Large-scale duplication events underpin population-level flexibility in tRNA gene copy number in *Pseudomonas fluorescens* SBW25

Zahra Khomarbaghi[†], Wing Y. Ngan[†], Gökçe B. Ayan, Sungbin Lim, Gunda Dechow-Seligmann, Pabitra Nandy, Jenna Gallie[*]

Microbial Evolutionary Dynamics Research Group, Department of Theoretical Biology, Max Planck Institute for Evolutionary Biology, Plön, 24306, Germany

[†] Joint authors: these authors contributed equally to this work.
* To whom correspondence should be addressed. Email: gallie@evolbio.mpg.de

Present Address: Zahra Khomarbaghi, Division of Evolution and Genomic Sciences, School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, M13 9PL, United Kingdom

ABSTRACT

The complement of tRNA genes within a genome is typically considered to be a (relatively) stable characteristic of an organism. Here we demonstrate that bacterial tRNA gene set composition can be more flexible than previously appreciated, particularly regarding tRNA gene copy number. We report the high-rate occurrence of spontaneous, large-scale, tandem duplication events in laboratory populations of the bacterium *Pseudomonas fluorescens* SBW25. The identified duplications are up to ~1 Mb in size (~15 % of the wildtype genome) and are predicted to change the copy number of up to 917 genes, including several tRNA genes. The observed duplications are inherently unstable: they occur, and are subsequently lost, at extremely high rates. We propose that this unusually plastic type of mutation provides a mechanism by which tRNA gene set diversity can be rapidly generated, while simultaneously preserving the underlying tRNA gene set in the absence of continued selection. That is, if a tRNA set variant provides no fitness advantage, then high-rate segregation of the duplication ensures the maintenance of the original tRNA gene set. However, if a tRNA gene set variant is beneficial, the underlying duplication fragment(s) may persist for longer and provide raw material for further, more stable, evolutionary change.

INTRODUCTION

The complement of tRNA genes encoded by an organism (the 'tRNA gene set') is an important determinant of translational efficiency (1, 2) and disease trajectories (3–5). In bacteria, tRNA gene set composition varies considerably, yet the tRNA gene set of any given bacterial strain is typically treated as a stable entity. For example, the tRNA gene sets of *Pseudomonas fluorescens* strains SBW25 and Pf0-1 are reported to respectively contain 66 and 73 canonical tRNA genes, encoding 39

38 and 40 distinct species of mature tRNA (each with a unique anticodon-amino acid combination) (6–8).
39 In this work we highlight that, at the population level, the tRNA gene set of a bacterial strain may be
40 more flexible than previously appreciated, particularly with regard to the number of gene copies
41 encoding each type of tRNA.

42 The composition of tRNA gene sets differs across the bacterial phylogeny; extant bacterial
43 tRNA gene sets differ in both the tRNA species that they encode, and in the number of gene copies
44 coding for each tRNA species (7, 9–15). Variations in tRNA species encoded are possible because of
45 wobble base pairing, the process by which some tRNAs can – either naturally (16) or through post-
46 transcriptional modifications (17, 18) – translate multiple codons. Ultimately, wobble base pairing
47 allows all 61 sense codons to be translated by far fewer than the number of directly corresponding
48 tRNA species, with most bacteria encoding between 25 and 46 different types of tRNA (7, 9, 19). In
49 contrast to this apparently reductionist strategy, the tRNA species that are present within a bacterial
50 genome may be encoded by multiple – and often, identical – gene copies (7, 10).

51 The content of a tRNA gene set has a strong influence on the composition of the mature
52 tRNA pool, which in turn is a major determinant of the speed and accuracy of translation (1, 2, 20).
53 The capacity of a tRNA set for efficient translation depends on a number of additional factors, many of
54 which co-evolve with tRNA set composition. Prominent examples include (i) which post-transcriptional
55 tRNA modification enzymes are present (affecting anticodon-codon matching patterns) (15, 18, 21,
56 22), (ii) codon usage (which influences the demand for specific tRNAs) (23–29), and (iii)
57 environmental conditions that affect anticodon-codon base pairing properties (*e.g.*, pH, temperature)
58 (30, 31). Evolutionary optimization of translational efficiency is expected to shape the evolution of
59 each of these factors, particularly under conditions where efficient translation is most important (*e.g.*,
60 during rapid bacterial growth) (32–35). For example, bacteria with faster growth rates tend to carry
61 tRNA gene sets with fewer tRNA species, encoded by higher gene copy numbers (11). Empirically,
62 multiple copies of several *Escherichia coli* tRNA genes have been demonstrated to be advantageous
63 only under conditions that support rapid growth (and hence, rapid translation) (36).

64 The molecular mechanisms by which tRNA genes are gained and lost from tRNA gene sets
65 have been inferred by phylogenetic analyses. Such computational studies have provided evidence for
66 the evolution of extant bacterial tRNA gene sets via a number of routes, including: (i) mutation of
67 tRNA genes to encode different tRNA species ('anticodon switching'), (ii) loss of tRNA genes by
68 deletion, and (iii) acquisition of new tRNA genes, either by within-genome duplication events, or by
69 horizontal gene transfer (10, 12–14, 37–39). In addition, we recently provided an empirical example of
70 a bacterial tRNA gene set undergoing adaptive evolution; elimination of tRNA-Ser(CGA) from the
71 mature tRNA pool of the fast-growing, model bacterium *P. fluorescens* SBW25 was compensated by
72 within-genome duplication events that serve to increase the copy number of the tRNA gene *serTGA*
73 (and corresponding levels of tRNA-Ser(UGA) in the mature tRNA pool) (6). Others have reported that
74 translational challenges imposed by (i) novel codon usage patterns, or (ii) the fitness costs associated
75 with antibiotic resistance mutations can be readily alleviated by similar duplication (and, sometimes,
76 downstream amplification) events that affect the copy numbers of remaining tRNA genes (40, 41).

77 Together, these studies demonstrate a considerable degree of evolutionary flexibility in the
78 composition of bacterial tRNA gene sets.
79 In the above-mentioned study of the tRNA gene set of *P. fluorescens* SBW25 (6), the
80 adaptive increase in *serTGA* copy number was shown to be mediated through the tandem duplication
81 of large sections of the *P. fluorescens* SBW25 chromosome (ranging in size from ~45 kb to ~290 kb).
82 In addition to the 91 bp target gene (*serTGA*), these duplication events affect the dosage of up to a
83 further 278 genes, and generate considerable alterations in chromosome size and architecture.
84 Similar types of large-scale, tandem duplications are a pervasive phenomenon in bacterial genomes
85 (42–46) and – despite their large size and widespread gene dosage effects – they can provide a net
86 selective advantage in response to a range of selective pressures (47–53). Notably, these types of
87 mutations have been reported to arise spontaneously, and at extremely high rates; estimates range
88 from frequencies of 10^{-1} to 10^{-5} duplications per cell per generation (43, 54–57), several orders of
89 magnitude higher than those expected for single nucleotide polymorphisms (SNPs) (58). Given their
90 pervasiveness and ease of occurrence, such large-scale, tandem duplications challenge the
91 prevailing view of fixed, stable tRNA gene sets, particularly in large, rapidly growing bacterial
92 populations.
93 In this work, we firstly show that large-scale, tandem duplication events affecting the *P.*
94 *fluorescens* SBW25 tRNA gene content are not limited to the genomic region surrounding *serTGA*
95 (located at ~4.16 Mb on the SBW25 chromosome). Specifically, we demonstrate that a fitness defect
96 caused by deleting a locus containing four tRNA genes (two copies of each of *gluTTC* and *glyGCC*) is
97 repeatedly and rapidly compensated by large-scale, tandem duplication events encompassing the
98 remaining *glyGCC* gene (at genomic position ~2.38 Mb). Secondly, we provide empirical evidence
99 that these duplication fragments arise, and are lost, at very high frequencies; we show that in each
100 independently evolving mutant population, multiple distinct duplication fragments (and hence, tRNA
101 gene sets) rapidly arise and compete.
102

103 **MATERIALS AND METHODS**

104 **Strains and growth conditions.** A list of strains is provided in Supplementary Table S1; construction
105 details can be found in Supplementary Text S1. Unless otherwise stated, *P. fluorescens* SBW25 was
106 grown in King's Medium B (KB) (59) for ~16 hours at 28˚C with shaking. For growth curves, strains
107 were grown in either liquid KB, or minimal (M9) medium (60) supplemented with 4 % (w/v) glucose.
108 During engineering procedures, *E. coli* strains were grown in Lysogeny Broth (LB) for 16-18 hours at
109 37˚C with shaking.

110 **Photography.** Colonies were grown for the time indicated, at room temperature (~20-22˚C) or 28˚C,
111 on KB agar. Colonies were visualized under a Leica MS5 dissection microscope and photographed
112 with a VisiCam® 1.3 (VWR International). Photographs were cropped and, where noted in figure
113 legends, the exposure and/or brightness uniformly altered in Preview (v11.0) to improve visibility.

114 **Genotype growth curves.** Seven colonies per genotype were inoculated into individual wells of a 96-

115 well pre-plate (each containing 200 µL KB), and grown overnight (28˚C, 200 rpm). From each well, 2
116 µL of grown culture was transferred into 198 µL fresh KB, covered with a Breathe-Easy® membrane,
117 and grown at 28˚C in an Epoch 2 Microplate Spectrophotometer (Agilent BioTek). Absorbance at
118 600nm (OD_{600}) was measured at five-minute intervals, with 5 seconds of 3 mm orbital shaking
119 preceding each read. Gen5™ Data Analysis Software (version 3.00.19) was used to calculate the
120 maximum growth rate (the calculated value of the mean slope, V_{max}) and lag time (time interval
121 between the line of maximum slope of the propagation phase and the absorbance baseline at time
122 zero) for each growth curve, with the following parameters: calculation window=1-8 hours (KB) or 1-10
123 hours (M9); sliding time interval per calculation ("*n*")=6 points). The data was then exported to Excel
124 for further processing and drawing of graphs (Figures 3 and 4). Where assumptions are satisfied,
125 parametric one-way ANOVA was used to test for differences in the mean maximum growth rate
126 between pairs of genotypes. Where assumptions are violated, non-parametric Dunn's tests, followed
127 by the Benjamini-Hochberg procedure to correct for multiple comparisons, were used to test for
128 differences in the median lag time between pairs of genotypes.

129 **Competition assays.** One-to-one competition experiments were used to determine the relative
130 fitness of 9 genotype pairs under conditions similar to those of a single growth cycle of the evolution
131 experiment (see below). Competitions were performed in two blocks, with between 5 and 8 replicates
132 per pair of genotypes. Single colonies of each genotype were grown on KB agar (48 hours, room
133 temperature) and, for each competition replicate, a single colony of each competitor was inoculated
134 into an overnight culture (4 mL KB, 28˚C, 200 rpm). Once grown, the two competitors for each
135 competition were mixed in a ~1:1 ratio. Ten µL of each mixture was used to inoculate 10 mL KB in a
136 50 ml Falcon tube, and grown for 24 hours (28˚C, 200 rpm). Cell samples taken from each
137 competition tube at the start (T_0) and end (T_{24}) of the incubation period were serially diluted and plated
138 on LB agar containing 60 µg mL^{-1} X-gal (48 hours, room temperature). The number of colonies per
139 competitor were recorded at T_0 and T_{24}, where differentiation between competing genotypes was
140 achieved through distinct colony morphologies and/or colours (in competitions including neutrally
141 marked SBW25-*lacZ* (61)). The change in the ratio of the competitors was calculated, and relative
142 fitness of the competitors determined (62). Parametric one-sample *t*-tests or, where normality
143 assumptions were violated, non-parametric Wilcoxon rank sum tests, were used to detect statistical
144 differences in the mean relative fitness of genotype pairs. The Bonferroni procedure was
145 subsequently used to correct for multiple comparisons.

146 **Evolution experiment.** SBW25 (wildtype) and ΔEGEG (quadruple tRNA gene deletion mutant) were
147 streaked from glycerol stocks onto KB agar and grown (48 hours, 28˚C). Five single colonies per
148 genotype were grown overnight (4 mL KB; 28˚C, 200 rpm). Grown cultures ("day 0") were vortexed
149 and 100 µL of each used to inoculate 9.9 mL KB in a 50 mL Falcon tube (28˚C, shaking, 24 hours).
150 Every 24 hours thereafter, 1 % (100 µL) of each culture was transferred to a fresh 9.9 mL KB, and a
151 sample of the population frozen in glycerol saline at -80˚C. The experiment was continued until day 28.
152 The later days (~day 21 onwards) of mutant lines M4 and M5 were discovered to contain external
153 contaminants (see Supplementary Figure S1), and so were removed from downstream population

154 analyses (*i.e.*, Figure 7 and Figure 8). Extensive checks for external and cross contamination events
155 were performed on the remaining 8 lines (including day 28 population-level whole genome re-
156 sequencing, population colony morphology checks, and population growth curve analyses).

157 **Population growth curves.** Forty populations of interest (10 evolving lines, at each of 4 time points)
158 were retrieved from glycerol stocks by re-suspending large scrapings in 200 µL sterile Ringer's
159 solution. Each re-suspension was used to inoculate 4 wells of 96-well pre-plates, and grown overnight
160 (*i.e.*, 4 technical replicates; 2 µL re-suspension into 198 µL KB per replicate). Subsequently, grown
161 pre-cultures were transferred to a new 96-well plate (2 µL pre-culture into 198 µL fresh KB), and the
162 protocol continued as described above for the genotype growth curves. This protocol was repeated
163 three times per population (*i.e.*, 3 biological replicates), giving a total of 12 replicates per population of
164 interest (4 technical replicates x 3 biological replicates). The final population growth curves are
165 presented in Supplementary Figure S1.

