FULL TITLE: The transcriptomic and proteomic ramifications of segmental amplification

SHORT TITLE: Consequences of segmental amplification

Authors: Ryan K. Fritts[1], Christopher C. Ebmeier[2], Shelley D. Copley[1]

[1]Department of Molecular, Cellular, and Developmental Biology and Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder, Boulder, CO, United States

[2]Department of Biochemistry, University of Colorado Boulder, Boulder, CO, United States

*For correspondence: Shelley.Copley@colorado.edu

Abstract

Gene duplication and subsequent amplification can drive adaptation by rapidly increasing the cellular dosage of critical gene products. Gene duplication events often encompass large genomic segments surrounding the gene(s) under selection for higher dosage. Overexpression of co-amplified genes imposes a significant metabolic burden. While compensatory mutations can decrease inappropriate overexpression of co-amplified genes, it takes time for such mutations to arise. The extent to which intrinsic regulatory mechanisms can modulate the expression of co-amplified genes in the immediate aftermath of segmental amplification is largely unknown. To address the collateral effects of segmental amplification, we evolved replicate cultures of an *Escherichia coli* mutant under conditions that select for higher dosage of an inefficient enzyme whose weak activity limits growth rate. Segmental amplifications encompassing the gene encoding the weak-link enzyme were detected in all populations. These amplifications ranged in size, from 9 to 125 kb, and copy number, from 2 to 12 copies. We performed RNA-seq and label-free quantitative proteomics to quantify the expression of amplified genes present at 2, 6, and 12 copies. mRNA expression generally scales with gene copy number, but protein expression scales poorly with both gene copy number and mRNA expression. We propose mechanisms to explain these discrepancies for several cases in which expression is either higher or lower than expected based on gene copy number. We also show that segmental amplifications can have system-wide consequences by indirectly altering expression of non-amplified genes. Overall, our findings indicate that the fitness benefit derived from segmental

36 amplification depends on the combined effects of amplicon size, copy number, and

37 gene content.

38

Introduction

Organisms frequently encounter new and/or changing environments. Successful adaptation to a new niche may require an increase in the expression of particularly important gene products. Gene amplification allows organisms to rapidly increase the cellular dosage of newly needed proteins. Gene amplification has been shown to be adaptive under various contexts for bacteria [1–9] and eukaryotes [10–16] and is implicated in the development of cancer cells [17–19].

Beyond serving as a mechanism for rapidly increasing the dosage of a gene product, gene amplification can also be the starting point for evolution of new molecular functions. The Innovation-Amplification-Divergence (IAD) model describes an evolutionary scenario by which a gene encoding a new function can arise [2,20] (Fig. 1A). Most, if not all, proteins have weak side activities, termed promiscuous activities, in addition to their native activities. If a promiscuous activity becomes important for fitness (innovation), natural selection for increased expression often leads to gene amplification. Following amplification, mutations in a new gene copy can improve the weak newly advantageous activity while the ancestral function is maintained in another gene copy (divergence). This process has been responsible for the evolution of vast superfamilies of enzymes, transporters, transcriptional regulators, and signaling proteins [21–24].

Gene amplification does not neatly increase the copy number of a single gene under selection for higher expression. Rather, segmental amplifications often amplify large genomic regions containing dozens to hundreds of genes surrounding the gene(s) under selection for higher dosage. Segmental amplifications in bacteria and yeast often

62 reach above 10 copies and can span over 100 kb during adaptation to stresses such as
63 starvation or antibiotic exposure [4–9,13,25,26]. Despite the importance of segmental
64 amplification, little attention has been given to the consequences of co-amplifying
65 neighboring genes (Fig 1B). Co-amplification of neighboring genes would be expected
66 to inappropriately increase levels of the encoded proteins and cause physiological
67 perturbations and/or energetic burdens. These collateral effects could have important
68 effects on population dynamics in communities of organisms that are competing to
69 adapt in the face of an environmental challenge or opportunity.
70
71 **Fig 1.** The expression and function of co-amplified neighboring genes can affect the
72 fitness benefit derived from segmental amplification**.** (A) The Innovation-Amplification-
73 Divergence (IAD) model. A, native function; b, promiscuous function. (B) The
74 expression of all amplified genes determines the fitness benefit gained in the immediate
75 aftermath of segmental amplification.
76
77 Previous investigations of the levels of mRNAs and proteins produced from co-
78 amplified genes in aneuploid yeast strains with one or two extra copies of whole
79 chromosomes [27–31] and in cancer cells [32,33] have shown that mRNA and protein
80 levels often, but not always, scale with gene copy number. Alterations in expression of
81 co-amplified genes in these settings may be due to either intrinsic regulatory
82 mechanisms that counteract inappropriate gene expression or to compensatory
83 mutations that arose over many generations. A study of the expression of individually
84 duplicated yeast genes found that protein levels were attenuated in some cases,

85 although the precise molecular mechanisms that dampened protein expression were
86 not clear [34]. The expression of high-copy-number co-amplified genes in the immediate
87 aftermath of segmental amplification has not been addressed. Consequently, we lack
88 an understanding of the intrinsic capacity of regulatory mechanisms to buffer against the
89 collateral effects of segmental amplification.

90 We have developed a model system in *Escherichia coli* to characterize the
91 immediate ramifications of segmental amplification on the transcriptome and proteome.
92 The *E. coli* enzymes ArgC and ProA are nonhomologous dehydrogenases that catalyze
93 similar reactions in arginine and proline biosynthesis, respectively (Fig 2). We previously
94 deleted *argC* from the *E. coli* genome and selected for recruitment of another enzyme to
95 replace the missing ArgC activity during growth on minimal medium [35]. A strain with a
96 point mutation in *proA*, causing substitution of Glu383 to Ala near the active site of
97 ProA, was able to restore arginine production. E383A ProA (ProA*) catalyzes both its
98 compromised native reaction in proline synthesis and the weak promiscuous activity
99 required for arginine synthesis (Fig 2). The poor activity of ProA* compromises growth
100 in minimal media in which synthesis of proline and arginine is required for growth,
101 leading to strong selective pressure for amplification of *proA*.* We previously observed
102 segmental amplifications encompassing *proA** ranging in size from 5-165 kb and
103 reaching copy numbers between 3-20 copies in experimentally evolved *E. coli*
104 populations [4,36,37].
105

106 **Fig 2.** Reactions catalyzed by the dehydrogenases ArgC and ProA in *E. coli*. The acetyl
107 group of *N*-acetylglutamylphosphate (red) is the only difference between the two
108 substrates.
109
110 Many of the genes neighboring the *proBA** operon are pseudogenes or
111 uncharacterized cryptic prophage genes. Thus, amplification of *proBA** in its native
112 locus does not provide an ideal system for probing the effects of co-amplification of
113 genes near the gene under selection for higher copy number. For this work, we
114 relocated the *proBA** operon to a new site neighboring a gene that we hypothesized
115 might cause physiological perturbations if overexpressed following segmental
116 amplification. We evolved replicate populations of this *E. coli* strain for 145-200
117 generations in a turbidostat. Whole genome sequencing revealed that all populations
118 contained segmental amplifications of differing sizes and copy numbers. We examined
119 the transcriptomes and proteomes of the two fastest-growing populations to determine
120 the collateral effects of segmental amplification. The expression of many amplified
121 genes scales with gene copy number at the mRNA and protein levels. However, in
122 several cases, mRNA and protein levels either exceed or fall short of the level expected
123 based solely upon gene copy number. We discuss possible mechanisms for several
124 cases in which such discrepancies exist. We also found significant changes in levels of
125 hundreds of non-amplified genes, demonstrating the profound impact of segmental
126 amplification on cellular physiology.
127
128 **Results and Discussion**

129 **Segmental amplification increases the growth rate of an *E. coli* strain with the**
130 ***proBA** operon relocated next to *rpoS***

131 We deleted the *proBA** operon from its native locus and relocated it next to *rpoS,*
132 which encodes the alternative sigma factor RpoS/σS (also known as σ38) (Fig 3A). We
133 included the upstream region containing a C to T promoter mutation 45 bp upstream of
134 *proB* that increases transcription of *proBA** [36]. The *proBA** operon was oriented
135 convergently to *rpoS* so that transcriptional read-through from *proBA** would not
136 increase *rpoS* expression. A dual *rrnB* T1/T7 TE transcriptional terminator was placed
137 immediately after *proA**. As an alternative sigma factor, RpoS serves as a global
138 regulator by interacting with RNA polymerase to activate transcription of numerous
139 genes important for redirecting cellular resources from growth toward survival [38,39].
140 Consequently, expression of RpoS is typically associated with a slowing of growth
141 during stressful conditions or limited nutrient availability. Elevated expression of genes
142 that make up the RpoS regulon confers improved resistance to starvation, high
143 temperatures, low pH, oxidative stress and osmotic shock [38,39]. We hypothesized
144 that overexpression of RpoS following co-amplification with *proA** might perturb cellular
145 physiology because it would shift cells toward a transcriptional profile intended as a
146 stress response to halt growth while the cells are growing exponentially with adequate
147 nutrients.

