Large-scale duplication events underpin population-level flexibility in bacterial tRNA gene copy number

Zahra Khomarbaghi, Gökçe B. Ayan, Wing Y. Ngan, Gunda Dechow-Seligmann, Pabitra Nandy, Jenna Gallie*

Department of Evolutionary Theory, Max Planck Institute for Evolutionary Biology, Plön, 24306, Germany

* To whom correspondence should be addressed. Email: gallie@evolbio.mpg.de

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

ABSTRACT

The complement of tRNA genes within the genome is typically regarded as a 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 spontaneous, high-rate occurrence of large-scale, tandem duplication events in laboratory populations of the bacterium *Pseudomonas fluorescens* SBW25. The identified duplication fragments 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 duplication fragments 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 that rapidly generates tRNA gene set diversity, 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 the high-rate segregation of the duplication fragment ensures the maintenance of the original tRNA gene set. However, if a tRNA gene set variant is selectively beneficial, the underlying duplication fragments persist and provide the raw material for further, more stable, evolutionary changes.

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 tRNA genes, encoding 39 and 40 different types of mature tRNA (tRNA 'isotypes') (6–8). In this work we highlight that, at the population

level, tRNA gene set of a given bacterial strain can be more flexible than previously appreciated, particularly with regard to the number of gene copies encoding each tRNA isotype.

The composition of tRNA gene sets differs across the bacterial phylogeny; extant bacterial tRNA gene sets differ in both the isotypes that they encode (isotype patterns) and in the number of gene copies coding for each isotype (7, 9–15). Variations in isotype patterns are possible because of wobble base pairing, the process by which some isotypes can – either naturally (16) or through post-transcriptional modifications (17, 18) – translate multiple codons. Ultimately, wobble base pairing allows all 62 codons (61 sense codons plus the AUG initiator codon) to be translated by far fewer than 62 directly corresponding tRNA types, with most bacteria encoding between 25 and 46 isotypes (7, 9, 19). In contrast to the reductionist strategy observed for isotype patterns, many tRNA types are encoded by multiple, often identical, gene copies; tRNA gene copy numbers differ widely both between and within bacterial genomes, with most isotypes being encoded by between one and eight gene copies (7, 10).

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

The molecular mechanisms by which tRNA genes are gained and lost from tRNA gene sets have been inferred by phylogenetic analyses. Such computational studies have provided evidence for the evolution of extant bacterial tRNA gene sets via a number of routes, including: (i) mutation of tRNA genes to encode new isotypes ('anticodon switching'), (ii) loss of tRNA genes by deletion, and (iii) acquisition of new tRNA genes, either by within-genome duplication events, or by horizontal gene transfer (10, 12–14, 37–39). In addition, we recently provided an empirical example of a bacterial tRNA gene set undergoing adaptive evolution; an engineered, deleterious change in the isotype pattern of the fast-growing, model bacterium *P. fluorescens* SBW25 was compensated by within-genome duplication events that served to increase the copy number of a remaining tRNA gene (*serTGA*) (6). Others have reported that translational challenges imposed by (i) novel codon usage patterns, or (ii) the fitness costs associated with antibiotic resistance mutations can be readily alleviated by similar duplication (and, sometimes, downstream amplification) events that affect the copy numbers of remaining tRNA genes (40, 41). Together, these studies demonstrate a considerable degree of evolutionary flexibility in the composition of bacterial tRNA gene sets.

77 In the above-mentioned *P. fluorescens* SBW25 tRNA gene study (6), the adaptive increase in
78 *serTGA* copy number was shown to be mediated through the tandem duplication of large sections of
79 the *P. fluorescens* SBW25 chromosome (ranging in size from ~45 kb to 290 kb). In addition to the 91-
80 bp target *serTGA* gene, these duplication events affect the dosage of up to 278 other genes, and
81 generate considerable alterations in chromosome size and architecture. Similar types of large-scale,
82 tandem duplications are a pervasive phenomenon in bacterial genomes (42–46) and – despite their
83 large size and widespread gene dosage effects – can provide a net selective advantage in response
84 to a range of selective pressures (47–53). Notably, these types of mutations have been reported to
85 arise spontaneously, and at extremely high rates; estimates range from frequencies of 10^{-1} to 10^{-5}
86 duplications *per* cell *per* generation (43, 54–57), several orders of magnitude higher than those
87 expected for single nucleotide polymorphisms (SNPs) (58). Given their pervasiveness and ease of
88 occurrence, such large-scale, tandem duplications challenge the prevailing view of fixed, stable tRNA
89 gene sets, particularly in large, rapidly growing bacterial populations.
90 In this work, we firstly show that large-scale, tandem duplication events affecting the *P.*
91 *fluorescens* SBW25 tRNA gene content are not limited to the genomic region surrounding *serTGA*
92 (located at ~4.16 Mb on the SBW25 chromosome). Specifically, we demonstrate that a fitness defect
93 caused by deleting two of three *glyGCC* tRNA gene copies is repeatedly and rapidly compensated by
94 large-scale, tandem duplication events encompassing the remaining *glyGCC* gene (at genomic
95 position ~2.38 Mb). The identified duplication fragments are several times larger than those reported
96 for *serTGA*; they are up to~ 1 Mb in size, and encompass up to 917 genes. Secondly, we provide
97 empirical evidence that these duplication fragments arise, and are lost, spontaneously and at very
98 high frequencies; we show that in each independently evolving *glyGCC*-deficient population, multiple
99 distinct duplication fragments (and hence, tRNA gene sets) arise rapidly and compete for dominance.
100 We present evidence of duplication events occurring in evolving wildtype populations (i.e., in the
101 absence of positive selection). Our results demonstrate that large-scale, tandem duplication events
102 can rapidly generate population-level diversity in bacterial tRNA gene set composition, providing
103 variation upon which natural selection can act.

104 **MATERIALS AND METHODS**

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

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

115 **Growth curves and analysis.** Seven colonies *per* genotype were inoculated into individual wells of a
116 96-well plate (each containing 200 µL KB), and grown overnight (28˚C, 200 rpm). From each well, 2
117 µL of grown culture was transferred into 198 µL fresh KB, covered with a Breathe-Easy® membrane,
118 and grown at 28˚C in an Epoch 2 Microplate Spectrophotometer (BioTek). Absorbance at 600nm
119 (OD_{600}) was measured at five-minute intervals, with 5 seconds of 3 mm orbital shaking preceding
120 each read. Average maximum growth rates (V_{max}, reported as change in mOD min^{-1}) and lag times
121 were calculated using a sliding window of six time points during rapid growth (between 1 and 8 hours
122 in KB, and between 1 and 10 hours in M9) using Gen5 Software version 3.00.19. Where assumptions
123 were satisfied, a parametric one-way ANOVA was used to test for differences in the mean maximum
124 growth rate between pairs of genotypes. Where the assumptions were violated, a non-parametric
125 Dunn's test, followed by the Benjamini-Hochberg procedure to correct for multiple comparisons, was
126 used to test for differences in the median lag time between pairs of genotypes.

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

144 **Evolution experiment.** SBW25 (wildtype) and ΔEGEG (quadruple tRNA gene deletion mutant) were
145 streaked from glycerol stocks onto KB agar and grown (48 hours, 28˚C). Five single colonies *per*
146 genotype were grown overnight (4 mL KB; 28˚C, 200 rpm). Grown cultures ("day 0") were vortexed
147 and 100 µL of each used to inoculate 9.9 ml KB in a 50 ml Falcon tube (28˚C, shaking, 24 hours).
148 Every 24 hours thereafter, 1 % (100 µL) of each culture was transferred to a fresh 9.9 mL KB, and a
149 sample of the population frozen in glycerol saline at -80˚C. The experiment was continued until day 28.
150 An additional, medium control lineage – containing no initial inoculum – was also transferred daily.

151 **Whole genome re-sequencing.** Six isolates were purified from day 21 of the evolution experiment
152 (W1-1, M1-1, M2-1, M3-1, M4-1, M5-1), and stored at -80˚C. Genomic DNA was isolated from 1 mL of

153 overnight culture of each isolate (DNeasy Blood & Tissue Kit; Qiagen). Paired-end, 250 bp reads
154 were generated with an Illumina MiSeq instrument at The Max Planck Institute for Evolutionary
155 Biology (Ploen, Germany), according to standard procedures. Raw reads are available at NCBI
156 Sequence Read Archive (SRA accession number: PRJNA790786) (63). A minimum of 1.08 million
157 raw reads *per* strain were obtained and aligned to the SBW25 reference genome sequence
158 (NC_012660.1) (8) using *breseq* (64–66) and Geneious (v11.1.4). A minimum mean coverage of 36.4
159 reads *per* base pair was obtained. Where appropriate, predicted mutations were confirmed by
160 standard PCR or duplication junction PCR (see below), followed by Sanger sequencing (see
161 Supplementary Text S2). A detailed list of final mutation predictions is provided in Supplementary
162 Table S2, and the genes contained in each duplication fragment are listed in Supplementary Table S3.