166 **Genotype whole genome re-sequencing.** Six single genotypes (W1-1, M1-1, M2-1, M3-1, M4-1,
167 M5-1) were purified from day 21 of the evolution experiment, including 1 genotype from each of the 5
168 mutant lines (M1-M5), and 1 from a representative wildtype control line (W1). Genomic DNA was
169 isolated from 1 mL of overnight culture of each genotype (DNeasy Blood & Tissue Kit; Qiagen).
170 Paired-end, 250-bp reads were generated with an Illumina MiSeq instrument at The Max Planck
171 Institute for Evolutionary Biology (Ploen, Germany), via standard procedures. Raw reads are available
172 at NCBI Sequence Read Archive (SRA accession number: PRJNA790786) (63). A minimum of 1.08
173 million raw reads per genotype were obtained and aligned to the SBW25 reference genome sequence
174 (NC_012660.1) (8) using *breseq* (64–66) and Geneious Prime (v2023.2.1). A minimum mean
175 coverage of 36.4 reads per bp was obtained. Per genotype, a list of the raw differences predicted by
176 *breseq* was manually curated to give a list of final mutation predictions (see Supplementary Table S2
177 for lists and further details). Each mutation prediction was confirmed by standard PCR or duplication
178 junction PCR (see below) and, where indicated, Sanger sequencing (see Supplementary Text S2).

179 **Mature tRNA pool sequencing (YAMAT-seq).** Three independent replicates of 8 genotypes (*i.e.*, 24
180 samples) were used for YAMAT-seq (67) according to the method described in (6). Briefly, (i) per
181 sample, a single colony was chosen from KB agar (in the case of genotypes carrying duplication
182 fragments, rapidly-growing colonies were chosen in an attempt to mitigate the inherent instability of
183 the duplication fragment; see stability test method below), grown in overnight KB culture, diluted 1/10
184 in fresh KB, and grown to mid-exponential phase, (ii) total RNA was isolated from each rapidly-
185 growing liquid culture using a TRIzol Max Bacterial RNA isolation kit, (iii) amino acids were removed
186 from mature tRNAs by deacylation treatment with 20 mM Tris-HCl, (iv) Y-shaped, DNA/RNA hybrid
187 adapters (Eurofins; see Supplementary Table S1) were annealed to the conserved, exposed 5'-
188 NCCA-3' and 3'-inorganic phosphate-5' ends of non-acylated tRNAs, and ligated T4 RNA ligase 2, (v)
189 ligation products were reverse-transcribed into cDNA using SuperScript™ III reverse transcriptase, (vi)
190 cDNA products were amplified, and a unique barcode added, with eleven cycles of PCR using
191 Phusion™, (vii) products were quantified using a Bioanalyzer and fluorescent Nanodrop, and mixed in
192 equimolar amounts, (viii) the mixed sample was run on a 5 % polyacrylamide gel, the bands between

193 ~210 bp and ~280 bp (tRNAs plus adapters and barcodes) cut out and co-extracted overnight (by
194 soaking in deionized, ultrapure water (Illumina)), (ix) the extracted liquid sample was collected by
195 centrifugation through a filtered microcentrifuge tube (cellulose acetate membrane with pore size 0.45
196 µm; Merck CLS-8162), (x) the purified product was quantified and deep sequenced at the Max Planck
197 Institute for Evolutionary Biology (Ploen, Germany). Single-end, 150 bp reads were generated with a
198 NextSeq 550 Output v2.5 kit (Illumina) (NCBI Gene Expression Omnibus accession number
199 GSE196705) (68).

200 **YAMAT-seq analysis.** The raw YAMAT-seq sequencing reads were split into 24 samples by
201 extracting exact matches to each unique, 6-bp Illumina index. The resulting 24 set of raw reads, each
202 containing a minimum of 1,470,795 reads, were analysed using Geneious Prime (version 11.1.4).
203 Reads of the expected length (80-151 bp) were extracted (resulting in at least 99.8 % read retention
204 per sample). The extracted reads were aligned to a set of 42 reference tRNA sequences from SBW25
205 (see Supplementary Text S4) (Geneious settings: per read, up to 10 % mismatches, gaps of <3 bp,
206 up to five ambiguities; discard reads that align equally well to more multiple reference sequences).
207 The unused reads for each sample were *de novo* assembled, and the resulting contigs were checked
208 to ensure that none contained substantial numbers of tRNA reads. The within-sample proportion of
209 reads aligned to each tRNA species was calculated. Next, mean mature tRNA proportions were
210 calculated for each genotype across the 3 replicates (Supplementary Table S4). DESeq2 (69) was
211 used in R (version 3.6.0) to detect tRNA expression differences between pairs of genotypes
212 (Supplementary Table S5). DESeq2 corrects for multiple testing (Benjamini-Hochberg procedure) (70).

213 **Stability test**. Duplication fragment stability was estimated using colony morphology stability tests,
214 according to the method detailed in (71). Briefly, 10 genotypes (WT, rWT, ΔEGEG, Δ*glyGCC*, W1-1,
215 M1-1, M2-1, M3-1, M4-1, and M5-1) were streaked from glycerol stocks onto KB agar plates and
216 grown (48 hours, 28˚C). Five colonies per genotype were used to inoculate 4 mL KB cultures (28˚C,
217 200 rpm, overnight). For duplication-carrying genotypes, all chosen colonies were large (*i.e.*, were
218 expected to retain the duplication). Cultures were then dilution plated on KB agar (48 hours, room
219 temperature), and the number of large and small colonies per plate was enumerated. Between 41 and
220 295 colonies were recorded per replicate, with no more than ~200 colonies counted on any given
221 plate. For each replicate, the proportion of small colonies was calculated. A non-parametric Kruskal-
222 Wallis test was used to test for evidence of a difference in the median of at least 1 founder (p=6.25e-
223 07). To determine which genotypes showed a different median rate of small colony enumeration as
224 compared with SBW25 (wildtype), Dunn's test was used with Benjamini-Hochberg procedure to adjust
225 for multiple comparisons. A Welch non-parametric two-sample *t*-test was used to test for a difference
226 in the mean proportion in small colonies arising from the genotype carrying the small duplication
227 fragment (M3-1), and those carrying a large duplication fragment (M1-1, M2-1, M4-1, M5-1).

228 **Duplication junction PCRs.** Duplication junction PCRs were used in this work to (i) confirm the
229 computationally-predicted duplications in genotypes isolated from day-21 of the evolution experiment
230 (see Supplementary Text S2), and (ii) reveal the evolutionary trajectories of duplication fragments in

evolving populations (method described here). For each trajectory of interest, 3 independent rounds of duplication junction PCR were performed. Per round, PCR templates were prepared as follows: a large chunk of the relevant glycerol stock was melted at room temperature, and 10 µL used to inoculate a 4 mL KB culture (20 h at 28˚C, 200 rpm). Genomic DNA was extracted from 1 mL of culture (DNeasy Blood & Tissue Kit; Qiagen) and quantified on a Nanodrop. Each genomic DNA sample was used as a template for the following PCRs, at a final concentration of 1 ng μL^{-1}: (PCRs 1 and 2) the duplication junction PCR(s) of interest, targeting ~1 Mb duplications (with primer pair M1and2junct_f/M5_junct_r), and/or ~0.49 Mb duplications (with primer pair M3_junct_f/r); (PCR 3) a standard, control PCR targeting a genomic region unaffected by the duplication events (at genomic position ~0.01 Mb; primer pair *gly*Q_f_/r). PCRs were performed with GoTaq Flexi (Promega) for 30 cycles of: 30 sec at 95˚C, 30 sec at 58˚C or 59˚C, 72˚C for 1 min per kb of expected product. Five µL of each PCR product was mixed with loading dye and run on a 1 % agarose gel (SYBR Safe; Life Technologies) at 90 V for ~45 min, against 1 kb DNA ladder (Promega) and photographed under UV illumination (see Supplementary Figures S3 and S4 for examples). PCR product intensities were quantified using ImageJ2 (version 2.14.0/1.54f) (72), and duplication junction PCR intensities were normalized using the corresponding control PCR (see Supplementary Table S6). Final evolutionary trajectories are plotted in Figure 7 and Supplementary Figure S5A.

Population whole genome re-sequencing. Nine samples were used for population-level whole genome re-sequencing: 8 day-28 populations (from lines M1-M3 and W1-W5), and the non-evolved SBW25 wildtype genotype (which, as a control, was treated as a population sample in downstream computational analyses). For each of the 8 day-28 samples, the corresponding day-27 glycerol stock was grown for 1 cycle of the evolution experiment (in 9.9 mL KB: 180 µL thawed glycerol stock, consisting of ~100 µL cell culture and ~80 µL glycerol; 28˚C, 200 rpm, 24 hours). For the control sample, a single SBW25 colony picked from KB agar was grown under the same conditions. Genomic DNA was extracted from 2 mL of each of the nine overnight cultures (DNeasy Blood & Tissue Kit; Qiagen). Library preparation and sequencing was performed by Novogene Europe (Cambridge, UK), using Illumina technology on a NovaSeq 6000 platform. The 150-bp, paired-end raw reads are available at NCBI Sequence Read Archive (SRA accession number: PRJNA1035442) (63). Raw reads were quality-trimmed using fastp (settings: --detect_adapter_for_pe, --cut_right, minimum read length 50 bp) (73). The polymorphism setting in *breseq* (64–66) was used to align each set of trimmed reads to the reference sequence (SBW25; NC_012660.1). This gave a minimum mean coverage of 154.9 reads per bp, and a set of predicted mutations in each sample (Supplementary Tables S2 and S7). Geneious Prime (v2023.2.1) was used to (i) confirm and extend the set of the duplication fragments predicted by *breseq* (Supplementary Text S3), and (ii) rigorously check for contamination in evolving populations (none was detected).

RESULTS

Severe reduction in essential tRNA species' gene copy number slows growth

269 The *P. fluorescens* SBW25 genome is predicted to contain 66 canonical tRNA genes encoding 39
270 tRNA species, 14 of which are coded for by multiple (between two and five) gene copies (Figure 1A)
271 (6, 8, 74). This work initially focusses on two such multicopy tRNA species: tRNA-Gly(GCC) (encoded
272 by three identical *glyGCC* gene copies) and tRNA-Glu(UUC) (encoded by four identical *gluTTC* gene
273 copies). Two copies of each gene are found together at chromosomal position ~2.03 Mb (the *gluTTC-*
274 *glyGCC-gluTTC-glyGCC* locus, hereafter referred to as "*EGEG*") (Figure 1A-B). The third *glyGCC*
275 copy is a lone tRNA gene at genomic position ~2.38 Mb, while the final two *gluTTC* genes occur,
276 together with two copies of *alaGGC*, at ~4.65 Mb. Both of the focal tRNA species are essential: at
277 least one functional copy of each of *glyGCC* and *gluTTC* is expected to be required for translation and
278 growth. Specifically, tRNA-Gly(GCC) is predicted to be the sole decoder of codons 5'-GGC-3'
279 (cognate match) and 5'-GGU-3' (G_{34}:U_3 wobble match), while tRNA-Glu(UUC) is required to translate
280 5'-GAA-3' (cognate) and 5'-GAG-3' (G_{34}:U_3 wobble) (Figure 1C).
281 We began by constructing two tRNA gene deletion genotypes in the SBW25 background: (i)
282 ∆EGEG, in which the entire *EGEG* locus was removed from SBW25 (leaving two copies of *gluTTC*
283 and only a single copy of *glyGCC*), and (ii) ∆*glyGCC*, in which the lone *glyGCC* copy at position ~2.03
284 Mb was removed (leaving two copies of *glyGCC*, and all four copies of *gluTTC*) (Figure 1D;
285 Supplementary Text S1). The two *gluTTC* copies at position ~4.65 Mb were not initially targeted
286 because the tRNA genes with which they co-localize are expected to be essential for survival (*i.e.*,
287 they are the only two copies of *alaGGC*, encoding the essential tRNA-Ala(GGC)) (Figure 1A).
288 Deletion of *EGEG* results in an immediately obvious growth defect; compared with SBW25,
289 ∆EGEG forms smaller colonies on KB and M9 agar plates, and grows more slowly in liquid culture
290 (Figure 2). Further, wildtype-like growth is restored by the re-introduction of *EGEG* into ∆EGEG
291 (generating the reconstructed wildtype, rWT). The observed reduction in growth is presumably due to
292 a drop in levels of either tRNA-Glu(UUC), tRNA-Gly(GCC), or both. However, we note that there is at
293 least partial redundancy among the *glyGCC* gene copies; removal of the lone *glyGCC* gene has no
294 discernible effect on SBW25 growth under the conditions tested (∆*glyGCC* in Figure 2). This
295 observation is consistent with earlier reports that multiple copies of various tRNA genes (including
296 *glyGCC*) show a degree of redundancy in *E. coli* (36).

297 **The ∆EGEG growth defect is rapidly compensated by evolution**

298 In order to determine whether, and how, the ∆EGEG growth defect can be genetically compensated,
299 a serial transfer evolution experiment was performed. Ten independent lines were passaged through
300 repeated cycles of growth and dilution in shaken rich medium (KB), for 28 days (a total of ~200
301 generations over 28 days; 1 % daily bottleneck). Five evolutionary lines were founded by ∆EGEG
302 (lines M1-M5), and five by the wildtype (control lines W1-W5). Within 21 days (~150 generations)
303 each of the five mutant lines showed visibly improved growth, while the wildtype lines did not (from
304 this point on, wildtype line W1 is used as a representative control line) (see Supplementary Figure S1).
305 Single genotypes purified from day 21 of each mutant line (namely, genotypes M1-1, M2-1, M3-1, M4-
306 1, and M5-1) showed improved growth on agar plates and in liquid culture, with the growth profile of
307 one genotype (M3-1) approximating that of the wildtype (Figure 3A-D). A control genotype, W1-1,

308 isolated from day 21 of the representative wildtype line showed a growth profile comparable to that of
309 SBW25 (Figure 3A-D).
310 Three of the evolved mutant genotypes – M1-1, M3-1, and M5-1, each representing a
311 different class of compensatory mutation (see next section) – were chosen for one-to-one competition
312 experiments with (i) ΔEGEG, and (ii) a neutrally marked wildtype genotype (SBW25-*lacZ*). Each of the
313 evolved mutant genotypes outperformed ΔEGEG (two-tailed one-sample *t*-tests $p<0.0006$), but were
314 themselves outperformed by SBW25-*lacZ* (two-tailed one-sample *t*-tests $p<0.05$) (Figure 3E).
315 Together, these results demonstrate that the growth defect caused by *EGEG* deletion can be readily
316 compensated, at least in part, by evolution.

