The birth of a bacterial tRNA gene by large-scale, tandem duplication events

Gökçe B. Ayan[1], Hye Jin Park[1,2], Jenna Gallie[1]*

[1]Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, 24306 Plön, Germany
[2]Asia Pacific Center for Theoretical Physics, Pohang, 37673, Korea
*To whom correspondence should be addressed: gallie@evolbio.mpg.de

ABSTRACT

Organisms differ in the types and numbers of tRNA genes that they carry. While the evolutionary mechanisms behind tRNA gene set evolution have been investigated theoretically and computationally, direct observations of tRNA gene set evolution remain rare. Here, we report the evolution of a tRNA gene set in laboratory populations of the bacterium *Pseudomonas fluorescens* SBW25. The growth defect caused by deleting the single-copy tRNA gene, *serCGA*, is rapidly compensated by large-scale (45-290 kb) duplications in the chromosome. Each duplication encompasses a second, compensatory tRNA gene (*serTGA*) and is associated with a rise in tRNA-Ser(UGA) in the mature tRNA pool. We postulate that tRNA-Ser(CGA) elimination increases the translational demand for tRNA-Ser(UGA), a pressure relieved by increasing *serTGA* copy number. This work demonstrates that tRNA gene sets can evolve through duplication of existing tRNA genes, a phenomenon that may contribute to the presence of multiple, identical tRNA gene copies within genomes.

INTRODUCTION

Even though tRNAs perform the same canonical function in all organisms – decoding 61 sense codons into 20 amino acids – tRNA gene sets vary considerably across the tree of life (Fujishima and Kanai, 2014; Marck and Grosjean, 2002). Two aspects in which they vary are the types of tRNAs encoded, and the number of gene copies encoding each type. Bacterial tRNA complements typically contain 28-46 types of tRNA, encoded by 28-120 genes (Chan and Lowe, 2016). Elucidating the factors contributing to, and the molecular mechanisms behind, the evolution of these variations has been of long-standing interest in biology. There is general agreement that bacterial tRNA gene sets are, in conjunction with the rest of the translational machinery, shaped by selection for rapid and accurate protein synthesis ("translational efficiency"; reviewed in Gingold and Pilpel, 2011). Efficient translation is an important determinant of bacterial growth rate, with more efficient protein production enabling faster growth and division (Kurland et al., 1996; Kurland and Ehrenberg, 1987).

tRNAs mainly contribute to translational efficiency during elongation, the stage of translation where codons are sequentially matched to aminoacylated tRNAs (reviewed in Gingold and Pilpel, 2011; Rodnina, 2018). Codon-tRNA matching occurs by a trial and error process; tRNAs – in the form of ternary complexes (Bensch et al., 1991) – are stochastically sampled from the available pool. The speed with which a matching tRNA is selected depends on the absolute and relative concentration of tRNAs

that match the codon; codons matched by more abundant tRNAs are expected, on average, to be translated more quickly than those matched by rarer tRNAs (reviewed in Plotkin and Kudla, 2011). Given that the formation of codon-tRNA matches by stochastic sampling is the rate-limiting step of elongation (Varenne et al., 1984), any factors affecting the matching process are likely to influence the evolution of bacterial tRNA gene sets. Examples include variations in codon-tRNA matching patterns, and codon bias. Each of these is discussed in more detail below.

Codon-tRNA matching patterns are complex; some tRNAs match, and hence translate, more than one synonymous codon (Crick, 1966; Ikemura, 1981). This expanded translational capacity is the result of relaxed base pairing between the first base of the tRNA anticodon (tRNA position 34) and third base of the codon (codon position 3). In this binding position, a number of non-standard, "wobble" pairings are permitted (reviewed in Agris et al., 2018). Most wobble pairings involve post-transcriptionally, enzymatically modified bases in the anticodon stem-and-loop region of tRNAs (Boccaletto et al., 2018; Machnicka et al., 2016). These post-transcriptional modifications affect the binding capacity and/or accuracy of a large fraction of bacterial tRNAs (Björk and Hagervall, 2014; Manickam et al., 2015). Hence, the set of post-transcriptional modification pathways active within a bacterial species affects the codon-tRNA matching pattern and, in turn, is expected to influence tRNA gene set composition. Indeed, various post-transcriptional modification pathways have been shown to co-vary with tRNA repertoires (Diwan and Agashe, 2018; Grosjean et al., 2010; Novoa et al., 2012).

Codon bias refers to the preferential use of some synonymous codons over others. A number of hypotheses exist regarding the evolutionary origins and consequences of preferred codons (reviewed in Novoa and Ribas de Pouplana, 2012; Plotkin and Kudla, 2011), one of which is the optimization of codon-tRNA matching during elongation (Berg and Kurland, 1997; Bulmer, 1991, 1987; Higgs and Ran, 2008; Rocha, 2004). There are several lines of support for this hypothesis. Firstly, different codons have been demonstrated to be translated at different rates in yeast, with more frequent codons generally being translated more quickly (Gardin et al., 2014). Secondly, codon and tRNA abundances have been observed to co-vary across many bacterial genomes (reviewed in Ikemura, 1985), particularly under conditions where rapid translation is required (*e.g.*, during rapid growth (Dong et al., 1996; Emilsson and Kurland, 1990), in highly expressed genes (Ikemura, 1981), or among bacteria with faster growth rates (Sharp et al., 2010)). Thirdly, a number of studies have demonstrated an increase in protein expression when codon-tRNA co-variation is strengthened, either by optimizing synonymous codon use (Sørensen et al., 1989; Zhou et al., 2004), or by the addition of exogeneous tRNAs (Gu et al., 2004; Misra and Reeves, 1985).

As outlined above, two factors expected to affect tRNA gene set evolution are the presence of tRNA post-transcriptional modification pathways, and codon bias; post-transcriptional modifications heavily influence codon-tRNA matching patterns (affecting which tRNA types are encoded), while codon bias dictates the translational demand for individual tRNA types (influencing tRNA abundances and hence tRNA gene copy number). Overall, theoretical and computational studies are consistent with

the optimization of translational efficiency by the streamlining of tRNA gene sets; bacterial growth rate correlates with fewer tRNA types encoded by more gene copies (Ran and Higgs, 2010; Rocha, 2004).

In addition to factors influencing the evolution of tRNA gene sets, the mechanisms behind their evolution are also an area of interest. Hypothetically, tRNA gene sets can evolve by several different mechanisms. Surplus tRNA genes may be lost (by deletion), while tRNA genes may be acquired from external sources (by horizontal gene transfer), or from within the genome (by duplication events). Additionally, the identity of existing tRNA genes may be altered by the acquisition of anticodon mutations (anticodon switching). Thus far, most evidence for the above routes of tRNA gene set evolution is indirect: phylogenetic analyses provide evidence of the flexibility of bacterial tRNA gene sets by loss, gain (both by horizontal gene transfer and duplication), and anticodon switch events (Diwan and Agashe, 2018; McDonald et al., 2015; Tremblay-Savard et al., 2015; Wald and Margalit, 2014; Withers et al., 2006). An anticodon switch event has also been directly observed in laboratory yeast; *Saccharomyces cerevisiae* populations in which the gene encoding tRNA-Arg(CCU) had been removed were repeatedly rescued by a C→T mutation in one of eleven gene copies encoding tRNA-Arg(UCU) (Yona et al., 2013). While the aforementioned study demonstrates the power of experimental evolution to provide insight into the evolution of tRNA gene sets, there remains a shortage of empirical studies directly investigating the evolution of bacterial tRNA gene sets and translation.

To address this, we (i) engineer a sub-optimal bacterial tRNA gene set, (ii) compensate the defect using experimental evolution, and (iii) determine the genetic and molecular bases of compensation. More specifically, we delete the single-copy tRNA gene, *serCGA*, from the bacterium *Pseudomonas fluorescens* SBW25. We compensate the resulting growth defect during a 13-day serial transfer evolution experiment, and show that the genetic basis of compensation is large-scale (45-290 kb), tandem duplications encompassing a second tRNA gene (*serTGA*). Using a bacterial adaptation of YAMAT-seq (a method of tRNA pool deep-sequencing originally developed for use in human cell lines (Shigematsu et al., 2017)), we demonstrate that each duplication event is accompanied by an increase in tRNA-Ser(UGA) in the mature tRNA pool. Finally, we develop a model that combines our experimental results with the predicted effects of codon-tRNA matching patterns and codon bias, to provide a molecular explanation of how the observed tRNA pool changes may affect translation and growth. Together, our results provide a direct observation of the evolution of a bacterial tRNA gene set by gene duplication, and lend empirical support for the optimization of translation by codon-tRNA matching.

RESULTS

The *P. fluorescens* SBW25 tRNA gene set

Isolated from the leaf of a sugar beet plant, *P. fluorescens* SBW25 is a non-pathogenic bacterium that is frequently used as a model system in evolutionary biology. The SBW25 genome (Silby et al., 2009) is predicted by GtRNAdb 2.0 (Chan and Lowe, 2016) to contain 67 tRNA genes (Figure 1A; supplemental file 1). The RNA product of one of these genes (*cysGCA-2*) is predicted to form a secondary structure

that deviates significantly from the cloverleaf structure typical of canonical tRNAs (Chan and Lowe, 2019). Hence, the SBW25 tRNA gene set consists of 66 canonical tRNA genes. These encode 39 different tRNA types, 14 of which are encoded by multiple (between two and five) gene copies. Twelve of these 14 types are encoded by gene copies that are identical in sequence; only tRNA-Asn(GUU) and tRNA-fMet(CAU) are encoded by multiple gene copies with different sequences.

The 39 different tRNA types identified in SBW25 must together be capable of translating all 61 sense codons. To investigate SBW25 codon-tRNA matching patterns, a combination of current wobble rules (reviewed in Agris et al., 2018) and tRNA post-transcriptional modification prediction tools (Boccaletto et al., 2018; Machnicka et al., 2016; Panwar and Raghava, 2014) were applied. Evidence for the translation of multiple synonymous codons was found for 28 of 39 tRNA types (supplemental file 1). The proposed SBW25 codon-tRNA matching patterns indicate that the 39 tRNA types consist of a core set of 33 essential types (that together are theoretically capable of translating all 61 codons), plus six non-essential types (whose cognate codons are predicted to also be translated by an essential tRNA type) (Figure 1A; supplemental file 1). The six non-essential types are candidates for sub-optimal tRNA gene set construction; their apparent functional redundancy suggests that they can be eliminated, while their retention indicates that elimination may be detrimental.

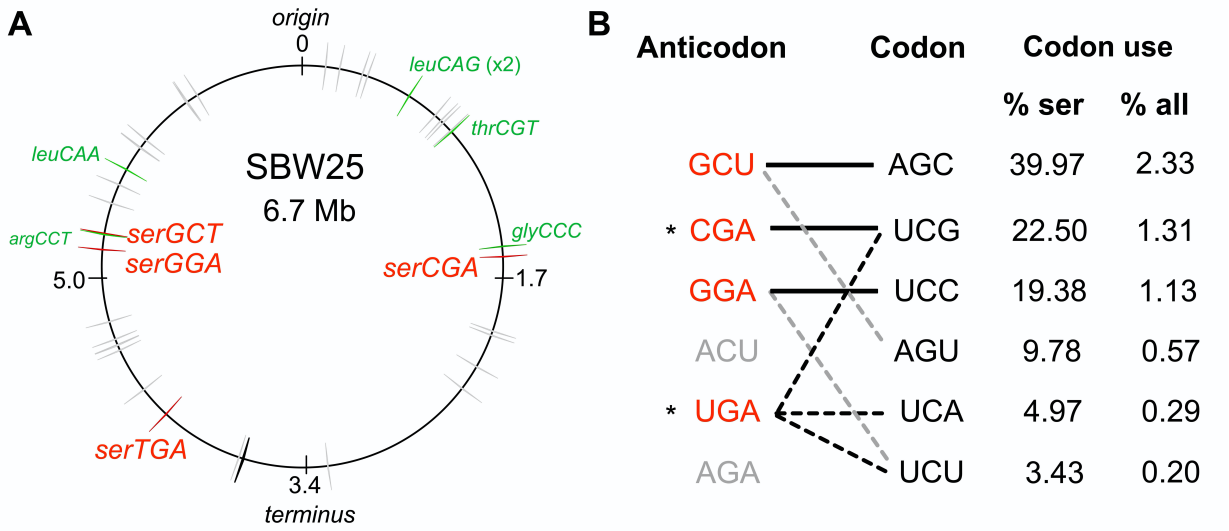

Figure 1. The tRNA gene set and serine translation system in *P. fluorescens* SBW25. (**A**) Genomic location of the 66 canonical (grey arrows) and one non-canonical (*cysGCA-2*; black arrow) tRNA genes. Four tRNA genes encode seryl-tRNAs (red arrows). One of these, *serCGA*, is predicted to encode a non-essential tRNA type. Six other tRNA genes encoding the remaining five non-essential tRNA types (green arrows). Replication origin and terminus are indicated. (**B**) The predicted translational relationship between seryl-tRNAs and serine codons. The six theoretically possible seryl-tRNA anticodons are listed on the left (red=present in SBW25, grey=absent, *=theoretically capable of translating codon UCG), and six cognate codons are listed in column 2. Connections signify a theoretical match (solid black lines=Watson-Crick pairing; black dotted lines=wobble pairing through post-transcriptional modification; grey dotted line=G-U wobble pairing). Columns 3 and 4 list codon use as a percentage of serine and all codons, respectively (Chan and Lowe, 2016). Anticodons and codons are 5'→3'.