163 **Mature tRNA pool sequencing (YAMAT-seq).** Three independent replicates of eight genotypes (*i.e.*,
164 24 samples) were used for YAMAT-seq (67) according to the method described in (6). Briefly, (i) per
165 sample, a single colony was chosen from KB agar (in the case of genotypes carrying duplication
166 fragments, rapidly-growing colonies were chosen in an attempt to mitigate the inherent instability of
167 the duplication fragment; see **Stability test** method below), grown in overnight KB culture, diluted
168 1/10 in fresh KB, and grown to mid-exponential phase, (ii) total RNA was isolated from each rapidly-
169 growing liquid culture using a TRIzol Max Bacterial RNA isolation kit, (iii) amino acids were removed
170 from mature tRNAs by deacylation treatment with 20 mM Tris-HCl, (iv) Y-shaped, DNA/RNA hybrid
171 adapters (Eurofins) were annealed to the conserved, exposed 5'-NCCA-3' and 3'-inorganic
172 phosphate-5' ends of non-acylated tRNAs, and ligated T4 RNA ligase 2, (v) ligation products were
173 reverse-transcribed into cDNA using SuperScript™ III reverse transcriptase, (vi) cDNA products were
174 amplified, and a unique barcode added, with eleven cycles of PCR using Phusion™, (vii) products
175 were quantified using a Bioanalyzer and fluorescent Nanodrop, and mixed in equimolar amounts, (viii)
176 the mixed sample was run on a 5 % polyacrylamide gel, the bands between ~210 bp and ~280 bp
177 (tRNAs plus adapters and barcodes) cut out and extracted overnight in deionized water, (ix) the
178 purified product was quantified and subjected to deep sequencing at the Max Planck Institute for
179 Evolutionary Biology (Ploen, Germany). Single-end, 150 bp reads were generated with a NextSeq
180 550 Output v2.5 kit (Illumina), and are available at NCBI Gene Expression Omnibus (GEO accession
181 number GSE196705) (68).

182 **YAMAT-seq analysis.** The raw YAMAT-seq sequencing reads were split into 24 samples by
183 extracting exact matches to each unique, 6 bp Illumina index. The resulting 24 set of raw reads, each
184 containing a minimum of 1,470,795 reads, were analysed using Geneious (version 11.1.4). Reads of
185 the expected length (80 bp - 151 bp) were extracted (resulting in at least 99.8 % read retention *per*
186 sample). The extracted reads were assembled to a set of 42 reference tRNA sequences from SBW25
187 (6, 7) (Supplementary Text S3). During assembly, up to 10 % mismatches, gaps of <3 bp, and up to
188 five ambiguities were allowed *per* read. Reads that aligned equally well to more than one reference
189 sequence were discarded in order to minimize misallocation of reads to similar reference sequences.
190 The unused reads for each sample were *de novo* aligned, and the resulting contigs were checked to
191 ensure that none contained substantial numbers of tRNA reads. The within-sample proportion of

192 reads aligned to each tRNA isotype was calculated. Next, mean mature tRNA proportions were
193 calculated for each strain across the three replicates (Supplementary Table S4). DESeq2 (69) was
194 used in R (version 3.6.0) to detect tRNA expression differences between pairs of strains
195 (Supplementary Table S5). DESeq2 corrects for multiple testing using the Benjamini-Hochberg
196 procedure (70).

197 **Stability test**. Genotype stability was estimated using colony morphology stability tests, according to
198 the method detailed in (71). Briefly, 10 genotypes (WT, rWT, ΔEGEG, Δ*glyGCC*, W1-1, M1-1, M2-1,
199 M3-1, M4-1, and M5-1) were streaked from glycerol stocks onto KB agar plates and grown (48 hours,
200 28˚C). Five colonies *per* genotype were used to inoculate 4 ml KB cultures (28˚C, 200 rpm, overnight).
201 Grown cultures were dilution plated on KB agar (48 hours, room temperature), and the number of
202 large and small colonies *per* plate was enumerated. Between 41 and 295 colonies were recorded *per*
203 replicate, with no more than 200 colonies counted on any given plate. For each replicate, the
204 proportion of small colonies was calculated. A non-parametric Kruskal-Wallis test was used to test for
205 evidence of a difference in the median of at least one founder (p=6.25e-07). To determine which
206 strains showed a different median rate of small colony enumeration as compared with SBW25
207 (wildtype), Dunn's test was used with Benjamini-Hochberg procedure to adjust for multiple
208 comparisons. A Welch non-parametric two-sample *t*-test was used to test for a difference in the mean
209 proportion in small colonies arising from the genotype carrying the small duplication fragment (M3-1),
210 and those carrying a large duplication fragment (M1-1, M2-1, M4-1, M5-1).

211 **Duplication junction PCR and estimating duplication fragment frequencies.** Large-scale, tandem
212 duplication events result in the emergence of a unique piece of DNA (the 'duplication junction'), which
213 can be amplified by PCR specifically in isolates (or populations) carrying the relevant junction.
214 Primers targeting each of the five identified duplication junctions were used for duplication junction
215 PCRs on isolates and/or populations. PCR templates were prepared by (i) growing isolates or
216 populations in overnight KB culture, (ii) pelleting by centrifugation, and (iii) re-suspending in twice the
217 volume of sterile, deionized water. PCRs were performed with GoTaq Flexi (Promega) for 30 cycles of:
218 15 sec at 94˚C, 15 sec at 58˚C or 59˚C, 72˚C for 1 min *per* kb of expected product (see
219 Supplementary Text S2 for full details). Five µl of each PCR product was mixed with loading dye and
220 run on a 1 % agarose gel (SYBR Safe; Life Technologies) at 90 V for ~45 min, against 1 kb DNA
221 ladder (Promega) and photographed under UV illumination. To improve visibility of low intensity bands,
222 the colour of gel images was subsequently inverted in Preview (version 11.0) or ImageJ (version 1.53)
223 (72). Where indicated, the relative densities (*i.e.*, intensities) of duplication junction bands from a
224 single gel image were quantified using ImageJ (see Supplementary Table S6). Intensities were
225 converted to a percentage of a positive control from the same image, and relative intensities were
226 plotted using the pheatmap package (version 1.0.12) in R (version 3.6.0). Gel images are provided on
227 Zenodo (https://zenodo.org/record/7276661#.Y3JJHYKZONh).

228

229

230 **RESULTS**

231 **Reducing the gene copy number of essential tRNA types slows SBW25 growth**

232 The *P. fluorescens* SBW25 genome is predicted to contain 66 canonical tRNA genes encoding 39
233 isotypes, 14 of which are coded for by multiple (between two and five) gene copies (Figure 1A) (6, 8,
234 73). Here, we focus on one such multicopy isotype: tRNA-Gly(GCC), which is encoded by three
235 identical *glyGCC* gene copies. One copy occurs as a lone tRNA gene at chromosomal position ~2.38
236 Mb, and two copies are located in a four-gene tRNA locus (*gluTTC-glyGCC-gluTTC-glyGCC*,
237 hereafter referred to as "*EGEG*") at ~2.03 Mb (Figure 1A-B). Given that tRNA-Gly(GCC) is predicted
238 to be the sole decoder of codons 5'-GGC-3' (cognate match) and 5'-GGU-3' (G_{34}:U_3 wobble match), at
239 least one functional *glyGCC* gene copy is required for translation and growth (Figure 1C). With this in
240 mind, two genotypes were constructed in which the SBW25 *glyGCC* gene copy number was
241 sequentially reduced: (i) Δ*glyGCC*, in which the lone *glyGCC* copy was removed from SBW25
242 (leaving two copies of *glyGCC*), and (ii) ΔEGEG, in which the entire *EGEG* locus was removed from
243 SBW25 (leaving only a single copy of *glyGCC*) (Figure 1D; Supplementary Text S1). In addition to
244 changes in *glyGCC* copy number, ΔEGEG carries a reduction in *gluTTC* copy number (from four to
245 two). The remaining two *gluTTC* copies were not manipulated because they are interspersed with
246 other tRNA genes that are expected to be essential for survival (the only two copies of *alaGGC*,
247 encoding essential isotype tRNA-Ala(GGC)) (Figure 1A).

248 Growth assays showed that the deletion of one *glyGCC* copy had no detectable effect on
249 growth in a rich, undefined medium (KB) or a minimal, defined medium (M9+glucose), while the
250 elimination of two *glyGCC* copies resulted in an immediately obvious growth defect in both media
251 (particularly in KB). Specifically, compared with SBW25, ΔEGEG forms smaller colonies on KB and
252 M9 agar plates, and grows more slowly in liquid culture (Figure 2). Re-introduction of *EGEG* into
253 ΔEGEG – generating the reconstructed wildtype (rWT) – restores wildtype-like growth. These results
254 demonstrate a degree of redundancy in SBW25 *glyGCC* gene copy number; at least one of the three
255 *glyGCC* copies can be lost without significantly impacting growth rate. This observation is consistent
256 with earlier reports that multiple copies of various *E. coli* tRNA genes (including *glyGCC*) show a
257 degree of redundancy that is most beneficial when nutrients are plentiful (36).

258 **The ΔEGEG growth defect is rapidly compensated by evolution**

259 To determine whether the ΔEGEG growth defect can be genetically compensated, a serial transfer
260 evolution experiment was performed. Ten independent lineages were passaged through repeated
261 cycles of growth and dilution in shaken rich medium (KB), for 28 days (a total of ~200 generations
262 over 28 days; 1 % daily bottleneck). Five evolutionary lineages were founded by ΔEGEG (lineages
263 M1-M5), and five by the wildtype (control lineages W1-W5). Within 21 days (~150 generations) each
264 of the five mutant lineages showed visibly improved growth. Isolates were purified from day 21 of
265 each mutant lineage (giving M1-1, M2-1, M3-1, M4-1, and M5-1). All five isolates showed improved
266 growth on agar plates and in liquid culture, with the growth profile of one isolate (M3-1) approximating
267 that of the wildtype (Figure 3A-D). A control isolate from day 21 of the wildtype lineage W1 (W1-1)
268 showed a growth profile comparable to that of SBW25 (Figure 3A-D).