148 We used a single founder colony of the Δ*argC proBA** relocation strain to start
149 eight replicate populations in a turbidostat. The cultures were grown in M9 minimal
150 medium with glucose as the sole carbon source to impose strong selection for
151 mutations that increase ProA* dosage. After 30-60 generations, we observed

152 improvements in growth rate for all populations (Fig 3B). Populations 3 and 8 showed
153 an additional slight increase in growth rate around generation 180. We ended the
154 experiment after 145-200 generations to limit the opportunity for compensatory
155 mutations that might alleviate harmful effects of segmental amplification to arise.
156 Similarly, we wanted to limit the chance for amplicon remodeling to remove
157 unnecessary copies of amplified genes [37].
158
159 **Fig 3.** Evolution of an *E. coli* strain with the *proBA** operon relocated next to *rpoS.* (A)
160 The *proBA** operon with a mutant constitutive promoter was deleted from its native
161 locus and relocated directly next to the *rpoS* gene. (B) Growth rate increased during
162 laboratory evolution of the *proBA*-rpoS* relocation strain in M9/glucose in a turbidostat.
163
164 We sequenced genomic DNA (gDNA) isolated from each population at end of the
165 evolution experiment. All populations had segmental amplifications encompassing
166 *proBA** and *rpoS.* The amplicons ranged in size from 9-125 kb and in copy number from
167 2-12 copies (Fig 4A and 4B). Copy numbers were confirmed by quantitative PCR
168 (qPCR) using primers targeting *proA**. Most of the segmental amplifications were
169 mediated by the transposable insertion sequence (IS) elements IS186 and IS1, leading
170 to similar amplification junctions across populations. IS186 transposition into *serV*
171 served as an upstream junction point in 7 of the 8 populations, while IS186 transposition
172 into the intergenic region between *ygcE* and *queE* served as a downstream junction in 6
173 of the 8 populations (Fig 4A and Table A in S1 Text). IS1 transposition was detected at
174 amplification junctions in populations 3 and 7 (Table A in S1 Text). As neither IS

175 element was present at these sites in the ancestral strain, it appears that these sites
176 may be hotspots for IS element insertion.

177 We detected two amplicons at frequencies ≥5% in 6 of the 8 populations based
178 on the read coverage data (Fig 4A). The presence of two amplicons could be due to two
179 clades with distinct amplicons in the population or to an amplicon nested inside a larger
180 amplicon in a single clade. In populations 1 and 7, we suspect that low-frequency
181 clades with ~88 kb amplicons are being outcompeted by clades with smaller amplicons
182 (~56 and ~33 kb). In populations 4 and 8, low-frequency clades with ~9 and ~67 kb
183 amplicons, respectively, might have arisen more recently than the clades with larger
184 amplicons (~88 and 125 kb) and thus might still represent a small fraction of the
185 population. Alternatively, the ~9 and ~67 kb amplicons might have arisen by remodeling
186 of the larger amplicons in clades that are in the process of overtaking the original
187 clades.

188

189 **Fig 4.** Segmental amplifications occurred during evolution of all *proBA*-rpoS* relocation
190 populations. (A) Sizes and copy numbers of amplicons detected in the eight
191 populations. In populations with two amplicons, dark gray bars represent the higher-
192 frequency amplicon and light gray bars represent the lower-frequency amplicon. Only
193 amplicons estimated to be present at population frequencies of ≥5% are shown. (B)
194 Growth rate, *proA** copy number, and amplification size for each population at the end
195 of the evolution experiment. Only the most abundant amplicon is shown for each
196 population for simplicity. (C) The nested amplicons in population 3. (D) The 125 kb
197 amplicon in population 8.

198

199 For populations 2 and 3, which contain two amplicons with frequencies at or
200 above 20%, we sought to determine whether there were nested amplicons or competing
201 clades with different amplicons. To distinguish between these possibilities, we carried
202 out PCR reactions designed to amplify products that span putative amplification
203 junctions (S1 and S2 Figs). Population 2 had two distinct clades; almost all colonies (19
204 out of 20) had either the ~54 or the ~88 kb amplicon (S1 Fig). In contrast, the majority of
205 colonies isolated from population 3 (13 out of 20) showed PCR products for both the
206 ~40 and ~76 kb amplicons, indicating that most cells contained a smaller 12-copy
207 amplification nested inside a larger duplication (Fig 4C and S2 Fig). Of the remaining
208 isolates tested from population 3, four had the ~40 kb amplicon and three had the ~76
209 kb amplicon.

210 The growth rate of a cell with a particular segmental amplification is a complex
211 function of amplicon size, copy number and the nature of the amplified genes. While co-
212 amplification of most genes would be expected to impose a fitness cost unless
213 expression is attenuated by some regulatory mechanism, in rare cases co-amplification
214 of a gene might be beneficial. Despite having the largest amplicon (~125 kb) at the
215 lowest copy number (~6), population 8 has the second-fastest growth rate. It is possible
216 that one or more co-amplified genes in population 8 provide a fitness benefit when
217 overexpressed. One of the co-amplified genes unique to population 8 is *eno*, which
218 encodes enolase. In addition to its roles in glycolysis and gluconeogenesis, enolase is
219 part of the RNA degradosome [40,41]. An increase in the expression of enolase upon
220 co-amplification could provide a modest fitness benefit due to its roles in glycolysis

221 and/or RNA degradation. Another unique aspect of population 8 is the insertion of IS186
222 into *rlmM,* generating the downstream amplification junction (Table A in S1 Text). High-
223 throughput growth assays of an *E. coli* transposon mutant library indicate that
224 insertional disruption of *rlmM,* which encodes 23S rRNA 2'-*O*-ribose C2498
225 methyltransferase, provides a slight fitness benefit in M9 minimal medium with single
226 carbon sources [42].

227 Populations 4 and 6 contain very similar amplicons with *proA** at nearly the same
228 copy number, yet population 4 grows faster. Although population 4 has a slightly higher
229 *proA** copy number, other factors may also contribute to the difference in growth rate.
230 Population 4 has a 9 kb amplicon at a frequency of ~5% (Fig 4A). This was the smallest
231 amplicon we detected across all populations and it contains only five co-amplified
232 genes. The small fitness burden imposed by this small amplicon should allow this clade
233 to grow faster than the rest of the population, potentially contributing to the elevated
234 growth rate observed for population 4.

235 The slower growth rate of population 6 relative to that of population 4 may also
236 be due to a slow-growing clade in the population. Population 6 is the only population in
237 which we identified an additional amplification elsewhere in the genome. Population
238 genome sequencing data show that population 6 has a 6.5 kb amplification containing
239 *argE*, Δ*argC*::*kan*^{R} and *argB* present at ~3 copies (S3 Fig). Identical 8 bp sequences in
240 the flanking genes *ppc* and *argH* appear to have mediated this amplification. Because
241 ArgB (*N*-acetylglutamate kinase) produces *N*-acetylglutamylphosphate, a substrate for
242 ProA* in the Δ*argC proA** strain, amplification of *argB* might help push material through
243 the bottleneck in the arginine synthesis pathway caused by the inefficiency of ProA*. In

244 addition, amplification of *argE,* which encodes an enzyme downstream of the
245 bottleneck, might help pull material through the pathway. We speculate that the small
246 6.5 kb amplicon might have arisen first in the population and enhanced arginine
247 synthesis due to the increased copy number of *argB,* but that this amplification was less
248 beneficial than amplification of *proA*.* Later amplification of *proBA** might have occurred
249 in either the same clade or in a different clade. The presence of a slower-growing clade
250 that has only the small amplicon containing *argE* and *argB* may contribute to the
251 relatively low growth rate of population 6.

252 We focused on populations 3 and 8 for transcriptomic and proteomic analysis
253 because they are the fastest-growing populations and are two of the most disparate
254 populations in terms of amplification size and copy number (Fig 4C and D). Most cells in
255 population 3 appear to have a 39.6 kb amplification present at 12 copies nested inside a
256 75.7 kb duplication (Fig 4C), while population 8 primarily consists of a clade with a 6-
257 fold amplification of a 125 kb region, the largest amplicon we detected (Fig 4D). We
258 detected a 7 bp duplication in *prlF*, which encodes the antitoxin of the PrlF-YhaV toxin-
259 antitoxin system in population 8 at the end of the experiment. This mutation, which was
260 fixed in the population, has been reported to alleviate toxic overexpression of fusion
261 proteins by hyperactivating the Lon protease [43,44]. While the molecular mechanism
262 underlying this effect is not clear, this mutation might compensate for protein
263 overexpression, as Lon protease plays an important role in global protein homeostasis
264 by degrading misfolded proteins in addition to particular client proteins. Because our
265 goal was to characterize the immediate effects of segmental amplification on the
266 transcriptome and proteome in the absence of compensatory mutations, we sequenced

267 *prlF* from population 8 samples at earlier timepoints and determined that the mutation
268 was not detected in isolates screened at generation 125. This finding suggests that the
269 *prlF* mutation is highly beneficial, as it swept to fixation in approximately 60 generations.
270 qPCR on genomic DNA isolated from population 8 at generation 125 indicated that
271 *proA** was amplified to 5.2 copies, compared to 5.8 copies at generation 186 (S4 Fig).
272

273 **Expression of most amplified genes is elevated in the aftermath of segmental**
274 **amplification**

275 We performed RNA-seq and label-free quantitative proteomics on replicate
276 cultures started from single colonies isolated from population 3 at generation 199 and
277 from population 8 at generation 125 (i.e. prior to the *prlF* mutation). We verified that the
278 population 3 isolate contained the nested ~40 kb amplification inside a larger ~76 kb
279 duplication by PCR using primers near the amplification junctions. Transcriptomic and
280 proteomic analyses on the population 3 isolate with a nested duplication allowed us to
281 characterize how mRNA and proteins levels changed at two different amplification
282 levels, 2 and 12 copies, within the same genome. The ancestral *rpoS-proBA** relocation
283 strain was used as the reference strain for determining expression fold changes.