With the aim of constructing a sub-optimal tRNA gene set, we focussed our attention on *serCGA*, a single-copy gene encoding the non-essential tRNA type, tRNA-Ser(CGA) (Figure 1A). According to the predicted codon-tRNA matching pattern, the cognate codon of tRNA-Ser(CGA), UCG, can also be translated by the essential tRNA type, tRNA-Ser(UGA) (Figure 1B; supplemental file 1). In several Gram-negative bacteria, tRNA-Ser(UGA) is post-transcriptionally modified at U_{34} to 5-methoxycarbonylmethoxyuridine ($mcmo^5U_{34}$; Boccaletto et al., 2018), resulting in the expansion of translational capacity from codon UCA to include UCG and UCU (Takai et al., 1999a, 1996) (see Figure 1 – supplemental figure 1C). The $mcmo^5U_{34}$ modification is performed by the CmoA/B/M enzymatic pathway (Björk and Hagervall, 2014; Sakai et al., 2016). Homologues of these enzymes are present in SBW25; protein BLAST searches of *E. coli* MG1655 CmoA, CmoB, and CmoM against the SBW25 proteome give significant hits to Pflu1067, Pflu1066, and Pflu0633, respectively (BLASTp e-values<1e-50; Altschul et al., 1990).

CmoA/B/M-mediated expansion of tRNA-Ser(UGA) translational capacity to include codon UCG is expected to rescue a *serCGA* deletion mutant. Such a rescue event would require all UCG (and UCA) codons to be translated by tRNA-Ser(UGA). Given that UCG is a relatively high use codon in SBW25 – accounting for 22.5 % of serine and 1.31 % of all codons (Figure 1B) – *serCGA* deletion is expected to considerably increase translational demand for tRNA-Ser(UGA). Hence, while a *serCGA* deletion mutant may survive, significant translation and growth defects are anticipated.

Deletion of *serCGA* limits rapid growth

To test the prediction that *serCGA* deletion results in a viable strain with a growth defect, the entire 90-bp, single-copy *serCGA* gene was removed from *P. fluorescens* SBW25 by two-step allelic exchange (see supplemental file 2 for construction details). This process changed the tRNA gene set from 66 tRNA genes of 39 types, to 65 tRNA genes of 38 types. The engineering process was performed two independent times, yielding biological replicate strains Δ*serCGA*-1 and Δ*serCGA*-2. A third round of allelic exchange gave rise to an engineering control strain, SBW25-eWT. Successful deletion of *serCGA* demonstrates that it is not essential for survival; at least one other tRNA can translate codon UCG.

serCGA deletion results in an immediately obvious growth defect on King's medium B (KB), a rich medium that supports rapid growth and hence rapid translation (King et al., 1954). Δ*serCGA*-1 produces visibly smaller colonies than SBW25 on KB agar (Figure 2A), and shows a reduction in maximum growth rate in liquid KB (two-sample *t*-test $p=1.18\times10^{-6}$, Figures 2B, 2C). Further, both independent deletion mutants lose when grown in direct, 1:1 competition with the neutrally marked SBW25-*lacZ* in liquid KB (one sample *t*-tests $p<0.0001$; Figure 2D). The growth defect is much less pronounced in minimal media, which supports slower growth; SBW25 and Δ*serCGA*-1 colonies are similar sizes on M9 agar (Figure 2E), and no negative effects were detected in liquid M9 (Figures 2F, 2G). However, direct 1:1 competitions between Δ*serCGA* and SBW25-*lacZ* indicate a slight negative effect of *serCGA* deletion in liquid M9 (one sample *t*-tests $p=0.06882$ and 0.00908; Figure 2H).

The results in this section demonstrate that, in accordance with the predictions of the previous section, *serCGA* deletion generates a viable, but sub-optimal, tRNA gene set in SBW25. Further, the growth defect caused by *serCGA* deletion is more pronounced in KB than M9, indicating that while the product of *serCGA* is not essential for translation, it contributes to translational speed.

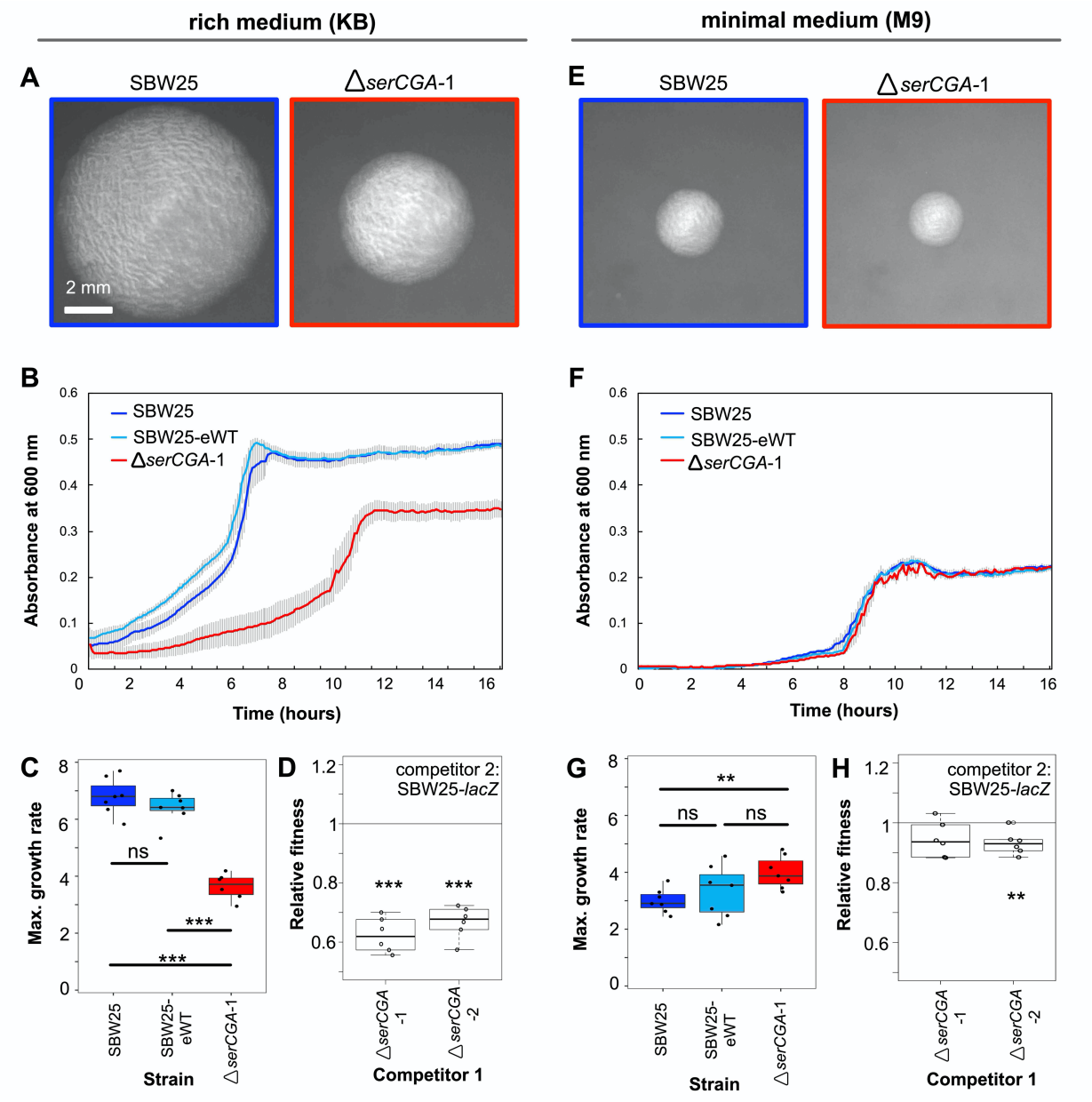

Figure 2. Deletion of *serCGA* leads to a growth defect in a rich medium. (**A**) 45-hour colonies on KB agar. (**B**) Growth (absorbance at 600 nm) in KB. Lines=mean of six (Δ*serCGA*-1) or seven (SBW25 and SBW25-eWT) replicates; error bars=one standard error. (**C**) Maximum growth rate (change in mOD min^{-1}) in KB, calculated with a sliding window of 9 points between hours 2 and 12. (**D**) Relative fitness values from direct 1:1 competitions between competitor 1 (Δ*serCGA*-1 or Δ*serCGA*-2) and the neutrally marked wild type strain, SBW25-*lacZ* (six replicates *per* competition), in KB. Relative fitness of 1=no difference, <1=SBW25-*lacZ* wins, >1=competitor 1 wins. (**E**) 45-hour colonies on M9 agar, at same magnification and time as in panel A. (**F**) Growth in M9. Lines=mean of seven replicates, error bars=one standard error. (**G**) Maximum growth rate (change in mOD min^{-1})

in M9, calculated with a sliding window of 9 points between hours 2 and 12. (**H**) Relative fitness values from direct 1:1 competitions between competitor 1 (Δ*serCGA*-1 or Δ*serCGA*-2) and SBW25-*lacZ* (six replicates *per* competition), in M9. Parametric two-tailed two-sample *t*-tests (panel C, G) and parametric two-tailed one sample *t*-tests (panel D, H) ***$p<0.001$, **$p<0.01$, *$p<0.05$, ns=not significant ($p>0.05$).

The growth defect is repeatedly and rapidly compensated during experimental evolution

In order investigate whether and how the growth defect exhibited by the *serCGA* deletion mutant can be compensated for genetically, a serial transfer evolution experiment was performed. This experiment consisted of eight independent lineages: W1-W4 were control lines, each founded by a wild type strain (W1 and W2 by SBW25, W3 and W4 by SBW25-eWT), while M1-M4 were founded by the *serCGA* deletion mutant (M1 and M2 by Δ*serCGA*-1, M3 and M4 by Δ*serCGA*-2). Each lineage was started from a single colony and maintained in liquid KB for 15 days, with daily transfer of 1 % to fresh medium. Samples of each population were frozen daily.

After 13 days (~90 generations), all four mutant lineages (M1, M2, M3, and M4) showed visibly improved growth. Plating of Day 13 populations on KB agar revealed that many colonies from lineages M1-M4 were larger than those of the founding *serCGA* deletion mutant (Figure 3A). Notably, lineage M2 showed two phenotypically distinct types of large colonies: a phenotypically standard type, and an opaque type. The opaque type closely resembles the previously reported switcher phenotype, in which on-off switching of colanic acid based capsules generates opaque-translucent colony bistability (Beaumont et al., 2009; Gallie et al., 2019, 2015; Remigi et al., 2019). No obvious change was observed in the size of colonies derived from Day 13 of the four wild type lineages (Figure 3A).

Five colonies were isolated from the Day 13 mutant lineages for further analysis. These included one standard-looking, large colony from each mutant lineage (these isolates are hereafter referred to as M1-L, M2-L, M3-L, and M4-L), and a second large, opaque colony from lineage M2 (hereafter M2-Lop). Two representative colonies were isolated from Day 13 of different wild type lineages (hereafter W1-L and W3-L). Growth analyses in liquid KB showed improved growth profiles in all five isolates from the mutant lineages (Figures 3B-D). In line with the observed improvement, each of the five mutant lineage isolates outcompeted the *serCGA* deletion mutant in direct, 1:1 competition (one sample *t*-tests $p<0.01$; Figure 3E). Indeed, no fitness difference was detected in 1:1 competitions between two mutant lineage isolates and neutrally-marked SBW25-*lacZ*, providing evidence for high levels of compensation in these genotypes (M1-L and M4-L; one sample *t*-tests $p>0.05$; Figure 3E). The other three mutant lineage isolates were outcompeted by SBW25-*lacZ*, indicating partial compensation (M2-L, M2-Lop, and M3-L; one sample *t*-tests $p<0.01$; Figure 3E). No changes were observed in the growth or fitness of W1-L, the control isolate from Day 13 of wild type lineage 1 (Figure 3A-E).

The results in this section demonstrate that the growth defect caused by the deletion of *serCGA* was repeatedly and rapidly compensated, to varying degrees, in isolates from each of the four mutant lineages on Day 13 of the evolution experiment.