269 Three of the evolved isolates from mutant lineages (M1-1, M3-1, and M5-1) were chosen for
270 direct, 1:1 competition experiments with ΔEGEG and a neutrally marked wildtype strain (SBW25-*lacZ*).
271 Each of the three evolved isolates outperformed ΔEGEG (two-tailed one-sample *t*-tests $p<0.0006$),
272 but were themselves outperformed by SBW25-*lacZ* (two-tailed one-sample *t*-tests $p<0.05$) (Figure 3E).
273 Together, these results demonstrate that the growth defect caused by *EGEG* elimination can be
274 readily compensated – at least in part – by evolution.

275 **Compensation of *EGEG* loss occurs via large-scale, tandem duplications spanning *glyGCC***
276 To identify the mutations behind compensation, whole genome re-sequencing (Illumina) was
277 performed on each of six isolates from day 21 of the evolution experiment: one isolate from each of
278 five independent mutant lineages (M1-1, M2-1, M3-1, M4-1, and M5-1), plus a control isolate from
279 wildtype lineage W1 (W1-1). *Per* sample, over one million, 250-bp long raw sequencing reads were
280 aligned to the SBW25 reference genome (8), resulting in a minimum mean coverage of 36.4 reads
281 *per* genomic base. A full list of predicted mutations per isolate is provided in Supplementary Table S2.
282 In each mutant lineage isolate, a large, direct, tandem duplication was identified at around
283 ~2.38 Mb of the SBW25 chromosome (Figure 4A-C). In addition, in isolate M5-1, a non-synonymous
284 point mutation was identified in *rplC* (located outside of the duplicated region; see Figure 4B). While
285 the *rplC* mutation affects the translational apparatus (via amino acid change T43I in the 50S
286 ribosomal protein L3), no evidence was found for a role in the compensated phenotype of M5-1; in the
287 absence of the accompanying duplication fragment, the *rplC* mutation did not significantly improve the
288 growth of ΔEGEG (see Supplementary Figure S1). No evidence of any mutations was found in the
289 wildtype control isolate (W1-1).
290 The duplicated region in each mutant lineage isolate was more precisely defined using
291 computational analyses, PCR, and Sanger sequencing (Table 1, Supplementary Text S2). The
292 duplication fragments fall into two classes: large (~1 Mb) and smaller (~0.49 Mb) duplication
293 fragments. Large duplication fragments were identified in four isolates (M1-1, M2-1, M4-1, and M5-1).
294 This class of fragments occurs between long stretches of repetitive DNA (intergenic repeats and/or
295 *rhs* genes) at ~1.85, Mb and ~2.86 Mb of the SBW25 chromosome, and encompasses over 900
296 putative genes. The smaller duplication fragment was identified in only one isolate, M3-1, between
297 regions with no obvious homology at ~2.3 Mb and ~2.8 Mb of the SBW25 chromosome. This smaller
298 fragment encompasses a predicted 426 genes (see Supplementary Table S3). Notably, each class of
299 fragment contains a different set of tRNA genes: the large duplication fragments encompass seven
300 tRNA (giving a total of 69 tRNA genes), while the smaller fragment contains one tRNA gene (63 tRNA
301 genes) (Figure 4D).
302 To identify which part of the duplicated region compensates for loss of the *EGEG* locus, we
303 focused on the smaller duplication fragment because (i) this smaller region is duplicated in all five
304 evolved genotypes (see Figure 4B), and (ii) it affords the greatest growth advantage of all observed
305 duplication fragments (see Figure 3B). Of the 426 putative genes in this 0.49 Mb fragment, only one
306 has an obvious role in translation: tRNA gene *glyGCC*. Importantly, this is the sole remaining gene
307 copy encoding one of the tRNA isotypes targeted by the original engineering process (see Figure 1).
308 These results suggest that the observed fitness changes are primarily caused by differences in

309 *glyGCC* gene copy number: the ΔEGEG growth defect likely results from a reduction in *glyGCC*
310 copies (from three to one copy), while the regain in fitness by the evolved strains is most likely the
311 result of a (partial) recovery in *glyGCC* copy number (from one to two copies).
312
313 **tRNA gene set changes affect the proportional contribution of tRNA-Gly(GCC) to the mature**
314 **tRNA pool**
315 We sought to test whether the above tRNA gene set changes are reflected in mature tRNA pools
316 using YAMAT-seq, a deep-sequencing technique that quantifies the relative proportions of tRNA
317 isotypes in the mature tRNA pool (6, 67). YAMAT-seq was performed on 24 samples, including three
318 replicates of each of eight genotypes: wildtype SBW25, two *glyGCC* deletion genotypes (Δ*glyGCC*,
319 ΔEGEG), three evolved genotypes containing compensatory duplication fragments (M1-1, M3-1, and
320 M5-1), and two control genotypes (engineering control rWT, evolution control W1-1). Illumina
321 sequencing resulted in a minimum of 1.47 million raw reads *per* sample; these were aligned to a set
322 of reference sequences, consisting of all unique tRNA gene sequences predicted in SBW25 (6, 7)
323 (see Supplementary Text S3). Each of the expected 39 tRNA isotypes was detected in every sample,
324 in varying proportions (Supplementary Table S4; see Figure 5A for a representation of the wildtype
325 tRNA pool). Notably, tRNA-Gly(GCC) – encoded by the three *glyGCC* genes of interest – is the most
326 highly represented of all isotypes in the wildtype, accounting for 11.0 ± 1.0 % (mean ± SE) of the
327 mature tRNA pool (Figure 5A). Four isotypes (tRNA-Ile(CAU), tRNA-Phe(GAA), tRNA-Thr(CGU), and
328 tRNA-Glu(UUC)) were detected only at very low levels (<0.1 % of the pool). It is probable that various
329 post-transcriptional modifications present in these tRNA isotypes reduced the efficacy of the
330 YAMAT-seq protocol (specifically, the reverse transcription reaction; see Methods) (6, 67, 74). Due to
331 unreliably low read numbers, these four isotypes were removed from downstream statistical analyses.
332 DESeq2 analyses were used to detect differences in the relative proportions of all 39 tRNA
333 isotypes between genotype pairs of interest (Figure 5B-D, Supplementary Table S5) (6, 36, 69). The
334 predominant – and in many cases, only statistically significant – effect of the engineering and
335 evolution experiments was on isotype tRNA-Gly(GCC) (Figure 5B-E). Deletion of either *glyGCC* or the
336 *EGEG* locus (see Figure 1B) saw a reduction in the relative contribution of tRNA-Gly(GCC) to the
337 mature tRNA pool. Specifically, compared with 11.0 ± 1.0 % of the wildtype tRNA pool,
338 tRNA-Gly(GCC) accounts for only 6.9 ± 0.43 % and 3.6 ± 0.34 % of the Δ*glyGCC* and ΔEGEG mature
339 tRNA pools, respectively. The severe reduction in tRNA-Gly(GCC) in ΔEGEG was partially reversed
340 by adaptive evolution; isolates M1-1, M3-1, and M5-1 show tRNA-Gly(GCC) levels of 5.1 ± 0.34 %,
341 6.3 ± 0.30 %, and 5.8 ± 0.42 %, respectively. No such changes were observed in the tRNA pools of
342 the engineering or evolution controls (rWT, W1-1) (Figure 5B-E). Unfortunately, due to the unreliably
343 low read numbers obtained for tRNA-Glu(UUC) (see above), we are unable to comment on the effect
344 of *EGEG* deletion on this isotype.
345 In addition to tRNA-Gly(GCC), the 1 Mb duplication carried by isolates M1-1 and M5-1
346 changes the number of genes coding for a further four tRNA isotypes: tRNA-Ala(UGC) (from five to
347 six gene copies), tRNA-Asp(GUC) (from four to six gene copies), tRNA-Ile(GAU) (from five to six gene
348 copies), and tRNA-Val(UAC) (from three to five gene copies). While the large, 1 Mb duplication

349 fragment increases the contribution of each of these isotypes to the mature tRNA pool, these changes
350 are less pronounced than those observed for tRNA-Gly(GCC) (and very few are statistically significant)
351 (Figures 5C & 5F). One reason for these weaker effects is that each of the additional four isotypes is
352 encoded by multiple, dispersed tRNA genes, only some of which lie within the duplicated region.
353 Finally, we note that in addition to incurring no detectable growth or fitness effects (see
354 Supplementary Figure S1), the *rplC* mutation had no major effects on the composition of the mature
355 tRNA pool under the tested conditions; no statistically significant differences were detected between
356 isolates M1-1 (1 Mb duplication) and M5-1 (1 Mb duplication + *rplC* mutation) (Figure 5D).
357 Results thus far show that changes in *glyGCC* gene copy number are mirrored by changes in
358 the proportional contribution of tRNA-Gly(GCC) to the mature tRNA pool. Specifically, reductions in
359 *glyGCC* gene copy number result in progressively lower proportions of tRNA-Gly(GCC), which
360 eventually manifest as a reduction in growth rate. Growth rate is subsequently restored by large-scale,
361 tandem duplication events that likely serve to elevate *glyGCC* gene copy number, thereby partially
362 reinstating tRNA-Gly(GCC) levels.

