **16**(3): p. e1008615.
50. Cao, S., et al., *Positive Selection during Niche Adaptation Results in Large-Scale and Irreversible Rearrangement of Chromosomal Gene Order in Bacteria.* Mol Biol Evol, 2022. **39**(4): p. msac069.
51. Royce, L., et al., *Identification of mutations in evolved bacterial genomes.* Methods Mol Biol, 2013. **985**: p. 249-67.
52. Reams, A.B., et al., *Duplication frequency in a population of Salmonella enterica rapidly approaches steady state with or without recombination.* Genetics, 2010. **184**(4): p. 1077-94.
53. Sipola, A., P. Marttinen, and J. Corander, *Bacmeta: simulator for genomic evolution in bacterial metapopulations.* Bioinformatics, 2018. **34**(13): p. 2308-2310.
54. Cury, J., et al., *Simulation of bacterial populations with SLiM.* Peer Community Journal, 2022. **2**: p. e7.
55. Brown, T., et al., *SimBac: simulation of whole bacterial genomes with homologous recombination.* Microb Genom, 2016. **2**(1): p. mgen.0.000044.
56. Kuhl, M.A., B. Stich, and D.C. Ries, *Mutation-Simulator: fine-grained simulation of random mutations in any genome.* Bioinformatics, 2021. **37**(4): p. 568-569.
57. Khomarbaghi, Z., et al., *Large-scale duplication events underpin population-level flexibility in tRNA gene copy number in Pseudomonas fluorescens SBW25.* Nucleic Acids Res, 2024. **52**(5): p. 2446-2462.
58. Deatherage, D.E. and J.E. Barrick, *Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq.* Methods Mol Biol, 2014. **1151**: p. 165-88.
59. Barrick, J.E., et al., *Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq.* BMC Genomics, 2014. **15**(1): p. 1039.
60. Morales-Soto, N., et al., *Preparation, imaging, and quantification of bacterial surface motility assays.* J Vis Exp, 2015(98).
61. O'Toole, G.A., *Microtiter dish biofilm formation assay.* J Vis Exp, 2011(47).
62. Shigematsu, M., et al., *YAMAT-seq: an efficient method for high-throughput sequencing of mature transfer RNAs.* Nucleic Acids Res, 2017. **45**(9): p. e70.

63. Raval, P.K., et al., *The layered costs and benefits of translational redundancy.* Elife, 2023. **12**.
64. Love, M.I., W. Huber, and S. Anders, *Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2.* Genome Biol, 2014. **15**(12): p. 550.
65. Ralser, M., et al., *An efficient and economic enhancer mix for PCR.* Biochem Biophys Res Commun, 2006. **347**(3): p. 747-51.
66. Ayan, G.B., H.J. Park, and J. Gallie, *The birth of a bacterial tRNA gene by large-scale, tandem duplication events.* Elife, 2020. **9**.
67. Hegeman, G.D. and S.L. Rosenberg, *The evolution of bacterial enzyme systems.* Annu Rev Microbiol, 1970. **24**: p. 429-62.
68. Brandis, G. and D. Hughes, *The SNAP hypothesis: Chromosomal rearrangements could emerge from positive Selection during Niche Adaptation.* Plos Genetics, 2020. **16**(3).
69. Cao, S., et al., *Positive Selection during Niche Adaptation Results in Large-Scale and Irreversible Rearrangement of Chromosomal Gene Order in Bacteria.* Mol Biol Evol, 2022. **39**(4).
70. Rainey, P.B., et al., *The evolutionary emergence of stochastic phenotype switching in bacteria.* Microb Cell Fact, 2011. **10 Suppl 1**(Suppl 1): p. S14.
71. Mukherjee, A., G. Dechow-Seligmann, and J. Gallie, *Evolutionary flexibility in routes to mat formation by Pseudomonas.* Mol Microbiol, 2022. **117**(2): p. 394-410.
72. Remigi, P., et al., *Ribosome Provisioning Activates a Bistable Switch Coupled to Fast Exit from Stationary Phase.* Mol Biol Evol, 2019. **36**(5): p. 1056-1070.
73. Gallie, J., et al., *Repeated Phenotypic Evolution by Different Genetic Routes in Pseudomonas fluorescens SBW25.* Mol Biol Evol, 2019. **36**(5): p. 1071-1085.
74. Gallie, J., et al., *Bistability in a metabolic network underpins the de novo evolution of colony switching in Pseudomonas fluorescens.* PLoS Biol, 2015. **13**(3): p. e1002109.
75. Dasgupta, N., et al., *fleQ, the gene encoding the major flagellar regulator of Pseudomonas aeruginosa, is sigma70 dependent and is downregulated by Vfr, a homolog of Escherichia coli cyclic AMP receptor protein.* J Bacteriol, 2002. **184**(19): p. 5240-50.
76. Ohnishi, K., et al., *Gene fliA encodes an alternative sigma factor specific for flagellar operons in Salmonella typhimurium.* Mol Gen Genet, 1990. **221**(2): p. 139-47.
77. Starnbach, M.N. and S. Lory, *The filA (rpoF) gene of Pseudomonas aeruginosa encodes an alternative sigma factor required for flagellin synthesis.* Molecular Microbiology, 1992. **6**(4): p. 459-469.
78. Blanco-Romero, E., et al., *Genome-wide analysis of the FleQ direct regulon in Pseudomonas fluorescens F113 and Pseudomonas putida KT2440.* Sci Rep, 2018. **8**(1): p. 13145.