284 Levels of mRNAs and small regulatory RNAs (sRNAs) expressed from amplified
285 genes generally scale linearly with gene copy number (Fig 5A and S5 Fig, R^2 = 0.63).
286 On average, RNA levels increased 2.1±0.1-, 5.2±0.2-, and 10.0±0.6-fold (mean ± SEM)
287 for genes present at 2, 6 and 12 copies, respectively (Fig 5A). Protein levels increased
288 on average by 2.3±0.1-, 5.0±0.2-, and 10.5±1.3-fold for genes present at 2, 6, and 12
289 copies, respectively (Fig 5B). However, protein fold-change values varied widely,

290 especially for 6- and 12-copy genes. Levels of proteins expressed from amplified genes
291 correlate relatively weakly with both gene copy number (R^2 = 0.47) and mRNA level (R^2
292 = 0.35) (S5 Fig). Similar results have been observed in aneuploid yeast [31]. To some
293 extent the poor correlation may be because estimation of protein abundance by mass
294 spectrometry is less precise than estimation of RNA levels by RNA-Seq. However, in
295 many cases, posttranscriptional and posttranslational regulatory mechanisms likely
296 modulate RNA stability, translation and protein degradation in ways that attenuate the
297 impact of increased gene copy number.
298
299 **Fig 5.** Effect of segmental amplification on expression of RNA and protein from
300 amplified genes. (A) RNA fold-change for amplified genes. (B) Protein fold-change for
301 amplified genes. (C) Protein-fold change as a function of mRNA fold change for
302 amplified genes. Asterisks indicate statistically significant differences between groups
303 based on non-parametric Kruskal-Wallis test with Dunn's multiple comparisons test (***
304 p <0.001; **** p <0.001). Amplified genes that are discussed in the main text are
305 indicated by different colors.
306
307 **Expression of *rpoS* is attenuated in both populations 3 and 8**
308 The most glaring discrepancy between gene copy number and mRNA and
309 protein levels in our dataset is the minimal change in *rpoS* mRNA and RpoS protein
310 when *rpoS* is present at 12 copies in population 3 (Fig 5A and 5B). The junction
311 between the smaller amplicon and the larger duplicated region in population 3 (Fig 4C)
312 lies between *rpoS* and *nlpD*; the nested amplification separates *rpoS* from six of seven

313 potential promoters by moving them hundreds of kb upstream of the *rpoS* start codon
314 (Fig 6A). Thus, population 3 avoids problematic overexpression of *rpoS* even when the
315 gene copy number is very high. The expression of *rpoS* is also not proportional to the 6
316 copies of the gene in population 8 (Fig 5C). RpoS levels are tightly regulated by a
317 combination of multiple transcriptional regulators, sRNAs and an adaptor-mediated
318 proteolysis pathway [38]. The combined activity of these regulatory mechanisms may
319 keep RpoS levels in check in the absence of the canonical stresses to which it normally
320 responds.
321
322 **Fig 6.** Mechanisms that lead to discrepancies between gene copy number and
323 transcript or protein levels. (A) The nested amplification junction present in population 3
324 separates *rpoS* from 6 of 7 potential promoters. (B) Levels of the transcriptional
325 activator CysB, which is encoded by an unamplified gene, may not be sufficient to
326 activate transcription at all copies of the amplified *cysJIH* and *cysDNC* operons. (C)
327 Lower-than-expected FtsB levels might result from misfolding and degradation of
328 excess protein in the absence of enough of its interaction partners FtsQ and FtsL. (D)
329 Disproportionate expression of genes involved in the Csr system results from both
330 amplification of some but not all genes and complex interactions between RNA and
331 protein components of the circuit.
332
333 **Amplification of *cysJIH-DNC* may outstrip the availability of the transcriptional**
334 **activator CysB**

335 The *cysJIH* and *cysDNC* operons, which encode proteins involved in assimilatory
336 sulfate reduction, were amplified to 6 copies in population 8, but levels of the encoded
337 mRNAs increased by an average of only 3.6-fold (Fig 5C and S1 Dataset). Expression
338 of these operons requires the transcriptional activator CysB, which is encoded outside
339 the amplified region and is thus present at only a single copy (Fig 6B). The relatively
340 modest transcriptional response to amplification of the *cysJIH* and *cysDNC* operons
341 may be due to insufficient CysB to initiate transcription at all of the extra copies of the
342 *cysJIH* and *cysDNC* operons.
343

344 **Dosage imbalance may lead to degradation of a protein produced at a higher level**
345 **than its interaction partners**

346 FtsB is an essential cell division protein that forms a 1:1:1 complex with FtsQ and FtsL
347 [45]. *ftsB* is located within the amplified region in population 8 and is present at ~6
348 copies, but *ftsQ* and *ftsL* are ~2.8 Mb away from *ftsB* and are not amplified. While *ftsB*
349 expression increased 5.5-fold at the transcript level, it increased only 1.9-fold at the
350 protein level (Fig 5C and S1 Dataset). Proteins that form complexes are often unstable
351 in the absence of their interaction partners due to improper folding and subsequent
352 degradation by proteases [46–49]. The imbalance between the levels of FtsB and FtsQL
353 may result in misfolding and degradation of excess FtsB (Fig 6C).
354

355 **The structure of the Csr circuit seems to buffer it against dosage imbalances**
356 **following segmental amplification**

357 In population 8, *csrA*, *csrB* and *barA* are present within the 6-copy amplicon. The
358 levels of CsrA and BarA proteins are increased by 5.6- and 5.2-fold, respectively.
359 However, the level of the sRNA CsrB is disproportionally increased by 16-fold. These
360 genes are components of the carbon storage regulator (Csr) system, a complex
361 regulatory circuit consisting of both proteins and sRNAs that coordinates expression of
362 genes involved in diverse processes such as glycogen synthesis, glycolysis and
363 flagellar motility [50]. *csrC*, *csrD* and *uvrY*, which are also part of the Csr system, were
364 not amplified, but the levels of the sRNA CsrC and the protein UvrY are increased by
365 3.4- and 1.9, fold respectively, and the level of CsrD is modestly decreased by 40% (Fig
366 6D and S1 Dataset).
367 The striking increases in the levels of the CsrB and CsrC sRNAs in population 8
368 are likely due to a combination of effects mediated by the intricate network of regulatory
369 interactions in the Csr circuit (Fig 6D). The 5.2-fold increase in the level of the histidine
370 kinase BarA and the 1.9-fold increase in the level of the downstream response
371 regulation UvrY would be expected to increase transcription of *csrB* and *csrC.* In
372 addition, the decreased expression of CsrD, which recruits RNase E to degrade CsrB
373 and CsrC sRNAs, should also contribute to increased levels of CsrB and CsrC.
374 The substantial perturbations to sRNA and protein levels in this complex
375 regulatory circuit have only modest effects on expression of genes regulated by CsrA.
376 CsrA blocks translation of *glgCAP* mRNAs, which encode glycogen synthesis enzymes.
377 However, we see little change in GlgCAP protein levels (S2 Dataset). Conversely, CsrA
378 promotes translation of *flhDC* mRNAs, which encode the master regulator of flagellar
379 synthesis genes. We did not detect FlhDC proteins in our proteomics data, despite their

380 elevated transcript levels (S2 Dataset). The effect of the elevated expression of CsrA is
381 likely blunted by the elevated levels of both CsrB and CsrC, which sequester CsrA in an
382 inactive conformation. In addition, the effect of a modestly elevated level of active CsrA
383 may be insignificant in the context of other forms of regulation that influence *glgCAP*
384 and *flhDC* expression [51]. Overall, the intricate wiring of the Csr regulatory circuit,
385 replete with different forms of activation and inhibition, seems to buffer it against
386 stoichiometric imbalances between interacting sRNAs and proteins. This intrinsic
387 buffering capacity should be useful for acclimating to environmental changes while
388 avoiding dysregulation, but it also seems to have the incidental effect of limiting severe
389 physiological perturbations following segmental amplification.
390

391 **Segmental amplification has system-wide effects**

392 The segmental amplifications in populations 3 and 8 resulted in significant
393 changes hundreds of transcripts (Fig 7), including mRNAs and sRNAs encoded by
394 genes that were not present in the amplified regions. Expression of four classes of non-
395 amplified genes was increased in both populations 3 and 8: colanic acid capsular
396 polysaccharide synthesis and export genes; ribosomal RNA (rRNA) and ribosomal
397 protein genes; sugar transport/uptake genes; and flagellar synthesis and assembly
398 genes (S3 Dataset).
399

400 **Fig 7.** The global transcriptomic consequences of segmental amplification in population
401 3 (A) and population 8 (B). Dashed red lines indicate $\log_2$ fold change in RNA
402 expression of ≥1 or ≤-1 and $-\log_{10}$ adjusted *p*-values ≥2 (*p*-values <0.01). Amplified

403 genes that are differentially expressed ($\log_2$ fold-change of ≥1) in populations 3 and 8
404 are indicated by blue and pink points, respectively. Amplified genes *proA** and *rpoS* and
405 classes of non-amplified genes are indicated by different colors.
406
407 The increased expression of rRNA and ribosomal protein genes is likely an
408 indirect effect of the amplification of *proA**, which alleviates the shortage of proline and
409 arginine in the ancestral strain and should dampen the stringent response that
410 diminishes transcription of ribosomal genes [52]. The increased expression of colanic
411 acid polysaccharide synthesis genes appears to be caused by overexpression of the
412 transcriptional activator RcsA. Although *rcsA* is not among the co-amplified genes, *rcsA*
413 transcript levels are increased by 2.6- and 4-fold and protein levels are increased by
414 8.2- and 4.6-fold in populations 3 and 8, respectively (S2 Dataset and S3 Dataset).
415 RcsA forms a heterodimer with RscB that activates transcription of colanic acid
416 synthesis genes and of *rcsA* itself (S6 Fig) [53,54]. Degradation of RcsA by Lon and
417 HslUV proteases prevents excess accumulation of RcsA [53–56]. Inappropriate
418 overexpression of co-amplified genes following segmental amplification may divert
419 proteases away from client proteins like RcsA and toward overexpressed and potentially
420 misfolded proteins. The resulting higher level of RcsA would then increase transcription
421 of colanic acid synthesis genes as well as *rcsA* itself. As mentioned above, a mutation
422 that is known to hyperactivate Lon protease arose in population 8 after generation 125
423 and quickly rose to fixation. The highly beneficial nature of this mutation supports the
424 notion that increasing protein degradation alleviates the fitness cost associated with
425 improper overexpression of amplified genes.