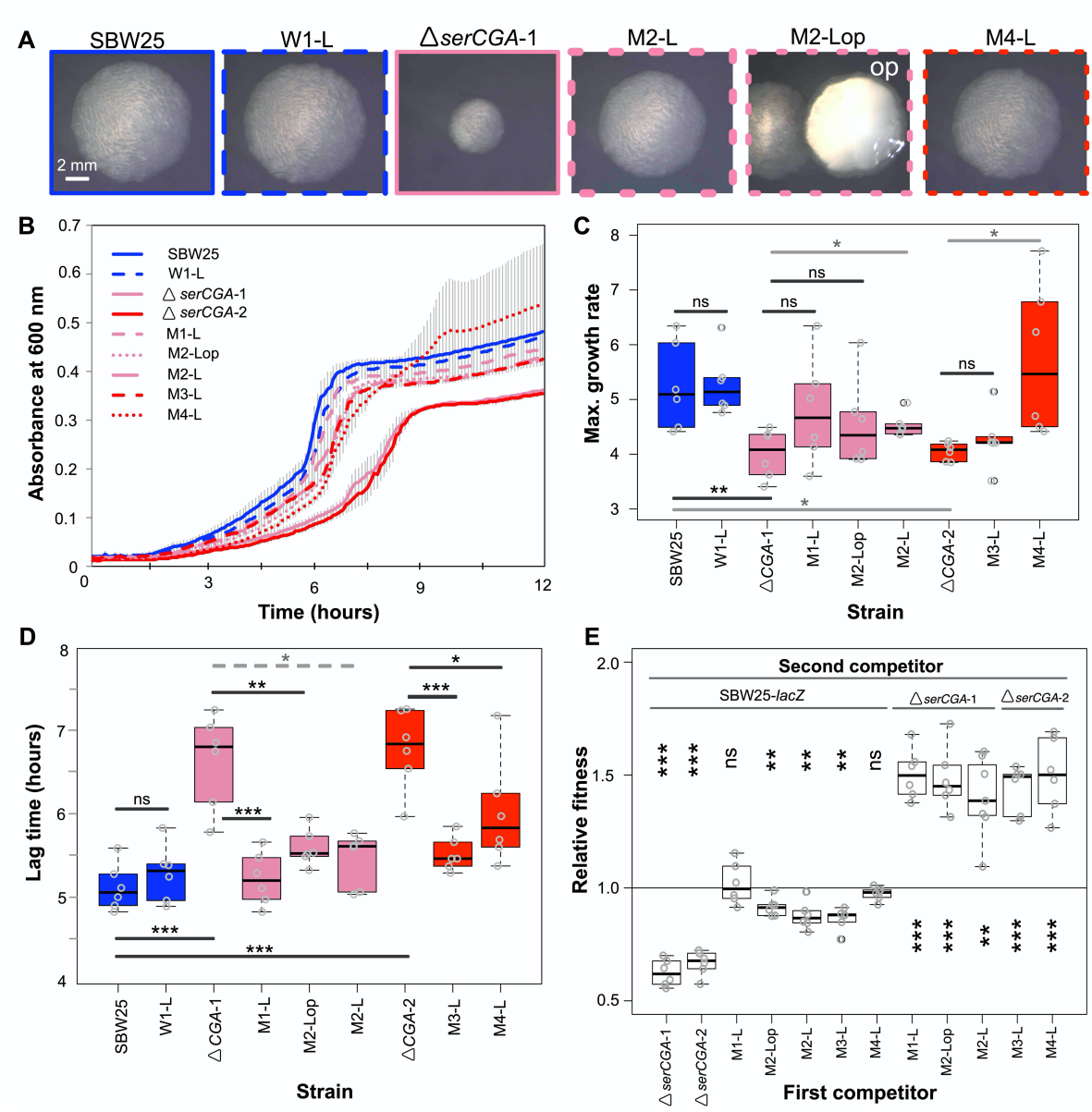

Figure 3. Loss of *serCGA* is repeatedly and rapidly compensated by experimental evolution. (**A**) Colony morphology of founder (solid outlines) and evolved (dotted outlines) isolates on KB agar (30 h, 28˚C). Lineage M2 yielded two large colony morphotypes: standard (left), and opaque ("op", right). Image border colours match line colours in panel B. (**B**) 12-hour growth curves in liquid KB for founder (Day 0, solid lines) and evolved (Day 13, dotted lines) isolates. Lines=mean of six replicates, error bars=1 standard error. (**C, D**) Box plots showing the maximum growth rate (change in mOD min^{-1}) and lag time (hours) of founding and evolved strains from the evolution experiment, grown in liquid KB (n=6; maximum growth rates and lag times calculated using a sliding window of 9 points between 2 and 12 hours). Statistically significant differences were determined using parametric *t*-tests (solid black lines), non-parametric *t*-tests (solid grey lines), or Mann-Whitney-Wilcoxon rank sum tests (dotted grey line). (**E**) Box plots of the relative fitness of competitor 1 (*x*-axis) and competitor 2 (horizontal bars at top). Direct, 1:1 competitions were performed in liquid KB for 24 hours (28˚C, shaking). Six replicate competitions were performed for each set of strains. Relative fitness >1 means competitor 1 wins, <1 means competitor 2 wins. The first two competitions are also presented in Figure 2G. Statistically significant deviations of relative fitness from 1 were determined using parametric two-tailed one-sample *t*-tests. ***$p<0.001$, **$p<0.01$, *$p<0.05$, ns=not significant ($p>0.05$).

The genetic basis of compensation is large duplications spanning *serTGA*

To determine the genetic basis of Δ*serCGA* compensation, Illumina whole genome sequencing was performed on the seven Day 13 isolates from the previous section: two control isolates from two wild type lineages (W1-L, W3-L), and five isolates from four mutant lineages (M1-L, M2-L, M2-Lop, M3-L, M4-L). In each of the five mutant lineage isolates, a large, direct, tandem duplication was identified at around 4.16 Mb in the SBW25 chromosome (Figure 4A-B, Figure 4 – supplemental figure 1, supplemental file 3). No evidence of any such duplications was found in either of the wild type control isolates. In addition to the large duplications, one synonymous point mutation was identified in one mutant lineage: in the carbohydrate metabolism gene *edd* of M2-L, codon 17 is changed from CGC to CGA (both encoding arginine; supplemental file 3). This mutation was not identified in any other isolate, including the second isolate from the M2 lineage (M2-Lop). Overall, the synonymous *edd* mutation is not considered likely to contribute to the compensatory phenotype of M2-L.

A combination of computational analyses, PCR, and Sanger sequencing was used to determine the precise region of duplication in four out of five isolates. The duplications range in size from 45 kb (in M1-L) to 290 kb (in M4-L), and occur between 4.05 Mb – 4.34 Mb of the chromosome (Table 1). The precise location of the duplication in the fifth isolate (M3-L) could not be determined due to highly repetitive flanking DNA.

Strain name	Dup. size (bp)	Junction name	Junction side 1		Junction side 2	
			Base	Region	Base	Region
M1-L	45,043	M1junct	4,164,966	*murB*	4,119,923	intergenic repeat
M2-L	191,833	M2junct1	4,310,940 – 4,311,029	intergenic repeat	4,119,235 – 4,119,352	intergenic repeat
M2-Lop	182,877	M2junct2	4,224,306	*nuoL*	4,042,455	*pflu3649*
M3-L	~192,000	M3junct1	~4,310,800	intergenic repeat	~4,119,100	intergenic repeat
M4-L	290,335	M4junct1	4,339,314	Intergenic	4,048,979	*pflu3655*

Table 1. Duplication junctions in five isolates from the mutant lineages reveal duplication fragments of 45-290 kb. Base positions refer to the SBW25 wild type genome sequence (Silby et al., 2009). For details of the genes contained within each duplication segment see supplemental file 4.

Next, we sought to identify the region(s) within the duplications responsible for the observed gain in fitness. Closer examination revealed a ~45 kb segment that is duplicated in all five strains (4,119,923 – 4,164,966). This segment is predicted to contain 45 genes, including 44 protein-coding genes (none of which are obviously linked to translation), and one tRNA gene: *serTGA* (Figure 4A, supplemental file 4). Close manual inspection revealed no evidence of any point mutations in any copy of the *serTGA* gene or promoter sequence, in any of the evolved isolates (supplemental file 5), meaning that each duplication strain contains a second, wild-type copy of *serTGA*. In addition, isolate M4-L carries additional copies of four other tRNA genes (*argTCT*, *hisGTG, leuTAA, hisGTG*; supplemental

file 4). Notably, this experiment has revealed three different tRNA gene sets that each provide a similar level of fitness in KB (see Figure 3E): SBW25 encodes 66 canonical tRNA genes of 39 tRNA types, four of five compensated isolates carry 66 tRNA genes of 38 types, and the fifth compensated isolate (M4-L) carries 70 tRNA genes of 38 types.

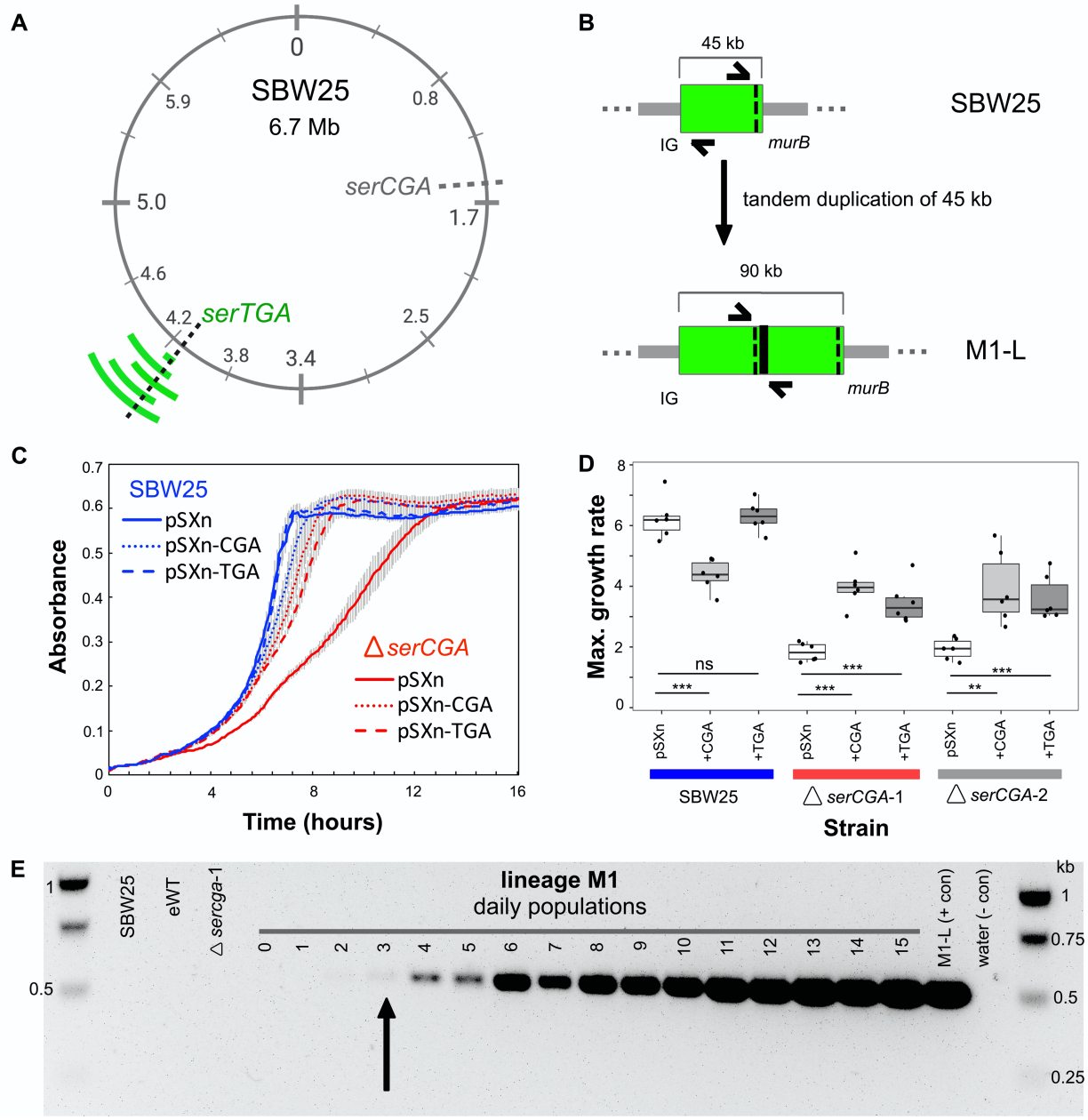

Figure 4. Direct, tandem duplications spanning *serTGA* compensate for *serCGA* loss. (A) Five isolates from the mutant Lines have unique, large, tandem duplications between 4.05 and 4.34 Mb of the SBW25 chromosome (green arcs; moving outwards: M1-L, M2-L, M2-Lop, M3-L, M4-L). The duplications contain a shared 45 kb region with *serTGA* (dotted black line) (see also Figure 4 – supplemental figure 1). (**B**) Cartoon depiction of the duplication event in M1-L, resulting in two copies of a 45 kb fragment (green) and an emergent junction (thick black line). The junction can be PCR-amplified using primers to either side (black arrows). IG=intergenic, black dotted line=*serTGA*. (**C**) 12-hour growth curves in LB+Gm (20 µg ml^{-1}) for Δ*serCGA*-1 (red) and SBW25 (blue) expressing *serCGA* or *serTGA* from the pSXn plasmid. Lines=mean of six replicates, error bars=1 standard error.

(**D**) Maximum growth speed (change in mOD min^{-1}; calculated with a sliding window of points between 0 and 23 hours) of SBW25, Δ*serCGA*-1, Δ*serCGA*-2 carrying empty pSXn, pSXn-*serCGA* ("+CGA") and pSXn-*serTGA* ("+TGA"). Parametric two-tailed, two-sample *t*-tests ***$p<0.001$, **$p<0.01$, *$p<0.05$, ns=not significant ($p>0.05$). (**E**) The duplication junction in lineage M1 was first definitively amplified from lineage M1 population on Day 3 (black arrow). Gel photograph colours were inverted using Preview to better detect faint PCR products. See Figure 4 – supplemental figure 2 for the history of other junctions.

Given that the duplicated *serTGA* gene encodes tRNA-Ser(UGA), the tRNA type that can theoretically perform the function of tRNA-Ser(CGA) (see Figure 1 – supplemental figure 1C), it is a logical candidate for the underlying cause of compensation. However, there are also 44 protein-coding genes in the shared duplication segment, any of which could contribute to the compensatory effect. We therefore tested whether a plasmid-based increase in *serTGA* expression can compensate for *serCGA* loss. To this end, the *serCGA* and *serTGA* genes were individually amplified from SBW25, and each ligated into the expression vector pSXn (giving pSXn-CGA and pSXn-TGA; supplemental file 2). This placed the expression of the tRNA gene under the control of an isopropyl-ß-D-1-thiogalactopyranoside (IPTG)-inducible *tac* promoter (de Boer et al., 1983; Owen and Ackerley, 2011).