79. Ni, B., et al., *Evolutionary Remodeling of Bacterial Motility Checkpoint Control.* Cell Rep, 2017. **18**(4): p. 866-877.
80. Shepherd, M.J., A.P. Pierce, and T.B. Taylor, *Evolutionary innovation through transcription factor rewiring in microbes is shaped by levels of transcription factor activity, expression, and existing connectivit.* Plos Biology, 2023. **21**(10).
81. McDonald, M.J., et al., *Adaptive divergence in experimental populations of Pseudomonas fluorescens. IV. Genetic constraints guide evolutionary trajectories in a parallel adaptive radiation.* Genetics, 2009. **183**(3): p. 1041-53.
82. Lind, P.A., A.D. Farr, and P.B. Rainey, *Evolutionary convergence in experimental Pseudomonas populations.* ISME J, 2017. **11**(3): p. 589-600.
83. Spiers, A.J., et al., *Cellulose Expression in Pseudomonas fluorescens SBW25 and Other Environmental Pseudomonads.* Cellulose - Medical, Pharmaceutical and Electronic Applications, 2013: p. 1-26.
84. Spiers, A.J., et al., *Biofilm formation at the air-liquid interface by the Pseudomonas fluorescens SBW25 wrinkly spreader requires an acetylated form of cellulose.* Mol Microbiol, 2003. **50**(1): p. 15-27.
85. Spiers, A.J. and P.B. Rainey, *The Pseudomonas fluorescens SBW25 wrinkly spreader biofilm requires attachment factor, cellulose fibre and LPS interactions to maintain strength and integrity.* Microbiology (Reading), 2005. **151**(Pt 9): p. 2829-2839.
86. Herren, C.D., et al., *The BarA-UvrY two-component system regulates virulence in avian pathogenic Escherichia coli O78:K80:H9.* Infect Immun, 2006. **74**(8): p. 4900-9.
87. Workentine, M.L., et al., *The GacS-GacA two-component regulatory system of Pseudomonas fluorescens: a bacterial two-hybrid analysis.* FEMS Microbiol Lett, 2009. **292**(1): p. 50-6.
88. Pernestig, A.K., et al., *The Escherichia coli BarA-UvrY two-component system is needed for efficient switching between glycolytic and gluconeogenic carbon sources.* J Bacteriol, 2003. **185**(3): p. 843-53.
89. Tomenius, H., et al., *The Escherichia coli BarA-UvrY two-component system is a virulence determinant in the urinary tract.* BMC Microbiol, 2006. **6**: p. 27.
90. Parab, L., et al., *Chloramphenicol reduces phage resistance evolution by suppressing bacterial cell surface mutants.* bioRxiv, 2023: p. 2023.08.28.552763.
91. Kassen, R., M. Llewellyn, and P.B. Rainey, *Ecological constraints on diversification in a model adaptive radiation.* Nature, 2004. **431**(7011): p. 984-8.
92. Bailey, S.F., et al., *Competition both drives and impedes diversification in a model adaptive radiation.* Proceedings of the Royal Society B-Biological Sciences, 2013. **280**(1766).