317 **Compensation of *EGEG* loss occurs via large-scale, tandem duplications spanning *glyGCC***
318 In order to identify the mutation(s) behind compensation, whole genome re-sequencing was
319 performed on the six genotypes isolated from day 21 of the evolution experiment (M1-1, M2-1, M3-1,
320 M4-1, M5-1, and W1-1). Per sample, over one million, 250-bp long Illumina sequencing reads were
321 aligned to the SBW25 reference genome (8), resulting in a minimum mean coverage of 36.4 reads
322 per genomic base. *Breseq* (64) and downstream analyses were used to predict the mutations in each
323 evolved genotype (Supplementary Table S2).
324 In each of the five genotypes isolated from a mutant line, a large, direct, tandem duplication
325 was identified at around ~2.38 Mb of the SBW25 chromosome (Figure 4A-C). No evidence of any
326 such duplication fragments was found in the representative wildtype control genotype, strongly
327 suggesting that these large duplications play an important role in the observed compensatory effect.
328 In one mutant genotype (M5-1), an additional, non-synonymous SNP was identified in *rplC* (located
329 outside of the duplicated region; see Figure 4B). While the *rplC* mutation affects the translational
330 apparatus (via amino acid change T43I in the 50S ribosomal protein L3), no evidence was found for a
331 role in the compensated phenotype of M5-1: in the absence of the accompanying duplication, the *rplC*
332 mutation did not improve the growth of ΔEGEG (see Supplementary Figure S2).
333 The duplicated region in each ΔEGEG-derived genotype was more precisely defined using
334 computational analyses and duplication junction PCRs (Table 1, Supplementary Text S2). While each
335 duplication fragment has distinct endpoints, the fragments fall into two broad classes based on size:
336 ~1 Mb and ~0.49 Mb duplications. ~1 Mb duplications were identified in four evolved genotypes (M1-1,
337 M2-1, M4-1, and M5-1). This duplication type typically occurs between long stretches of repetitive
338 DNA at ~1.86 Mb and ~2.86 Mb of the SBW25 chromosome, and encompasses over 900 putative
339 genes (Supplementary Table S3). The ~0.49 Mb duplication was identified in one genotype, M3-1,
340 between positions ~2.30 Mb and ~2.79 Mb of the SBW25 chromosome. At approximately half the size
341 of the ~1 Mb duplication, the ~0.49 Mb duplication is predicted to contain only 426 genes
342 (Supplementary Table S3). Notably, both types of duplication contain distinct subsets of tRNA genes:
343 the ~1 Mb fragments encompass seven tRNA genes (resulting in 69 tRNA genes in the evolved
344 genome), while the ~0.49 Mb fragment contains one tRNA gene (giving 63 tRNA genes) (Figure 4D).
345 Both ~1 Mb and ~0.49 Mb duplications encompass many hundreds of genes, with each
346 constituent gene having a particular effect – positive, negative, or neutral – on growth. Since the net

347 effect of each duplication fragment is compensatory, at least one gene in each fragment is presumed
348 to have a sizeable positive effect on ΔEGEG growth. To gain insight into the gene(s) that may
349 contribute to this positive effect, we focused on the ~0.49 Mb fragment, for two reasons: (i) this region
350 is duplicated in all five evolved genotypes, and (ii) it affords the greatest growth advantage of all
351 observed duplication fragments. Of the 426 putative genes in the ~0.49 Mb fragment, only one has an
352 obvious role in translation: *glyGCC*, the sole remaining gene encoding tRNA-Gly(GCC) in ΔEGEG
353 (see Figure 1). Interestingly, none of duplicated genes present a clear link to the co-deleted *gluTTC*
354 genes. Hence, we hypothesize that changes in *glyGCC* gene copy number are the major determinant
355 of the observed fitness effects.
356
357 **The observed tRNA gene set changes affect tRNA-Gly(GCC) levels**
358 We sought to test whether the above *glyGCC* gene copy number changes are reflected in mature
359 tRNA pools, using YAMAT-seq (a deep-sequencing technique that quantifies the relative proportions
360 of tRNA species in the mature tRNA pool) (6, 67). YAMAT-seq was performed on 24 samples,
361 including three replicates of each of eight genotypes: wildtype SBW25, the two engineered genotypes
362 with reductions in *glyGCC* copy number (ΔEGEG, Δ*glyGCC*), three evolved mutant genotypes that
363 reflect the identified compensatory mutations (M1-1, M3-1, and M5-1), and two control genotypes
364 (engineering control rWT, evolution control W1-1). Illumina sequencing resulted in a minimum of 1.47
365 million raw tRNA reads per sample; these were aligned to a set of reference sequences, consisting of
366 all the unique tRNA gene sequences predicted for SBW25 (6, 7) (see Supplementary Text S4). Each
367 of the expected 39 tRNA species was detected in every sample, in varying proportions
368 (Supplementary Table S4). Notably, tRNA-Gly(GCC) – encoded by three *glyGCC* genes – is the most
369 highly represented of all tRNA species in the wildtype, accounting for 11.0 ± 1.0 % (mean ± SE) of the
370 mature tRNA pool (Figure 5A). Four tRNA species (tRNA-Ile(CAU), tRNA-Phe(GAA), tRNA-Thr(CGU),
371 and tRNA-Glu(UUC)) were detected only at very low levels (<0.1 % of the pool). It is probable that
372 various post-transcriptional modifications present in these tRNA species reduced the efficacy of the
373 YAMAT-seq protocol (specifically, the reverse transcription reaction; see Methods) (6, 67, 75). Due to
374 unreliably low read numbers, these four tRNA species were removed from downstream analyses.
375 DESeq2 analyses were used to detect differences in the relative proportions of all 39 tRNA
376 species between genotype pairs of interest (Figure 5B-D, Supplementary Table S5) (6, 36, 69).
377 Among the measurable tRNA species, the predominant – and in many cases, only statistically
378 significant – effect of the engineering and evolution experiments was on tRNA-Gly(GCC) (Figure 5B-
379 E). Deletion of either *glyGCC* or the *EGEG* locus saw a reduction in the relative contribution of
380 tRNA-Gly(GCC) to the mature tRNA pool. Specifically, compared with 11.0 ± 1.0 % of the wildtype
381 tRNA pool, tRNA-Gly(GCC) accounts for only 6.9 ± 0.43 % and 3.6 ± 0.34 % of the Δ*glyGCC* and
382 ΔEGEG mature tRNA pools, respectively. The severe reduction in tRNA-Gly(GCC) in ΔEGEG was
383 partially reversed by adaptive evolution; genotypes M1-1, M3-1, and M5-1 show tRNA-Gly(GCC)
384 levels of 5.1 ± 0.34 %, 6.3 ± 0.30 %, and 5.8 ± 0.42 %, respectively. No such changes were observed
385 in the tRNA pools of the engineering or evolution control genotypes (rWT, W1-1) (Figure 5B-E).

386 Unfortunately, due to the unreliably low read numbers obtained, we are unable to comment on the
387 effect of *EGEG* deletion, or subsequent evolution, on tRNA-Glu(UUC).
388 In addition to tRNA-Gly(GCC), the ~1 Mb duplication carried by genotypes M1-1 and M5-1
389 changes the number of genes coding for a further four tRNA species: tRNA-Ala(UGC) (from five to six
390 gene copies), tRNA-Asp(GUC) (from four to six gene copies), tRNA-Ile(GAU) (from five to six gene
391 copies), and tRNA-Val(UAC) (from three to five gene copies). While the ~1 Mb duplication fragment
392 increases the contribution of each of these tRNA species to the mature tRNA pool, these changes are
393 less pronounced than those observed for tRNA-Gly(GCC) (and very few are statistically significant)
394 (Figures 5C, 5F). One reason for these weaker effects could be that each of the additional four tRNA
395 species is encoded by multiple, dispersed tRNA gene copies, only some of which lie within the
396 duplicated region. The degree to which each of these dispersed gene copies contributes to the final
397 level of the mature tRNA species may vary considerably, and remains unknown (76, 77). Finally, we
398 note that in addition to incurring no detectable growth or fitness effects (see Supplementary Figure
399 S2), the *rplC* mutation had no major effects on the relative proportions of tRNA species in the mature
400 tRNA pool, under the tested conditions; no statistically significant differences were detected between
401 genotypes M1-1 (~1 Mb duplication) and M5-1 (~1 Mb duplication + *rplC* mutation) (Figure 5D).
402 The results in this section show that changes in *glyGCC* gene copy number are mirrored by
403 changes in the proportional contribution of tRNA-Gly(GCC) to the mature tRNA pool. Specifically,
404 reductions in *glyGCC* gene copy number result in progressively lower proportions of tRNA-Gly(GCC),
405 which we hypothesize eventually manifest as a reduction in growth rate. Growth rate is subsequently
406 restored by large-scale, tandem duplication events, one effect of which is to elevate *glyGCC* gene
407 copy number and (partially) reinstate tRNA-Gly(GCC) levels.

408 **Duplication fragments are unstable, leading to genetically mixed populations**

409 Large-scale, tandem duplications are typically highly unstable; they are known to be lost at high rates
410 via homologous recombination events during DNA replication (often leaving no trace in the segregant
411 genome) (55, 78, 79). The duplication fragments observed in this work are no exception. When grown
412 in overnight culture and plated on agar plates, each of the duplication-carrying genotypes repeatedly
413 gives rise to colonies of two distinct types: (i) a majority of larger, wildtype-like colonies that retain the
414 duplication fragment, and (ii) a sizeable minority of smaller, ΔEGEG-like colonies in which the initial
415 duplication fragment has been lost (Figure 6A-B; see also Supplementary Figure S2).
416 The degree of duplication fragment loss was quantified using a stability test, in which the
417 proportion of small colonies (*i.e.*, segregants) arising from independent cultures founded by each of
418 the five duplication-carrying, evolved mutant genotypes was determined. Between 3 % and 34 % of
419 colonies arising from duplication genotype cultures showed the small, ΔEGEG-like phenotype (Figure
420 6C). The proportion of small colonies observed was dependent on the size of the initial duplication
421 fragment, with small and large duplication fragments giving rise to respective averages of 5.0 % ±
422 1.0% and 18 % ± 1.5 % segregants (mean ± SE; non-parametric two-sample *t*-test for difference
423 between means p=3.5e-07***). The higher proportion of segregants in populations founded by
424 duplication genotypes carrying the larger, ~1 Mb fragments presumably results from a combination of

425 the higher number of recombination events occurring across a more extensive region of homology,
426 and the lesser fitness advantage afforded by the ~1 Mb duplications (see Figure 3).
427 The results in this section demonstrate that the adaptive duplication events are inherently
428 unstable; they, along with their compensatory tRNA gene copies, are lost at high rates. In growing
429 populations founded by duplication-carrying genotypes, this high rate of loss leads to the presence of
430 at least two genetically distinct sub-populations of cells: (i) those carrying the initial duplication
431 fragment, and (ii) those in which the fragment has been lost (*i.e.*, ΔEGEG) (Figure 6D).

432 **Multiple duplication fragments rapidly arise and compete within evolving populations**

433 The ease with which duplications arise and are lost suggests complex evolutionary dynamics within
434 the mutant lines. To begin investigating, we tracked ~1 Mb and ~0.49 Mb duplication fragments
435 through one of the evolutionary lines in which they were identified by the day 21 genotype sequencing
436 (M2 and M3, respectively). This was done by normalizing the intensity of a duplication junction PCR
437 band (which is expected to change as the population evolves) with that of a product amplified from a
438 genomic region unaffected by the duplication events (which is expected to remain constant) (see
439 Methods, Supplementary Figure S3, and Supplementary Table S6). Across an evolving population,
440 changes in the frequency of the target duplications are expected to be reflected by changes in the
441 normalized intensity values (we note that the normalized intensities are not intended to provide
442 precise frequency estimates, but to reveal patterns of change).
443 The normalized intensity profile of the ~1 Mb fragment across line M2 is shown in Figure 7A,
444 and that of the ~0.49 Mb fragment across line M3 is in Figure 7B. Both duplications were detected
445 early in their respective lines (~days 4-6), and spread rapidly to peak frequency (~days 12-18).
446 Thereafter, the ~1Mb fragment steadily declined to low levels, while the smaller (and hence fitter and
447 more stable) ~0.49 Mb fragment was maintained at high levels until the end of the experiment.
448 The above results hint at the presence of multiple adaptive genotypes in at least some
449 evolving populations. To further explore this possibility, normalized duplication junction PCRs were
450 used to test for the co-existence of ~1 Mb and ~0.49 Mb duplication fragments in evolving lines M1,
451 M2, M3, as well as a representative control line (W1). (Lines M4 and M5 were excluded due to the
452 discovery of external contaminants in the later stages of the experiment, see Supplementary Figure
453 S1.) The results presented in Figure 7C show that both the ~1 Mb and ~0.49 Mb duplications are
454 indeed detectable in all three evolving mutant lines (and not in W1). The co-existence of both
455 fragment types is most prominent during the early stages of the experiment (~days 7-21); in general,
456 the ~1 Mb fragment declines in the later stages of the experiment, while the ~0.49 Mb fragment
457 remains at relatively constant levels. Notably, distinct subtypes of the smaller duplication fragment –
458 each with subtly different endpoints and hence variably sized duplication junction PCR products –
459 were detected in lines M1-M3 (see Supplementary Figures S4 and S5A for more detail). Further,
460 since each duplication junction PCR tests only for a subset of hypothetically possible duplication
461 fragments, we note that there may be other duplications present that are not captured by these PCRs.
462 The results in this section provide evidence of complex evolutionary dynamics of duplications
463 in the evolving mutant populations; both ~1 Mb and ~0.49 Mb fragments, each typically occurring
464 between sets of long repeats, arise readily and repeatedly within evolving ΔEGEG populations. Over

465 time, the larger ~1 Mb fragments are outcompeted by their smaller (and fitter/more stable) ~0.49 Mb
466 counterparts.