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

364 Large-scale, tandem duplications are typically highly unstable; they are known to be lost at high rates
365 via homologous recombination events during DNA replication (often leaving no trace in the segregant
366 genome) (55, 75, 76). The duplication fragments observed in this work are no exception. When grown
367 in overnight culture and plated on agar plates, each of the duplication-carrying strains repeatedly
368 gives rise to colonies of two distinct types: (i) a majority of larger, wildtype-like colonies that retain the
369 duplication fragment, and (ii) a sizeable minority of smaller, ΔEGEG-like colonies in which the initial
370 duplication fragment has been lost (Figure 6A-B).
371 The degree of duplication fragment loss was quantified using a stability test, in which the
372 proportion of small colonies (*i.e.*, segregants) arising from independent cultures founded by each
373 duplication-carrying strain was determined. Between 3 % and 34 % of colonies arising from
374 duplication genotype cultures showed the small, ΔEGEG-like phenotype (Figure 6C). The proportion
375 of small colonies observed was dependent on the size of the initial duplication fragment, with small
376 and large duplication fragments giving rise to respective averages of 5.0 % ± 1.0% and 18 % ± 1.5 %
377 segregants (mean ± se; non-parametric two sample *t*-test for difference between means p=3.5e-07***).
378 The higher proportion of segregants in populations founded by duplication strains carrying the larger
379 fragment results from (i) a higher number of recombination events occurring across the more
380 extensive region of homology, and (ii) the lesser fitness advantage afforded by the 1 Mb duplication
381 fragment (see Figure 3).
382 The stability test outlined above demonstrates that any sufficiently large population founded
383 by any of the five duplication-carrying strains will contain not only the duplication strain, but also a
384 variable – and sometimes considerable – fraction of segregants. This includes populations grown
385 during the phenotypic and genotypic assays in the earlier sections of this work (*e.g.*, cultures from
386 which DNA was extracted for genome sequencing; cultures from which RNA was extracted for
387 YAMAT-seq). We note here that loss of the duplication fragment by some population members only

388 dilutes its effect(s), and hence any significant differences reported between a duplication-carrying
389 strain and ΔEGEG will hold regardless of any segregants that may have been present in the
390 duplication genotype cultures. For example, the proportional increase in tRNA-Gly(GCC) in the
391 mature tRNA pools of cells carrying the duplication fragments (see Figure 5) would be expected to
392 increase in the absence of segregation effects.
393 The results in this section demonstrate that the adaptive duplication events are inherently
394 unstable; they, along with their compensatory tRNA gene copies, are lost at high rates. In growing
395 populations founded by duplication-carrying strains, this high rate of loss leads to the presence of at
396 least two genetically distinct sub-populations of cells: (i) those carrying the initial duplication fragment,
397 and (ii) those in which the fragment has been lost (*i.e.*, ΔEGEG) (Figure 6D).

398 **Multiple duplication fragments rapidly arise and compete within evolving populations**
399 The ease with which duplications fragments arise and are subsequently lost suggests complex,
400 rapidly changing evolutionary dynamics within the evolving mutant lineages. In particular, the
401 repeated and rapid generation of ΔEGEG-like genotypes from duplication genotypes provides the
402 opportunity for distinct duplication fragments to arise independently and then compete. To investigate,
403 we quantified duplication fragments of interest across our evolving populations. Duplication junction
404 PCRs – each of which has the potential to capture a range of similar duplication fragments with subtly
405 different endpoints – were performed on population samples taken at 48-hour intervals across each
406 28-day lineage, and the intensity of the resulting bands were measured relative to a positive control
407 (see Supplementary Table S6).
408 Initially, we followed the dynamics of four duplication fragments (M2f, M3f, M4f, and M5f) in
409 the mutant lineages in which they were initially identified (M2, M3, M4, and M5, respectively) (Figure
410 7A; Supplementary Figure S2). Each of the three large duplication fragments (M2f, M4f, and M5f; ~1
411 Mb) was first detected early in the course of its respective lineage (between days 4 and 6). The
412 fragments then rose in frequency before beginning to wane at around day 16, with two of the three
413 (M4f, M5f) dropping below the limit of detection before the end of the experiment. In comparison, the
414 smaller duplication fragment (M3f; ~0.49 Mb) was first detected slightly later (day 8) and, following an
415 initial sweep, was maintained at relatively high levels.
416 While the above results demonstrate that duplication fragments readily arise and spread, it is
417 notable that none of the observed duplication fragments appears to have fixed at any point in its
418 respective lineage; the highest estimated frequency was ~82 % (M2f in lineage M2, on days 14 and
419 18; see Supplementary Table S6). In addition to reflecting the inherent instability of the large-scale
420 duplication fragments (see Figure 6), this lack of fixation suggests that multiple duplication fragments
421 may arise and compete within each lineage. To investigate, each lineage was tested for the co-
422 existence of large and small duplication fragments; duplication junction PCRs to detect M5f (large
423 fragment) and M3f (smaller fragment) were performed on population samples from days 8, 14, and 28
424 of all five mutant lineages (M1-M5) and a wildtype control lineage (W1) (Figure 7B; Supplementary
425 Figure S3).
426 These experiments provide three lines of evidence for complex dynamics between large-scale
427 duplication fragments in the evolving populations: (i) both the large and smaller duplication fragments

428 are detected simultaneously in the earlier stages of adaptation (*i.e.*, days 8 and 14) in all mutant
429 lineages, (ii) multiple subtypes of the smaller duplication fragments – with subtly different endpoints,
430 and hence variably sized duplication junction PCR products (see Supplementary Figure S3B) – exist
431 within all mutant lineages, and (iii) mutations other than the identified large and smaller duplication
432 fragments are likely to exist in the adapted mutant populations (because the sum of the estimated
433 frequencies of the large and small fragments is still far below 1 in the majority of day 14 and day 28
434 populations). Importantly, there is also evidence that the large-scale, tandem duplication fragments
435 can exist, but not sweep, in the absence of selection; a small duplication fragment was detected at
436 low levels (~1 %) on day 8 and day 14 of the wildtype control lineage W1 (Figure 7B; Supplementary
437 Figure S3B).

438 **DISCUSSION**

439 Various studies have reported adaptive changes in tRNA gene copy number in laboratory populations
440 of bacteria, in response to changing translational requirements (6, 40, 41). In these studies, changes
441 in gene copy number were underpinned by large-scale, tandem duplication (6, 41) or amplification (40)
442 events. Large-scale, tandem duplications (and/or amplifications) are a class of mutation that is both
443 pervasive and mechanistically reversible: large-scale duplications typically arise, and are lost,
444 extremely rapidly (43, 44, 75, 77, 78), in a manner reminiscent of contingency loci (79, 80). Their
445 transient nature means that their contributions to population diversity, and downstream effects on
446 evolution, can be underappreciated (44). In this work, we demonstrate the contribution of large-scale
447 duplications to the initial stages of bacterial adaptation to novel translational requirements,
448 highlighting that tRNA gene copy number is more fluid than previously appreciated (particularly in
449 large populations of rapidly growing bacteria).

450 **Adaptive changes in bacterial tRNA gene copy number and their effect on mature tRNA pools**
451 In the first part of our work, we provide an example of tRNA gene content of a bacterium adapting by
452 large-scale duplication events. Specifically, deletion of *P. fluorescens* SBW25 tRNA gene locus
453 *gluTTC-glyGCC-gluTTC-glyGCC* (*EGEG*) leads to a drop in the proportional contribution of
454 tRNA-Gly(GCC) – a high-use, essential tRNA isotype – to the mature tRNA pool. The accompanying
455 translational defect is repeatedly compensated by readily-occurring, large-scale, tandem duplications
456 that encompass up to 1 Mb (~15 %) of the wildtype chromosome. These duplication events affect the
457 copy number of hundreds of genes, including those encoding several tRNA isotypes. Despite their
458 large size, the proposed compensatory element of each duplication fragment is predicted to be only a
459 single, 76-bp *glyGCC* gene (and promoter). Indeed, the duplication fragments lead to partial
460 restoration of tRNA-Gly(GCC) levels and, in the case of the smaller (~0.49 Mb) duplication fragment,
461 growth patterns that approximate those of the wildtype. Presumably, the higher degree of
462 compensation observed in M3-1 is directly related to the smaller size of the M3-1 duplication fragment;
463 once generated, smaller duplication fragments surrounding *glyGCC* are logically expected to provide
464 greater net fitness (as a result of carrying less surplus DNA, altering the dosages of fewer additional
465 genes, and/or incurring smaller effects on chromosomal architecture).

466 Compensation of translational defects by large-scale, tandem duplications is not limited to the
467 SBW25 *glyGCC* gene. We previously reported similar large-scale duplications (up to ~290 kb) in a
468 distinct region of the SBW25 chromosome (depicted in Figure 4B) (6). In this earlier case, the
469 duplication fragments served to increase the gene copy number of *serTGA* – and hence
470 tRNA-Ser(UGA) levels – in response to elimination of tRNA-Ser(CGA), a non-essential but relatively
471 high-use tRNA isotype. Furthermore, similar large-scale duplications that change tRNA gene copy
472 number have been observed throughout the *E. coli* chromosome, and postulated to compensate for
473 changes in codon use incurred by horizontal gene transfer events (41). These experiments
474 demonstrate that large-scale duplication events can underpin adaptive gene copy number changes
475 for various tRNA isotypes, in different areas of the genome, across species. Taken together with their
476 pervasive nature and high rate of occurrence, such mutations may play a widespread role in bacterial
477 adaption to novel translational requirements.