93. Prak, E.T. and H.H. Kazazian, Jr., *Mobile elements and the human genome.* Nat Rev Genet, 2000. **1**(2): p. 134-44.
94. Bourque, G., et al., *Ten things you should know about transposable elements.* Genome Biology, 2018. **19**.
95. Krasileva, K.V., *The role of transposable elements and DNA damage repair mechanisms in gene duplications and gene fusions in plant genomes.* Curr Opin Plant Biol, 2019. **48**: p. 18-25.
96. Shi, Q., et al., *DNA damage differentially activates regional chromosomal loci for Tn7 transposition in Escherichia coli.* Genetics, 2008. **179**(3): p. 1237-50.
97. Hasanin, M.S., et al., *Sustainable bacterial cellulose production by Achromobacter using mango peel waste.* Microbial Cell Factories, 2023. **22**(1).
98. Pooja, R., K. Vadodaria, and S. Vidhya, *Synthesis of bacterial cellulose and herbal extract for the development of wound dressing.* Materials Today-Proceedings, 2019. **15**: p. 284-293.
99. Mokhena, T.C. and M.J. John, *Cellulose nanomaterials: new generation materials for solving global issues.* Cellulose, 2020. **27**(3): p. 1149-1194.
100. Ottenhall, A., J. Illergård, and M. Ek, *Water Purification Using Functionalized Cellulosic Fibers with Nonleaching Bacteria Adsorbing Properties.* Environmental Science & Technology, 2017. **51**(13): p. 7616-7623.
101. Elsacker, E., et al., *Mechanical characteristics of bacterial cellulose-reinforced mycelium composite materials.* Fungal Biol Biotechnol, 2021. **8**(1): p. 18.
102. Lee, J.S. and J. Han, *Exploring the potential of bacterial cellulose paste as a fat replacer for low-fat plant-based hamburger patties.* Food Res Int, 2024. **176**: p. 113832.
103. Amorim, L.F.A., et al., *Sustainable bacterial cellulose production by low cost feedstock: evaluation of apple and tea by-products as alternative sources of nutrients.* Cellulose, 2023. **30**(9): p. 5589-5606.
104. Tapias, Y.A.R., et al., *Bacterial cellulose films production by Kombucha symbiotic community cultured on different herbal infusions.* Food Chemistry, 2022. **372**.
105. Guan, Q.F., et al., *Growing Bacterial Cellulose-Based Sustainable Functional Bulk Nanocomposites by Biosynthesis: Recent Advances and Perspectives.* Accounts of Materials Research, 2022. **3**(6): p. 608-619.
106. Rainey, P.B., M.J. Bailey, and I.P. Thompson, *Phenotypic and genotypic diversity of fluorescent pseudomonads isolated from field-grown sugar beet.* Microbiology (Reading), 1994. **140 (Pt 9)**: p. 2315-31.
107. Kusmierska, A. and A.J. Spiers, *New Insights into the Effects of Several Environmental Parameters on the Relative Fitness of a Numerically Dominant Class of Evolved Niche Specialist.* Int J Evol Biol, 2016. **2016**: p. 4846565.