Each plasmid construct was inserted into SBW25, Δ*serCGA*-1, and Δ*serCGA*-2, and the growth of the resulting strains was analysed in the absence of the inducer (to achieve lower-level, leaky expression of the tRNA gene). Expression of either *serCGA* or *serTGA* was shown to improve the growth of the *serCGA* deletion mutants in rich medium; addition of pSXn-CGA or pSXn-TGA increases the maximum growth rate of Δ*serCGA*, while addition of empty pSXn does not (Figure 4C-D; two-sample *t*-tests). Contrastingly, expression of neither tRNA improved growth of SBW25 (Figure 4C), with *serCGA* expression actually leading to a decrease in SBW25 maximum growth rate (Figure 4D; one-sided two-sample *t*-test $p=0.000158$). While it should be noted that tRNA-Ser(UGA) levels resulting from pSXn-based expression are likely to exceed those resulting from an additional chromosomal copy of *serTGA*, the result that pSXn-based *serTGA* expression specifically improves the growth rate of Δ*serCGA* demonstrates that *serTGA* can provide a degree of compensation for *serCGA* loss. Other genes in the shared 45 kb fragment may nevertheless contribute to compensation, *via* unidentified mechanisms.

Thus far, our results provide strong evidence that the growth defect caused by *serCGA* deletion is repeatedly and rapidly compensated by one of several large-scale duplications encompassing *serTGA*.

Large-scale duplications arise quickly and are heterogeneous

The duplication-carrying strains were isolated from all mutant lineages on Day 13, demonstrating that the large-scale duplications occur repeatedly and rapidly. To more closely investigate the rapidity with which the duplications arose, a PCR was performed to identify the first time point at which the emergent M1-L duplication junction, M1junct1 (see supplemental file 2), could be amplified from lineage M1 (see Figure 4B). The lineage M1 population samples frozen daily during the evolution

experiment were revived, and used as templates for the PCR. The M1-L duplication junction was first visible by PCR on Day 3, and grew stronger as the experiment progressed (Figure 4E). Similar PCRs for the remaining duplication strains detected the presence of the relevant duplication junctions on Day 2 (M2-Lop), Day 4 (M3-L), and Day 5 (M2-L, M4-L) (Figure 4 – supplemental figure 2).

A notable feature of the duplication fragments is that they are heterogeneous. That is, each of the five compensated isolates carries a unique duplication fragment with distinct endpoints (Table 1). Further, there is evidence of within-lineage heterogeneity; two distinct duplication fragments were identified in isolates from a single lineage (M2-L, M2-Lop; Table 1). To investigate within-lineage heterogeneity more closely, additional large colonies were isolated from Day 13 of each mutant lineage and tested by PCR for the presence of the duplication junction(s) previously identified in the lineage. The results provide evidence for genotypic heterogeneity in compensated isolates from three of the four mutant lineages (M2, M3, and M4; Figure 4 – supplemental figure 2). Three differently-sized PCR products were detected among the four isolates from lineage M2, demonstrating the presence of at least three distinct duplication fragments in the Day 13 population. A mixture of compensated genotypes was also detected in lineages M3 and M4, with duplication junctions M3junct1 and M4junct1 amplifying in only three (of four) and two (of four) isolates, respectively. No PCR products were detected for the remaining isolates in these lineages, indicating that these three isolates either carry a duplication junction that was not tested for, or compensate by a different mechanism.

The results in this section demonstrate that (i) strains carrying duplication fragments arise early within the mutant lineages of the evolution experiment (within 2-5 days, or ~7-35 generations), and (ii) a degree of heterogeneity exists in duplication fragments, both between and within mutant lineages. These observations suggest that a mixture of duplication strains – and, by extension, tRNA gene sets – arise and compete within each mutant lineage.

Duplication events increase the proportion of tRNA-Ser(UGA) in the mature tRNA pool

Next, we sought to quantify the effect of *serCGA* deletion and subsequent *serTGA* duplication on the mature tRNA pool. To this end, YAMAT-seq – an established method of deep-sequencing tRNA pools in human cells (Shigematsu et al., 2017) – was adapted for use in SBW25. Briefly, YAMAT-seq involves (i) isolation of total RNA from growing cells, (ii) removal of amino acids from mature tRNAs to expose the universally conserved 5'-CCA-3' tRNA ends, (iii) ligation of Y-shaped, DNA/RNA hybrid adapters to the exposed 5'-CCA-3' ends, (iv) reverse transcription and amplification of adapter-tRNA complexes, (v) high throughput sequencing, and (vi) computational and statistical analyses.

YAMAT-seq was performed on three replicates of nine strains: wild type (SBW25), the two independent *serCGA* deletion mutants (Δ*serCGA*-1, Δ*serCGA*-2), and six isolates from Day 13 of the evolution experiment (W1-L, M1-L, M2-L, M2-Lop, M3-L, M4-L). High throughput sequencing of the reverse-transcribed tRNA pools resulted in an average of 1,177,340 raw reads *per* sample. More than 99.99 % of the raw reads fell within the range of lengths expected for tRNA-containing reads (80-150 bp), indicating good adapter binding specificity. The >80 bp reads for each sample were aligned to a set

of 42 reference tRNA sequences, consisting of all unique tRNA gene sequences predicted in the SBW25 chromosome (including *cysGCA*-2; supplemental file 6; Chan and Lowe, 2016). At the conclusion of the alignment process, an average of 1,050,749 reads *per* sample were aligned to the reference sequences (~89 % read retention; supplemental file 7). Within each sample, the reference sequences with the highest and lowest (above zero) read counts consistently varied by a factor of ~10,000. For example, in sample 1, 109,963 reads aligned to Gly-GCC, and 13 to Ile2-CAU (supplemental file 7).

All samples showed reads aligned to 40 or 41 (of 42) reference sequences. The reference sequences without reads were Ser-CGA (in Δ*serCGA* and derivatives, as expected) and Cys-GCA-2 (in all 27 samples). Together with the prediction of a non-standard secondary structure for *cysGCA-2* (Chan and Lowe, 2019, 2016), our inability to detect the Cys-GCA-2 sequence in any sample strongly indicates that Cys-GCA-2 does not contribute to translation. Overall, the YAMAT-seq results support the tRNA gene set predicted for SBW25: all 39 predicted tRNA types were detectable, including the 33 essential and the six non-essential tRNA types.

While the YAMAT-seq results provide a useful overview of the relative abundances of tRNAs in a mature tRNA pool, within-strain comparisons of different tRNA types should be interpreted cautiously. As outlined by Shigematsu et al. (2017), variations in tRNA structural components and post-transcriptional modifications can adversely affect the relative efficiency of the reverse transcription reaction for some tRNA types, reducing their apparent proportions. In our results, three tRNA types consistently align a very low proportion of reads (<0.0001): Phe-GAA, Glu-UUC, and Ile2-CAU. A much higher proportion of reads was expected in particular for Phe-GAA and Glu-UUC, both of which are the sole tRNA types responsible for decoding all synonymous codons for their respective amino acids (accounting for 3.66 % and 5.43 % of genome wide codons, respectively; Chan and Lowe, 2016). Given their low read numbers, and non-central role in our experiment, these three tRNA types were removed from downstream statistical analyses.

The strength of the YAMAT-seq procedure lies in comparing changes in tRNA expression across genotypes. That is, any issues with efficiencies are expected to remain relatively constant across the strains in this experiment, allowing changes in relative tRNA type abundances to be detected. To this end, DESeq2 (Love et al., 2014) was used to compare normalized expression levels of 36 tRNA types – all SBW25 tRNA types except for Phe-GAA, Glu-UUC, and Ile2-CAU – in pairs of strains. Firstly, the effect of deleting *serCGA* was investigated by comparing tRNA sequences from each of the two independent *serCGA* deletion strains with those in SBW25 (Figure 5A). The absence of tRNA-Ser(CGA) from both deletion mutants demonstrates that (i) tRNA-Ser(CGA) is encoded solely by the deleted *serCGA* gene, and (ii) under the conditions tested, tRNA-Ser(CGA) is indeed a non-essential tRNA type in SBW25 (*i.e.*, can be eliminated without causing death). The *serCGA* deletion mutants also showed consistently lower expression of tRNA-Thr(CGU) (DESeq2 adjusted $p<0.001$), a result that may reflect a close metabolic relationship between threonine and serine (Sawers, 1998). No significant differences were detected between the two deletion mutants (DESeq2 adjusted $p>0.1$).

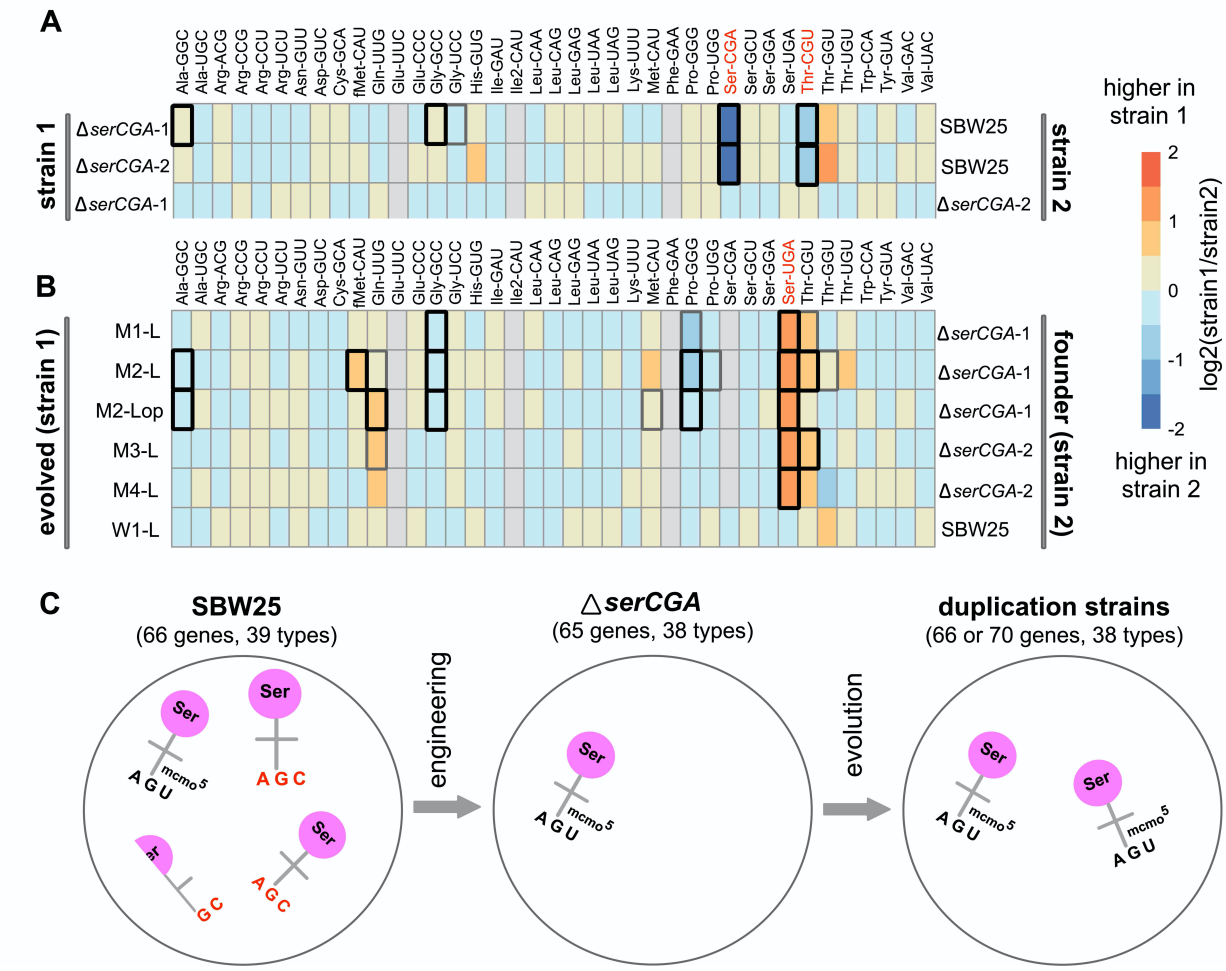

Figure 5. Heatmaps showing differences in the expression of tRNA types between strains. The log2-fold.change(strain1/strain2) difference in expression for 36 tRNA types (with 38 different primary sequences) was determined for pairs of strains using DESeq2. (**A**) tRNA expression levels in the *serCGA* deletion mutants compared with SBW25, demonstrating a consistently lower expression of tRNA-Ser(CGA) and tRNA-Thr(CGU) upon deletion of *serCGA*. tRNAs in red show a consistent difference in all comparisons except Δ*serCGA*-1 vs. Δ*serCGA*-2 (row 3), the control comparison for which no significant differences in tRNA expression were detected. (**B**) tRNA-Ser(UGA) is higher in the mature tRNA pool in each of the five *serTGA* duplication isolates compared with the deletion mutant (with no significant differences in expression detected in the wild type control Line, row 6). Some tRNA types were removed from the DESeq2 analysis (filled grey boxes): Glu-UUC, Ile2-CAU, Phe-GAA, and – in some comparisons – Ser-CGA, consistently gave low read numbers for which DESeq2 results are not meaningful. Box borders represent statistical significance: thin grey=adjusted p>0.01, thick grey=0.01>adjusted p >0.001, black=adjusted p<0.001. tRNAs written in red show a consistent difference in all comparisons except for the control (row 6). (**C**) Cartoon depicting the major effects of *serCGA* deletion (loss of tRNA-Ser(CGA)) and subsequent *serTGA* duplication (two-fold increase of tRNA-Ser(UGA)) on the relative proportions of seryl-tRNAs in mature tRNA pool of SBW25.