478 **Rapid, spontaneous duplication events generate diversity in the population tRNA gene set**
479 As outlined in the introduction, large-scale duplications typically arise spontaneously (during DNA
480 replication), at high rates (43, 46, 54–57). For instance, it has been reported that – without invoking
481 any growth advantage – around 10 % of cells in an overnight *Salmonella* culture carry a large
482 duplication of some region of the chromosome, with between 0.005 % and 3 % of cells carrying a
483 duplication of a given locus (42, 43, 55). Notably, our experiment also captures evidence of a large-
484 scale duplication arising in the absence of positive selection: a duplication fragment was detected at
485 low frequency in evolving wildtype population W1, on days 8 and 14 (see Figure 7B). This duplication
486 fragment likely provided no (or negligible) selective advantage in the wildtype background, and did not
487 spread through the population (instead dropping below the limit of detection by day 28). Similar
488 duplication fragments were first detected in the ΔEGEG lineages between days 4 and 6, where they
489 provided a fitness advantage and hence rose in frequency.
490 In addition to revealing that duplication fragments occurred in the early stages of the evolution
491 experiment, the population-level duplication junction PCRs demonstrate the co-existence of multiple
492 distinct types of duplication fragment in all five mutant lineages (see Figure 7B). More specifically, at
493 least two distinct types of duplication fragments – with different tRNA gene contents (see Figure 4D) –
494 were present in each mutant lineage on days 8 and 14. These data are consistent with the hypothesis
495 that large-scale duplication fragments arise spontaneously, at high rates, thus generating population-
496 level diversity in chromosomal content and tRNA gene copy numbers.
497 While multiple distinct duplication fragments were identified during this work, it is notable that
498 three of the five compensated isolates carry a very similar ~1 Mb duplication fragment with endpoints
499 *pflu1686* and *pflu2597* (see Table 1). This fragment does not confer a particularly high degree of
500 compensation (see Figure 3B), suggesting that its prevalence is instead due to an especially high rate
501 of occurrence (81). Consistent with this hypothesis, the 1 Mb fragment endpoint often occur within two
502 *rhs* genes that share a long stretch (~3 kb) of near-perfect homology, and hence provide ample
503 opportunity for duplications to form by RecA-mediated unequal recombination events during DNA
504 replication (43, 76). While duplications can form by mechanisms that do not require homology or
505 RecA, these typically occur at much lower frequencies (45, 76, 78, 82). Similar, lower-rate

506 mechanisms are presumed to have underpinned the emergence of the less prevalent, yet higher
507 effect, duplication fragment in isolate M3-1 (0.49 Mb in size, with no obvious endpoint homology).
508
509 **Unstable duplications are an evolutionary stepping stone**
510 Large-scale, tandem duplication fragments are inherently unstable; they are prone to loss through
511 RecA-mediated recombination during DNA replication (55, 76). The duplication fragments described
512 in this work are similarly unstable – the stability test showed that, despite the net fitness advantage
513 that they provide, duplication fragments are lost by between 3 and 34 % of cells in overnight culture
514 (see Figure 6). The rate of loss was higher for strains carrying the larger (~1 Mb) versus the smaller
515 (~0.49 Mb) duplication fragments (median loss rates of 19 % and 5 %, respectively). This was
516 presumably a result of the larger duplication fragment (i) carrying twice the length of homologous DNA
517 for recombination (*i.e.*, the duplication fragment), and (ii) conferring a lesser fitness advantage (44).
518 Our large-scale duplications are inherently unstable and are therefore unlikely to represent a
519 long-term evolutionary solution, despite providing a fitness advantage. That is, if the serial transfer
520 evolution experiment were to be continued, it is expected that the characterized duplication fragments
521 would eventually be displaced by a more stable type of mutation. Evidence to this effect is already
522 visible in the duplication junction PCRs across the five mutant lineages (see Figure 7). As described
523 above, all five mutant lineages contain both the larger (1 Mb) and smaller (0.49 Mb) class of
524 duplication fragment on days 8 and 14 of the evolution experiment. However, by day 28 the larger
525 (and also less fit/stable) duplication fragment drops below the limit of detection in three (of five)
526 mutant lineages (M3, M4, and M5). Furthermore, in two of the five lineages (M3 and M5), the
527 frequency of both the larger and smaller duplication fragments shows a large reduction at some point
528 between days 14 and 28, strongly suggesting the presence of at least one other adaptive mutation in
529 the later stages of these lineages (see Figure 7B). The observed duplication junction PCR patterns
530 support the hypothesis that such alternative mutations occur at lower rates, and hence take longer to
531 arise (on average), than the characterized duplication fragments. However, once present, these
532 alternative mutations presumably afford greater fitness, leading to displacement of the characterized
533 duplication fragments in the population.
534 The alternative mutation(s) described above currently remain unidentified. However, they
535 could conceivably be SNPs (*e.g.*, in the promoter of a compensatory tRNA gene, or a post-
536 transcriptional tRNA modification gene), and/or progressively smaller (and hence, more stable and
537 more advantageous), duplication fragments. Such smaller duplication fragments could potentially
538 arise directly from ΔEGEG (through lower-rate recombination events between chromosomal regions
539 with low or no homology (76)), or from duplication-carrying strains (through the partial segregation of
540 a duplication fragment). Indeed, an example of the latter was recently reported in an experiment
541 examining the fate of an advantageous, large-scale (~1.66 Mb), tandem duplication fragment in
542 *Salmonella*; during 2,000 generations of positive selection, the initial duplication fragment repeatedly
543 shrank to ~0.66 Mb, via the loss of an internal ~1 Mb segment (83, 84).
544 Both types of alternative mutation listed above (SNPs and smaller duplication fragments)
545 represent distinct mechanisms of adaptation to changing translational requirements: SNPs are

expected to affect the mature tRNA pool without changing the underlying tRNA gene content of the organism, while adaptive duplication fragments contain tRNA genes and hence change tRNA gene set composition. Presumably, the smallest possible adaptive duplication fragment is approximately ~150-200 bp (encompassing merely the compensatory tRNA gene and promoter), and would endow a new, comparatively stable tRNA gene set on the adapted strain. The long(er) term evolutionary fate of the duplication fragments, and hence the new tRNA gene copies, remains to be seen, and this information 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 number PRJNA790786.

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 and are accessible at: https://zenodo.org/record/7276661#.Y3JJHYKZONh

SUPPLEMENTARY DATA

Supplementary Data are available at NAR online.

ACKNOWLEDGEMENT

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

FUNDING

This work was supported by the German Research Foundation (DFG) [AOBJ:679030 to JG], the Max Planck Society [all authors], and the International Max Planck Research School for Evolutionary Biology [ZK and WYN]. Funding for open access charge: The Max Planck Society.

AUTHOR CONTRIBUTIONS

Conceived and designed the experiments: GBA, JG, 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.

579 **CONFLICT OF INTEREST**

580 The authors declare no conflict of interest exists.

581

582 **REFERENCES**

583 1. Gingold,H. and Pilpel,Y. (2011) Determinants of translation efficiency and accuracy. *Mol Syst Biol*,
584 **7**, 481.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

619 17. Shepherd,J. and Ibba,M. (2015) Bacterial transfer RNAs. *FEMS Microbiology Reviews*, **39**, 280–
620 300.

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

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

626 20. Varenne,S., Buc,J., Lloubes,R. and Lazdunski,C. (1984) Translation is a non-uniform process.
627 Effect of tRNA availability on the rate of elongation of nascent polypeptide chains. **180**, 549–
628 576.

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

685 46. Tomanek,I., Grah,R., Lagator,M., Andersson,A.M.C., Bollback,J.P., Tkačik,G. and Guet,C.C.
686 (2020) Gene amplification as a form of population-level gene expression regulation. *Nat Ecol*
687 *Evol*, **4**, 612–625.

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

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

692 49. Lind,P.A., Tobin,C., Berg,O.G., Kurland,C.G. and Andersson,D.I. (2010) Compensatory gene
693 amplification restores fitness after inter-species gene replacements. *Mol Microbiol*, **75**, 1078–
694 1089.

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

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

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

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

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

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

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

713 57. Reams,A.B., Kofoid,E., Duleba,N. and Roth,J.R. (2014) Recombination and annealing pathways
714 compete for substrates in making *rrn* duplications in *Salmonella enterica*. *Genetics*, **196**, 119–
715 135.

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

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

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

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

724 62. Lenski,R.E., Rose,M.R., Simpson,S.C. and Tadler,S.C. (1991) Long-term experimental evolution
725 in *Escherichia coli*. I. Adaptation and divergence during 2000 generations. *Evolution*, **138**,
726 1315–1341.

727 63. Leinonen,R., Sugawara,H. and Shumway,M. (2011) The Sequence Read Archive. *Nucl Acids Res*,
728 **39**, D19–D21.

729 64. Deatherage,D.E. and Barrick,J.E. (2014) Identification of mutations in laboratory-evolved
730 microbes from next-generation sequencing data using *breseq*. *Methods Mol Biol*, **1151**, 165–
731 188.

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

735 66. Barrick,J.E., Colburn,G., Deatherage,D.E., Traverse,C.C., Strand,M.D., Borges,J.J.,
736 Knoester,D.B., Reba,A. and Meyer,A.G. (2014) Identifying structural variation in haploid
737 microbial genomes from short-read resequencing data using *breseq*. *BMC Genomics*, **15**,
738 1039.

739 67. Shigematsu,M., Honda,S., Loher,P., Telonis,A.G., Rigoutsos,I. and Kirino,Y. (2017) YAMAT-seq:
740 An efficient method for high-throughput sequencing of mature transfer RNAs. *Nucl Acids Res*,
741 **45**, e70.

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

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

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

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

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

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

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

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

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

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

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

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

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

769 81. Yampolsky,L.Y. and Stoltzfus,A. (2001) Bias in the introduction of variation as an orienting factor
770 in evolution. *Evol Dev*, **3**, 73–83.

771 82. Reams,A.B., Kofoid,E., Kugelberg,E. and Roth,J.R. (2012) Multiple pathways of duplication
772 formation with and without recombination (RecA) in *Salmonella enterica*. *Genetics*, **192**, 397–
773 415.