467 **Smaller duplication fragments arise and displace larger duplication fragments**

468 To further investigate the evolutionary dynamics of duplications, we performed population-level whole
469 genome re-sequencing on the day 28 populations of mutant lines M1-M3 (day 28 wildtype populations
470 W1-W5, and an ancestral SBW25 sample were included as controls). For each of the nine samples,
471 150-bp paired-end Illumina reads were aligned to the SBW25 reference sequence (8) using *breseq*
472 (on polymorphism detection settings, see Methods) (64), giving a minimum mean coverage of 154
473 reads per genomic base. The raw *breseq* output, and a list of curated mutation predictions, are
474 provided in Supplementary Table S7.

475 From the *breseq* output and further computational analyses (see Supplementary Text S3),
476 eleven distinct duplication fragments were identified across the three day-28 ΔEGEG-derived
477 populations (Table 2). No such duplications were observed in any of the control samples. Ten of
478 these eleven duplication fragments are novel; the eleventh is identical to the ~0.49 Mb fragment
479 previously identified in the genotype M3-1 (isolated from the day-21, line M3 population). Each of the
480 three day-28 mutant populations was found to contain at least three distinct, co-existing duplication
481 fragments, including multiple (relatively) large fragments (between ~0.28 Mb and ~0.77 Mb in size,
482 often with endpoints in repetitive DNA regions), and one much smaller fragment (between ~2.2 kb and
483 ~4.3 kb, with no obvious endpoint homology) (Figure 8A). Notably, each mutant population is
484 predicted to contain a ~0.49 Mb duplication fragment between genomic positions ~2.30 Mb and ~2.79
485 Mb (*i.e.*, is expected to give a PCR product with the duplication junction primers used in Figure 7C).

486 The smallest duplication fragment is predicted in line M3. It is just 2,170-bp long and contains
487 only three complete genes: *pgsA* (predicted to be involved in phospholipid biosynthesis), *glyGCC*
488 (tRNA gene), and *pflu2192* (a putative pseudogene of phage origin) (Figure 8B). When all 15 of the
489 duplication fragments identified in this study are considered, only two genes are duplicated in every
490 fragment: *glyGCC* and *pflu2192* (Figure 8B, Supplementary Table S3). Taken together, the population
491 sequencing and duplication junction PCRs provide strong support for the hypothesis that the major
492 compensatory element of the duplication fragments is *glyGCC*; over time, progressively smaller
493 duplication fragments encompassing *glyGCC* (and pseudogene *pflu2192)* arise and spread in the
494 evolving, ΔEGEG-derived populations.

495

496 **DISCUSSION**

497 Various studies have reported adaptive changes in tRNA gene copy number in laboratory populations
498 of bacteria, in response to changing translational needs (6, 40, 41). In these studies, changes in gene
499 copy number are underpinned by large-scale, tandem duplications (6, 41) or amplifications (40). Such
500 mutations are both pervasive and, often, mechanistically reversible: large-scale duplications typically
501 arise, and are lost, extremely rapidly (43, 44, 78, 80, 81), in a manner that is reminiscent of
502 contingency loci in pathogenic bacteria (82, 83). Their transient nature means that their contributions
503 to population diversity, and downstream effects on evolution, can be underappreciated (44). In this

work, we (i) demonstrate the contribution of large-scale duplications to population-level diversity in tRNA gene set content, and (ii) show that such duplications may provide the raw material for longer-term evolutionary changes in basal tRNA gene set content.

Spontaneous duplications rapidly generate population-level tRNA set diversity

As outlined in the introduction, large-scale duplications typically arise spontaneously (during DNA replication), at high rates, in a wide range of organisms (43, 46, 54–57). For instance, it has been reported that – without invoking any growth advantage – around 10 % of cells in an overnight *Salmonella* culture carry a large duplication of some region of the chromosome, with between 0.005 % and 3 % of cells carrying a duplication of a given locus (42, 43, 55). In short, a growing population of initially isogenic cells is expected to rapidly consist of a mixture of genotypes that vary with respect to the copy numbers of many genes (including tRNA genes).

We have demonstrated that similar large-scale, tandem duplications readily occur in at least two regions of the *P. fluorescens* SBW25 chromosome: either side of the putative replication terminus, centered around genomic positions ~2.38 Mb (this work) and ~4.16 Mb (6) (see Figure 4B). In this work we have reported 15 distinct duplications at position ~2.38 Mb, including five duplications identified in genotypes isolated from day 21 of the evolution experiment, and a further ten duplications predicted in the day 28 population sequencing (see Tables 1-2 and Supplementary Table S3). These 15 duplications range from ~2.2 kb to ~1 Mb in size, and each one increases the initial tRNA gene set (of ∆EGEG) from 62 tRNA genes to either 63 or 69 tRNA genes (see Figure 4D). In our earlier work, we observed five duplications at position ~4.16 Mb (6). These five duplications are up to ~290 kb, and each adds one tRNA gene copy to the initial genome (∆*serCGA*) (6). The tRNA gene set changes incurred by large-scale duplications have phenotypic consequences: the additional tRNA gene copies generally lead to higher proportional contributions of the corresponding tRNA species to the mature tRNA pool (see Figure 5 and (6)). Taken together, these results show that large-scale duplications readily occur at multiple locations in the SBW25 chromosome, thereby providing a mechanism for the rapid generation of population-level diversity in mature tRNA pool content.

The duplications reported in this work are detected at unequal rates. For instance, examples of similar ~1 Mb and ~0.49 Mb duplication fragments are found in the early stages of each evolving mutant line (see Table 2, Figure 7). This observation is consistent with literature showing that some chromosomal regions – particularly those delineated by long stretches of homologous DNA (*e.g., rrn* operons) – give rise to duplication events at especially high rates (79). Indeed, the prevalent ~1 Mb and ~0.49 Mb fragment classes are both delineated by long stretches of homologous DNA: the ~1 Mb duplication fragments are typically flanked by two *rhs* genes at positions ~1.86 Mb and ~2.86 Mb (~3 kb of homologous sequence), while the ~0.49 Mb fragments are generally delineated by intergenic repeat regions at ~2.30 Mb and ~2.79 Mb (~1.1 kb of homology).

In addition to the above mechanisms, tandem duplications can arise – via alternative mechanisms, and at much lower rates – between regions with little or no DNA homology (79). Such alternative mechanisms presumably contribute to the generation of the smaller duplication fragments detected at the later stages of evolution in this work; these smaller fragments, many of which have no obvious endpoint homology, could conceivably arise at low rates either directly from the founder

544 genotype (ΔEGEG), or from duplication-carrying genotypes (through the partial segregation of a
545 duplication fragment). Indeed, an example of the latter was recently reported in an experiment
546 examining the fate of an advantageous, large-scale (~1.66 Mb), tandem duplication fragment in
547 *Salmonella*; during 2,000 generations of positive selection, the initial duplication fragment repeatedly
548 shrank to ~0.66 Mb, via the loss of an internal ~1 Mb segment (84, 85).
549
550 **Unstable but adaptive duplications as evolutionary stepping stones**
551 Generally speaking, large-scale, tandem duplication fragments are inherently unstable; they are prone
552 to loss through RecA-mediated recombination during DNA replication (55, 79). The duplication
553 fragments described in this work are similarly precarious – the stability test showed that, despite the
554 net fitness advantage that they provide, duplications are lost by between 3 and 34 % of cells in
555 overnight culture (see Figure 6). The rate of segregation was higher for genotypes carrying the larger
556 (~1 Mb) versus the smaller (~0.49 Mb) duplications (median loss rates of 19 % and 5 %, respectively).
557 This difference presumably results from a combination of the larger duplications (i) carrying twice the
558 length of homologous DNA for recombination (*i.e.*, the duplication fragment), and (ii) conferring a
559 lesser fitness advantage (see Figure 3) (44).
560 The inherent instability of the large-scale duplications renders them unlikely to represent a
561 long-term evolutionary solution. That is, over time, inherently unstable duplication fragments are
562 expected to be displaced by a more stable type of mutation (*e.g.*, SNPs affecting tRNA gene
563 expression, or smaller/more stable duplication fragments). Indeed, we already see evidence of such
564 displacement over the course of our 28-day evolution experiment; the large, ~1 Mb fragments
565 detected at the early stages of each mutant line tested (M1-M3) falls to very low or undetectable
566 levels by the end of the experiment (see the population-level duplication junction PCRs in Figure 7).
567 Furthermore, in one mutant line (M3), the co-existing smaller/fitter ~0.49 Mb fragment also appears to
568 drop in frequency towards the end of the experiment, hinting at the presence of at least one additional,
569 higher-effect mutation in this population. Indeed, subsequent population-level genome re-sequencing
570 detected the presence of much smaller (and, presumably, fitter and mechanistically more stable)
571 duplication fragments in the later stages of all three mutant lines; each day 28 mutant population
572 contains at least one duplication fragment of <5 kb (see Table 2, Figure 8).
573 The small (<5 kb) duplications fragments provide considerable insight into the mechanism
574 behind the compensatory effects of the duplications. Together, the three smallest fragments reduce
575 the core, overlapping region of all 15 reported duplication fragments to just 1,064 bp, which
576 encompasses only two complete genes: tRNA gene *glyGCC*, and predicted pseudogene *pflu2192*
577 (see Figure 8B). This observation strongly supports the hypotheses that (i) the ΔEGEG growth defect
578 is primarily caused by a drop in tRNA-Gly(GCC), and (ii) this defect is readily compensated by
579 duplicating *glyGCC* and thus (partially) restoring tRNA-Gly(GCC) levels. We note that this does not
580 rule out a more minor role for tRNA-Glu(UUC), the second tRNA species targeted by the initial *EGEG*
581 deletion. For instance, it is conceivable that a secondary, lower-effect wave of adaptation could target
582 *gluTTC*. Alternatively, selection pressure on *gluTTC* may be stronger under different experimental
583 conditions.

The detection of progressively shorter duplication fragments raises the interesting question of what would occur if the evolution experiment were continued. One possibility is that the duplications would continue to reduce in size. Presumably, the smallest possible adaptive duplication is approximately ~150-200 bp long (encompassing merely the *glyGCC* gene and promoter). Mechanistically speaking, such a short duplication fragment would be expected to be relatively stable, and hence a genotype carrying such a duplication could be considered to encode a new, comparatively stable tRNA gene set. The long-term evolutionary fate of the duplication fragments – and hence the new tRNA gene copies – remains to be seen, and will provide insight into the extent to which within-genome duplication events contribute to the evolution of extant bacterial tRNA gene sets.

DATA AVAILABILITY

Whole genome sequencing data has been deposited with NCBI Short Read Archive (SRA) under accession numbers PRJNA790786 (isolate sequencing) and PRJNA1035442 (population sequencing).

YAMAT-seq data has been deposited with NCBI Gene Expression Omnibus (GEO) under accession number GSE196705.

The remaining raw data relating to this project have been deposited on Zenodo (doi: 10.5281/zenodo.7276660).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENTS

The authors thank Frederic Bertels for useful comments on the manuscript.

FUNDING

This work was supported by the German Research Foundation (DFG) [GA 2895/2-1 to JG], the Max Planck Society [all authors], and the International Max Planck Research School for Evolutionary Biology [WYN]. Funding for open access charge: The Max Planck Society.

AUTHOR CONTRIBUTIONS

Multi-author contributions are listed alphabetically. Conceived and designed the experiments: GBA, JG, PN, SL, WYN, ZK. Performed the experiments: all authors. Analyzed the data: all authors. Wrote the paper: JG. Commented on the paper: all authors. Conceived the research: JG.

618 **CONFLICT OF INTEREST**

619 The authors declare no conflict of interest exists.

620

621 **REFERENCES**

622 1. Gingold,H. and Pilpel,Y. (2011) Determinants of translation efficiency and accuracy. *Mol*
623 *Syst Biol*, **7**, 481.

624 2. Rak,R., Dahan,O. and Pilpel,Y. (2018) Repertoires of tRNAs: The couplers of genomics
625 and proteomics. *Annu Rev Cell Dev Biol*, **34**, 239–264.

626 3. Abbott,J.A., Francklyn,C.S. and Robey-Bond,S.M. (2014) Transfer RNA and human
627 disease. *Front Genet*, **5**, 158.

628 4. Lant,J.T., Berg,M.D., Heinemann,I.U., Brandl,C.J. and O'Donoghue,P. (2019) Pathways to
629 disease from natural variations in human cytoplasmic tRNAs. *J Biol Chem*, **294**,
630 5294–5308.

631 5. Knight,J.R.P., Garland,G., Poyry,T., Mead,E., Vlahov,N., Sfakianos,A., Grosso,S., De-
632 Lima-Hedayioglu,F., Mallucci,G.R., Von Der Haar,T., *et al.* (2020) Control of
633 translation elongation in health and disease. *Dis Model Mech*, **13**, dmm043208.

634 6. Ayan,G.B., Park,H.J. and Gallie,J. (2020) The birth of a bacterial tRNA gene by large-
635 scale, tandem duplication events. *eLife*, **9**, e57947.

636 7. Chan,P.P. and Lowe,T.M. (2016) GtRNAdb 2.0: An expanded database of transfer RNA
637 genes identified in complete and draft genomes. *Nucl Acids Res*, **44**, D184-189.

638 8. Silby,M.W., Cerdeño-Tárraga,A.M., Vernikos,G.S., Giddens,S.R., Jackson,R.W.,
639 Preston,G.M., Zhang,X.X., Moon,C.D., Gehrig,S.M., Godfrey,S.A.C., *et al.* (2009)
640 Genomic and genetic analyses of diversity and plant interactions of *Pseudomonas*
641 *fluorescens*. *Genome Biol*, **10**, R51.

642 9. Ikemura,T. and Ozeki,H. (1983) Codon usage and transfer RNA contents: Organism-
643 specific codon-choice patterns in reference to the isoacceptor contents. *Cold Spring*
644 *Harb Symp Quant Biol*, **47**, 1087–1097.

645 10. Marck,C. and Grosjean,H. (2002) tRNomics: Analysis of tRNA genes from 50 genomes
646 of Eukarya, Archaea, and Bacteria reveals anticodon-sparing strategies and domain-
647 specific features. *RNA*, **8**, 1189–1232.