108. Jyot, J., N. Dasgupta, and R. Ramphal, *FleQ, the major flagellar gene regulator in Pseudomonas aeruginosa, binds to enhancer sites located either upstream or atypically downstream of the RpoN binding site.* J Bacteriol, 2002. **184**(19): p. 5251-60.
109. Barembruch, C. and R. Hengge, *Cellular levels and activity of the flagellar sigma factor FliA of Escherichia coli are controlled by FlgM-modulated proteolysis.* Mol Microbiol, 2007. **65**(1): p. 76-89.
110. Santos, F.B. and L.E. Del-Bem, *The Evolution of tRNA Copy Number and Repertoire in Cellular Life.* Genes, 2023. **14**(1).
111. Pal, A. and D.I. Andersson, *Bacteria can compensate the fitness costs of amplified resistance genes via a bypass mechanism.* Nature Communications, 2024. **15**(1).
112. Anderson, P. and J. Roth, *Spontaneous tandem genetic duplications in Salmonella typhimurium arise by unequal recombination between rRNA (rrn) cistrons.* Proc Natl Acad Sci U S A, 1981. **78**(5): p. 3113-7.
113. Lo, Y.L., et al., *Characterization of the role of global regulator FliA in the pathophysiology of Pseudomonas aeruginosa infection.* Res Microbiol, 2018. **169**(3): p. 135-144.
114. Cao, S., et al., *Positive Selection during Niche Adaptation Results in Large-Scale and Irreversible Rearrangement of Chromosomal Gene Order in Bacteria.* Molecular Biology and Evolution, 2022. **39**(4).
115. Daubin, V. and G.J. Szollosi, *Horizontal Gene Transfer and the History of Life.* Cold Spring Harb Perspect Biol, 2016. **8**(4): p. a018036.
116. Kurt Lienau, E., et al., *The mega-matrix tree of life: using genome-scale horizontal gene transfer and sequence evolution data as information about the vertical history of life.* Cladistics, 2011. **27**(4): p. 417-427.
117. Soucy, S.M., J. Huang, and J.P. Gogarten, *Horizontal gene transfer: building the web of life.* Nat Rev Genet, 2015. **16**(8): p. 472-82.
118. Balaban, N.Q., et al., *A problem of persistence: still more questions than answers?* Nature Reviews Microbiology, 2013. **11**(8): p. 587-591.
119. Balaban, N.Q., et al., *Bacterial persistence as a phenotypic switch.* Science, 2004. **305**(5690): p. 1622-1625.
120. Lewis, K., *Persister cells, dormancy and infectious disease.* Nat Rev Microbiol, 2007. **5**(1): p. 48-56.
121. Biselli, E., S.J. Schink, and U. Gerland, *Slower growth of Escherichia coli leads to longer survival in carbon starvation due to a decrease in the maintenance rate.* Mol Syst Biol, 2020. **16**(6): p. e9478.