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

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

779

780

781 **TABLES AND FIGURES**

782

Strain	Dup. size (bp)	# Genes	#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)*

783 **Table 1. Details of duplication fragments identified in five compensated isolates.** Base positions
784 refer to the SBW25 wildtype genome sequence (NCBI accession number: NC_012660.1). For a
785 detailed list of the genes contained within each duplication fragment, see Supplementary Table S3.

786

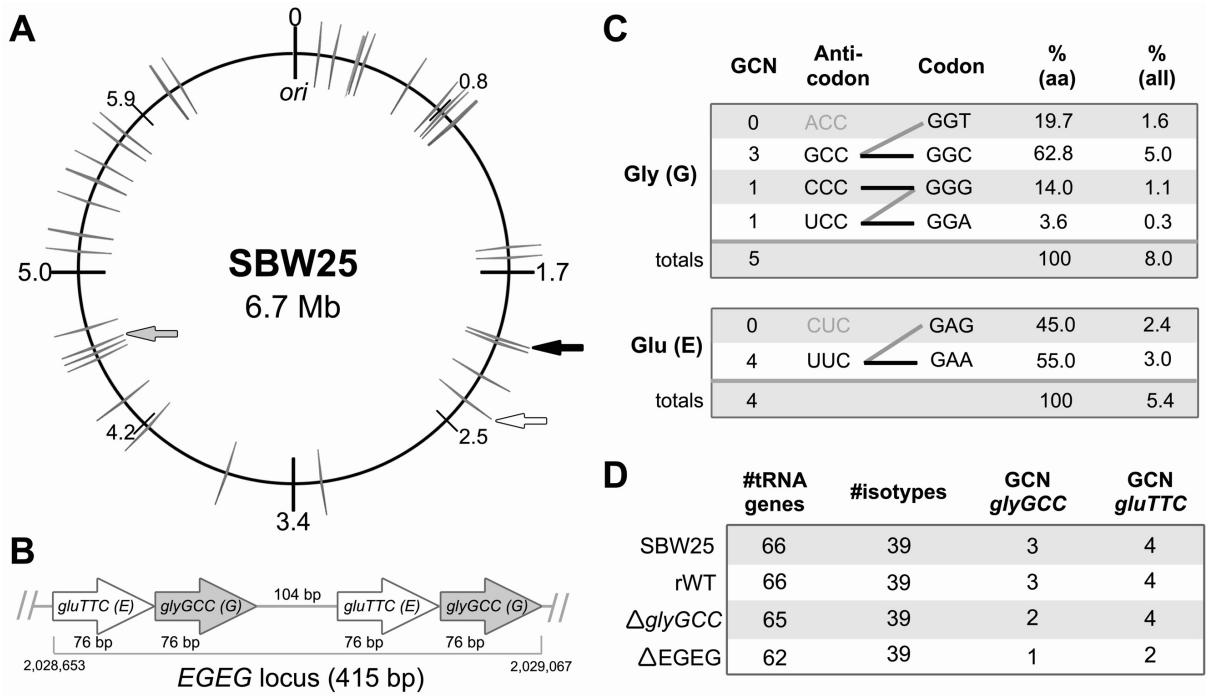
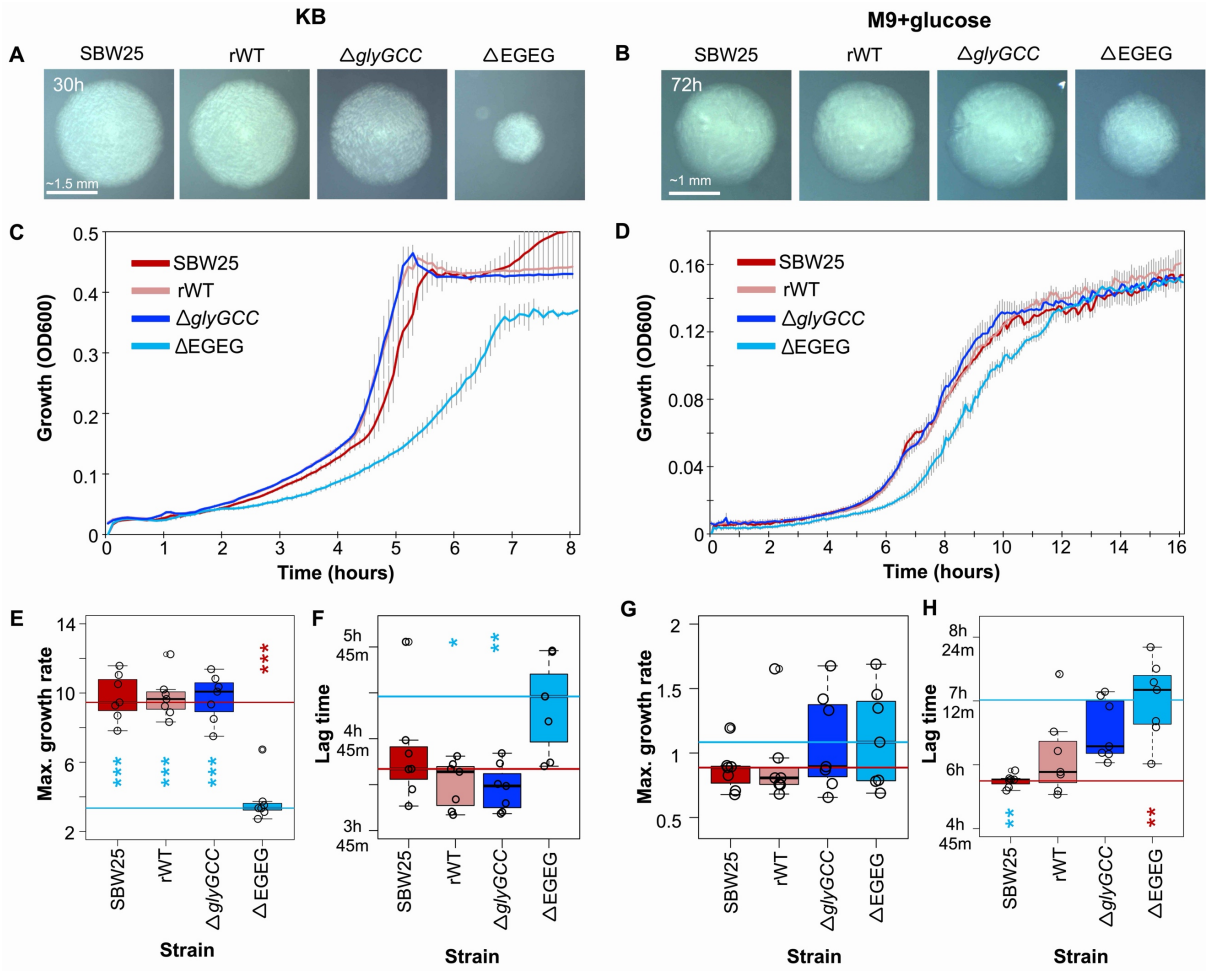


Figure 1. Translation of glycine (G) and glutamic acid (E) in *P. fluorescens* SBW25. (**A**) Cartoon of the SBW25 chromosome depicting loci predicted by GtRNAdb2.0 to encode canonical tRNAs. Grey lines=tRNA loci (each representing between one and four tRNA genes); black arrow=*gluTTC-glyGCC-gluTTC-glyGCC* ("*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) (6). Codons are listed with their cognate anticodons (black=isotype 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 (7). (**D**) Details of the tRNA gene sets carried by engineered genotypes.



802

803 **Figure 2. Deletion of the four-gene tRNA locus *EGEG* reduces SBW25 growth in rich,**
804 **undefined (KB) and poorer, defined (M9+glucose) media.** (**A**-**B**) Colonies of SBW25,
805 reconstructed wildtype (the engineering control, rWT), and the two engineered genotypes Δ*glyGCC*
806 and ΔEGEG at room temperature for 30 hours on KB agar, or 72 hours on M9 agar. Scale bar applies
807 to all four images in each panel. (**C**-**D**) Growth (measured as absorbance at 600 nm) of the same
808 strains in liquid KB for 8 hours, or M9 for 16 hours. Lines show the mean of 7 replicates at 5-minute
809 intervals; error bars are +/- 1 SE. (**E**-**H**) Maximum growth rates (change in absorbance; mOD min^{-1})
810 and lag times of growth profiles in panel C (KB; using a sliding window of 6 points between 1 h and 10
811 h), and panel D (M9; using a sliding window of six points between 1 h and 8 h). In all four panels, red
812 and blue lines are the medians of SBW25 and ΔEGEG, respectively. Non-parametric Dunn's tests,
813 followed by the Benjamini-Hochberg procedure to correct for multiple comparisons, were used to test
814 for a difference in medians between pairs of strains in each panel. Stars indicate statistically
815 significant *p*-values for a difference from SBW25 (red stars) or ΔEGEG (blue stars).