648 11. Rocha,E.P.C. (2004) Codon usage bias from tRNA's point of view: Redundancy,
649 specialization, and efficient decoding for translation optimization. *Genome Res*, **14**,
650 2279–2286.

651 12. Withers,M., Wernisch,L. and Dos Reis,M. (2006) Archaeology and evolution of transfer
652 RNA genes in the *Escherichia coli* genome. *RNA*, **12**, 933–942.

653 13. Wald,N. and Margalit,H. (2014) Auxiliary tRNAs: Large-scale analysis of tRNA genes
654 reveals patterns of tRNA repertoire dynamics. *Nucl Acids Res*, **42**, 6552–6566.

655 14. Tremblay-Savard,O., Benzaid,B., Lang,B.F. and El-Mabrouk,N. (2015) Evolution of tRNA
656 repertoires in *Bacillus* inferred with OrthoAlign. *Mol Biol Evol*, **32**, 1643–1656.

657 15. Diwan,G.D. and Agashe,D. (2018) Wobbling forth and drifting back: The evolutionary
658 history and impact of bacterial tRNA modifications. *Mol Biol Evol*, **35**, 2046–2059.

659 16. Crick,F.H.C. (1966) Codon-anticodon pairing: The wobble hypothesis. *J Mol Biol*, **19**,
660 548–555.

661 17. Shepherd,J. and Ibba,M. (2015) Bacterial transfer RNAs. *FEMS Microbiology Reviews*,
662 **39**, 280–300.

663 18. Agris,P.F., Eruysal,E.R., Narendran,A., Väre,V.Y.P., Vangaveti,S. and Ranganathan,S.V.
664 (2018) Celebrating wobble decoding: Half a century and still much is new. *RNA Biol*,
665 **15**, 537–553.

666 19. Alkatib,S., Scharff,L.B., Rogalski,M., Fleischmann,T.T., Matthes,A., Seeger,S.,
667 Schöttler,M.A., Ruf,S. and Bock,R. (2012) The contributions of wobbling and
668 superwobbling to the reading of the genetic code. *PLoS Genet*, **8**, e1003076.

669 20. Varenne,S., Buc,J., Lloubes,R. and Lazdunski,C. (1984) Translation is a non-uniform
670 process. Effect of tRNA availability on the rate of elongation of nascent polypeptide
671 chains. **180**, 549–576.

672 21. Ran,W. and Higgs,P.G. (2010) The influence of anticodon-codon interactions and
673 modified bases on codon usage bias in bacteria. *Mol Biol Evol*, **27**, 2129–2140.

674 22. Novoa,E.M., Pavon-Eternod,M., Pan,T. and Ribas De Pouplana,L. (2012) A role for
675 tRNA modifications in genome structure and codon usage. *Cell*, **149**, 202–213.

676 23. Ikemura,T. (1981) Correlation between the abundance of *Escherichia coli* transfer RNAs
677 and the occurrence of the respective codons in its protein genes: A proposal for a
678 synonymous codon choice that is optimal for the *E. coli* translational system. *J Mol*
679 *Biol*, **151**, 389–409.

680 24. Ikemura,T. (1985) Codon usage and tRNA content in unicellular and multicellular
681 organisms. *Mol Biol Evol*, **2**, 13–34.

682 25. Dong,H., Nilsson,L. and Kurland,C.G. (1996) Co-variation of tRNA abundance and
683 codon usage in *Escherichia coli* at different growth rates. *J Mol Biol*, **260**, 649–663.

684 26. Sharp,P.M., Bailes,E., Grocock,R.J., Peden,J.F. and Sockett,R.E. (2005) Variation in the
685 strength of selected codon usage bias among bacteria. *Nucl Acids Res*, **33**, 1141–
686 1153.

687 27. Higgs,P.G. and Ran,W. (2008) Coevolution of codon usage and tRNA genes leads to
688 alternative stable states of biased codon usage. *Mol Biol Evol*, **25**, 2279–2291.

689 28. Plotkin,J.B. and Kudla,G. (2011) Synonymous but not the same: The causes and
690 consequences of codon bias. *Nat Rev Genet*, **12**, 32–42.

691 29. Novoa,E.M. and Ribas de Pouplana,L. (2012) Speeding with control: Codon usage,
692 tRNAs, and ribosomes. *Trends Genet*, **28**, 574–581.

693 30. Serra,M.J., Baird,J.D., Dale,T., Fey,B.L., Retatagos,K. and Westhof,E. (2002) Effects of
694 magnesium ions on the stabilization of RNA oligomers of defined structures. *RNA*, **8**,
695 307–323.

696 31. Nikolova,E.N. and Al-Hashimi,H.M. (2010) Thermodynamics of RNA melting, one base
697 pair at a time. *RNA*, **16**, 1687–1691.

698 32. Kurland,C.G. and Ehrenberg,M. (1987) Growth-optimizing accuracy of gene expression.
699 *Ann Rev Biophys Chem*, **16**, 291–317.

700 33. Emilsson,V. and Kurland,C.G. (1990) Growth rate dependence of transfer RNA
701 abundance in *Escherichia coli*. *EMBO J*, **9**, 4359–4366.

702 34. Kurland,C.G., Hughes,D. and Ehrenberg,M. (1996) Limitations of translational accuracy.
703 In Neidhardt,F.C. (ed), *Escherichia coli and Salmonella: Cellular and Molecular*
704 *Biology*. ASM Press, Washington, D.C, Vol. 1, pp. 979–1004.

705 35. Sharp,P.M., Emery,L.R. and Zeng,K. (2010) Forces that influence the evolution of codon
706 bias. *Philos Trans R Soc Lond B Biol Sci*, **365**, 1203–1212.

707 36. Raval,P.K., Ngan,W.Y., Gallie,J. and Agashe,D. (2022) The layered costs and benefits of
708 translational redundancy. 10.1101/2022.06.16.495782.

709 37. Saks,M.E., Sampson,J.R. and Abelson,J. (1998) Evolution of a transfer RNA gene
710 through a point mutation in the anticodon. *Science*, **279**, 1665–1670.

711 38. McDonald,M.J., Chou,C.H., Swamy,K.B.S., Huang,H.D. and Leu,J.Y. (2015) The
712 evolutionary dynamics of tRNA-gene copy number and codon-use in *E. coli*. *BMC*
713 *Evol Biol*, **15**, 163.

714 39. Callens,M., Scornavacca,C. and Bedhomme,S. (2021) Evolutionary responses to codon
715 usage of horizontally transferred genes in *Pseudomonas aeruginosa*: gene retention,
716 amelioration and compensatory evolution. *Microb Genom*, **7**.

717 40. Nilsson,A.I., Zorzet,A., Kanth,A., Dahlström,S., Berg,O.G. and Andersson,D.I. (2006)
718 Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA
719 genes. *Proc Natl Acad Sci USA*, **103**, 6976–6981.

720 41. Bedhomme,S., Amorós-Moya,D., Valero,L.M., Bonifaci,N., Pujana,M.-À. and Bravo,I.G.
721 (2019) Evolutionary changes after translational challenges imposed by horizontal
722 gene transfer. *Genome Biol Evol*, **11**, 814–831.

723 42. Anderson,R.P. and Roth,J.R. (1977) Tandem genetic duplications in phage and bacteria.
724 *Annu Rev Microbiol*, **31**, 473–505.

725 43. Anderson,P. and Roth,J. (1981) Spontaneous tandem genetic duplications in *Salmonella*
726 *typhimurium* arise by unequal recombination between rRNA (*rrn*) cistrons. *Proc Natl*
727 *Acad Sci USA*, **78**, 3113–3117.

728 44. Andersson,D.I. and Hughes,D. (2009) Gene amplification and adaptive evolution in
729 Bacteria. *Annu Rev Genet*, **43**, 167–195.

730 45. Elliott,K.T., Cuff,L.E. and Neidle,E.L. (2013) Copy number change: Evolving views on
731 gene amplification. *Future Microbiol*, **8**, 887–899.

732 46. Tomanek,I., Grah,R., Lagator,M., Andersson,A.M.C., Bollback,J.P., Tkačik,G. and
733 Guet,C.C. (2020) Gene amplification as a form of population-level gene expression
734 regulation. *Nat Ecol Evol*, **4**, 612–625.

735 47. Andersson,D.I., Susan Slechta,E. and Roth,J.R. (1998) Evidence that gene amplification
736 underlies adaptive mutability of the bacterial *lac* operon. *Science*, **282**, 1133–1135.

737 48. Riehle,M.M., Bennett,A.F. and Long,A.D. (2001) Genetic architecture of thermal
738 adaptation in Escherichia coli. **98**, 525–530.

739 49. Lind,P.A., Tobin,C., Berg,O.G., Kurland,C.G. and Andersson,D.I. (2010) Compensatory
740 gene amplification restores fitness after inter-species gene replacements. *Mol*
741 *Microbiol*, **75**, 1078–1089.

742 50. Paulander,W., Andersson,D.I. and Maisnier-Patin,S. (2010) Amplification of the gene for
743 isoleucyl-tRNA synthetase facilitates adaptation to the fitness cost of mupirocin
744 resistance in *Salmonella enterica*. *Genetics*, **185**, 305–312.

745 51. Wannarat,W., Wannarat,W., Motoyama,S., Masuda,K., Kawamura,F. and Inaoka,T.
746 (2014) Tetracycline tolerance mediated by gene amplification in Bacillus subtilis. **160**,
747 2474–2480.

748 52. Toussaint,J.P., Farrell-Sherman,A., Feldman,T.P., Smalley,N.E., Schaefer,A.L.,
749 Greenberg,E.P. and Dandekar,A.A. (2017) Gene duplication in Pseudomonas
750 aeruginosa improves growth on adenosine. *J Bacteriol*, **199**, e00261-17.

751 53. Blount,Z.D., Maddamsetti,R., Grant,N.A., Ahmed,S.T., Jagdish,T., Baxter,J.A.,
752 Sommerfeld,B.A., Tillman,A., Moore,J., Slonczewski,J.L., *et al.* (2020) Genomic and
753 phenotypic evolution of *Escherichia coli* in a novel citrate-only resource environment.
754 *eLife*, **9**, e55414.

755 54. Straus,D.S. and Straus,L.D.A. (1976) Large overlapping tandem genetic duplications in
756 *Salmonella typhimurium*. *J Mol Biol*, **103**, 143–153.

757 55. Reams,A.B., Kofoid,E., Savageau,M. and Roth,J.R. (2010) Duplication frequency in a
758 population of *Salmonella enterica* rapidly approaches steady state with or without
759 recombination. *Genetics*, **184**, 1077–1094.

760 56. Sun,S., Ke,R., Hughes,D., Nilsson,M. and Andersson,D.I. (2012) Genome-wide
761 detection of spontaneous chromosomal rearrangements in Bacteria. *PLoS ONE*, **7**,
762 e42639.

763 57. Reams,A.B., Kofoid,E., Duleba,N. and Roth,J.R. (2014) Recombination and annealing
764 pathways compete for substrates in making *rrn* duplications in *Salmonella enterica*.
765 *Genetics*, **196**, 119–135.

766 58. Westra,E.R., Sünderhauf,D., Landsberger,M. and Buckling,A. (2017) Mechanisms and
767 consequences of diversity-generating immune strategies. *Nat Rev Immunol*, **17**, 719–
768 728.

769 59. King,E.O., Ward,M.K. and Raney,D.E. (1954) Two simple media for the demonstration of
770 pyocyanin and fluorescin. *J Lab Clin Med*, **44**, 301–307.

771 60. Sambrook,J., Fritsch,E.F. and Maniatis,T. (1989) Molecular cloning: a laboratory manual
772 2nd. ed. Cold Spring Harbor laboratory press, Cold Spring Harbor.

773 61. Zhang,X.X. and Rainey,P.B. (2007) Construction and validation of a neutrally-marked
774 strain of *Pseudomonas fluorescens* SBW25. *J Microbiol Methods*, **71**, 78–81.

775 62. Lenski,R.E., Rose,M.R., Simpson,S.C. and Tadler,S.C. (1991) Long-term experimental
776 evolution in *Escherichia coli*. I. Adaptation and divergence during 2000 generations.
777 *Evolution*, **138**, 1315–1341.

778 63. Leinonen,R., Sugawara,H. and Shumway,M. (2011) The Sequence Read Archive. *Nucl*
779 *Acids Res*, **39**, D19–D21.

780 64. Deatherage,D.E. and Barrick,J.E. (2014) Identification of mutations in laboratory-evolved
781 microbes from next-generation sequencing data using *breseq*. *Methods Mol Biol*,
782 **1151**, 165–188.

783 65. Deatherage,D.E., Traverse,C.C., Wolf,L.N. and Barrick,J.E. (2015) Detecting rare
784 structural variation in evolving microbial populations from new sequence junctions
785 using *breseq*. *Front Genet*, **5**, 468.

786 66. Barrick,J.E., Colburn,G., Deatherage,D.E., Traverse,C.C., Strand,M.D., Borges,J.J.,
787 Knoester,D.B., Reba,A. and Meyer,A.G. (2014) Identifying structural variation in
788 haploid microbial genomes from short-read resequencing data using *breseq*. *BMC*
789 *Genomics*, **15**, 1039.

790 67. Shigematsu,M., Honda,S., Loher,P., Telonis,A.G., Rigoutsos,I. and Kirino,Y. (2017)
791 YAMAT-seq: An efficient method for high-throughput sequencing of mature transfer
792 RNAs. *Nucl Acids Res*, **45**, e70.

793 68. Edgar,R., Domrachev,M. and Lash,A.E. (2002) Gene Expression Omnibus: NCBI gene
794 expression and hybridization array data repository. *Nucl Acids Res*, **30**, 207–210.

795 69. Love,M.I., Huber,W. and Anders,S. (2014) Moderated estimation of fold change and
796 dispersion for RNA-seq data with DESeq2. *Genome Biol*, **15**, 550.

797 70. Anders,S. and Huber,W. (2010) Differential expression analysis for sequence count data.
798 *Genome Biol*, **11**, R106.