Appendix

Appendix 1

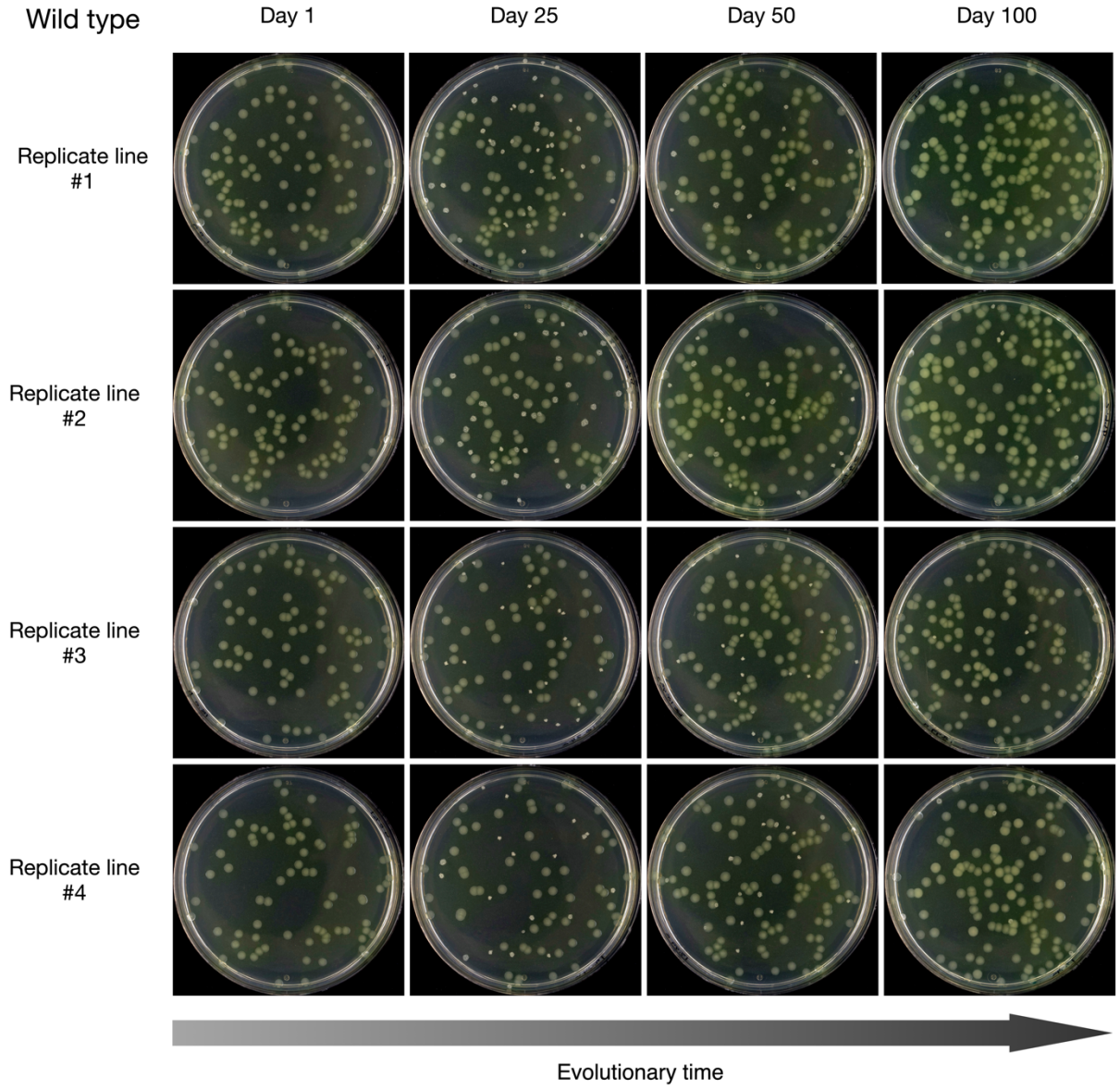