816

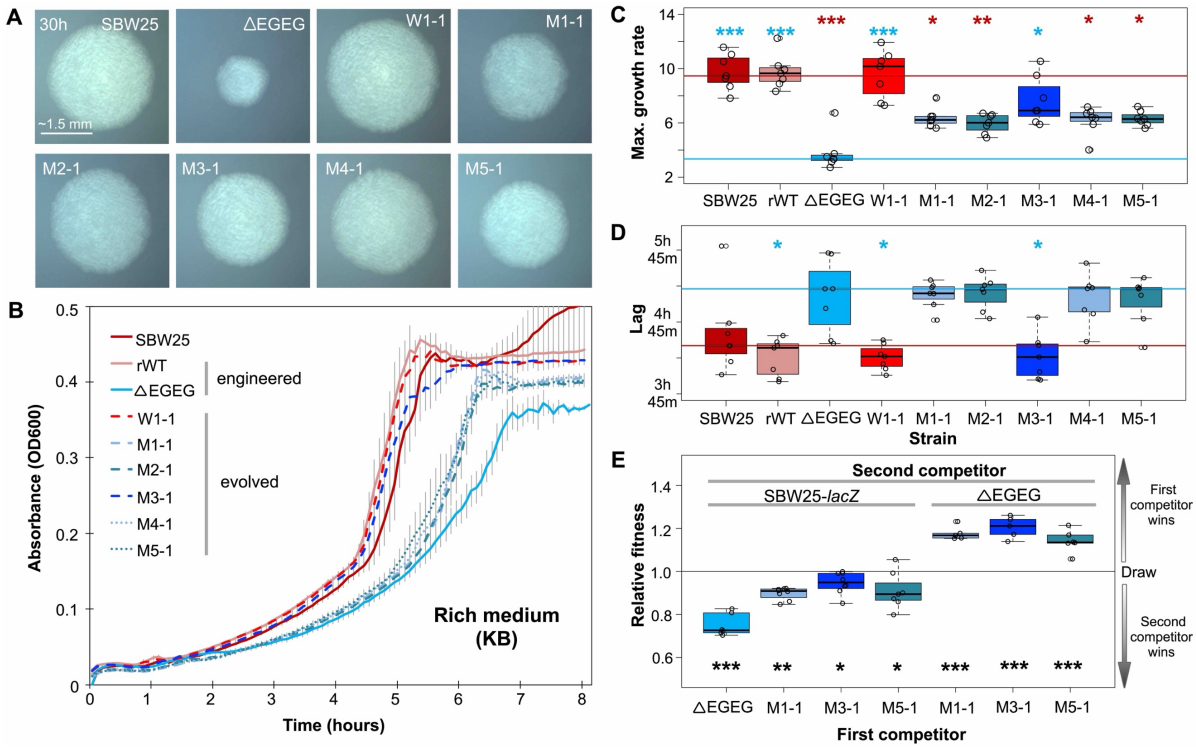


Figure 3. The growth defect caused by deleting four-gene tRNA locus *gluTTC-glyGCC-gluTTC-glyGCC* (*EGEG*) undergoes repeated compensation during a serial transfer experiment. (**A**) Colonies of founders SBW25, ΔEGEG, and 6 isolates from day 21 of the serial transfer evolution experiment, grown on KB 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, using a sliding window of 6 points during rapid growth (1 h - 8 h). 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 strains. Stars indicate statistically significant *p*-values for a difference from SBW25 (red stars) or ΔEGEG (blue stars). 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, 1:1 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 ***$p<0.001$, **$p<0.01$, *$p<0.05$.

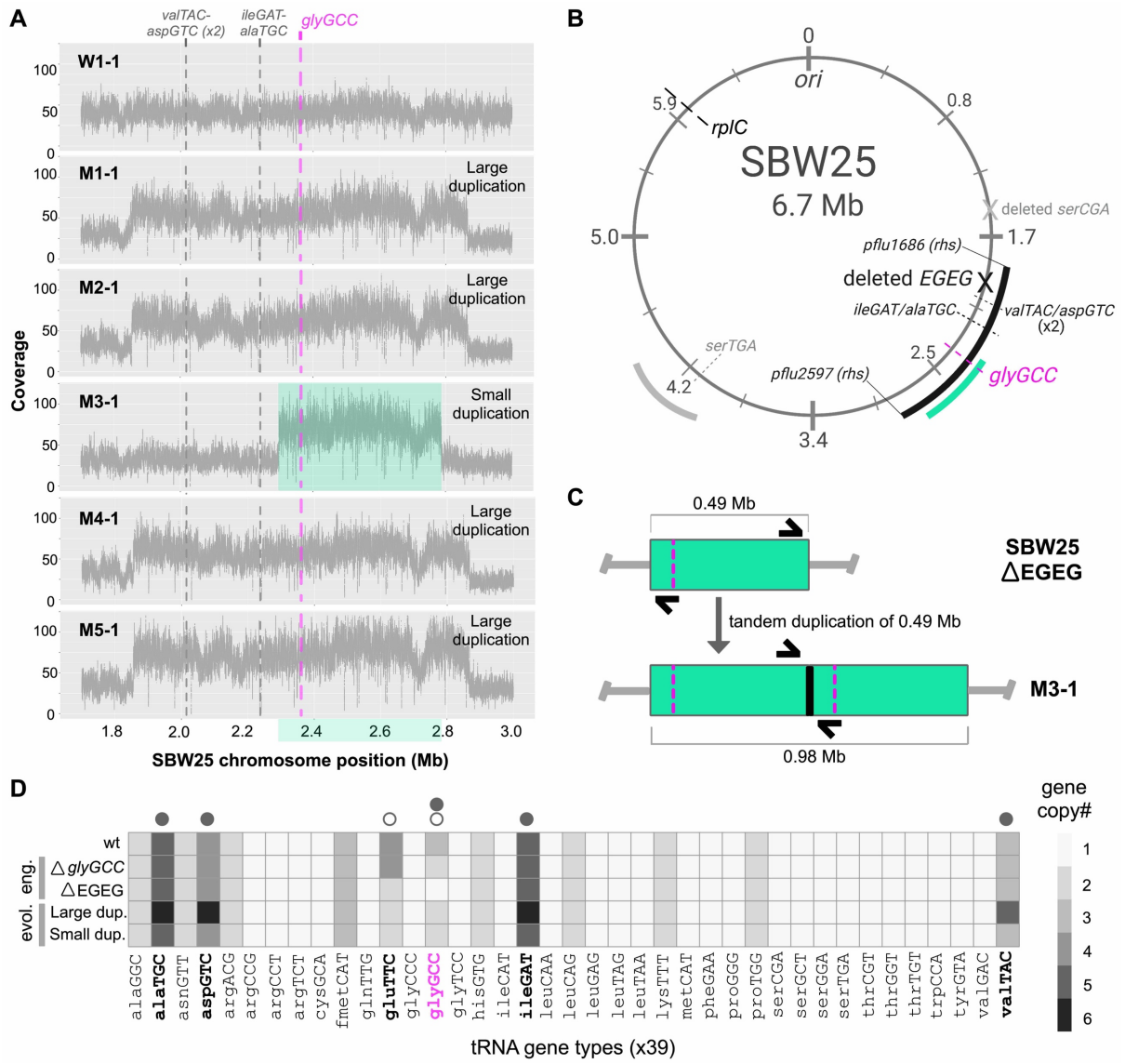


Figure 4. Large-scale, tandem duplications compensate for deletion of *EGEG*. (A) Number of raw sequencing reads covering 1.8 - 3.0 Mb of the reference genome. A jump to two-fold coverage is evident in mutant lineage isolates (M1-1 to M5-1), and not in the control isolate (W1-1). The ~0.49 Mb region 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 fragment (in M1-1, M2-1, M4-1, M5-1); green arc=~0.49 Mb duplication fragment (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 isotypes between wildtype (wt), engineered (eng.), and evolved (evol.) genotypes. Open/closed circles=isotypes whose gene copy number changed by engineering/evolution.