Next, the effect of evolution on the *serCGA* deletion mutant was investigated by pairwise comparisons between each Day 13 isolate and its corresponding ancestor (Figure 5B). Importantly, no differences in tRNA pools were detected in the wild type control lineage (W1-L versus SBW25;

DESeq2 adjusted $p>0.1$). Contrastingly, a single consistent, statistically significant difference was observed across the five mutant lineage isolates: the expression of tRNA-Ser(UGA) was 2.06 to 2.60-fold higher than in the Δ*serCGA* ancestor (DESeq2 adjusted $p<0.0001$). A number of other tRNA types co-vary with the rise in tRNA-Ser(UGA) (Figure 5B). While none of these differences are statistically significant in all strains, they are consistently in the same direction (*i.e.*, increase or decrease). Thus, while the main effect of *serCGA* deletion and *serTGA* duplication is the loss of tRNA-Ser(CGA) and elevation of tRNA-Ser(UGA) respectively, other more subtle effects are likely to exist. Finally, pairwise comparisons between the duplication strains reveal some tRNA pool differences between the different duplication strains (Figure 5 – supplemental figure 1). Interestingly, no differences were detected between strain M4-L and any of the other duplication strains (DESeq2 adjusted $p>0.3$), indicating that the four additional tRNA genes duplicated in M4-L are not expressed. It is possible that the duplication junction in M4-L – which lies 112-bp upstream of the duplicated *argTCT-hisGTG-leuTAG-hisGTG* tRNA genes – truncates a promoter, leading to little or no expression of the duplicated copies. This result highlights that tRNA gene copy number does not always correlate with mature tRNA level.

The YAMAT-seq data shows the major effects of our engineering and evolution on the tRNA pool in SBW25 (Figure 5C). In the wild type strain, the proportion of tRNA-Ser(CGA) is around 2.5-fold higher than that of tRNA-Ser(UGA) (0.015 and 0.0059, respectively). The dominant effect of *serCGA* deletion is elimination of tRNA-Ser(CGA), with the proportions of the other types, including tRNA-Ser(UGA), remaining relatively stable. The subsequent large-scale duplications spanning *serTGA* generate an approximately two-fold increase in the relative abundance of tRNA-Ser(UGA).

A model for how elevation of tRNA-Ser(UGA) increases translational efficiency

Thus far, our results show that the growth defect caused by tRNA-Ser(CGA) elimination can be compensated by large-scale, duplication events that serve to increase the proportion tRNA-Ser(UGA) in the mature tRNA pool. In this final section, we develop a model to provide a molecular explanation of how elevating tRNA-Ser(UGA) levels can compensate for tRNA-Ser(CGA) loss.

As described in the introduction, tRNA pool composition is an important determinant of translational speed. During elongation, the codon occupying the ribosomal A site is matched to a corresponding tRNA by stochastically sampling from the available pool of ternary complexes. Ternary complexes consist of a mature tRNA, an elongation factor (EF-Tu), and GTP (Bensch et al., 1991). Given that EF-Tu binds all correctly aminoacylated tRNA types with approximately equal affinity (Louie et al., 1984), the relative levels of each ternary complex are expected to reflect mature tRNA proportions. Overall, the average time taken to match any given codon is dependent on the availability of the corresponding mature tRNA type(s); codons matched by a higher proportion of mature tRNAs will, on average, be translated more quickly than those matched by rarer mature tRNAs (Varenne et al., 1984). To illustrate, consider the serine codons and seryl-tRNAs from our experiments; codon UCA is translated by tRNA-Ser(UGA), while codon UCG can theoretically be translated by tRNA-Ser(CGA) *or*

tRNA-Ser(UGA) (see Figure 1B). When both tRNA types are present, UCG is matched by a higher proportion of tRNAs than UCA, and hence is expected to be translated more quickly (on average).

The above logic can be extended to provide relative numerical estimates of codon translation times, where codon-tRNA matching patterns and tRNA proportions are known. The average time required to translate any particular codon (τ_{codon}) can be approximated by the inverse proportion of matching tRNAs in the pool (Equation 1; where p_{tRNA} is the proportion of matching tRNAs).

$$\tau_{codon} = \frac{1}{p_{tRNA}} \qquad \textbf{Equation 1}$$

Accordingly, the average time taken to translate UCA can be estimated by the inverse proportion of tRNA-Ser(UGA) in the mature tRNA pool (Equation 2a), while UCG translation time can be estimated as the inverse proportion of tRNA-Ser(UGA) *plus* tRNA-Ser(CGA) (Equation 2b). For simplicity, these equations assume that tRNA-Ser(CGA) and tRNA-Ser(UGA) translate UCG with equal efficiency (see Discussion).

$$\tau_{UCA} = \frac{1}{p_{UGA}} \qquad \textbf{Equation 2a}$$

$$\tau_{UCG} = \frac{1}{(p_{CGA} + p_{UGA})} \qquad \textbf{Equation 2b}$$

The proportions of tRNA-Ser(CGA) and tRNA-Ser(UGA) measured during YAMAT-seq (see supplemental file 7) can be substituted into Equations 2a and 2b to obtain relative measures of UCA and UCG translation times, in various genetic backgrounds (Table 2). Accordingly to these calculations, *serCGA* deletion increases the time taken to translate UCG by a factor of four (mean UCG translation times of 49.4 and 207 in wild type and *serCGA* deletion tRNA gene sets, respectively). Subsequent *serTGA* duplication elevates the proportion of tRNA-Ser(UGA), partially restoring UCG translation time (mean UCG translation times of 207, 86.9, and 49.4 in the *serCGA* deletion strains, wild types, and *serTGA* duplication strains, respectively).

The cumulative impact of changing UCG translation speeds on translation and, ultimately, growth logically depends on how frequently the UCG codon is used. In SBW25, UCG is a relatively high use codon, occurring approximately 4.5 times more frequently than UCA (see Figure 1B; Chan and Lowe, 2016). To account for UCG codon bias, we estimated the average time required by each strain to translate an mRNA that reflects the relative UCG/UCA codon usage of SBW25. Using the same principle as in Equations 1, 2a, and 2b, the average time taken to translate the mRNA (τ_{mRNA}) is calculated as the sum of the time taken to translate one UCA codon, plus the time taken to translate 4.5 UCG codons (Equation 3).

$$\tau_{mRNA} = 1\left(\frac{1}{p_{UGA}}\right) + 4.5\left(\frac{1}{(p_{CGA} + p_{UGA})}\right)$$ **Equation 3**

The estimated average time required for the three tRNA gene complements to translate the mRNA can be calculated using the tRNA proportions measured during YAMAT-seq (Table 2). Using this method, on average a *serCGA* deletion mutant is expected to require approximately three times longer to translate the mRNA than the wild type (mean translation times of 1135 and 397, respectively). *serTGA* duplication is estimated to restore average translation time to near-wild type levels (mean translation times of 478 and 397, respectively).

Strain	**YAMAT-seq proportion**		τ_{UCA}	τ_{UCG}	τ_{mRNA}	τ_{mRNA} **relative to tRNA gene set 1**
	tRNA-Ser(UGA)[a]	tRNA-Ser(CGA)[b]				
tRNA gene set 1: wild type (1 serCGA*, 1* serTGA*)*						
SBW25	0.0059	0.015	169	47.8	385	0.970
W1-L	0.0056	0.014	179	51.0	408	1.03
Mean			174	49.4	397	1
tRNA gene set 2: serCGA *deletion (0* serCGA*, 1* serTGA*)*						
Δ*serCGA*-1	0.0050	0	200	200	1100	2.77
Δ*serCGA*-2	0.0047	0	213	213	1170	2.95
Mean			207	207	1135	2.86
tRNA gene set 3: serCGA *deletion and* serTGA *duplication (0* serCGA*, 2* serTGA*)*						
M1-L	0.013	0	76.9	76.9	423	1.07
M2-Lop	0.012	0	83.3	83.3	458	1.15
M2-L	0.012	0	83.3	83.3	458	1.15
M3-L	0.011	0	90.9	90.9	500	1.26
M4-L	0.010	0	100	100	550	1.39
Mean			86.9	86.9	478	1.20

Table 2. Numerical estimates of relative translation times using the mature tRNA pool measurements obtained during YAMAT-seq. Proportions of tRNA-Ser(UGA)[a] and tRNA-Ser(CGA)[b], as measured by YAMAT-seq for each strain (see supplemental file 7), were substituted into Equation 2a to estimate UCA translation time (τ_{UCA}) Equation 2b to estimate UCG translation time τ_{UCG}, and Equation 3 to estimate translation time of an artificial mRNA representing the relative codon use of UCA and UCG in SBW25 (τ_{mRNA}). Each calculation is performed for the nine strains listed, which can be separated into three tRNA gene sets. Mean tRNA proportions and translation times are provided below the last strain in each tRNA gene set. tRNA proportions are given to 2 significant figures (s.f.), and calculated translation times to 3 s.f.

Overall, the predictions of the above model are consistent with our experimental results: the growth defect caused by *serCGA* deletion is compensated to near-wild type levels by the large-scale duplications encompassing *serTGA* (see Figure 3). Together, our results indicate that tRNA-Ser(CGA) elimination exerts increased translational demand on tRNA-Ser(UGA), and that this pressure is at least partially relieved by elevating tRNA-Ser(UGA) through increased *serTGA* copy number.

DISCUSSION

The evolutionary and molecular mechanisms by which different tRNA gene sets emerge have been of consistent, long-standing interest (periodically reviewed in Gingold and Pilpel, 2011; Ikemura, 1985;

Rak et al., 2018). A multitude of theoretical studies have focused on various aspects of tRNA gene set evolution, highlighting roles for post-transcriptional modifications and codon bias (Bulmer, 1987; Higgs and Ran, 2008; Ikemura, 1981; Novoa et al., 2012; Ran and Higgs, 2010; Rocha, 2004; Sharp et al., 2010). Phylogenetic analyses have provided evidence of bacterial tRNA gene set evolution by four main routes: (i) tRNA gene loss through deletion events, (ii) tRNA gene acquisition through horizontal gene transfer, (iii) tRNA gene acquisition by within-genome duplication events, and (iv) tRNA gene changes as a result of anticodon switching (Diwan and Agashe, 2018; Marck and Grosjean, 2002; McDonald et al., 2015; Tremblay-Savard et al., 2015; Wald and Margalit, 2014; Withers et al., 2006). In this work, we provide direct, empirical evidence for one of these routes: tRNA gene acquisition by within-genome duplication events. Loss of the single-copy tRNA gene *serCGA* was compensated by large-scale, tandem duplications that increase the copy number of tRNA gene *serTGA*, leading to elevation of tRNA-Ser(UGA) in the mature tRNA pool.

Retention of *serCGA* in *P. fluorescens* SBW25 wild type

The observation that increasing tRNA-Ser(UGA) proportions can compensate for tRNA-Ser(CGA) loss can be explained on the molecular level: we hypothesize that elimination of tRNA-Ser(CGA) places translational strain on tRNA-Ser(UGA) at UCG codons, and that this pressure can be at least partially relieved by elevating tRNA-Ser(UGA) levels. While this hypothesis is consistent with our experimental results, it raises the question of why *P. fluorescens* SBW25 might retain a copy of *serCGA* in the natural, plant environment; given that selection for translational efficiency favours fewer tRNA types encoded by more tRNA gene copies (Ran and Higgs, 2010; Rocha, 2004), one might expect *serCGA* elimination in favour of multiple *serTGA* gene copies. However, *serCGA* is frequently retained in bacterial genomes; a study of 319 bacteria from different genera indicates a *serCGA* retention rate of ~70 %, with retention correlating with higher UCG usage (Wald and Margalit, 2014). What is the advantage of retaining tRNA-Ser(CGA) over simply encoding more tRNA-Ser(UGA)?

One possible explanation for *serCGA* retention is that, while both tRNA-Ser(CGA) and tRNA-Ser(UGA) *can* translate UCG codons, tRNA-Ser(CGA) may do so with greater efficiency (*i.e.*, more quickly and/or more accurately). Indeed, changing the anticodon of *E. coli* tRNA-Ser(UGA) from UGA to CGA has been shown to increase the efficiency of UCG translational *in vitro* (Takai et al., 1999b). Further, any difference in translational efficiencies may depend on environmental conditions; temperature, acidity and ion concentration all alter the stability of RNA base pairings (Nikolova and Al-Hashimi, 2010; Serra et al., 2002). Overall more efficient translation of UCG by tRNA-Ser(CGA) in some environments could conceivably offset the cost of *serCGA* retention, particularly in bacteria with high UCG usage.

The possibility that restoration of *serCGA* may further increase the fitness of the *serTGA* duplication strains could be investigated by continuing the evolution experiment past Day 13. This is because, in addition to increasing fitness, the duplication of *serTGA* has provided a possible route by

which *serCGA* could be regained, and thus the original *P. fluorescens* SBW25 tRNA gene set restored. Specifically, one of the two copies of *serTGA* could acquire a T→C transition at tRNA position 34, changing the gene from *serTGA* to *serCGA* (*i.e.*, an anticodon switch event). In order for such a mutation to spread through the population, it must encode a functional tRNA. This requires the new, hypothetical tRNA-Ser(CGA) to be recognized by seryl-tRNA ligase (SerRS), the enzyme that adds serine to all types of serine-carrying tRNAs in the cell (for a review of tRNA ligases function, see Ibba and Söll, 2000). Recognition of seryl-tRNAs by SerRS depends not on the tRNA sequence or the anticodon, but rather on the characteristic three-dimensional shape of seryl-tRNAs (Lenhard et al., 1999). Therefore, even though *serTGA* and the original *serCGA* encode tRNAs with very different sequences (see Figure 1 – supplemental figure 1A), it is plausible that the new, hypothetical *serCGA* could form a functional tRNA (see Figure 1 – supplemental figure 1B). Indeed, a UGA→CGA anticodon switch alters the translational capacity of *E. coli* tRNA-Ser(UGA) from codon UCA to UCG *in vitro* (Takai et al., 1999a, 1999b). Notably, none of the five genome sequenced, mutant lineage, Day 13 isolates shows evidence of anticodon switch events in either copy of *serTGA* (see supplemental file 5); whether anticodon switching occurs across a longer evolutionary time scale remains to be seen.