426 The reasons for increased transcription of sugar transport genes and flagellar
427 genes following segmental amplification are less obvious. Amplification of *proA**
428 improves production of arginine and proline, allowing the cells to grow faster. The
429 increased transcription of sugar transport genes and flagellar genes may reflect the shift
430 in physiology associated with increased growth rate.
431

432 **Summary**

433 The strong selection for increased ProA* activity in our evolving populations
434 resulted in rapid segmental amplification encompassing *proA** and many neighboring
435 genes. Transcription of most amplified genes correlates well with copy number.
436 However, there are notable exceptions. Expression of *rpoS* is attenuated in population 3
437 by loss of several of its promoters. Excessive RpoS would be expected to be
438 problematic because RpoS promotes a transcriptional program that slows growth and
439 alters resource allocation when cells are starving or stressed. This response would be
440 inappropriate when glucose is abundant and proline and arginine synthesis have been
441 improved due to amplification of *proA*.* It is tempting to speculate that the loss of *rpoS*
442 promoters allowed the small amplicon nested within the originally duplicated region to
443 rise to 12 copies.
444 We also observed other mechanisms for attenuation of mRNA expression that do
445 not rely on a fortuitous amplification junction. Transcription of amplified genes in two
446 operons encoding sulfate assimilation genes is likely attenuated due to inadequate
447 levels of a transcriptional activator encoded by a gene that was not amplified. Such a
448 mismatch could occur in many situations when only some genes in a regulon are

449 amplified. Regulons often include operons that are scattered across the genome and
450 are controlled by transcriptional activators that may not be in close proximity to the
451 operons they control. While attenuation in this case appears to be due to outstripping
452 the capacity of a transcriptional activator, we expect that the opposite situation might
453 also arise; i.e. expression of amplified genes could be exacerbated if the capacity of a
454 transcriptional repressor is outstripped.

455 The complex regulation of many genes in *E. coli* may contribute to attenuation of
456 expression of amplified genes. *rpoS* expression is regulated by multiple transcription
457 factors and sRNAs, which allows integration of multiple signals in the decision to
458 activate the stress response. Likewise, the Csr system integrates multiple inputs in a
459 complex regulatory circuit in order to appropriately regulate expression of downstream
460 genes. While these systems evolved to enable *E. coli* to respond appropriately to
461 changing environments, a propitious side effect is the ability to buffer the effects of
462 segmental amplification.

463 Post-transcriptional mechanisms also appear to contribute to attenuation of
464 protein levels produced from amplified genes. For example, it is well accepted that
465 subunits of protein complexes that are produced in excess are often degraded, probably
466 because they misfold due to exposure of surfaces that are normally involved in protein-
467 protein interactions [34,46–49]. For example, excess FtsB might be degraded due to
468 insufficient levels of its interacting partner proteins FtsQ and FtsL. Some proteins such
469 as RpoS are subject to multiple layers of post-transcriptional regulation by sRNAs and
470 RNA-binding proteins that modulate translation and adaptor-mediated proteolysis and
471 collectively prevent inopportune overexpression.

472 Finally, we note that segmental amplification can have global consequences.
473 Approximately 84% and 43% of genes whose expression is increased by ≥2-fold at the
474 transcript level in populations 3 and 8, respectively, are not amplified. While co-
475 amplification of genes neighboring a gene under selection for increased copy number
476 has previously been recognized to contribute to fitness, our work adds a previously
477 unappreciated nuance; segmental amplification can also alter expression of mRNAs
478 and proteins from hundreds of unamplified genes. Thus, the fitness effects of
479 segmental amplification are due to the combination of the desirable effect of amplifying
480 the gene under selection for increase dosage and the collateral effects (both detrimental
481 and beneficial) of altered expression of co-amplified genes within the amplicon as well
482 as genes elsewhere in the genome.
483

484 **Materials and Methods**

485 **Strains and growth conditions**

486 *E. coli* strains, plasmids and primers used in this work are listed in Tables B-D in
487 S1 Text. *E. coli* strains were routinely cultured in lysogeny broth (LB) or on LB agar
488 plates containing 1.5% agar at 37 °C. M9 salts minimal medium (12.8 g/L
489 Na_2HPO_2•$7H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L NH_4Cl) was supplemented with
490 glucose at a final concentration of 0.2% (v/v). Antibiotics were used at the following final
491 concentrations unless otherwise indicated: ampicillin, 100-150 µg/mL; kanamycin, 20
492 µg/mL; chloroamphenicol, 20 µg/mL; and gentamicin, 15 µg/ml.
493

494 **Strain construction**

495 All strains used in this work are derived from *E. coli* BW25113. The Δ*argC*::*kan*^{R}
496 M2-*proBA**-*yfp* Δ*fimAICDFGH* Δ*csgBAC* Δ82 bp *rph-pyrE E. coli* strain AM327 was
497 previously constructed from the Keio collection Δ*argC*::*kan*^{R} *E. coli* BW25113 strain as
498 described in [4]. The Δ*fimAICDFGH* Δ*csgBAC* deletions help prevent biofilm formation
499 during laboratory evolution in a turbidostat. The 82 bp deletion in *rph* and upstream of
500 *pyrE* fixes a defect in pyrimidine synthesis due to a frameshift in *rph* in *E. coli* BW25113
501 and other K-12 strains [57,58]. The M2 promotor mutation is a C > T point mutation 45
502 bp upstream of *proB* that increases transcription of the *proBA* operon as described in
503 [59]. The AM187 strain (GenBank accession number CP037857.2) and its derivative
504 AM327 contain a 1.8-Mb inversion in the genome that occurred between two IS3
505 elements [37].
506 For this study, we deleted the M2-*proBA**-*yfp* region in the *E. coli* AM327
507 genome using λ-Red recombineering [60]. We transformed the temperature-sensitive
508 plasmid pSIM6 [61] containing the λ-Red recombinase genes (*exo*, *beta* and *gam*)
509 under control of a heat-inducible promoter into AM327 by electroporation. After recovery
510 in super optimal broth with catabolite repression medium (SOC) for ~3 h at 30 °C, cells
511 were plated on LB ampicillin and incubated at 30 °C to select for transformants. We
512 replaced *proBA** and *yfp* in AM327 with a *sacB-gtm*^{R} selection-counterselection
513 cassette amplified from the plasmid pJQ200SK [56]. We amplified ~1 kb fragments from
514 the regions upstream and downstream of the *proBA* operon using primers that
515 introduced ~20 bp ends complementary to the ends of the *sacB-gtm*^{R} cassette. We
516 assembled a linear *proBA*-yfp* replacement fragment from these three fragments using
517 splicing by overlap extension (SOE) PCR. The AM327 strain harboring pSIM6 was

518 inoculated into LB ampicillin, grown to an OD_{600} of 0.2-0.6 at 30 °C and then incubated
519 at 42 °C for 15 min to induce expression of the λ Red genes. Cells were then
520 centrifuged at 4,500 x g at 4 °C, washed with ice-cold deionized water and immediately
521 electroporated with ~0.5-2.0 µg of the *proBA** replacement linear fragment. Following
522 electroporation, cells were inoculated into fresh SOC and allowed to recover at 30 °C for
523 ~3 h before being spread onto LB gentamicin. Gentamicin-resistant colonies were
524 screened for integration of *sacB-gtm*^{R} into the *proBA* locus by colony PCR.
525 We constructed a fragment to delete the *sacB-gtm*^{R} cassette by amplifying ~1 kb
526 regions upstream and downstream of the *proBA* operon with overlapping ends. The two
527 PCR products were assembled together using SOE PCR. Cells containing the *sacB-*
528 *gtm*^{R} selection-counterselection marker in place of *proBA** and pSIM6 were inoculated
529 into LB ampicillin, grown to an OD_{600} of 0.2-0.6 at 30 °C and subjected to a second
530 round of λ-Red recombineering as described above. Cells were electroporated with
531 ~0.5-2.0 µg of the *sacB-gtm*^{R} replacement fragment. Following electroporation, cells
532 were inoculated into fresh SOC and allowed to recover at 30 °C for ~3 h before being
533 spread onto salt-free (i.e. no NaCl) LB plates containing 10% (w/v) sucrose and
534 incubated at 30 °C for counterselection. Colonies that grew on salt-free LB sucrose
535 plates were patched onto LB gentamicin plates to confirm gentamicin sensitivity due to
536 loss of the *sacB-gtm*^{R} cassette and screened for markerless deletion of *proBA** by
537 colony PCR and Sanger sequencing.
538 To relocate the M2-*proBA** operon next to *rpoS* in the *E. coli* genome, we first
539 introduced a *tetR-ccdB-cat* selection-counterselection cassette at the target site
540 between *rpoS* and *ygbN.* We amplified the *tetR-ccdB-cat* cassette from pDLM3 [62]. We