799 71. Mukherjee,A., Dechow-Seligmann,G. and Gallie,J. (2022) Evolutionary flexibility in
800 routes to mat formation by *Pseudomonas*. *Mol Microbiol*, **117**, 394–410.

801 72. Schneider,C.A., Rasband,W.S. and Eliceiri,K.W. (2012) NIH Image to ImageJ: 25 years
802 of image analysis. *Nat Methods*, **9**, 671–675.

803 73. Chen,S., Zhou,Y., Chen,Y. and Gu,J. (2018) fastp: an ultra-fast all-in-one FASTQ
804 preprocessor. *Bioinformatics*, **34**, i884–i890.

805 74. Chan,P.P., Lin,B.Y., Mak,A.J. and Lowe,T.M. (2021) tRNAscan-SE 2.0: improved
806 detection and functional classification of transfer RNA genes. *Nucl Acids Res*, **49**,
807 9077–9096.

808 75. Warren,J.M., Salinas-Giegé,T., Hummel,G., Coots,N.L., Svendsen,J.M., Brown,K.C.,
809 Drouard,L. and Sloan,D.B. (2021) Combining tRNA sequencing methods to
810 characterize plant tRNA expression and post-transcriptional modification. *RNA Biol*,
811 **18**, 64–78.

812 76. Thornlow,B.P., Armstrong,J., Holmes,A.D., Howard,J.M., Corbett-Detig,R.B. and
813 Lowe,T.M. (2020) Predicting transfer RNA gene activity from sequence and genome
814 context. *Genome Res*, **30**, 85–94.

815 77. Scheepbouwer,C., Aparicio-Puerta,E., Gomez-Martin,C., Verschueren,H., Van
816 Eijndhoven,M., Wedekind,L.E., Giannoukakos,S., Hijmering,N., Gasparotto,L., Van
817 Der Galien,H.T., *et al.* (2023) ALL-tRNAseq enables robust tRNA profiling in tissue
818 samples. *Genes Dev*, **37**, 243–257.

819 78. Adler,M., Anjum,M., Berg,O.G., Andersson,D.I. and Sandegren,L. (2014) High fitness
820 costs and instability of gene duplications reduce rates of evolution of new genes by
821 duplication-divergence mechanisms. *Mol Biol Evol*, **31**, 1526–1535.

822 79. Reams,A.B. and Roth,J.R. (2015) Mechanisms of gene duplication and amplification.
823 *Cold Spring Harb Perspect Biol*, **7**, a016592.

824 80. Anderson,R.P. and Roth,J.R. (1979) Gene duplication in Bacteria: Alteration of gene
825 dosage by sister-chromosome exchanges. *Cold Spring Harb Symp Quant Biol*, **43**,
826 1083–1087.

827 81. Hastings,P.J., Lupski,J.R., Rosenberg,S.M. and Ira,G. (2009) Mechanisms of change in
828 gene copy number. *Nat Rev Genet*, **10**, 551–564.

829 82. Moxon,E.R. and Thaler,D.S. (1997) The tinkerer's evolving tool-box. *Nature*, **387**, 659–
830 662.

831 83. Moxon,E.R., Rainey,P.B., Nowak,M.A. and Lenski,R.E. (1994) Adaptive evolution of
832 highly mutable loci in pathogenic bacteria. *Curr Biol*, **4**, 24–33.

833 84. Brandis,G. and Hughes,D. (2020) The SNAP hypothesis: Chromosomal rearrangements
834 could emerge from positive selection during niche adaptation. *PLoS Genet*, **16**,
835 e1008615.

836 85. Cao,S., Brandis,G., Huseby,D.L. and Hughes,D. (2022) Positive selection during niche
837 adaptation results in large-scale and irreversible rearrangement of chromosomal
838 gene order in Bacteria. *Mol Biol Evol*, **39**, msac069.

839

840

841 **TABLES AND FIGURES**

842

Geno type	Dup. size (bp)	# Dup. genes	# Dup. tRNA genes	Junct. name	Junction side 1		Junction side 2	
					Base	Locus	Base	Locus
M1-1	1,016,166	917	69	M1junct	2,864,810	*pflu2597 (rhs)*	1,848,645	Intergenic (no repeat)
M2-1	1,012,338	911	69	M2junct	2,865,243–2,865,277	*pflu2597 (rhs)*	1,852,904–1,853,090	*pflu1686 (rhs)*
M3-1	489,709	426	63	M3junct	2,786,207	*pflu2547* (no repeat)	2,296,499	Intergenic (repeat)
M4-1	~1,012,000	911	69	M4junct	~2,865,500	*pflu2597 (rhs)*	~1,853,600	*pflu1686 (rhs)*
M5-1	1,012,339	911	69	M5junct	2,866,737–2,866,785	*pflu2597 (rhs)*	1,854,339–1,854,447	*pflu1686 (rhs)*

843 **Table 1. Details of duplications identified in five compensated genotypes isolated from the five**
844 **evolving mutant populations on day 21 of the evolution experiment.** Base positions refer to the
845 SBW25 wildtype genome sequence (NCBI accession number: NC_012660.1). For a list of the genes
846 contained within each duplication fragment, see Supplementary Table S3.

847

848

849

Pop.	Frag.	Rank[a]	Dup. size (bp)	# Genes	Junction side 1		Junction side 2	
					Base	Locus	Base	Locus
M1 day 28	F1	3	767,013	670	3,142,292 – 3,142,294	*pflu2872*	2,375,280 – 2,375,282	*pgsA* (*pflu2191*)
	F2	4	490,582	427	2,786,879 – 2,786,885	IG repeat	2,296,298 – 2,296,304	IG repeat
	F3	2	444,617	403	2,712,554	*pflu2497*	2,267,938	IG (*pflu2087/8)*
	F4	1	4,273	7	2,379,967 – 2,379,968	*pflu2197*	2,375,695 – 2,375,696	*pgsA* (*pflu2191*)
M2 day 28	F1	2	490,215	427	2,787,135 – 2,787,168	IG repeat	2,296,921 – 2,296,954	IG repeat
	F2	4	318,344	278	2,666,045 – 2,666,046	*pflu2447*	2,347,702 – 2,347,703	IG repeat (*pflu2165/7*)
	F3	1	285,798	256	2,525,738	*soxB* (*pflu2319*)	2,239,941	IG *rrn* (23S/5S)
	F4	3	3,631	6	2,379,196	*pflu2196*	2,375,566	*pgsA* (*pflu2191*)
M3 day 28	F1*	1	489,709	426	2,786,206 – 2,786,207	*pflu2547*	2,296,498 – 2,296,499	IG repeat (*pflu2166/8*)
	F2	3	434,971	360	2,809,371	*pflu2553*	2,374,401	*uvrC* (*pflu2190*)
	F3	2	2,170	3	2,376,758	*ribD* (*pflu2193*)	2,374,589	*uvrC* (*pflu2190*)

850 **Table 2. Details of the duplications computationally predicted in day 28 populations from**
851 **evolving mutant lines M1, M2, and M3.** Base positions refer to the SBW25 wildtype genome
852 sequence (NCBI accession number: NC_012660.1). For a list of the genes contained within each
853 predicted duplication, see Supplementary Table S3. [a]Duplications are ranked within each population
854 in descending order of prevalence (according to the adjusted number of Illumina reads perfectly
855 matching the complete proposed duplication junction sequence; see Supplementary Text S3).
856 *Fragment M3d28-F1 is identical to the duplication identified in day 21 genotype M3-1 (see Table 1).

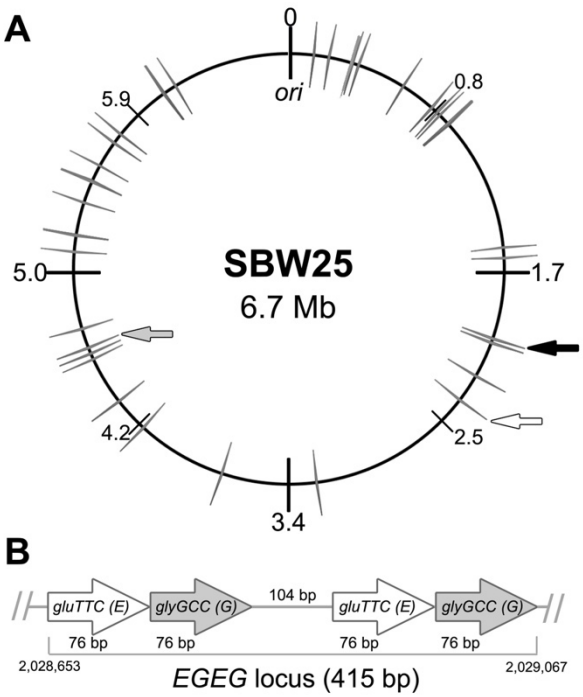
857

858

859

860

861



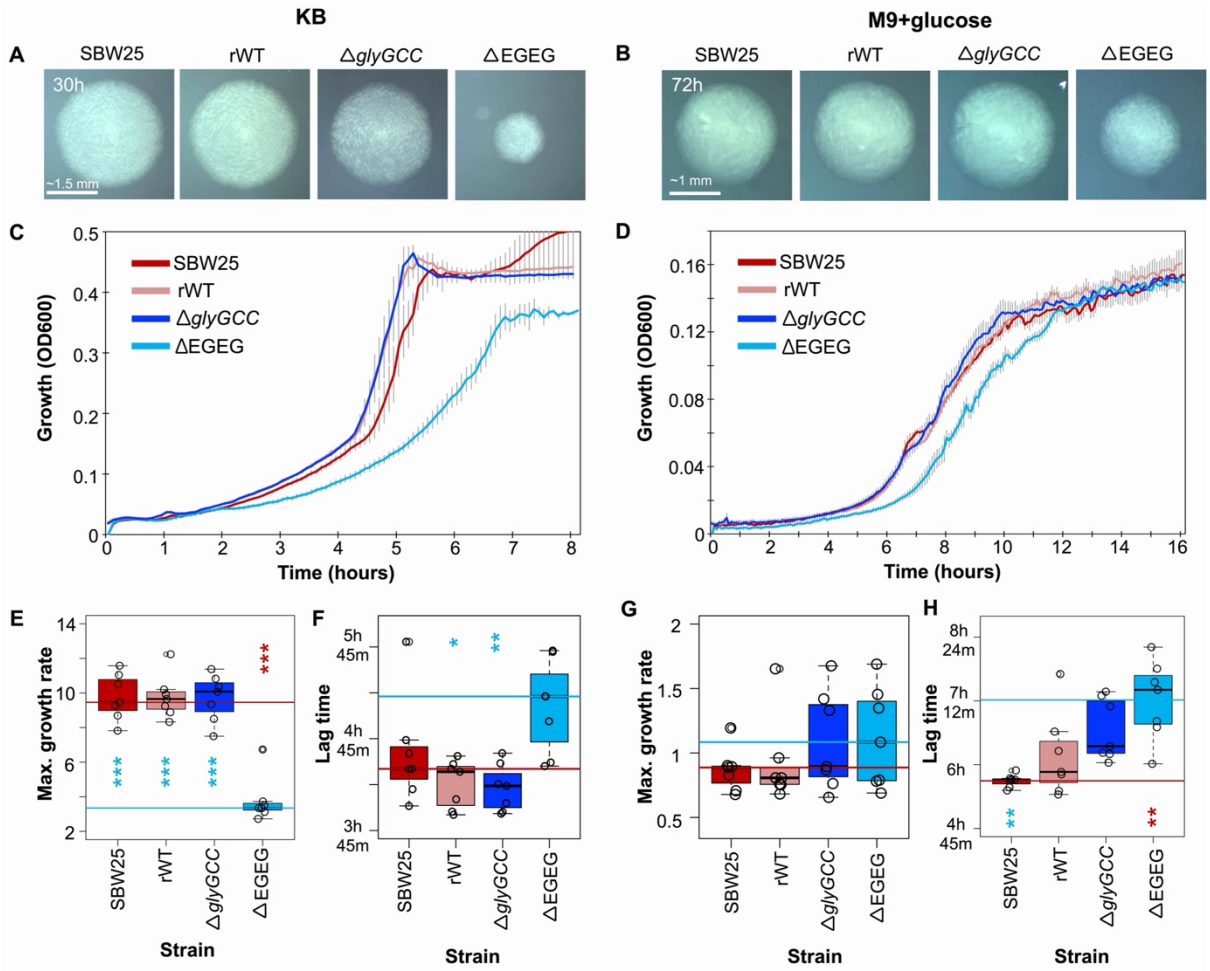
C

	GCN	Anti-codon	Codon	% (aa)	% (all)
Gly (G)	0	ACC	GGT	19.7	1.6
	3	GCC	GGC	62.8	5.0
	1	CCC	GGG	14.0	1.1
	1	UCC	GGA	3.6	0.3
totals	5			100	8.0
Glu (E)	0	CUC	GAG	45.0	2.4
	4	UUC	GAA	55.0	3.0
totals	4			100	5.4

D

	#tRNA genes	#tRNA species	GCN *glyGCC*	GCN *gluTTC*
SBW25	66	39	3	4
rWT	66	39	3	4
Δ*glyGCC*	65	39	2	4
ΔEGEG	62	39	1	2

Figure 1. Translation of glycine (G) and glutamic acid (E) in *P. fluorescens* SBW25. (**A**) Cartoon of the SBW25 chromosome depicting loci predicted by GtRNAdb2.0 to encode canonical tRNAs. Grey lines=tRNA loci (each representing between one and four tRNA genes); black arrow=*gluTTC-glyGCC-gluTTC-glyGCC* ("*EGEG*") tRNA locus deleted by genetic engineering; white arrow=*glyGCC* lone tRNA locus (encoding the remaining copy of *glyGCC*); grey arrow=*gluTTC-alaGGC-gluTTC-alaGGC* tRNA locus (encoding the remaining two copies of *gluTTC*). (**B**) Details of the *EGEG* tRNA locus indicated by the black arrow in panel A. (**C**) Putative translational relationships between anticodons and codons for glycine (top) and glutamic acid (bottom). Codons are listed with their cognate anticodons (black=tRNA species encoded in genome; grey=not encoded). Predicted anticodon-codon matching patterns are indicated by joining lines (black=cognate match; grey=wobble match). Percentage contribution of each codon to (i) the relevant amino acid (% aa), and (ii) all amino acids, is provided to 1 d.p. from GtRNAdb2.0 (tRNAscan-SE version 2.0.2; February 2019) (7). (**D**) Details of the tRNA gene sets carried by engineered genotypes.