The origin of large-scale duplications, and new tRNA gene copies

Large-scale, tandem duplication events similar to those seen in this work are a well-documented adaptive solution to various phenotypic challenges in phage, bacteria and yeast (reviewed in Anderson and Roth, 1977; Elliott et al., 2013; Reams and Roth, 2015). Extensive work has shown that large-scale tandem duplications occur at extremely high rates in bacteria; for example, in an unselected overnight culture, around 10 % of *Salmonella* cells reportedly carry a duplication of some sort, with 0.005-3 % carrying a duplication of a particular locus (Anderson and Roth, 1981, 1977; Reams et al., 2010). These rates are orders of magnitude higher than those typically reported for single nucleotide polymorphisms (SNPs; Westra et al., 2017), and are consistent with the early detection of duplication fragments in our evolution experiment: strains carrying large-scale duplications were isolated from every mutant lineage by Day 13 (~90 generations), and the duplication fragments they contained were first detected in the relevant population between Days 2 and 5 (Figure 4E; Figure 4 – supplemental figure 2).

Large-scale duplications arise through unusual exchange of DNA between two separate parts of the bacterial chromosome, with the separating distance determining the size of the duplication (reviewed in Anderson and Roth, 1977; Elliott et al., 2013; Reams and Roth, 2015). A variety of mechanisms have been reported to underpin duplication formation, including (i) RecA-mediated, unequal recombination, (ii) RecA independent unequal recombination, and (iii) errant topoisomerase or gyrase activity (Reams et al., 2014; Reams and Roth, 2015; Shyamala et al., 1990). The first of these, RecA-mediated unequal recombination, occurs between direct repeats some distance apart (*e.g.* rRNA operons, *rhs* genes), with longer repeats generally leading to higher rates of recombination (Anderson and Roth, 1981). Two of the five mutant lineage isolates in this work (M2-L, M3-L) have endpoints in

~1.5 kb direct, imperfect repeats at 4.12 Mb and 4.31 Mb of the SBW25 chromosome, suggesting that the ~192 kb duplication fragment they contain arose by RecA-mediated, unequal recombination between these regions. The duplication fragments in the other three mutant lineage isolates (M1-L, M2-Lop, M4-L) show no obvious signs of homology between their endpoints (Table 1), indicating an alternative mechanistic origin. Overall, the diversity in duplication fragment endpoints – location, and degree of homology – in this study are indicative of a range of mechanistic origins.

While duplication formation does not necessarily require sequence homology, duplications that occur at the highest rates typically result from unequal recombination between long (>200 bp) repeats (Anderson and Roth, 1981). The *P. fluorescens* SBW25 genome contains many of these types of repeats dispersed around the chromosome, including five nearly identical rRNA operons, and three highly similar *rhs* genes (Silby et al., 2009). In addition, there are hundreds of smaller repeats throughout the genome, including REPINs and tRNA genes (Bertels and Rainey, 2011; Silby et al., 2009). Errant recombination between any of these repeated sequences could plausibly generate large-scale duplications, meaning that almost any region of the SBW25 chromosome – and therefore many tRNA genes – could conceivably be duplicated *via* homologous recombination. It has previously been noted that the region surrounding the SBW25 replication terminus appears to be more susceptible to evolutionary change than the rest of the chromosome (Silby et al., 2009). This variable region extends approximately 1.4 Mb either side of the terminus, engulfing 28 tRNA genes (including *serTGA*; see Figure 1B and supplemental file 4). It seems probable that, while the copy number of many tRNA genes could feasibly be elevated by large-scale duplications, the tRNA genes surrounding the SBW25 replication terminus may be more prone to evolutionary change by duplication.

Given that dispersed repeats and large-scale duplications are a widespread feature of bacterial genomes (reviewed in Brazda et al., 2020; Reams and Roth, 2015), similar within-genome duplication may be capable of generating changes in tRNA gene copy number in many bacteria. Presumably, different duplication fragments – and therefore, tRNA genes – arise at varying rates (depending on the presence and length of direct repeats in the surrounding area), and have varying degrees of evolutionary success (depending on fragment size, and the dosage effects of genes in the fragment).

The evolutionary fate of the large-scale duplications, and new tRNA gene copies

Large-scale duplications in bacterial genomes are typically unstable (reviewed in Reams and Roth, 2010, 2015). That is, in addition to occurring at high rates, they are also lost – without a trace – at high rates. Duplications are lost in ~1 % of cells *per* generation in *Salmonella* cultures, in the absence of selection (Anderson and Roth, 1981). Due to their combined ease of gain and loss, it has been suggested that duplications serve as fleeting evolutionary solutions to transient phenotypic challenges (Sonti and Roth, 1989); they occur at high frequency without adversely affecting the long-term structure and integrity of the genome. The reported instability of large-scale duplications raises questions about the long-term fate of the duplications seen in during this experiment. That is, if the serial transfer

experiment were continued past Day 13, what would happen to the duplication strains and the tRNA genes that they contain?

Extension of the evolution experiment could have several outcomes. One is that duplication fragments encompassing *serTGA* may continue to arise and be lost, with duplication strains eventually reaching a stable level within the population (Reams et al., 2010). In this scenario, while individual large-scale duplications would continually rise and fall, they would remain the dominant evolutionary solution in the population. A second possibility is that progressively smaller (and inherently more stable) duplication fragments may arise, perhaps eventually resulting in a duplication fragment comprised only of *serTGA* and its promoter. Such smaller duplication fragments may arise either from the *serCGA* deletion mutant (by rarer duplication events between more proximate DNA sequences), or from a duplication strain (by remodelling of the existing duplication fragment). Once a smaller, more stable fragment is dominant in the population, one copy of *serTGA* could conceivably change to *serCGA* through an anticodon switch event (see earlier discussion). A third possible outcome of continuing the evolution experiment is that, over time, a more stable mutation may arise, independently of the duplication fragments. Possible examples include (i) point mutation(s) in the promoter of *serTGA*, leading to increased *serTGA* expression without requiring additional gene copies, (ii) mutation(s) that extend the translational capacity of seryl-tRNAs to translate (or, to better translate) UCG codons, and (iii) synonymous point mutation(s) in highly expressed UCG codon(s), lowering the translational demand for tRNA-Ser(UGA). Should any of these more stable mutations arise, they would be expected to displace large-scale duplications – and in some cases, the additional *serTGA* copy – as the dominant evolutionary strategy in the population.

Concluding remarks

The elimination of one tRNA type from *P. fluorescens* SBW25 was readily counteracted by large-scale duplication events that increased the gene copy number of a second, compensatory tRNA type. Hence, this work provides direct observation of the evolution of a bacterial tRNA gene set by the mechanism of tRNA gene duplication.

MATERIALS AND METHODS

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Source or reference**	**Identifiers**	**Additional information**
gene	*serCGA*	N/A	PFLUt39	Encodes tRNA-Ser(CGA)
gene	*serTGA*	N/A	PFLUt51	Encodes tRNA-Ser(UGA)

strain, strainbackground	*Pseudomonas fluorescens* SBW25	Rainey and Bailey, 1996; Silby et al., 2009		Wild type
Genetic reagent	SBW25-*lacZ*	Zhang and Rainey, 2007		Neutrally-marked SBW25 for competition experiments
Genetic reagent	Δ*serCGA*-1	This work		Bases 1624957–1625092, encompassing *serCGA*, removed. Biological replicate of Δ*serCGA*-2
Genetic reagent	Δ*serCGA*-2	This work		Bases 1624957–1625092, encompassing *serCGA*, removed. Biological replicate of Δ*serCGA*-1
Genetic reagent	SBW25-eWT	This work		Wild type SBW25 that has been through the engineering process
Genetic reagent	W1-L	This work		Evolution isolate from Day 13 of lineage W1 (founded by SBW25)
Genetic reagent	W3-L	This work		Evolution isolate from Day 13 of lineage W3 (founded by SBW25-eWT)
Genetic reagent	M1-L	This work		Evolution isolate from Day 13 of lineage M1 (founded by Δ*serCGA*-1)
Genetic reagent	M2-L	This work		Evolution isolate from Day 13 of lineage M2 (founded by Δ*serCGA*-1)
Genetic reagent	M2-Lop	This work		Second evolution isolate from Day 13 of lineage M2 (founded by Δ*serCGA*-1)
Genetic reagent	M3-L	This work		Evolution isolate from Day 13 of lineage M3 (founded by Δ*serCGA*-2)
Genetic reagent	M4-L	This work		Evolution isolate from Day 13 of lineage M4 (founded by Δ*serCGA*-2)
Recombinant DNA reagent	pSXn (plasmid)	Owen and Ackerley, 2011; Frederic Bertels		pSX with one copy of a 38-bp direct repeat removed
Recombinant DNA reagent	pSXn-CGA (plasmid)	This work		pSXn carrying *serCGA*
Recombinant DNA reagent	pSXn-TGA (plasmid)	This work		pSXn carrying *serTGA*

Commercial assay or kit	DNeasy Blood & Tissue Kit	Qiagen cat. no. 69506		
Commercial assay or kit	NextSeq 550 Output v2.5 kit	Illumina cat.no. 20024904		
Commercial assay or kit	TRIzol Max Bacterial RNA isolation kit	Life Technologies cat.no. 16096040		
Commercial assay or kit	DNA 7500 kit	Agilent Technologies		
Software, algorithm	GtRNAdb 2.0	Chan and Lowe, 2016		https://www.gtrnadb.ucsc.edu
Software, algorithm	tRNAscan-SE 2.0	Chan and Lowe, 2019		https://www.lowelab.ucsc.edu/tRNAscan-SE/
Software, algorithm	BLASTp	Altschul et al., 1990		https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins
Software, algorithm	Gen5™	BioTek		https://www.biotek.com/
Software, algorithm	Geneious v11.1.4	Geneious		https://www.geneious.com/home/
Software, algorithm	breseq v0.33.2	Deatherage et al., 2015; Deatherage and Barrick, 2014		https://barricklab.org/twiki/bin/view/Lab/ToolsBacterialGenomeResequencing
Software, algorithm	R v3.6.0			https://www.r-project.org/
other	SuperScript™ III reverse transcriptase	ThermoFisher Scientific cat.no.18080093		
other	T4 RNA ligase 2	New England BioLabs cat.no. M0239L		Also known as T4 Rnl2
other	Phusion™	ThermoFisher Scientific cat.no.M0531S		
other	5% Mini-PROTEAN® TBE Gels	Bio-Rad Laboratories cat.no.4565015		

1

Strains, growth conditions, and oligonucleotides. Full lists of strains, plasmids and oligonucleotides used are provided in supplemental file 2. The *serCGA* deletion was constructed twice, independently; Δ*serCGA*-1 and Δ*serCGA*-2 are biological replicates. Unless otherwise stated, *P. fluorescens* SBW25 cultures were grown in King's Medium B (KB; King et al., 1954) for ~16 hours at 28˚C with shaking. *E. coli* strains were grown in Lysogeny broth (LB) for 16-18 hours at 37˚C with shaking.

Growth curves. Strains were streaked from glycerol stocks on KB, M9, or LB+Gm (20 μg ml^{-1}) plates. After 48 hours incubation, six or seven colonies *per* strain (numbers of replicates based on previous work: Gallie et al., 2015; Lindsey et al., 2013) were grown in 200 μl of liquid KB, M9, or LB+Gm (20 μg ml^{-1}) in a 96-well plate. Two microliters of each culture were transferred to a fresh 198 μl of medium in a new 96-well plate, sealed with a plastic lid or a breathable rayon film (VWR), and grown at 28˚C in a BioTek Epoch 2 plate reader. Absorbance at 600 nm (OD_{600}) of each well was measured at 5-minute intervals, with five seconds of 3 mm orbital shaking before each read. Medium control wells were used to standardize other wells. The mean absorbance, and standard error, of the replicates at every time point were used to draw the growth curves in Figures 2, 3, and 4. Maximum growth rate and lag time were calculated using a sliding window of nine data points during the exponential growth window of the curve (Gen5™ software from BioTek) (see also source data files 1, 3, and 4).

Fitness assays. For each competition, six replicates were performed in three separate blocks. All competitions within a block were performed in parallel. The number of replicates is based on previous work (Beaumont et al., 2009; Gallie et al., 2019). Single colonies of each competitor were grown independently in shaken KB. Competition tubes were inoculated with ~$5x10^6$ cells of each competitor, and incubated at 28˚C (shaking, 24 hours). Competitor frequencies were determined by plating on KB agar or LB+X-gal (60 μg ml^{-1}) agar at 0 and 72 hours. Competing genotypes were readily distinguished by their distinctive morphologies (on KB agar) or colour (neutrally-marked SBW25-*lacZ* forms blue colonies on LB+X-gal; Zhang and Rainey, 2007) Relative fitness was expressed as the ratio of Malthusian parameters (Lenski, 1991) in Figures 2 and 3. Deviation of relative fitness from one was determined by two-tailed, parametric one-sample *t*-tests (see also source data file 2).

Evolution experiment. SBW25 (wild type), SBW25-eWT (engineering control), and the two independent tRNA-Ser(CGA) deletion mutants (Δ*serCGA*-1 and Δ*serCGA*-2) were streaked from glycerol stocks onto KB agar and grown at 28˚C for 48 hours. Two colonies from every strain were used to found one lineage each. This gave four independent wild type lineages (W1-W4), and four mutant lineages (M1-M4), plus a medium control. The numbers of parallel lineages were chosen based on the available laboratory resources. Each colony was inoculated into 4 ml KB in a 13 ml plastic tube, and incubated overnight at 28˚C (shaking). Each grown culture ("Day 0") was vortexed for 1 minute, and 100 μl used to inoculate 10 ml KB in a 50 ml Falcon tube (28˚C, shaking, 24 hours). Every 24 hours

thereafter, 1 % of each culture was transferred to a fresh 10 ml KB in a 50 ml Falcon tube, and a sample of the population frozen at -80˚C. The experiment was continued until Day 15. Populations were periodically dilution plated on KB agar to check for changes in colony size.