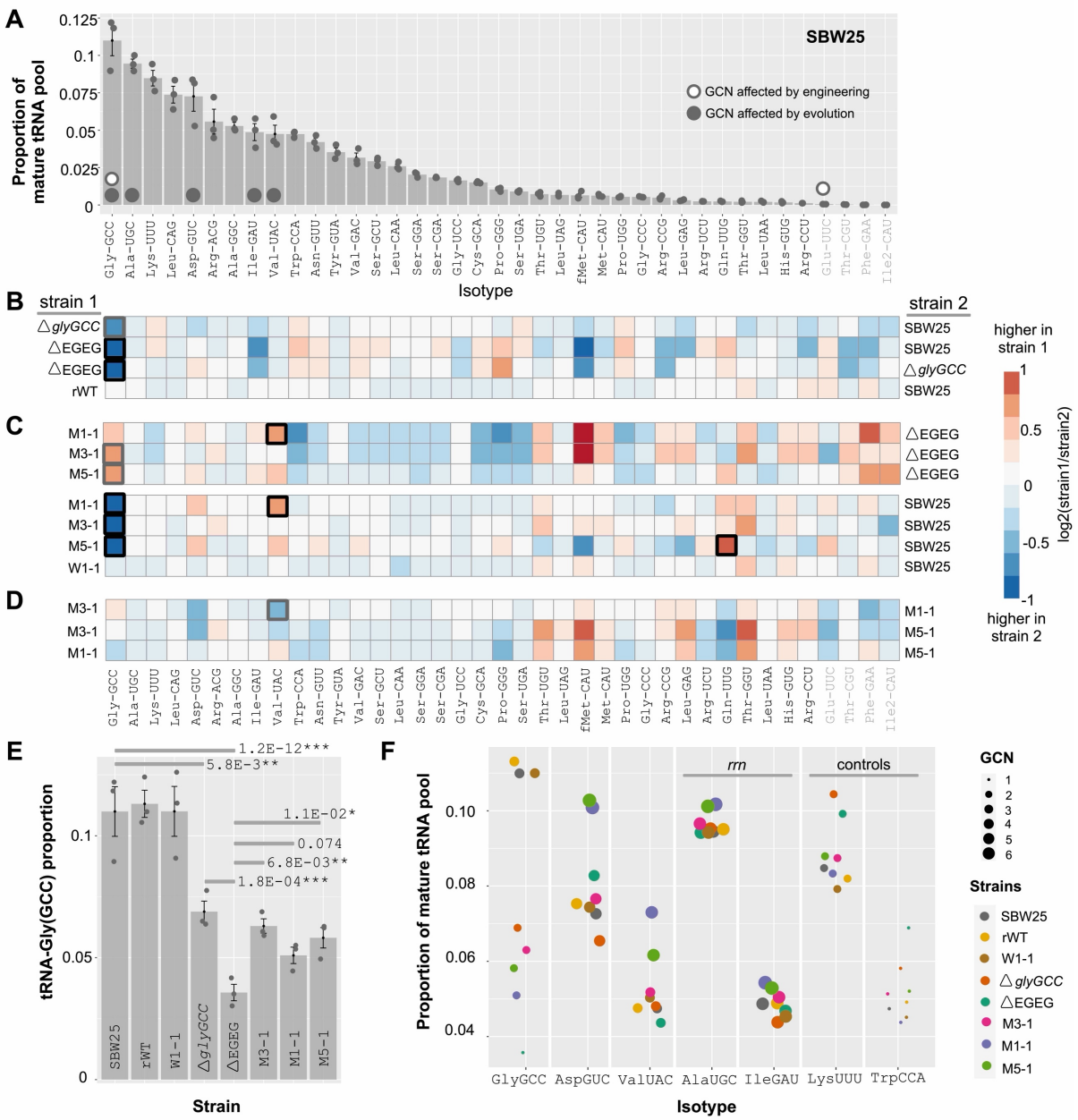


Figure 5. *glyGCC* gene copy number changes are mirrored in the mature tRNA pool. (**A**) Proportions of 39 isotypes in the SBW25 mature tRNA pool (highest to lowest). Light grey=excluded isotypes. Open/closed circles=isotypes whose gene copy number (GCN) changes by engineering/evolution in downstream experiments. (**B-D**) Heatmaps of differences in isotype proportions (expressed as log2-fold change(strain1/strain2); see Supplementary Table S5) between genotype pairs, to detect the effect of engineering (panel A) 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) proportions 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 isotypes, by genotype. GCN for first 5 isotypes varies by engineering and/or evolution (indicated by point size), while the final 2 isotypes are controls (constant GCN). *rrn*=isotypes whose genes co-localize with rRNA operons.

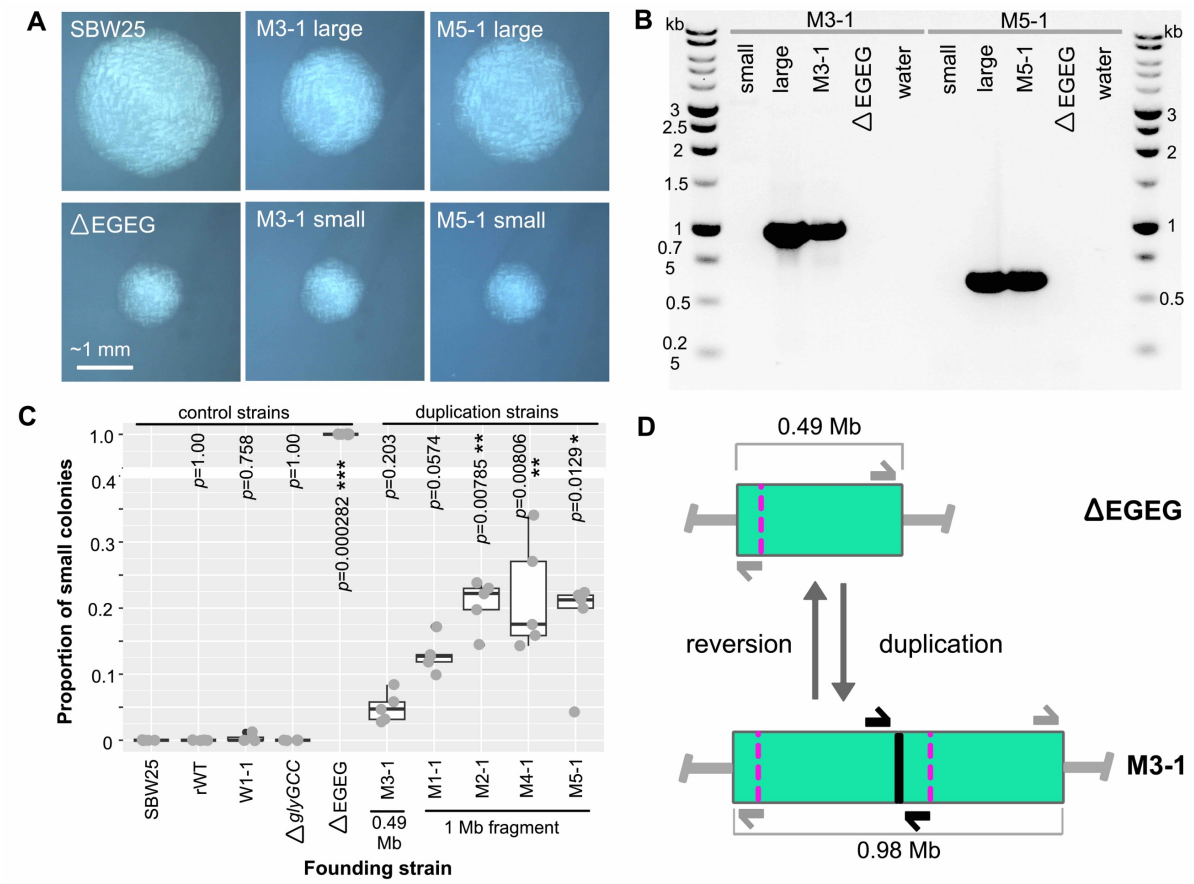


Figure 6. The large-scale, tandem duplication fragments are rapidly lost in overnight culture. (**A**) Two colony morphologies – large and small – are observed when plating from overnight cultures of duplication fragment-carrying strains (colonies from duplication genotypes M3-1 and M5-1 presented here). 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. (**C**) **D**) Cartoon depicting the gain and loss of duplication fragments leading to a mixed population of duplication strain (here, M3-1) and ΔEGEG.

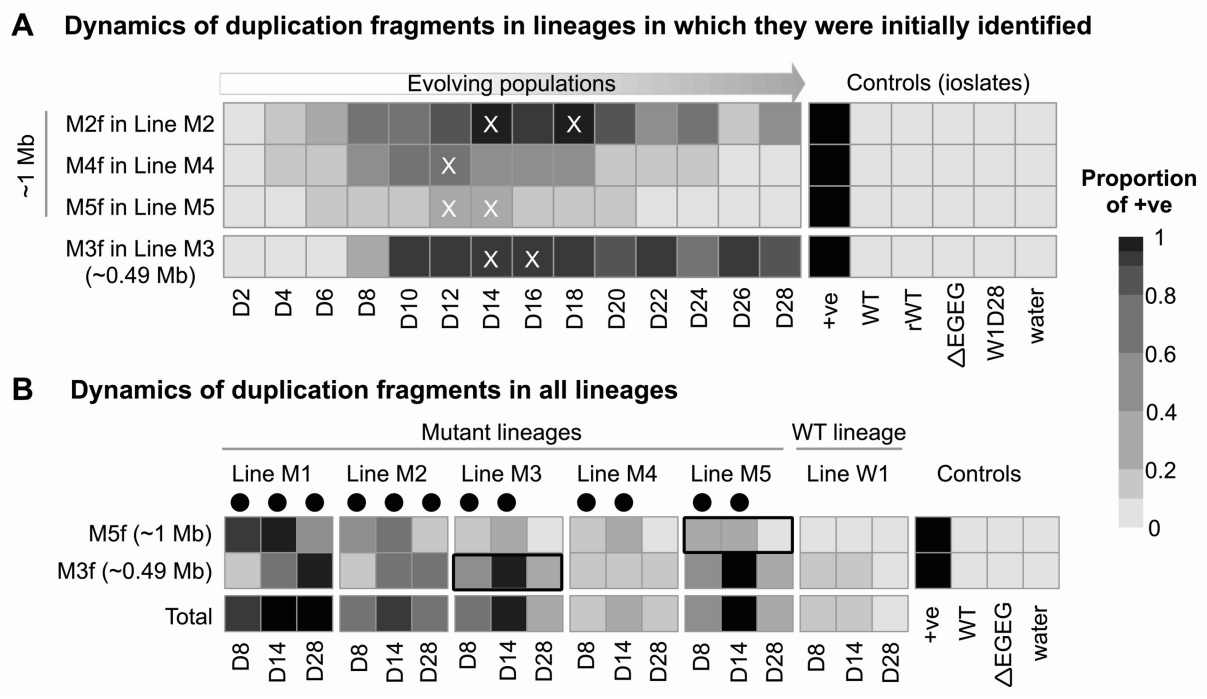


Figure 7. The complex evolutionary dynamics of the large-scale duplication fragments in evolving populations. Heatmaps showing the estimated proportion of individuals carrying large-scale duplication fragments of interest in evolving populations. Proportions were estimated from the intensity of duplication junction PCR bands on an agarose gel, relative to a positive control (using ImageJ; intensity measures are provided in Supplementary Table S6). Darker squares indicate higher proportions of individuals carrying the duplication fragment. (**A**) The dynamics of three large duplication fragments (M2f, M4f, and M5f; all ~1 Mb) and the smaller duplication fragment (M3f; ~0.49 Mb) across the mutant lineages in which they were initially identified (see Table 1). White crosses mark the peak frequency of each fragment. Gel images are provided in Supplementary Figure S2. (**B**) The dynamics of large (M5f; ~1 Mb) and smaller (M3f; ~0.49 Mb) duplication fragments in all evolving populations. 'Total' row provides the sum of M5f and M3f; thick black rectangles indicate the lineage in which the duplication fragment was initially identified; black circles mark populations in which both large and smaller duplication fragments were detected. Gel images are provided in Supplementary Figure S3.