541 amplified ~1 kb fragments upstream and downstream of the target site using primers
542 that introduced ~20 bp ends complementary to the ends of the *tetR-ccdB-cat* cassette.
543 These three fragments were assembled together using SOE PCR. The Δ*proBA** strain
544 RF72 harboring pSIM6 was inoculated into LB ampicillin, grown to an OD_{600} of 0.2-0.6
545 at 30 °C and subjected to λ-Red recombineering as described above. Cells were
546 electroporated with ~0.5-2.0 µg of the *tetR-ccdB-cat* integration fragment. Following
547 electroporation, cells were inoculated into fresh SOC and allowed to recover at 30 °C for
548 ~3 h or at room temperature overnight before being spread onto LB chloramphenicol
549 plates. The plates were incubated at 30 °C for 1-2 days. Chloramphenicol-resistant
550 colonies were screened for integration of *tetR-ccdB-cat* between *ygbN* and *rpoS* by
551 colony PCR. The *tetR-ccdB-cat* cassette was then replaced with M2-*proBA*.* We
552 amplified the M2-*proBA** operon (without *yfp*) from AM187 gDNA and ~1 kb fragments
553 upstream and downstream of the intergenic site between *rpoS* and *ygbN* from RF72
554 gDNA by PCR. These three fragments were assembled using SOE-PCR to generate
555 the *tetR-ccdB-cat* replacement cassette. The *ygbN-tetR-ccdB-cat-rpoS* strain harboring
556 pSIM6 was inoculated into LB ampicillin, grown to an OD_{600} of 0.2-0.6 at 30 °C and
557 subjected to λ-Red recombineering as described above. Cells were electroporated with
558 ~0.5-2.0 µg of the *tetR-ccdB-cat* replacement cassette. Following electroporation, cells
559 were inoculated into fresh SOC and allowed to recover at 30 °C for ~3 h or at room
560 temperature overnight before being spread onto LB anhydrotetracycline (2-6 ug/mL)
561 plates for counterselection. The plates were incubated at 30 °C overnight. Colonies that
562 grew on LB anhydrotetracycline plates were patched onto LB chloramphenicol plates to
563 confirm chloramphenicol sensitivity and screened for integration of M2-*proBA** into the

564 intergenic site between *ygbN* and *rpoS.* The resulting M2-*proBA**-*rpoS* strain was
565 streaked onto LB anhydrotetracycline plates and incubated at 40 °C overnight to cure
566 the temperature-sensitive pSIM6 plasmid. Colonies were patched onto fresh LB
567 anhydrotetracycline and LB ampicillin plates to confirm ampicillin sensitivity due to loss
568 of pSIM6. Genomic integration of M2-*proBA** next to *rpoS* was confirmed by Sanger
569 sequencing.
570
571 **Laboratory Evolution**
572 To initiate laboratory evolution of the *proBA**-*rpoS* relocation strain (RF90), cells
573 from a frozen glycerol stock were streaked onto LB kanamycin plates and the plates
574 were incubated at 37 °C overnight. A single founder colony was inoculated into LB
575 kanamycin and incubated at 37 °C overnight. The overnight culture was centrifuged at
576 8,000 x g at room temperature and the cells were washed and resuspended in M9
577 medium without glucose. A 125 µL aliquot of cells was inoculated into each of eight
578 glass vials containing 25 mL (1:200 dilution) of M9 minimal medium containing 0.2%
579 glucose and kanamycin (20 µg/mL). The vials were placed into the vial sleeves of an
580 eVOLVER turbidostat (FynchBio). Each vial contained a micro-flea magnetic stir bar (10
581 mm x 3 mm). The eVOLVER was set in 'turbidostat' operation mode at 37 °C with a stir
582 rate of 8, a lower OD_{600} threshold of 0.15, an upper OD_{600} threshold of 0.25 and a vial
583 volume of 25 mL. When cultures reached the upper OD_{600} threshold of 0.25, ~7-10 ml of
584 fresh M9 + glucose medium was pumped into vials and an equivalent volume of culture
585 was removed so that cultures were diluted to an OD_{600} of ~0.15. Cultures were evolved
586 in this manner for 145-200 generations (~14 days). Approximately every 40-60 h, ~7-10

587 ml samples of cultures expelled from the vials during dilution were collected and
588 centrifuged at 4,500 x g at room temperature. The cell pellets were resuspended in 1
589 mL of M9 medium without glucose. Aliquots (0.5 mL) of concentrated cells from each
590 population were mixed 1:1 with 50% glycerol and stored at -70 °C. The remaining 0.5 ml
591 cell suspensions were centrifuged at 8,000 x g at room temperature, the supernatants
592 were removed and the cell pellets were stored at -20 °C for extraction of genomic DNA
593 (gDNA).
594
595 **Whole Genome Sequencing and Analysis**
596 Cell pellets were thawed and gDNA was extracted using the PureLink™ Genomic DNA
597 Mini Kit (Invitrogen) according to the manufacturer's protocol for Gram-negative
598 bacteria. Purified gDNA samples were stored at -20 °C. Whole genome sequencing was
599 performed at SeqCenter (formerly the Microbial Genome Sequencing Center;
600 Pittsburgh, PA). Sample libraries were prepared with the Illumina DNA Prep kit and IDT
601 10 bp UDI indices and sequenced on an Illumina NextSeq 2000, producing 2x151 bp
602 paired-end reads. Demultiplexing, quality control, and adapter trimming was performed
603 with Illumina BCL Convert v3.9.1. The resulting fastq files were further processed using
604 fastp v0.20.1 [63] to trim and remove low quality reads with the following settings: -q 15
605 -u 20 -t 1. Mutations in evolved populations were identified using breseq v0.37.1 [64,65]
606 in polymorphism mode with bowtie 2.5.0 [66] and R 4.2.2 by mapping processed fastq
607 reads to the genome of the *E. coli proBA*-rpoS* strain. To visualize read coverage depth
608 of segmental amplifications encompassing *proBA**, read coverage depth for every 50 bp

609 for a 300 kb genome window was exported using the breseq BAM2COV subcommand:
610 -a seq id:2700000-3000000 -t -1 --resolution 6000.
611
612 **Quantitative PCR for measurement of *proA** copy number**
613 *proA** copy number was measured using qPCR of purified gDNA as described in
614 [4] with slight modifications. Briefly, PowerSYBR Green PCR master mix (Applied
615 Biosytems by Thermo Scientific) was combined with primers and gDNA in 48-well plates
616 (Thermo Scientific) and samples were analyzed using a StepOne Real-Time PCR
617 System (Applied Biosystems). The genes *gyrB* and *icd*, which remained at one copy in
618 evolved populations, were used as internal reference genes for quantification of *proA**
619 copy number. Threshold cycle (C_T) measurements for duplicate technical replicates of
620 each gDNA sample were averaged. The ΔC_T values for *proA** and single copy reference
621 genes were calculated by subtracting the C_T for gDNA from the ancestral *proBA*-rpoS*
622 strain from the C_T for gDNA from the evolved populations. The ΔC_T values of reference
623 genes *gyrB* and *icd* were averaged for each sample and $\Delta\Delta C_T$ values were then
624 calculated by subtracting the mean reference ΔC_T from *proA** ΔC_T for each sample.
625 *proA** copy number was calculated as $2^{-\Delta\Delta CT}$ for each sample. Primer sets used for
626 qPCR are listed in Table D in S1 Text.
627
628 **RNA extraction**
629 Glycerol stocks of ancestral *proBA*-rpoS E. coli* and evolved populations 3 (at
630 generation ~199) and 8 (at generation ~125) were streaked onto M9 plates containing
631 0.4% glucose. Single colonies of each strain were inoculated into 4 ml M9 containing

632 0.2% glucose and the cultures were grown at 37 °C overnight with shaking. The
633 overnight cultures were inoculated into fresh 25 ml M9 containing 0.2% glucose in 25
634 mL flasks at a 1:50 dilution for the ancestral strain and a 1:100 dilution for the evolved
635 populations. Four replicate cultures of each strain were grown at 37 ºC with shaking for
636 ~7 h to an OD_{600} of ~0.2-0.5. Cultures were centrifuged at 4,500 x g for 10 min at 4 °C.
637 Cell pellets were resuspended in 5 mL M9 salts without glucose. Four mL of each cell
638 suspension were flash frozen in a dry ice-ethanol (95%) bath and stored at -70 °C prior
639 to RNA extraction. The remaining 1 mL of each cell suspension was centrifuged at
640 8,000 x g for 2 min. The supernatant was removed and the cell pellets were flash frozen
641 in a dry ice-ethanol (95%) bath and stored at -20 °C prior to sample preparation for
642 proteomics.

643 Total RNA was extracted from cells using RNAprotect Bacteria Reagent and
644 RNeasy Mini Kit (Qiagen). Frozen 4-mL cell suspensions were thawed on ice, mixed 2:1
645 with RNAprotect Bacteria Reagent, incubated for 5 min at room temperature and
646 centrifuged at 4,500 x g for 10 min at room temperature. Cell pellets were resuspended
647 in TE buffer with lysozyme (1 mg/ml) to lyse the cells. RNA was purified using RNeasy
648 Mini Kits according to the manufacturer's instructions except that optional on-column
649 DNase digestion was performed by mixing 10 µL TURBO™ DNase (Invitrogen) with 70
650 µL buffer RDD (Qiagen), adding 80 µL of DNase solution directly to the spin column
651 membrane of each sample and incubating at room temperature for 15 min. RNA
652 samples were stored at -20 °C.

653

654 **RNA sequencing and differential expression analysis**

655 RNA sequencing was performed at SeqCenter (formerly the Microbial Genome
656 Sequencing Center; Pittsburgh, PA). Three replicate RNA samples were sequenced for
657 each strain. RNA samples were treated with DNase (Invitrogen). Sample libraries were
658 prepared with the Illumina Stranded Total RNA Prep Ligation with Ribo-Zerp Plus kit
659 and 10 bp IDT for Illumina indices and sequenced on a NextSeq 2000, producing 2x51
660 bp paired-end reads. Demultiplexing, quality control, and adapter trimming was
661 performed with Illumina BCL Convert v3.9.1.