Genome sequencing. Seven isolates were purified and stored from Day 13 of the evolution experiment (W1-L, W3-L, M1-L, M2-L, M2-Lop, M3-L, M4-L). Genomic DNA was isolated from 0.5 ml overnight culture of each using a Qiagen DNeasy Blood & Tissue Kit. DNA quality was checked by agarose gel electrophoresis. Whole genome sequencing was performed by the sequencing facility at the Max Planck Institute for Evolutionary Biology (Ploen, Germany). Paired-end, 150 bp reads were generated with an Illumina NextSeq 550 Output v2.5 kit. Raw reads are available at NCBI sequence read archive (SRA accession number: PRJNA558233) (Leinonen et al., 2011). A minimum of 4.5 million raw reads *per* strain were aligned to the SBW25 genome sequence (NCBI genome reference sequence NC_012660.1; Silby et al., 2009) using breseq (Deatherage and Barrick, 2014) and Geneious (v11.1.4). A minimum mean coverage of 94.7 reads *per* base pair was obtained. A full list of mutation predictions is provided in supplemental file 3.

Identification of duplication junctions. The duplication junctions in M1-L, M2-L, M2-Lop, M3-L and M4-L were identified using a combination of analysis of whole genome sequencing data, and laboratory-based techniques. The raw reads obtained from whole genome sequencing of each isolate were aligned to the SBW25 genome sequence (Silby et al., 2009) using breseq (Barrick et al., 2014; Deatherage et al., 2015; Deatherage and Barrick, 2014) and Geneious (v11.1.4). Coverage analyses were performed in Geneious, and coverage plots generated in R (v3.6.0) (Figure 4 – supplemental figure 1). Manual inspection of the Geneious alignment in coverage shift regions led to predicted junctions in all isolates except M3-L. Each predicted junction was checked by alignment to (i) raw reads, and (ii) previously unused sequences (using Geneious). Junction sequences were confirmed by PCR and Sanger sequencing (for primer details, see supplemental file 2).

Historical junction PCRs. Glycerol stock scrapings of the frozen daily populations, or large colony isolates, from each mutant lineage were grown in liquid KB. Washed cells were used as PCR templates, alongside positive and negative controls (see supplemental file 2 for primer details). The PCR products were run on a 1 % agarose gel against a 1 kb DNA ladder at 100 volts for 90 minutes. Gels were stained with SYBR Safe, and photographed under UV illumination. In order to better detect faint PCR products in the earlier days of the evolution experiment, the colours in each photograph were inverted using Preview (v11.0) (Figure 4E and Figure 4 – supplemental figure 2).

Expression of tRNA genes from the pSXn plasmid. Wild type copies of the *serCGA* and *serTGA* genes were individually ligated into the expression vector, pSXn, to give pSXn-CGA and pSXn-TGA.

The pSXn vector contains an IPTG-inducible *tac* promoter (de Boer et al., 1983; Owen and Ackerley, 2011). Together with the empty vector, the two constructs were separately placed into the SBW25, Δ*serCGA*-1, and Δ*serCGA*-2 backgrounds. This was achieved by transformation of the vector constructs into chemically competent cells (Gallie et al., 2015). The growth profiles of the nine resulting genotypes were obtained in liquid LB+Gm (20 μg ml^{-1}), in six replicates of six (see Growth Curves methods, Figures 4C-D, and source data file 4). No IPTG was added at any stage, in order to achieve lower-level, leaky expression of the tRNA gene from the uninduced *tac* promoter.

YAMAT-seq. Originally developed for use in human cell lines, YAMAT-seq (Shigematsu et al., 2017) is adapted in this work for use in *P. fluorescens* SBW25. Three independent replicates (based on replicate numbers reported in Shigematsu et al., 2017) of nine strains (*i.e.*, 27 samples) were grown to mid-exponential phase in 250 ml flasks containing 20 ml KB. Total RNA was isolated from 1.5 ml aliquots (TRIzol Max Bacterial RNA isolation kit). For each sample, 5 μg of total RNA was subjected to tRNA deacylation treatment, and Y-shaped adapters (Eurofins) were ligated to exposed 5'-CCA-3' tRNA ends using T4 RNA ligase 2. Ligation products were reverse transcribed into cDNA using SuperScript™ III reverse transcriptase, and amplified by eleven rounds of PCR with Phusion™. One of 27 sample-specific indices listed in supplemental file 7 was added to each of the 27 reactions. The quality and quantity of each PCR product was checked using an Agilent DNA 7500 kit on a Bioanalyzer, and samples combined in equimolar amounts into one tube. The mixture was run on a 5 % polyacrylaminde gel, and the bands between 180 and 250 bp excised. DNA was extracted in deionized water overnight, and agarose removed by centrifugation through filter paper. The final product was sequenced at the Max Planck Institute for Evolutionary Biology (Ploen, Germany). Single-end, 150 bp reads were generated with an Illumina NextSeq 550 Output v2.5 kit. YAMAT-seq data is available at NCBI Gene Expression Omnibus (GEO accession number GSE144791) (Edgar et al., 2002).

Analysis of YAMAT-seq data. Raw YAMAT-seq reads were sorted into 27 samples by extracting exact matches to each unique, 6-bp long Illumina index. Exact barcode matches were used in order to minimize the misallocation of reads as a result of barcode errors introduced during the sequencing process. The resulting 27 raw read files, each containing a minimum of 637,037 reads, were analysed using Geneious (version 11.1.4). Reads of the expected length (80-151 bp) were extracted (resulting in 99.99 % read retention). The extracted reads were assembled to a set of 42 reference tRNA sequences from SBW25 (supplemental file 6). During assembly, up to 10 % mismatches, gaps of <3 bp, and up to five ambiguities were allowed *per* read. Reads that aligned equally well to more than one reference sequence were discarded in order to minimize misallocation of reads to similar reference sequences. Finally, the unused reads for each sample were *de novo* aligned, and the resulting contigs checked to ensure that none contained substantial numbers of tRNA reads (particularly seryl-tRNAs). The above sorting and assembly parameters were initially based on those reported by Shigematsu et al. (2017), and

were subsequently refined for the SBW25 data. Following assembly, the within-sample proportion of reads aligned to each tRNA type was calculated, and mean tRNA proportions were calculated for each strain across the three replicates (supplemental file 7). DESeq2 (Love et al., 2014) was used in R (version 3.6.0) to detect tRNA expression differences between pairs of strains (see source data file 5). DESeq2 corrects for multiple testing with the Benjamini-Hochberg procedure (Anders and Huber, 2010). Three tRNA types (Glu-UUC, Ile2-CAU, Phe-GAA) were removed from the analyses due to very low read numbers (<0.01 % of the total reads *per* strain).

Statistical tests. Parametric two-tailed two-sample *t*-tests were performed to detect differences in maximum growth rate (V_{max}) and lag time in growth curves (Figures 2C, 2G, 3C, 3D, and 4D, source data files 1, 3, and 4) in cases where all assumptions were satisfied. Where equal variance or normality assumptions were violated, non-parametric Welch two-sample *t*-tests and Mann-Whitney-Wilcoxon rank sum tests were used, respectively (see Figures 3C, 3D, source data file 3). Parametric one-tailed one-sample *t*-tests were used to detect deviations of relative fitness values from 1 in competition assays (see Figures 2D, 2H, 3E, source data file 2). DESeq2 adjusted (for multiple comparisons) *p*-values were used to detect differences in tRNA expression during YAMAT-seq (Figure 5A, 5B, Figure 5 – supplemental figure 1, source data file 5). Analyses were performed in R (v3.6.0). Significance levels: ns=not significant ($p>0.05$), *$0.05<p<0.001$, **$0.01<p<0.001$, ***$p<0.001$.

ACKNOWLEDGEMENTS

The work was supported by the Max Planck Society (all authors). The authors thank Gunda Dechow-Seligmann for technical assistance, and Sven Künzel for assistance with troubleshooting the YAMAT-seq protocol. We thank Frederic Bertels for discussions during development of the YAMAT-seq analysis, and the gift of the pSXn plasmid. We thank Frederic Bertels, Anuradha Mukherjee, Devika Bhave and Bilal Haider for helpful comments on the manuscript.

COMPETING FINANCIAL INTERESTS

The authors hereby declare that there are no competing financial interests.

AUTHOR CONTRIBUTIONS

JG conceived the research. HJP and JG developed the mathematical model. GBA and JG performed the lab work and data analyses. JG wrote the manuscript, and all authors commented.

REFERENCES

Agris PF, Eruysal ER, Narendran A, Väre VYP, Vangaveti S, Ranganathan SV. 2018. Celebrating wobble decoding: Half a century and still much is new. *RNA Biol* **15**:537–553. doi:10.1080/15476286.2017.1356562

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* **215**:403–410. doi:10.1016/S0022-2836(05)80360-2

Anders S, Huber W. 2010. Differential expression analysis for sequence count data. *Genome Biol* **11**:R106. doi:10.1186/gb-2010-11-10-r106

Anderson P, Roth J. 1981. Spontaneous tandem genetic duplications in *Salmonella typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc Natl Acad Sci U S A* **78**:3113–3117. doi:10.1073/pnas.78.5.3113

Anderson RP, Roth JR. 1977. Tandem genetic duplications in phage and bacteria. *Annu Rev Microbiol* **31**:473–505. doi:10.1146/annurev.mi.31.100177.002353

Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ, Knoester DB, Reba A, Meyer AG. 2014. Identifying structural variation in haploid microbial genomes from short-read resequencing data using *breseq*. *BMC Genomics* **15**:1039. doi:10.1186/1471-2164-15-1039

Beaumont HJE, Gallie J, Kost C, Ferguson GC, Rainey PB. 2009. Experimental evolution of bet hedging. *Nature* **462**:90–93. doi:10.1038/nature08504

Bensch K, Pieper U, Ott G, Schirmer N, Sprinzl M, Pingoud A. 1991. How many EF-Tu molecules participate in aminocyl-tRNA binding? *Biochimie* **73**:1045–1050. doi:10.1016/0300-9084(91)90146-R

Berg OG, Kurland CG. 1997. Growth rate-optimised tRNA abundance and codon usage. *J Mol Biol*. doi:10.1006/jmbi.1997.1142

Bertels F, Rainey PB. 2011. Within-genome evolution of REPINs: A new family of miniature mobile DNA in bacteria. *PLoS Genet* **7**:e1002132. doi:10.1371/journal.pgen.1002132

Björk GR, Hagervall TG. 2014. Transfer RNA modification: Presence, synthesis, and function. *EcoSal Plus* **6**. doi:10.1128/ecosalplus.esp-0007-2013

Boccaletto P, MacHnicka MA, Purta E, Pitkowski P, Baginski B, Wirecki TK, De Crécy-Lagard V, Ross R, Limbach PA, Kotter A, Helm M, Bujnicki JM. 2018. MODOMICS: A database of RNA modification pathways. 2017 update. *Nucleic Acids Res* **46**:D303–D307. doi:10.1093/nar/gkx1030

Brazda V, Fojta M, Bowater RP. 2020. Structures and stability of simple DNA repeats from bacteria. *Biochem J* **477**:325-339. doi:10.1042/BCJ20190703.

Bulmer M. 1991. The selection-mutation-drift theory of synonymous codon usage. *Genetics* **129**:897–907.

Bulmer M. 1987. Coevolution of codon usage and transfer RNA abundance. *Nature* **325**:728–730. doi:10.1038/325728a0

Chan PP, Lowe TM. 2019. tRNAscan-SE: Searching for tRNA genes in genomic sequences. *Methods Mol Biol* **1962**:1–14. doi:10.1007/978-1-4939-9173-0_1

Chan PP, Lowe TM. 2016. GtRNAdb 2.0: An expanded database of transfer RNA genes identified in complete and draft genomes. *Nucleic Acids Res* **44**:D184-189. doi:10.1093/nar/gkv1309

Crick FHC. 1966. Codon-anticodon pairing: The wobble hypothesis. *J Mol Biol* **19**:548–555. doi:10.1016/S0022-2836(66)80022-0

de Boer HA, Comstock LJ, Vasser M. 1983. The *tac* promoter: a functional hybrid derived from the *trp* and *lac* promoters. *Proc Natl Acad Sci U S A* **80**:21–25. doi:10.1073/pnas.80.1.21

Deatherage DE, Barrick JE. 2014. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using *breseq*. *Methods Mol Biol* **1151**:165–188. doi:10.1007/978-1-4939-0554-6_12

Deatherage DE, Traverse CC, Wolf LN, Barrick JE. 2015. Detecting rare structural variation in evolving microbial populations from new sequence junctions using *breseq*. *Front Genet* **5**:468. doi:10.3389/fgene.2014.00468

Diwan GD, Agashe D. 2018. Wobbling forth and drifting back: The evolutionary history and impact of bacterial tRNA modifications. *Mol Biol Evol* **35**:2046–2059. doi:10.1093/molbev/msy110

Dong H, Nilsson L, Kurland CG. 1996. Co-variation of tRNA abundance and codon usage in *Escherichia coli* at different growth rates. *J Mol Biol* **260**:649–663. doi:10.1006/jmbi.1996.0428

Edgar R, Domrachev M, Lash AE. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Res* **30**:207–210. doi:10.1093/nar/30.1.207

Elliott KT, Cuff LE, Neidle EL. 2013. Copy number change: Evolving views on gene amplification. *Future Microbiol* **8**:887–899. doi:10.2217/fmb.13.53

Emilsson V, Kurland CG. 1990. Growth rate dependence of transfer RNA abundance in *Escherichia coli*. *EMBO J* **9**:4359–4366. doi:10.1002/j.1460-2075.1990.tb07885.x

Fujishima K, Kanai A. 2014. tRNA gene diversity in the three domains of life. *Front Genet* **5**:142. doi:10.3389/fgene.2014.00142