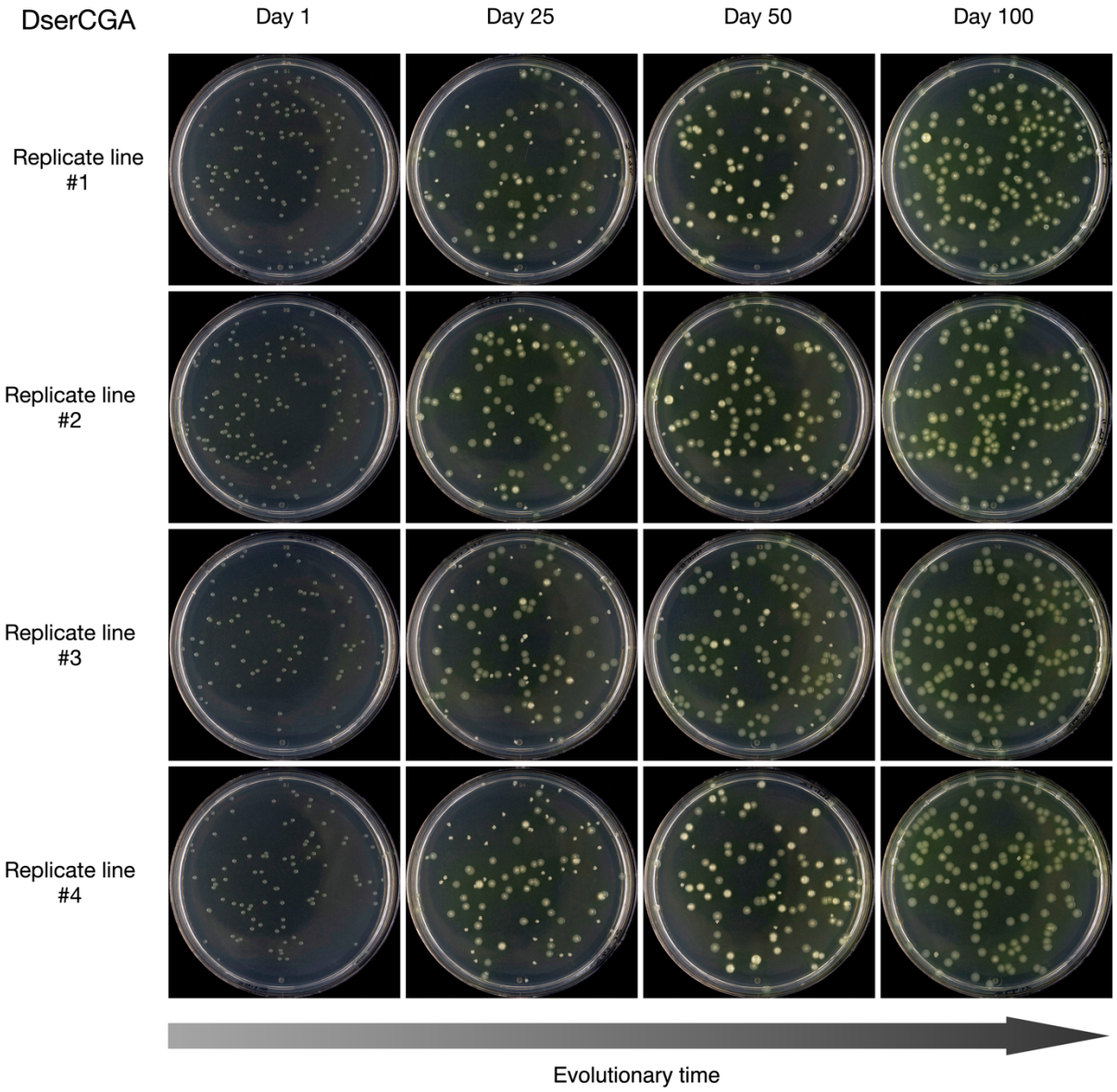
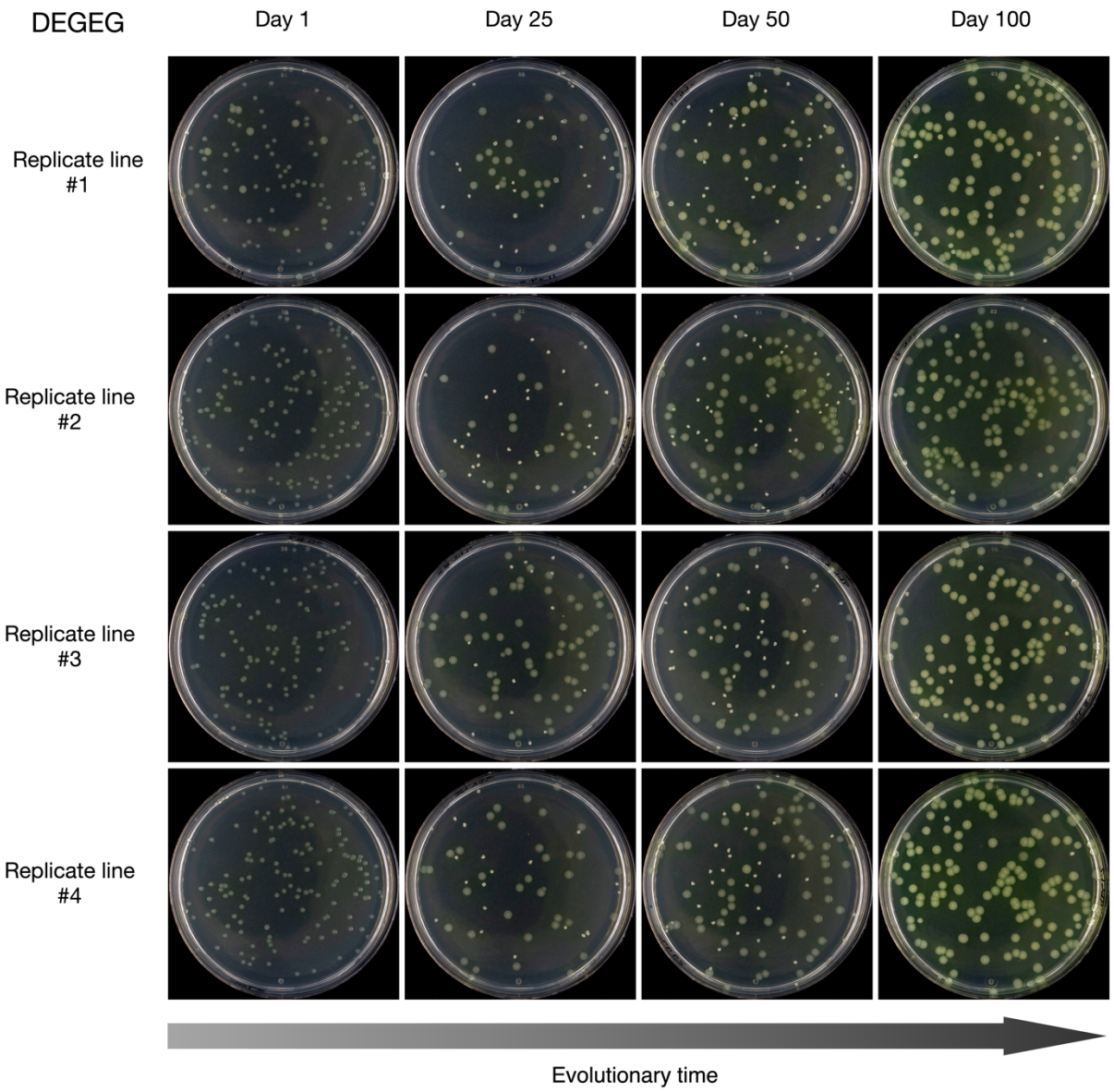