662 To map reads to the *E. coli* genome to generate read counts for each gene, fastq
663 files were processed with fastp v0.20.1 [63] to trim and remove low quality reads with
664 the following settings: -q 15 -u 20. Processed reads were aligned to the *E. coli* reference
665 genome using bowtie2 [66]. Read alignment files were converted into sorted.bam files
666 using samtools [67,68]. Aligned reads were mapped to *E. coli* genes using HTSeq [69]
667 to generate read counts with the settings: htseq-count -f bam –r pos -s no -t gene -i
668 Name -m intersection-nonempty --nonunique all.

669 Differential expression analysis was performed in R v4.3.1/RStudio (Posit team;
670 Boston, MA) with the following Bioconductor/R packages: edgeR [70,71], limma [72],
671 glimma, gplots, RColorBrewer and NFM. Read counts generated from HTSeq were
672 filtered to remove lowly expressed genes and normalized using trimmed mean of M
673 values with edgeR. Read counts were transformed into log counts per million (CPM)
674 with the voom function of limma for each sample. Differential expression analysis of
675 transformed logCPM data was performed with the contrasts.fit and eBayes functions in
676 limma. Differential expression data were exported and plotted in GraphPad Prism
677 v10.0.2 (GraphPad Software; Boston, MA)

678

679 **Label-free proteomics and differential expression analysis**

680 Frozen cell pellets were lysed in 50 mM Tris-HCl, pH 8.5 containing 5% (w/v)
681 sodium dodecyl sulfate, 10 mM tris(2-carboxyethylphosphine) and 40 mM
682 chloroacetamide by boiling for 10 minutes at 95 °C followed by probe sonication (1 sec
683 on, 1 sec off) for 1 minute. Each sample was digested using the SP3 method described
684 in [73]. Carboxylate-functionalized SpeedBeads (GE Life Sciences) were added to the
685 lysates. Addition of acetonitrile to 80% (v/v) caused proteins to bind to the beads. The
686 beads were washed twice with 80% (v/v) ethanol and twice with 100% acetonitrile.
687 Proteins were digested with Lys-C/Trypsin (Promega) at 37 °C overnight with rotation.
688 SpeedBeads were collected by centrifugation and placed on a magnet to hold the
689 SpeedBead pellet together so that the digested peptides in the supernatant could be
690 extracted. The peptides were then desalted using an Oasis HLB cartridge (Waters)
691 according to the manufacturer's instructions and dried in a SpeedVac.

692 The dried peptide samples were suspended in 3% (v/v) acetonitrile/0.1% (v/v)
693 trifluoroacetic acid and 1 µg was directly injected onto a C18 1.7 µm, 130 Å, 75 µm X
694 250 mm M-class column (Waters) using a Thermo Ultimate 3000 RSLCnano UPLC.
695 Peptides were eluted at 300 nL/minute using a gradient from 2% to 20% acetonitrile
696 over 100 min into a Q-Exactive HF-X mass spectrometer (Thermo Scientific). Precursor
697 mass spectra (MS1) were acquired at a resolution of 120,000 from 350 to 1550 m/z with
698 an automatic gain control (AGC) target of 3E6 and a maximum injection time of 50
699 milliseconds. Precursor peptide ion isolation width for MS2 fragment scans was 1.4 m/z,
700 and the top 12 most intense ions were sequenced. All MS2 spectra were acquired at a

701 resolution of 15,000 with higher energy collision dissociation (HCD) at 27% normalized
702 collision energy. An AGC target of 1E5 and 100 milliseconds maximum injection time
703 was used. Raw files were searched against the Uniprot *Escherichia coli* databases
704 UP000002032 and UP000000625 using Maxquant version 2.0.3.0 with cysteine
705 carbamidomethylation as a fixed modification. Methionine oxidation and protein N-
706 terminal acetylation were searched as variable modifications. All peptides and proteins
707 were thresholded at a 1% false discovery rate (FDR). Statistical analysis was performed
708 on $\log_2$-transformed iBAQ and LFQ intensities using limma [72].
709

710 **References**

711 1. Paulander W, Andersson DI, Maisnier-Patin S. Amplification of the gene for
712 isoleucyl–tRNA synthetase facilitates adaptation to the fitness cost of mupirocin
713 resistance in *Salmonella enterica*. Genetics. 2010;185: 305–312.
714 doi:10.1534/genetics.109.113514
715 2. Näsvall J, Sun L, Roth JR, Andersson DI. Real-time evolution of new genes by
716 innovation, amplification, and divergence. Science. 2012;338: 384 LP – 387.
717 doi:10.1126/science.1226521
718 3. Seaton SC, Elliott KT, Cuff LE, Laniohan NS, Patel PR, Neidle EL. Genome-wide
719 selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and
720 context-dependent variation in gene amplification. Mol Microbiol. 2012;83: 520–
721 535. doi:10.1111/j.1365-2958.2011.07945.x
722 4. Morgenthaler AB, Kinney WR, Ebmeier CC, Walsh CM, Snyder DJ, Cooper VS, et
723 al. Mutations that improve efficiency of a weak-link enzyme are rare compared to
724 adaptive mutations elsewhere in the genome. eLife. 2019;8: e53535.
725 doi:10.7554/eLife.53535
726 5. Girgis HS, DuPai CD, Lund J, Reeder J, Guillory J, Durinck S, et al. Single-
727 molecule nanopore sequencing reveals extreme target copy number
728 heterogeneity in arylomycin-resistant mutants. Proc Natl Acad Sci USA.
729 2021;118: e2021958118. doi:10.1073/pnas.2021958118
730 6. Steinrueck M, Guet CC. Complex chromosomal neighborhood effects determine
731 the adaptive potential of a gene under selection. eLife. 2017;6: e25100.
732 doi:10.7554/eLife.25100
733 7. Silva KPT, Sundar G, Khare A. Efflux pump gene amplifications bypass necessity
734 of multiple target mutations for resistance against dual-targeting antibiotic. Nat
735 Commun. 2023;14: 3402. doi:10.1038/s41467-023-38507-4

8. Saathoff M, Kosol S, Semmler T, Tedin K, Dimos N, Kupke J, et al. Gene amplifications cause high-level resistance against albicidin in gram-negative bacteria. PLoS Biol. 2023;21: e3002186. doi:10.1371/journal.pbio.3002186
9. Wang L, Asare E, Shetty AC, Sanchez-Tumbaco F, Edwards MR, Saranathan R, et al. Multiple genetic paths including massive gene amplification allow *Mycobacterium tuberculosis* to overcome loss of ESX-3 secretion system substrates. Proc Natl Acad Sci USA. 2022;119: e2112608119. doi:10.1073/pnas.2112608119
10. Brown CJ, Todd KM, Rosenzweig RF. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol Biol Evol. 1998;15: 931–942. doi:10.1093/oxfordjournals.molbev.a026009
11. Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, et al. The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast. PLoS Genet. 2008;4: e1000303. doi:10.1371/journal.pgen.1000303
12. Gu Z, Cavalcanti A, Chen F-C, Bouman P, Li W-H. Extent of gene duplication in the genomes of *Drosophila*, Nematode, and Yeast. Mol Biol Evol. 2002;19: 256–262. doi:10.1093/oxfordjournals.molbev.a004079
13. Payen C, Di Rienzi SC, Ong GT, Pogachar JL, Sanchez JC, Sunshine AB, et al. The dynamics of diverse segmental amplifications in populations of *Saccharomyces cerevisiae* adapting to strong selection. G3: Genes Genomes Genet. 2014;4: 399–409. doi:10.1534/g3.113.009365
14. Bass C, Zimmer CT, Riveron JM, Wilding CS, Wondji CS, Kaussmann M, et al. Gene amplification and microsatellite polymorphism underlie a recent insect host shift. Proc Natl Acad Sci USA. 2013;110: 19460–19465. doi:10.1073/pnas.1314122110
15. Vande Zande P, Zhou X, Selmecki A. The dynamic fungal genome: polyploidy, aneuploidy and copy number variation in response to stress. Annu Rev Microbiol. 2023. doi:10.1146/annurev-micro-041320-112443
16. Bergin SA, Zhao F, Ryan AP, Müller AC, Nieduszynski CA, Zhai B, et al. Systematic analysis of copy number variations in the pathogenic Yeast *Candida parapsilosis* identifies a gene amplification in RTA3 that is associated with drug resistance. mBio. 2022;13: e01777-22. doi:10.1128/mbio.01777-22
17. Albertson DG. Gene amplification in cancer. Trends Genet. 2006;22: 447–455. doi:10.1016/j.tig.2006.06.007
18. Beroukhim R, Mermel CH, Porter D, Wei G, Raychaudhuri S, Donovan J, et al. The landscape of somatic copy-number alteration across human cancers. Nature. 2010;463: 899–905. doi:10.1038/nature08822
19. Santarius T, Shipley J, Brewer D, Stratton MR, Cooper CS. A census of amplified and overexpressed human cancer genes. Nat Rev Cancer. 2010;10: 59–64. doi:10.1038/nrc2771
20. Bergthorsson U, Andersson DI, Roth JR. Ohno's dilemma: Evolution of new genes under continuous selection. Proc Natl Acad Sci USA. 2007;104: 17004–17009. doi:10.1073/pnas.0707158104
21. Copley SD. Evolution of new enzymes by gene duplication and divergence. FEBS J. 2020;287: 1262–1283. doi:10.1111/febs.15299