Gallie J, Bertels F, Remigi P, Ferguson GC, Nestmann S, Rainey PB. 2019. Repeated phenotypic evolution by different genetic routes in *Pseudomonas fluorescens* SBW25. *Mol Biol Evol* **36**:1071–1085. doi:10.1093/molbev/msz040

Gallie J, Libby E, Bertels F, Remigi P, Jendresen CB, Ferguson GC, Desprat N, Buffing MF, Sauer U, Beaumont HJE, Martinussen J, Kilstrup M, Rainey PB. 2015. Bistability in a metabolic network underpins the de novo evolution of colony switching in *Pseudomonas fluorescens*. *PLoS Biol* **13**:e1002109. doi:10.1371/journal.pbio.1002109

Gardin J, Yeasmin R, Yurovsky A, Cai Y, Skiena S, Futcher B. 2014. Measurement of average decoding rates of the 61 sense codons in vivo. *Elife* **3**:e03735. doi:10.7554/eLife.03735

Gingold H, Pilpel Y. 2011. Determinants of translation efficiency and accuracy. *Mol Syst Biol* **7**:481. doi:10.1038/msb.2011.14

Grosjean H, de Crécy-Lagard V, Marck C. 2010. Deciphering synonymous codons in the three domains of life: Co-evolution with specific tRNA modification enzymes. *FEBS Lett* **584**:252–264. doi:10.1016/j.febslet.2009.11.052

Gu W, Li M, Zhao WM, Fang NX, Bu S, Frazer IH, Zhao KN. 2004. tRNA Ser(CGA) differentially regulates expression of wild-type and codon-modified papillomavirus L1 genes. *Nucleic Acids Res* **32**:4448–4461. doi:10.1093/nar/gkh748

Higgs PG, Ran W. 2008. Coevolution of codon usage and tRNA genes leads to alternative stable states of biased codon usage. *Mol Biol Evol* **25**:2279–2291. doi:10.1093/molbev/msn173

Ikemura T. 1985. Codon usage and tRNA content in unicellular and multicellular organisms. *Mol Biol Evol* **2**:13–34. doi:10.1093/oxfordjournals.molbev.a040335

Ikemura T. 1981. Correlation between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: A proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol Biol* **151**:389–409. doi:10.1016/0022-2836(81)90003-6

King EO, Ward MK, Raney DE. 1954. Two simple media for the demonstration of pyocyanin and fluorescin. *J Lab Clin Med* **44**:301–307. doi:10.5555/uri:pii:002221435490222X

Kurland CG, Ehrenberg M. 1987. Growth-optimizing accuracy of gene expression. *Annu Rev Biophys Biophys Chem* **16**:291–317. doi:10.1146/annurev.bb.16.060187.001451

Kurland CG, Hughes D, Ehrenberg M. 1996. Limitations of translational accuracy. In: Neidhardt FC, editor. *Escherichia coli* and *Salmonella* - Cellular and Molecular Biology. Washington (DC). pp. 979–1004.

Leinonen R, Sugawara H, Shumway M. 2011. The sequence read archive. *Nucleic Acids Res* **39**:D19–D21. doi:10.1093/nar/gkq1019

Lenhard B, Orellana O, Ibba M, Weygand-Durašević I. 1999. tRNA recognition and evolution of determinants in seryl-tRNA synthesis. *Nucleic Acids Res* **27**:721–729. doi:10.1093/nar/27.3.721

Lenski RE. 1991. Quantifying fitness and gene stability in microorganisms. *Biotechnology* **15**:173–192. doi:10.1016/b978-0-409-90199-3.50015-2

Lindsey HA, Gallie J, Taylor S, Kerr B. 2013. Evolutionary rescue from extinction is contingent on a lower rate of environmental change. *Nature* **494**:463–467. doi:10.1038/nature11879

Louarn J, Cornet F, Francois V, Patte J, Louarn JM. 1994. Hyperrecombination in the terminus region of the *Escherichia coli* chromosome: Possible relation to nucleoid organization. *J Bacteriol* **176**:7524–7531. doi:10.1128/jb.176.24.7524-7531.1994

Louie A, Ribeiro NS, Reid BR, Jurnak F. 1984. Relative affinities of all *Escherichia coli* aminoacyl-tRNAs for elongation factor Tu-GTP. *J Biol Chem* **259**:5010–5016.

Love MI, Huber W, Anders S. 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* **15**:550. doi:10.1186/s13059-014-0550-8

Machnicka MA, Dunin-Horkawicz S, de Crécy-Lagard V, Bujnicki JM. 2016. tRNAmodpred: A computational method for predicting posttranscriptional modifications in tRNAs. *Methods* **107**:34–41. doi:10.1016/j.ymeth.2016.03.013

Manickam N, Joshi K, Bhatt MJ, Farabaugh PJ. 2015. Effects of tRNA modification on translational accuracy depend on intrinsic codon-anticodon strength. *Nucleic Acids Res* **44**:1871–1881. doi:10.1093/nar/gkv1506

Marck C, Grosjean H. 2002. tRNomics: Analysis of tRNA genes from 50 genomes of eukarya, archaea, and bacteria reveals anticodon-sparing strategies and domain-specific features. *RNA* **9**:1189–1232. doi:10.1017/S1355838202022021

McDonald MJ, Chou CH, Swamy KBS, Huang H Da, Leu JY. 2015. The evolutionary dynamics of tRNA-gene copy number and codon-use in *E. coli*. *BMC Evol Biol* **15**:163. doi:10.1186/s12862-015-0441-y

Misra R, Reeves P. 1985. Intermediates in the synthesis of TolC protein include an incomplete peptide stalled at a rare Arg codon. *Eur J Biochem* **152**:151–155. doi:10.1111/j.1432-1033.1985.tb09175.x

Nikolova EN, Al-Hashimi HM. 2010. Thermodynamics of RNA melting, one base pair at a time. *RNA* **16**:1687-1691. doi:10.1261/rna/2235010

Novoa EM, Pavon-Eternod M, Pan T, Ribas De Pouplana L. 2012. A role for tRNA modifications in genome structure and codon usage. *Cell* **149**:202–213. doi:10.1016/j.cell.2012.01.050

Novoa EM, Ribas de Pouplana L. 2012. Speeding with control: Codon usage, tRNAs, and ribosomes. *Trends Genet* **28**:574–581. doi:10.1016/j.tig.2012.07.006

Owen JG, Ackerley DF. 2011. Characterization of pyoverdine and achromobactin in *Pseudomonas syringae* pv. phaseolicola 1448a. *BMC Microbiol* **11**:218. doi:10.1186/1471-2180-11-218

Panwar B, Raghava GPS. 2014. Prediction of uridine modifications in tRNA sequences. *BMC Bioinformatics* **15**:326. doi:10.1186/1471-2105-15-326

Plotkin JB, Kudla G. 2011. Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet* **12**:32–42. doi:10.1038/nrg2899

Rainey PB, Bailey MJ. 1996. Physical and genetic map of the *Pseudomonas fluorescens* SBW25 chromosome. *Mol Microbiol* **19**:521–533. doi:10.1046/j.1365-2958.1996.391926.x

Rak R, Dahan O, Pilpel Y. 2018. Repertoires of tRNAs: The couplers of genomics and proteomics. *Annu Rev Cell Dev Biol* **34**:239–264. doi:10.1146/annurev-cellbio-100617-062754

Ran W, Higgs PG. 2010. The influence of anticodon-codon interactions and modified bases on codon usage bias in bacteria. *Mol Biol Evol* **27**:2129–2140. doi:10.1093/molbev/msq102

Reams AB, Kofoid E, Duleba N, Roth JR. 2014. Recombination and annealing pathways compete for substrates in making *rrn* duplications in *Salmonella enterica*. *Genetics* **196**:119–135. doi:10.1534/genetics.113.158519

Reams AB, Kofoid E, Savageau M, Roth JR. 2010. Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. *Genetics* **184**:1077–1094. doi:10.1534/genetics.109.111963

Reams AB, Roth JR. 2015. Mechanisms of gene duplication and amplification. *Cold Spring Harb Perspect Biol* **7**:a016592. doi:10.1101/cshperspect.a016592

Remigi P, Ferguson GC, McConnell E, De Monte S, Rogers DW, Rainey PB. 2019. Ribosome provisioning activates a bistable switch coupled to fast exit from stationary phase. *Mol Biol Evol* **36**:1056–1070. doi:10.1093/molbev/msz041

Rocha EPC. 2004. Codon usage bias from tRNA's point of view: Redundancy, specialization, and efficient decoding for translation optimization. *Genome Res* **14**:2279–2286. doi:10.1101/gr.2896904

Rodnina M V. 2018. Translation in prokaryotes. *Cold Spring Harb Perspect Biol* **10**:a032664. doi:10.1101/cshperspect.a032664

Sakai Y, Miyauchi K, Kimura S, Suzuki T. 2016. Biogenesis and growth phase-dependent alteration of 5-methoxycarbonylmethoxyuridine in tRNA anticodons. *Nucleic Acids Res* **44**:509–523. doi:10.1093/nar/gkv1470

Sawers G. 1998. The anaerobic degradation of L-serine and L-threonine in enterobacteria: Networks of pathways and regulatory signals. *Arch Microbiol* **171**:1–5. doi:10.1007/s002030050670

Serra MJ, Baird JD, Dale T, Fey BL, Retatagos K, Westhof E. 2002. Effects of magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA* **8**:307-323. doi:10.1017/s1355838202024226

Sharp PM, Emery LR, Zeng K. 2010. Forces that influence the evolution of codon bias. *Philos Trans R Soc B Biol Sci* **365**:1203–1212. doi:10.1098/rstb.2009.0305

Shigematsu M, Honda S, Loher P, Telonis AG, Rigoutsos I, Kirino Y. 2017. YAMAT-seq: An efficient method for high-throughput sequencing of mature transfer RNAs. *Nucleic Acids Res* **45**:e70. doi:10.1093/nar/gkx005

Shyamala V, Schneider E, Ames GF. 1990. Tandem chromosomal duplications: role of REP sequences in the recombination event at the join-point. *EMBO J* **9**:939–946. doi:10.1002/j.1460-2075.1990.tb08192.x

Silby MW, Cerdeño-Tárraga AM, Vernikos GS, Giddens SR, Jackson RW, Preston GM, Zhang XX, Moon CD, Gehrig SM, Godfrey SAC, Knight CG, Malone JG, Robinson Z, Spiers AJ, Harris S, Challis GL, Yaxley AM, Harris D, Seeger K, Murphy L, Rutter S, Squares R, Quail MA, Saunders E, Mavromatis K, Brettin TS, Bentley SD, Hothersall J, Stephens E, Thomas CM, Parkhill J, Levy SB, Rainey PB, Thomson NR. 2009. Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas fluorescens*. *Genome Biol* **10**:R51. doi:10.1186/gb-2009-10-5-r51

Sonti R V., Roth JR. 1989. Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. *Genetics* **123**:19–28.

Sørensen MA, Kurland CG, Pedersen S. 1989. Codon usage determines translation rate in *Escherichia coli*. *J Mol Biol* **207**:365–377. doi:10.1016/0022-2836(89)90260-X

Takai K, Okumura S, Hosono K, Yokoyama S, Takaku H. 1999a. A single uridine modification at the wobble position of an artificial tRNA enhances wobbling in an *Escherichia coli* cell-free translation system. *FEBS Lett* **447**:1–4. doi:10.1016/S0014-5793(99)00255-0

Takai K, Takaku H, Yokoyama S. 1999b. In vitro codon-reading specificities of unmodified tRNA molecules with different anticodons on the sequence background of *Escherichia coli* tRNA1(Ser). *Biochem Biophys Res Commun* **257**:662–667. doi:10.1006/bbrc.1999.0538

Takai K, Takaku H, Yokoyama S. 1996. Codon-reading specificity of an unmodified form of *Escherichia coli* tRNA1(Ser) in cell-free protein synthesis. *Nucleic Acids Res* **24**:2894–2899. doi:10.1093/nar/24.15.2894

Tremblay-Savard O, Benzaid B, Lang BF, El-Mabrouk N. 2015. Evolution of tRNA repertoires in *Bacillus* inferred with OrthoAlign. *Mol Biol Evol* **32**:1643–1656. doi:10.1093/molbev/msv029

Wald N, Margalit H. 2014. Auxiliary tRNAs: Large-scale analysis of tRNA genes reveals patterns of tRNA repertoire dynamics. *Nucleic Acids Res* **42**:6552–6566. doi:10.1093/nar/gku245

Westra ER, Sünderhauf D, Landsberger M, Buckling A. 2017. Mechanisms and consequences of diversity-generating immune strategies. *Nat Rev Immunol* **17**:719–728. doi:10.1038/nri.2017.78

Withers M, Wernisch L, Dos Reis M. 2006. Archaeology and evolution of transfer RNA genes in the *Escherichia coli* genome. *RNA* **12**:933–942. doi:10.1261/rna.2272306

Yona AH, Bloom-Ackermann Z, Frumkin I, Hanson-Smith V, Charpak-Amikam Y, Feng Q, Boeke JD, Dahan O, Pilpel Y. 2013. tRNA genes rapidly change in evolution to meet novel translational demands. *Elife* **2**:e01339. doi:10.7554/eLife.01339

Zhang XX, Rainey PB. 2007. Construction and validation of a neutrally-marked strain of *Pseudomonas fluorescens* SBW25. *J Microbiol Methods* **71**:78–81. doi:10.1016/j.mimet.2007.07.001

Zhou Z, Schnake P, Xiao L, Lal AA. 2004. Enhanced expression of a recombinant malaria candidate vaccine in *Escherichia coli* by codon optimization. *Protein Expr Purif* **34**:87–94. doi:10.1016/j.pep.2003.11.006