Fig S1.

Appendix 2


DserCGA
Day 1
Day 25
Day 50
Day 100
Replicate line #1
Replicate line #2
Replicate line #3
Replicate line #4
Evolutionary time

Appendix 3


DEGEG
Day 1
Day 25
Day 50
Day 100
Replicate line #1
Replicate line #2
Replicate line #3
Replicate line #4
Evolutionary time

Appendix 4

Use the Qiagen DNeasy Blood and Tissue kit to extract DNA from each sample as follows (use sterile filter tips all through the protocol):

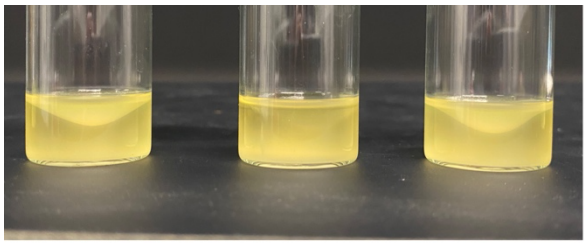
1. Re-suspend pellet (by gentle pipetting) in 180 ul Buffer ATL.
2. Add 20 ul proteinase K. Mix thoroughly by inversion and incubate at 56˚C in the shaking thermomixer (no additional vortexing) until the tissue is completely lysed **(*for 2 hours***).
3. Add 4 ul of RNase A and mix by inversion. Incubate at room temp for 30 min.
4. During the 30 mins incubation above, mix together 200 ul buffer AL and 200 ul of 100% ethanol in a 15 ml falcon tube (*per* sample). Mix well by inversion.
5. Mix each sample again by inverting several times.
6. Add 400 ul of the Buffer AL and ethanol mix, and mix immediately by thorough inversion.
7. Gently pipette the mixture from step 8 onto the DNeasy Mini spin column placed inside a 2 ml collection tube. Let stand on desktop for 1 minute.