782 22. Logeman BL, Wood LK, Lee J, Thiele DJ. Gene duplication and neo-
783 functionalization in the evolutionary and functional divergence of the metazoan
784 copper transporters Ctr1 and Ctr2. J Biol Chem. 2017;292: 11531–11546.
785 doi:10.1074/jbc.M117.793356
786 23. Voordeckers K, Brown CA, Vanneste K, van der Zande E, Voet A, Maere S, et al.
787 Reconstruction of ancestral metabolic enzymes reveals molecular mechanisms
788 underlying evolutionary innovation through gene duplication. PLoS Biol. 2012;10:
789 e1001446. doi:10.1371/journal.pbio.1001446
790 24. Rojas AM, Fuentes G, Rausell A, Valencia A. The Ras protein superfamily:
791 Evolutionary tree and role of conserved amino acids. J Cell Biol. 2012;196: 189–
792 201. doi:10.1083/jcb.201103008
793 25. Todd RT, Selmecki A. Expandable and reversible copy number amplification
794 drives rapid adaptation to antifungal drugs. eLife. 2020;9: e58349.
795 doi:10.7554/eLife.58349
796 26. Belikova D, Jochim A, Power J, Holden MTG, Heilbronner S. “Gene accordions”
797 cause genotypic and phenotypic heterogeneity in clonal populations of
798 *Staphylococcus aureus*. Nat Commun. 2020;11: 3526. doi:10.1038/s41467-020-
799 17277-3
800 27. Hose J, Yong CM, Sardi M, Wang Z, Newton MA, Gasch AP. Dosage
801 compensation can buffer copy-number variation in wild yeast. eLife. 2015;4:
802 e05462. doi:10.7554/eLife.05462
803 28. Torres EM, Springer M, Amon A. No current evidence for widespread dosage
804 compensation in *S. cerevisiae*. eLife. 2016;5: e10996. doi:10.7554/eLife.10996
805 29. Sheltzer JM, Torres EM, Dunham MJ, Amon A. Transcriptional consequences of
806 aneuploidy. Proc Natl Acad Sci USA. 2012;109: 12644–12649.
807 doi:10.1073/pnas.1209227109
808 30. Pavelka N, Rancati G, Zhu J, Bradford WD, Saraf A, Florens L, et al. Aneuploidy
809 confers quantitative proteome changes and phenotypic variation in budding yeast.
810 Nature. 2010;468: 321–325. doi:10.1038/nature09529
811 31. Dephoure N, Hwang S, O’Sullivan C, Dodgson SE, Gygi SP, Amon A, et al.
812 Quantitative proteomic analysis reveals posttranslational responses to aneuploidy
813 in yeast. eLife. 2014;3: e03023. doi:10.7554/eLife.03023
814 32. Myhre S, Lingjærde O-C, Hennessy BT, Aure MR, Carey MS, Alsner J, et al.
815 Influence of DNA copy number and mRNA levels on the expression of breast
816 cancer related proteins. Mol Oncol. 2013;7: 704–718.
817 doi:10.1016/j.molonc.2013.02.018
818 33. Geiger T, Cox J, Mann M. Proteomic changes resulting from gene copy number
819 variations in cancer cells. PLoS Genet. 2010;6: e1001090.
820 doi:10.1371/journal.pgen.1001090
821 34. Ascencio D, Diss G, Gagnon-Arsenault I, Dubé AK, DeLuna A, Landry CR.
822 Expression attenuation as a mechanism of robustness against gene duplication.
823 Proc Natl Acad Sci USA. 2021;118: e2014345118. doi:10.1073/pnas.2014345118
824 35. Yu McLoughlin S, Copley SD. A compromise required by gene sharing enables
825 survival: Implications for evolution of new enzyme activities. Proceedings of the
826 National Academy of Sciences. 2008;105: 13497–13502.
827 doi:10.1073/pnas.0804804105

828 36. Kershner JP, Yu McLoughlin S, Kim J, Morgenthaler A, Ebmeier CC, Old WM, et
829 al. A synonymous mutation upstream of the gene encoding a weak-link enzyme
830 causes an ultrasensitive response in growth rate. J Bacteriol. 2016;198: 2853–
831 2863. doi:10.1128/JB.00262-16
832 37. Morgenthaler AB, Fritts RK, Copley SD. Amplicon remodeling and genomic
833 mutations drive population dynamics after segmental amplification. Mol Biol Evol.
834 2022;39: msab289. doi:10.1093/molbev/msab289
835 38. Battesti A, Majdalani N, Gottesman S. The RpoS-mediated general stress
836 response in *Escherichia coli*. Annu Rev Microbiol. 2011;65: 189–213.
837 doi:10.1146/annurev-micro-090110-102946
838 39. Hengge-Aronis R. Signal transduction and regulatory mechanisms involved in
839 control of the σS (RpoS) subunit of RNA polymerase. Microbiol Mol Biol Rev.
840 2002;66: 373–395. doi:10.1128/mmbr.66.3.373-395.2002
841 40. Chandran V, Luisi BF. Recognition of enolase in the *Escherichia coli* RNA
842 degradosome. J Mol Biol. 2006;358: 8–15. doi:10.1016/j.jmb.2006.02.012
843 41. Tejada-Arranz A, de Crécy-Lagard V, de Reuse H. Bacterial RNA degradosomes:
844 Molecular machines under tight control. Trends Biochem Sci. 2020;45: 42–57.
845 doi:10.1016/j.tibs.2019.10.002
846 42. Wetmore KM, Price MN, Waters RJ, Lamson JS, He J, Hoover CA, et al. Rapid
847 quantification of mutant fitness in diverse bacteria by sequencing randomly bar-
848 coded transposons. mBio. 2015;6: 10.1128/mbio.00306-15.
849 doi:10.1128/mbio.00306-15
850 43. Snyder WB, Silhavy TJ. Enhanced export of beta-galactosidase fusion proteins in
851 *prlF* mutants is Lon dependent. J Bacteriol. 1992;174: 5661–5668.
852 doi:10.1128/jb.174.17.5661-5668.1992
853 44. Schmidt O, Schuenemann VJ, Hand NJ, Silhavy TJ, Martin J, Lupas AN, et al.
854 *prlF* and *yhaV* encode a new toxin-antitoxin system in *Escherichia coli*. J Mol Biol.
855 2007;372: 894–905. doi:10.1016/j.jmb.2007.07.016
856 45. Du S, Lutkenhaus J. Assembly and activation of the *Escherichia coli* divisome.
857 Mol Microbiol. 2017;105: 177–187. doi:10.1111/mmi.13696
858 46. Brennan CM, Vaites LP, Wells JN, Santaguida S, Paulo JA, Storchova Z, et al.
859 Protein aggregation mediates stoichiometry of protein complexes in aneuploid
860 cells. Genes Dev. 2019;33: 1031–1047. doi:10.1101/gad.327494.119
861 47. Mueller S, Wahlander A, Selevsek N, Otto C, Ngwa EM, Poljak K, et al. Protein
862 degradation corrects for imbalanced subunit stoichiometry in OST complex
863 assembly. Mol Biol Cell. 2015;26: 2596–2608. doi:10.1091/mbc.E15-03-0168
864 48. Marsh JA, Hernández H, Hall Z, Ahnert SE, Perica T, Robinson CV, et al. Protein
865 complexes are under evolutionary selection to assemble via ordered pathways.
866 Cell. 2013;153: 461–470. doi:10.1016/j.cell.2013.02.044
867 49. Taggart JC, Lalanne J-B, Li G-W. Quantitative control for stoichiometric protein
868 synthesis. Annu Rev Microbiol. 2021;75: 243–267. doi:10.1146/annurev-micro-
869 041921-012646
870 50. Pourciau C, Lai Y-J, Gorelik M, Babitzke P, Romeo T. Diverse mechanisms and
871 circuitry for global regulation by the RNA-binding protein CsrA. Front Microbiol.
872 2020. doi:10.3389/fmicb.2020.601352

873 51. Fitzgerald DM, Bonocora RP, Wade JT. Comprehensive mapping of the
874 *Escherichia coli* flagellar regulatory network. PLoS Genet. 2014;10: e1004649.
875 doi:10.1371/journal.pgen.1004649
876 52. Magnusson LU, Farewell A, Nyström T. ppGpp: a global regulator in *Escherichia*
877 *coli*. Trends Microbiol. 2005;13: 236–242.
878 doi:https://doi.org/10.1016/j.tim.2005.03.008
879 53. Ebel W, Trempy JE. *Escherichia coli* RcsA, a positive activator of colanic acid
880 capsular polysaccharide synthesis, functions to activate its own expression. J
881 Bacteriol. 1999;181: 577–584. doi:10.1128/jb.181.2.577-584.1999
882 54. Wall E, Majdalani N, Gottesman S. The complex Rcs regulatory cascade. Annu
883 Rev Microbiol. 2018;72: 111–139. doi:10.1146/annurev-micro-090817-062640
884 55. Kuo M-S, Chen K-P, Wu WF. Regulation of RcsA by the ClpYQ (HslUV) protease
885 in *Escherichia coli*. Microbiology. 2004;150: 437–446. doi:10.1099/mic.0.26446-0
886 56. Chang C-Y, Hu H-T, Tsai C-H, Wu W-F. The degradation of RcsA by ClpYQ
887 (HslUV) protease in *Escherichia coli*. Microbiol Res. 2016;184: 42–50.
888 doi:10.1016/j.micres.2016.01.001
889 57. Conrad TM, Joyce AR, Applebee MK, Barrett CL, Xie B, Gao Y, et al. Whole-
890 genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-term
891 laboratory evolution in lactate minimal media reveals flexible selection of adaptive
892 mutations. Genome Biol. 2009;10: R118. doi:10.1186/gb-2009-10-10-r118
893 58. Jensen KF. The *Escherichia coli* K-12 “wild types” W3110 and MG1655 have an
894 *rph* frameshift mutation that leads to pyrimidine starvation due to low *pyrE*
895 expression levels. J Bacteriol. 1993;175: 3401–3407.
896 doi:10.1128/JB.175.11.3401-3407.1993
897 59. Kristofich J, Morgenthaler AB, Kinney WR, Ebmeier CC, Snyder DJ, Old WM, et
898 al. Synonymous mutations make dramatic contributions to fitness when growth is
899 limited by a weak-link enzyme. PLoS Genet. 2018;14: e1007615.
900 doi:10.1371/journal.pgen.1007615
901 60. Datsenko KA, Wanner BL. One-step inactivation of chromosomal genes in
902 *Escherichia coli* K-12 using PCR products. Proceedings of the National Academy
903 of Sciences. 2000;97: 6640–6645. doi:10.1073/pnas.120163297
904 61. Datta S, Costantino N, Court DL. A set of recombineering plasmids for gram-
905 negative bacteria. Gene. 2006;379: 109–115. doi:10.1016/j.gene.2006.04.018
906 62. Bossi L, Ratel M, Laurent C, Kerboriou P, Camilli A, Eveno E, et al. NusG
907 prevents transcriptional invasion of H-NS-silenced genes. PLoS Genet. 2019;15:
908 e1008425. doi:10.1371/journal.pgen.1008425
909 63. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ
910 preprocessor. Bioinformatics. 2018;34: i884–i890.
911 doi:10.1093/bioinformatics/bty560
912 64. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved
913 microbes from next-generation sequencing data using breseq. In: Sun L, Shou W,
914 editors. Methods Mol Biol. New York, NY: Springer New York; 2014. pp. 165–188.
915 doi:10.1007/978-1-4939-0554-6_12
916 65. Barrick JE, Colburn G, Deatherage DE, Traverse CC, Strand MD, Borges JJ, et
917 al. Identifying structural variation in haploid microbial genomes from short-read