877

878 **Figure 2. Deletion of the four gene tRNA locus *EGEG* reduces SBW25 growth in rich, undefined**
879 **(KB) and poorer, defined (M9+glucose) media.** (**A**-**B**) Colonies of SBW25, reconstructed wildtype
880 (the engineering control, rWT), and the two engineered genotypes Δ*glyGCC* and ΔEGEG at room
881 temperature for 30 hours on KB agar, or 72 hours on M9 agar. Scale bar applies to all four images in
882 each panel. (**C**-**D**) Growth (measured as absorbance at 600 nm) of the same genotypes in liquid KB
883 for 8 hours, or M9 for 16 hours. Lines show the mean of 7 replicates at 5-minute intervals; error bars
884 are ± 1 SE. (**E-H**) Maximum growth rates (change in absorbance; mOD min^{-1}) and lag times of growth
885 profiles in panel C (KB), and panel D (M9). In all four panels, red and blue lines are the medians of
886 SBW25 and ΔEGEG, respectively. Non-parametric Dunn's tests, followed by the Benjamini-Hochberg
887 procedure to correct for multiple comparisons, were used to test for a difference in medians between
888 pairs of genotypes in each panel. Stars indicate statistically significant *p*-values for a difference from
889 SBW25 (red stars) or ΔEGEG (blue stars).

890

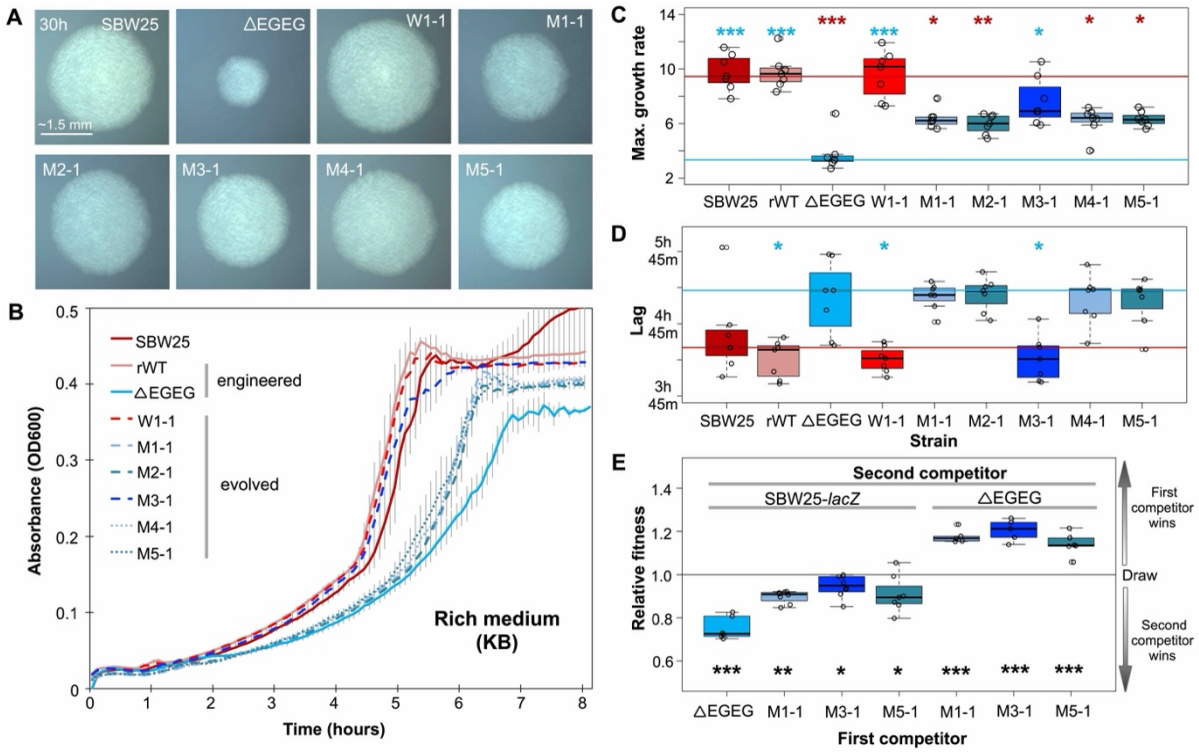


Figure 3. The growth defect caused by deleting four-gene tRNA locus *gluTTC-glyGCC-gluTTC-glyGCC* (*EGEG*) undergoes repeated compensation during a serial transfer experiment. (**A**) Colonies of founders SBW25, ΔEGEG, and six genotypes from day 21 of the serial transfer evolution experiment, grown on KB agar for 30 h at room temperature. Scale bar applies to all images. (**B**) Growth in liquid KB culture. Lines show the mean of 7 replicates at 5-minute intervals; error bars are ± 1 SE. (**C**,**D**) Maximum growth rate (change in absorbance; mOD min^{-1}) and lag time for each growth profile in panel B. In both panels, red and blue lines are the medians of SBW25 and ΔEGEG, respectively. Non-parametric Dunn's tests, followed by the Benjamini-Hochberg procedure to correct for multiple comparisons, were used to test for a difference in medians between pairs of genotypes. Stars indicate statistically significant *p*-values for a difference from SBW25 (red stars) or ΔEGEG (blue stars). The data for SBW25, rWT and ΔEGEG also appears in Figure 2. (**E**) Box plots of the relative fitness of competitor 1 (*x*-axis) and competitor 2 (horizontal bars at top). Direct, one-to-one competitions were performed in liquid KB for 24 hours (28˚C, shaking). Between 5 and 8 replicate competitions were performed for each genotype pair. Relative fitness >1 means that the first competitor wins, <1 means that the second competitor wins. Statistically significant deviations of relative fitness from 1 were determined using either one-tailed one-sample *t*-tests, followed by a Benjamini-Hochberg correction for multiple comparisons. Data points from all replicates are overlaid on boxplots, and ***$p<0.001$, **$p<0.01$, *$p<0.05$.

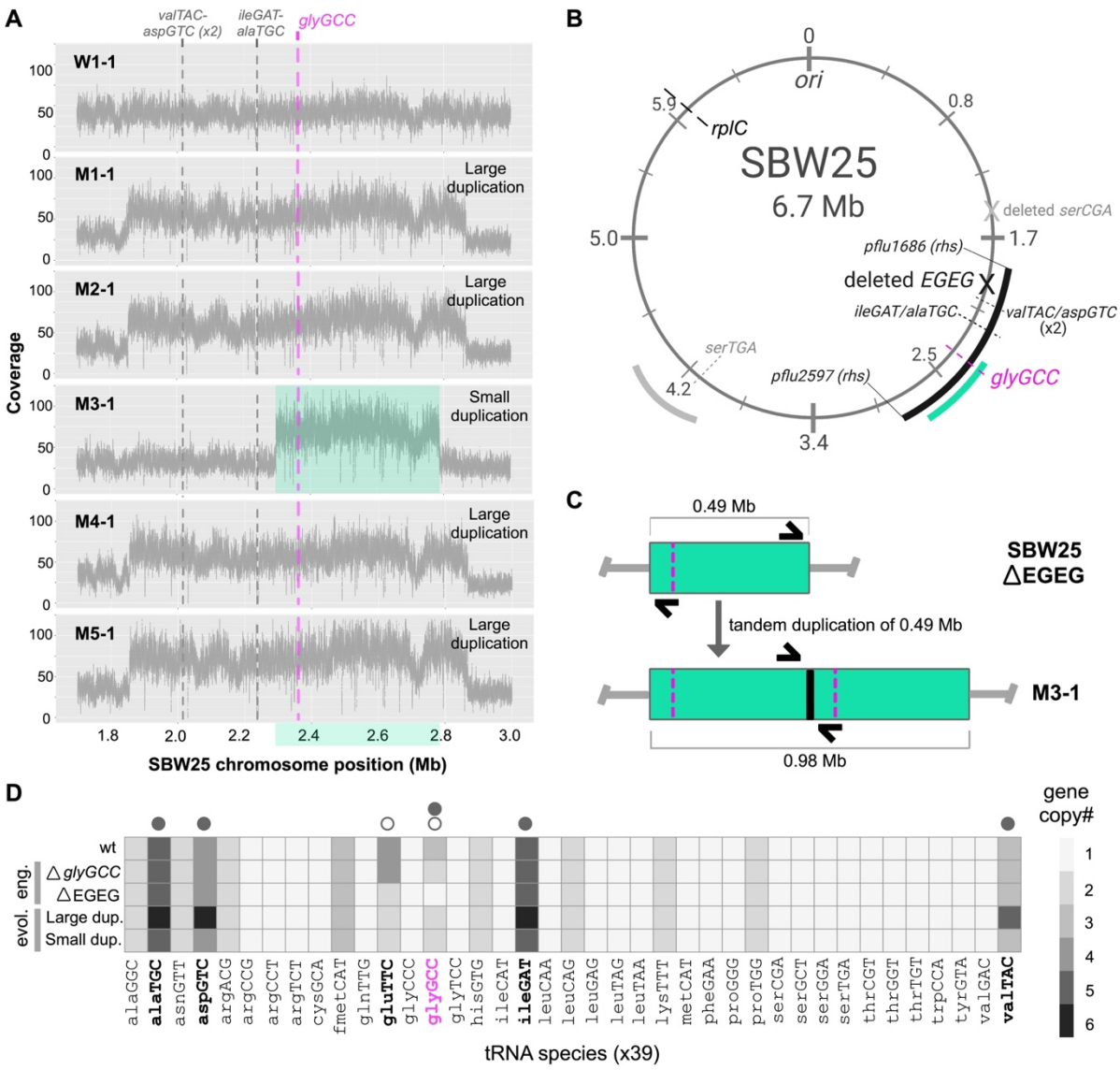


Figure 4. Large-scale, tandem duplications compensate for deletion of *EGEG*. (A) Number of raw sequencing reads covering 1.8 - 3.0 Mb of the reference genome. A jump to two-fold coverage is evident in the genotypes isolated from mutant populations (M1-1 to M5-1), and not in the control genotype isolated from the representative wildtype population (W1-1). The ~0.49 Mb region of double coverage shared by all mutants is highlighted in green; tRNA genes are shown by dotted lines. (**B**) Cartoon depicting the positions of duplicated regions on the SBW25 genome. Black arc=~1 Mb duplication (in M1-1, M2-1, M4-1, M5-1); green arc=~0.49 Mb duplication (in M3-1); grey arc=distinct ~0.4 Mb region in which duplication fragments have previously been reported (6). Duplicated tRNA genes are marked by dotted lines; deleted tRNA genes are marked with a cross. Image drawn using BioRender.com. (**C**) Depiction of the duplication event in M3-1, resulting in an additional ~0.49 Mb of DNA and double copies of all genes in the duplicated region. Thick black line=duplication junction, which can be amplified using a duplication junction PCR (primer positions in black; see Supplementary Text S2). (**D**) Heatmap showing differences in the gene copy numbers coding for 39 tRNA species between wildtype (wt), engineered (eng.), and evolved (evol.) genotypes. Open/closed circles=tRNA species whose gene copy number changed by engineering/evolution.