8. Centrifuge at 8,000 *x*g for 1 min. Discard flow through and collection tube.
9. Place the DNeasy Mini spin column in a new 2 ml collection tube, add 500 ul buffer AW1 and centrifuge at 8,000 *x*g for 1 min. Discard flow through and collection tube.
10. Place the DNeasy Mini spin column in a new 2 ml collection tube, add 500 ul buffer AW2.
11. Centrifuge at 17,000 *x*g for 3 min. Carefully discard the flow through, replace in same collection tube and spin for an additional minute. Carefully discard the flow through flow through and column.
12. Place the DNeasy Mini spin column in a sterile 1.5 ml eppi.
13. Add **50 ul** of sterile ddH_2O (heated to 60˚C). Incubate at 28oC for 6 min.
14. Centrifuge at 13,000 xg for 1 min to elute.

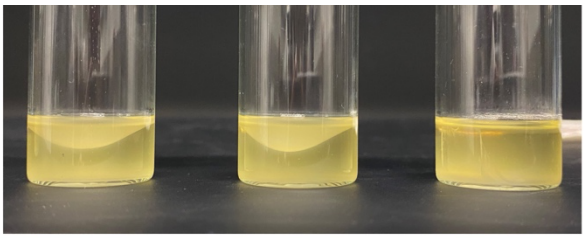
Nanodrop each sample and record quantity, 260:230 ratio, 260:280 ratio.

Use 3 ul of each sample plus 2 ul loading dye to run a 1% agarose gel (100V, 50 min, against 1 kb DNA ladder; 3uL ladder + 2uL loading; 300ng)

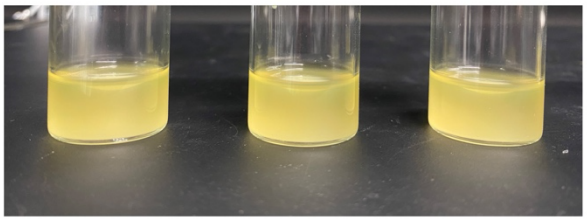
Appendix 5



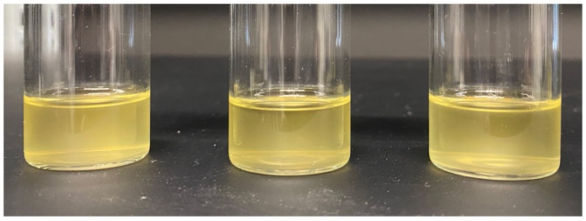
DelserCGA



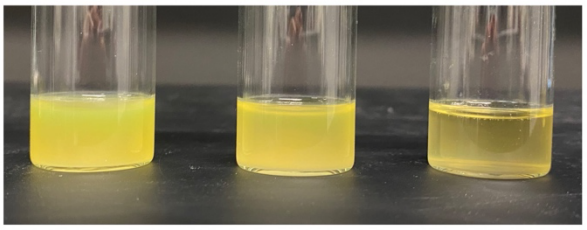
DelEGEG



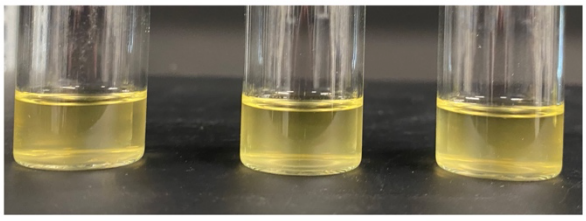
Wild type



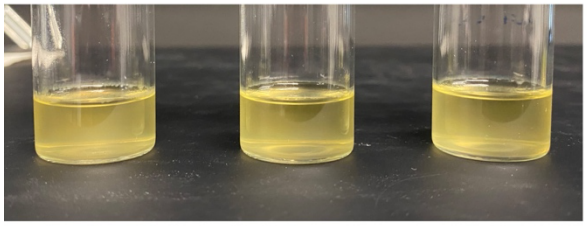
Large smooth 1



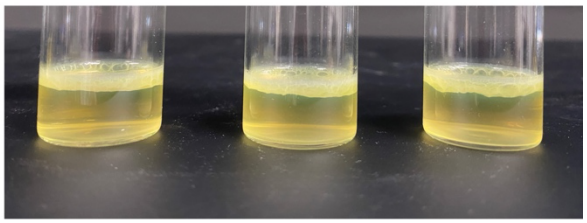
Round disc



Small smooth



Opaque



Wrinkly spreader