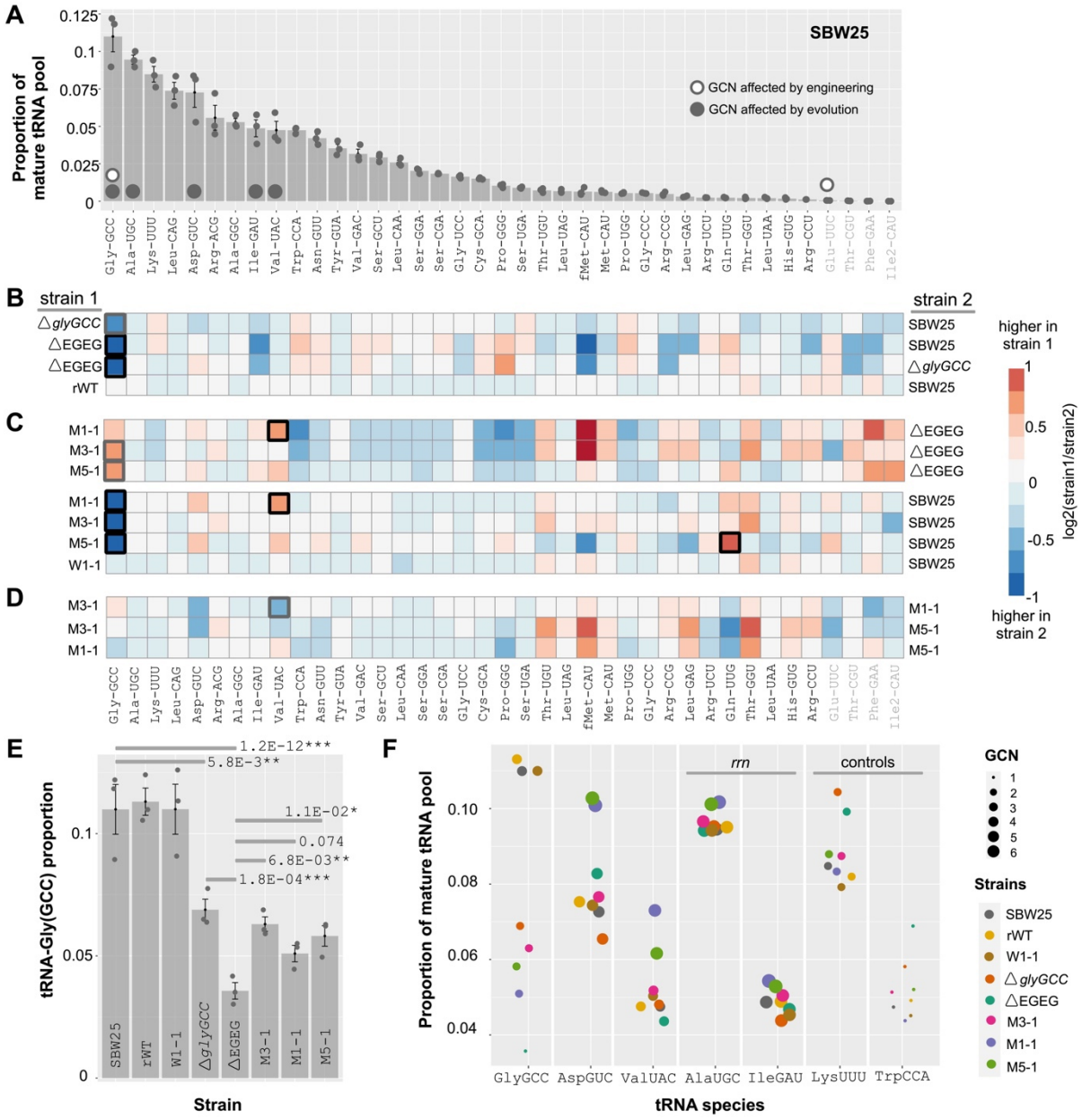


Figure 5. *glyGCC* gene copy number changes are mirrored in the mature tRNA pool. (**A**) Proportions of the 39 tRNA species in the SBW25 mature tRNA pool (highest to lowest). Light grey=excluded tRNA species. Open/closed circles=tRNA species whose gene copy number (GCN) changes by engineering/evolution in downstream experiments. (**B-D**) Heatmaps of differences in tRNA species proportions (expressed as log2-fold change(strain1/strain2); see Supplementary Table S5) between genotype pairs, to detect the effect of engineering (panel B) and evolution (panels C-D). Box outlines indicate statistical significance (DESeq2 adjusted p-values; thin line $p>0.01$; thick grey line $0.01<p>0.001$; thick black line $p<0.001$). (**E**) Bar graph of tRNA-Gly(GCC) proportion by genotype. Bars=means of 3 replicates ± 1 SE. DESeq2 adjusted *p*-values show tRNA-Gly(GCC) differences between genotypes pairs. (**F**) Scatter plot of mean proportions ± 1 SE of 7 tRNA species, by genotype. GCN for first 5 tRNA species varies by engineering and/or evolution (indicated by point size), while the final 2 are controls (constant GCN). *rrn*=tRNA species genes co-localize with rRNA operons.

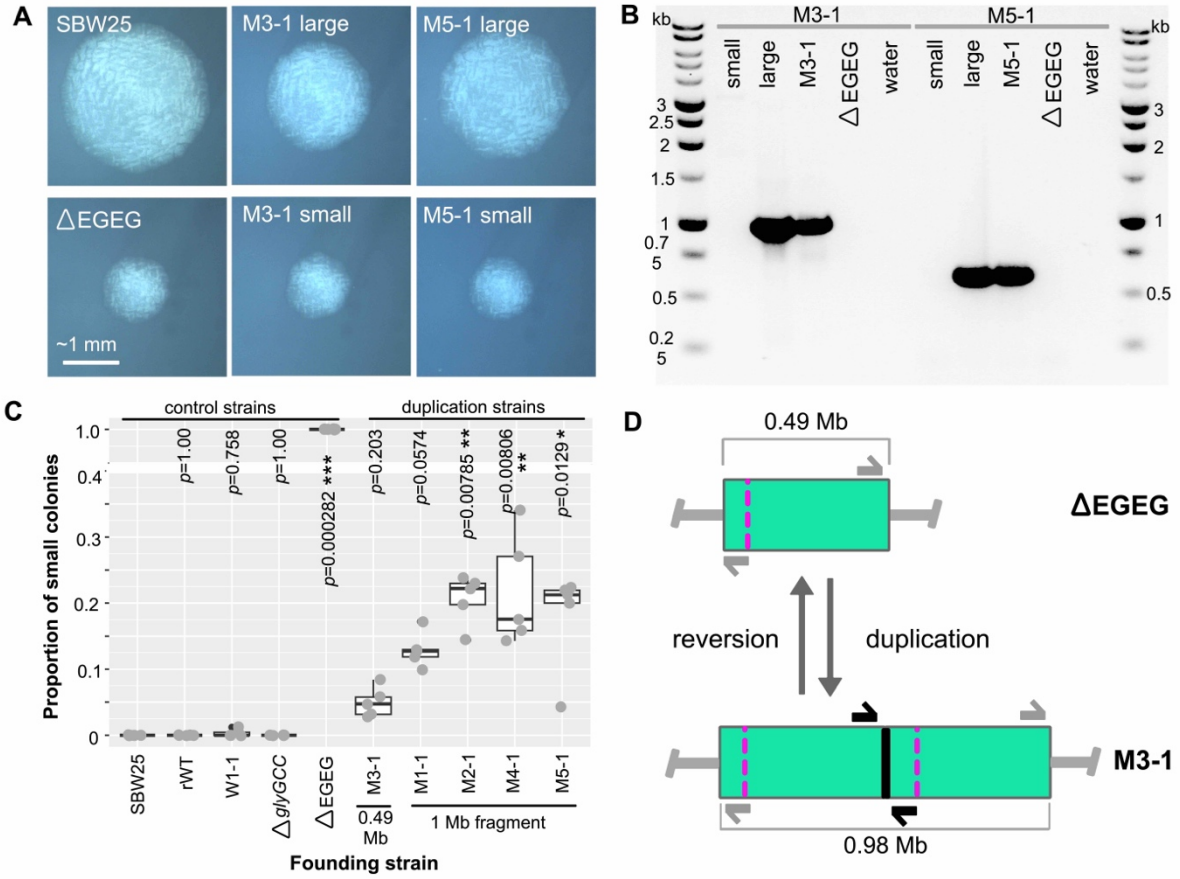


Figure 6. The large-scale, tandem duplications are rapidly lost in overnight culture. (**A**) Two colony morphologies – large and small – are observed when plating from overnight cultures of duplication-carrying genotypes (colonies from duplication genotypes M3-1 and M5-1 are presented here). All images taken under the same magnification after ~48 hours' incubation on KB agar at room temperature (~21˚C). Scale bar applies to all images in panel. (**B**) Small colonies no longer amplify the unique junction introduced by the relevant duplication fragment (illustrated by the thick black lines in M3-1 cartoon of panel D). (**C**) Duplication fragment loss was quantified as the proportion of small colonies when plating from an overnight culture. Five replicates were included per genotype, and at least 41 colonies were counted per replicate. Boxplots summarize the data (medians), and grey dots are individual replicates. Statistical significance of a difference in the median proportion of small colonies compared to SBW25 (wildtype) was calculated using a Dunn test (with Benjamini-Hochberg correction for multiple comparisons). Significance levels: no stars ($p>0.05$), *$0.05<p<0.001$, **$0.01<p<0.001$, ***$p<0.001$. (**D**) Cartoon depicting the gain and loss of duplication fragments, leading to a mixed population of the duplication genotype (here, M3-1) and ΔEGEG.

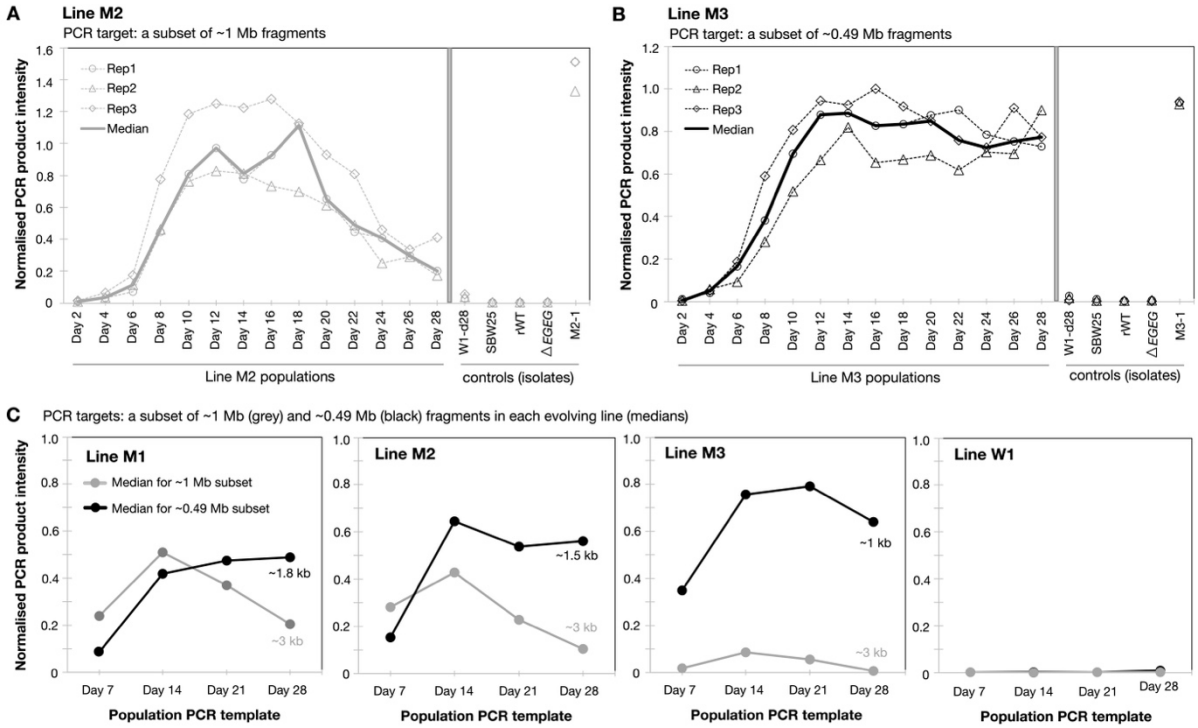


Figure 7. The complex evolutionary dynamics of large-scale duplication fragments in evolving mutant populations. Duplication junction PCRs were used to follow the dynamics of subsets of ~1 Mb and ~0.49 Mb duplications across evolving populations. The intensity of the PCR product obtained across each emergent duplication junction was measured relative to a PCR product amplified from a control region located outside of the genomic region affected by duplication events. Various control samples – in which (almost) all or none of the population is expected to carry duplication fragments – were included in each set of PCRs (see Methods for further details). The subsets of ~1 Mb (**A**; primer pair M1and2_junct_f/M5_junct_r) and ~0.49 Mb (**B**; primer pair M3_junct_f/r) duplications show different evolutionary trajectories in the lines in which they were originally identified; the ~1 Mb fragment rises and then wanes in line M2, while the ~0.49 Mb fragment rises to, and is maintained for longer at, a higher level in line M3. (**C**) Multiple duplication fragments arise and compete within evolving populations. Each PCR was performed in triplicate, and the median is presented here (for full data sets see Supplementary Figure S5A; Supplementary Table S6).

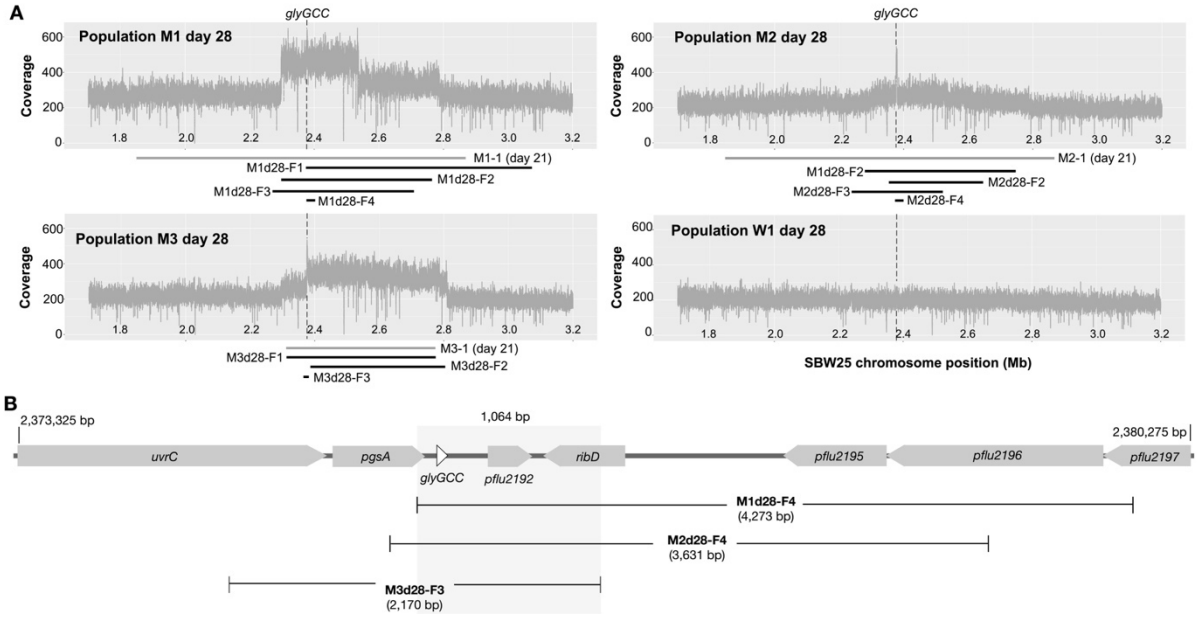


Figure 8. Eleven distinct duplication fragments are detected in mutant lines M1, M2, and M3 on day 28 of the evolution experiment. (**A**) Number of raw sequencing reads covering 1.7 Mb – 3.2 Mb of the reference genome, shown for the three mutant lines (M1, M2, M3) and one representative wildtype control line (W1) on day 28. Large changes in coverage can be seen in all mutant populations (but not the wildtype population). The black bars below each plot indicate approximate positions of each computationally predicted duplication (see Table 2); grey bars show the duplication identified in the corresponding day 21 isolate (see Table 1). (**B**) Structure of the ~7 kb genomic region containing the three smallest duplication fragments. Grey highlighting indicates the 1,064 bp that is duplicated in all three fragments. This region contains two complete genes: tRNA gene *glyGCC* and predicted pseudogene *pflu2192* (see Supplementary Table S3).