resequencing data using breseq. BMC Genomics. 2014;15: 1039. doi:10.1186/1471-2164-15-1039

66. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357–359. doi:10.1038/nmeth.1923

67. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25: 2078–2079. doi:10.1093/bioinformatics/btp352

68. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10: giab008. doi:10.1093/gigascience/giab008

69. Putri GH, Anders S, Pyl PT, Pimanda JE, Zanini F. Analysing high-throughput sequencing data in Python with HTSeq 2.0. Bioinformatics. 2022;38: 2943–2945. doi:10.1093/bioinformatics/btac166

70. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26: 139–140. doi:10.1093/bioinformatics/btp616

71. Lun ATL, Chen Y, Smyth GK. It's DE-licious: A recipe for differential expression analyses of RNA-seq experiments using quasi-likelihood methods in edgeR. Methods Mol Biol. 2016;1418: 391–416. doi:10.1007/978-1-4939-3578-9_19

72. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015;43: e47–e47. doi:10.1093/nar/gkv007

73. Hughes CS, Foehr S, Garfield DA, Furlong EE, Steinmetz LM, Krijgsveld J. Ultrasensitive proteome analysis using paramagnetic bead technology. Mol Syst Biol. 2014;10: 757. doi:10.15252/msb.20145625

946 **Supporting Information**

947

948 **S1 Fig. Population 2 contains two clades with distinct amplifications.** (A) Read
949 coverage map for population 2 shows 88 and 54 kb amplifications. Red arrows indicate
950 binding sites for primers used to interrogate potential amplification junctions. (B) Colony
951 PCR results for 20 evolved isolates using primers RF133 and RF135. (C) Colony PCR
952 results for 20 evolved isolates using primers RF133 and RF136.

953

954

955 **S2 Fig. The majority of cells in population 3 have an amplification nested inside a**
956 **duplication.** (A) Read coverage map for population 3 shows 76 and 40 kb
957 amplifications. Red arrows indicate binding sites for primers used to interrogate
958 potential amplification junctions. (B) Colony PCR results for 20 evolved isolates using
959 primers RF134 and RF135. (C) Colony PCR results for 20 evolved isolates using
960 primers RF134 and RF137.

961

962

963 **S3 Fig. Second amplification present in population 6.** The 6.5 kb amplification in
964 population 6, which encompasses genes for arginine biosynthesis enzymes and
965 *argC*::*kan*^{R}, was mediated by identical 8 bp sequences in *ppc* and *argH.* Read coverage
966 depth is mapped to genome coordinates of the ancestral *proBA*-rpoS* strain (RF90).

967

968

969 **S4 Fig. *proA** copy number in populations 3 and 8 during laboratory evolution.**
970 *proA** copy number was quantified by qPCR with gDNA isolated from populations 3 and
971 8.
972
973
974 **S5 Fig. Relationship between gene copy number and expression at the RNA and**
975 **protein levels**. (A) Correlation between mRNA fold change and gene copy number. (B)
976 Correlation between protein fold change and gene copy number. (C) Correlation
977 between protein fold change and mRNA fold change. R^2 values were calculated by
978 simple linear regression. Shaded areas indicate 95% confidence intervals. Data points
979 for *hycB* (<5 read counts/replicate) and HycE (<1 peptide/replicate) were removed from
980 graphs and not included in linear regression analysis because their levels in the
981 reference strain were very low and thus their foldchanges were not reliable.
982
983
984 **S6 Fig. Regulation of RcsA and colanic acid polysaccharide synthesis genes is**
985 **altered by segmental amplification.** (A) RcsAB heterodimers activate transcription of
986 *rcsA* (but not *rcsB*) and the colanic acid capsular polysaccharide operon. Lon and
987 HslUV proteases degrade RcsA. The ratio of *rcsA* to *rcsB* transcripts is 1:4 in the
988 ancestral strain. (B) After segmental amplification, overexpression and misfolding of
989 proteins expressed from amplified genes may divert proteases away from degrading
990 RcsA, leading to higher RcsA levels and elevated transcription of colanic acid synthesis
991 genes.

992

993

994

995 **S1 Text.**

996 **Table A. Amplicons present in evolved populations.**

997 **Table B. *E. coli* strains used in this work.**

998 **Table C. Plasmids used in this work.**

999 **Table D. Primers used in this work.**

1000

1001 **S1 Dataset. mRNA and protein fold change data for amplified genes in evolved**

1002 **populations 3 and 8**

1003

1004 **S2 Dataset. Label-free proteomics data and differential expression analysis for all**

1005 **genes in evolved populations 3 and 8.**

1006

1007 **S3 Dataset. RNA-sequencing transcriptomic data and expression analysis for all**

1008 **genes in evolved populations 3 and 8**

1009

1010 Supporting Text for
1011
1012 **The transcriptomic and proteomic ramifications of segmental amplification**
1013
1014

1015 Authors: Ryan K. Fritts and Shelley D. Copley*

1016 *For correspondence: Shelley.Copley@colorado.edu

1017

1018 S1 Text contains Tables A-D

1019

1020 **Table A. Amplicons present in evolved populations.**
1021

pop.	gen.	upstream junction position[a] (gene)	downstream junction position[a] (gene)	IS element involved	amplicon size (kb)	*proA** copy number
1	155	2808170 (*serV*)	2864360 (*ispD*)	*insL1* (IS186)	56.2	7.8
			2896344 (*ygcE/queE)*		88.2	
2	146	2808170 (*serV*)	2862506 (*umpG*)	*insL1* (IS186)	54.3	8.8
			2896283 (*ygcE/queE)*		88.1	
3	198	2820227 (*srlQ*)	2859849 (*nlpD*)	*insAB1* (IS1)	39.6	11.8
			2859931(*nlpD*)		39.7	
			2895963 (*ygcE/queE*)		75.7	
4	182	2851868 (*ygbJ*)	2860822 (*nlpD*)	none	9.0	7.7
		2808170 (*serV*)	2896344 (*ygcE/queE*)	*insL1* (IS186)	88.2	
5	157	2808170 (*serV*)	2861262 (*pcm*)	*insL1* (IS186)	53.1	7.2
6	159	2808170 (*serV*)	2896283 (*ygcE/queE*)	*insL1* (IS186)	88.1	7.0
7	172	2835653 (*hycE*)	2868717 (*cysD/iap*)	*insAB1* (IS1)	33.1	8.5
		2808170 (*serV*)	2896283 (*ygcE/queE*)	*insL1* (IS186)	88.1	
8	186	2808170 (*serV*)	2874835 (*casB*)	none	66.7	5.8
			2933283 (*rlmM*)	*insL1* (IS186)	125.1	

1022 [a]Position numbers based on reference genome for ancestral *proBA*-rpoS* strain (RF90).
1023

1024 **Table B. *E. coli* strains used in this work.**

1025

Strain	Genotype/description	Source/reference
JW3930 (AM008)	Δ*argC*::*kan*^{R} Keio collection strain; derived from *E. coli* BW25113	[1]
AM187	-45 C > T *proB* promoter mutation (M2)*, proA** (E383A), +*yfp* inserted 282 bp downstream *proA*, Δ*fimAICDFGH,* Δ*csgBAC;* derived from JW3930/AM008	[2,3]; Genbank accession number CP037857.2; M2 promoter mutation described in [4]
AM327	Δ82 bp of *rph* 90 bp upstream of *pyrE*, derived from AM187	[2]
RF70	Δ*proBA*-yfp*::*sacB-gtm*^{R}; derived from AM327	this study
RF72	Δ*proBA*-yfp;* derived from RF70	this study
RF83	Δ*proBA*-yfp*, *ygbN-tetR-ccdB-cat-rpoS*; derived from RF72	this study
RF90	Δ*proBA*-yfp,* M2-*proBA*-rpoS;* derived from RF83; ancestral strain for laboratory evolution	this study

1026

1027 **Table C. Plasmids used in this work.**

1028

Plasmid	Description	Source/reference
pJQ200SK	suicide vector containing sucrose-inducible *sacB* and constitutive *gtm*^{R} selection-counterselection cassette	[5]
pDLM3	plasmid containing anhydrotetracycline-inducible *tetR-ccdB-cat* selection-counterselection cassette	[6]
pSIM6	temperature-sensitive plasmid containing heat-inducible λ-Red genes (*exo*, *bet*, *gam*) for recombineering; *amp*^{R}	[7]

1029

1030 **Table D. Primers used in this work.**

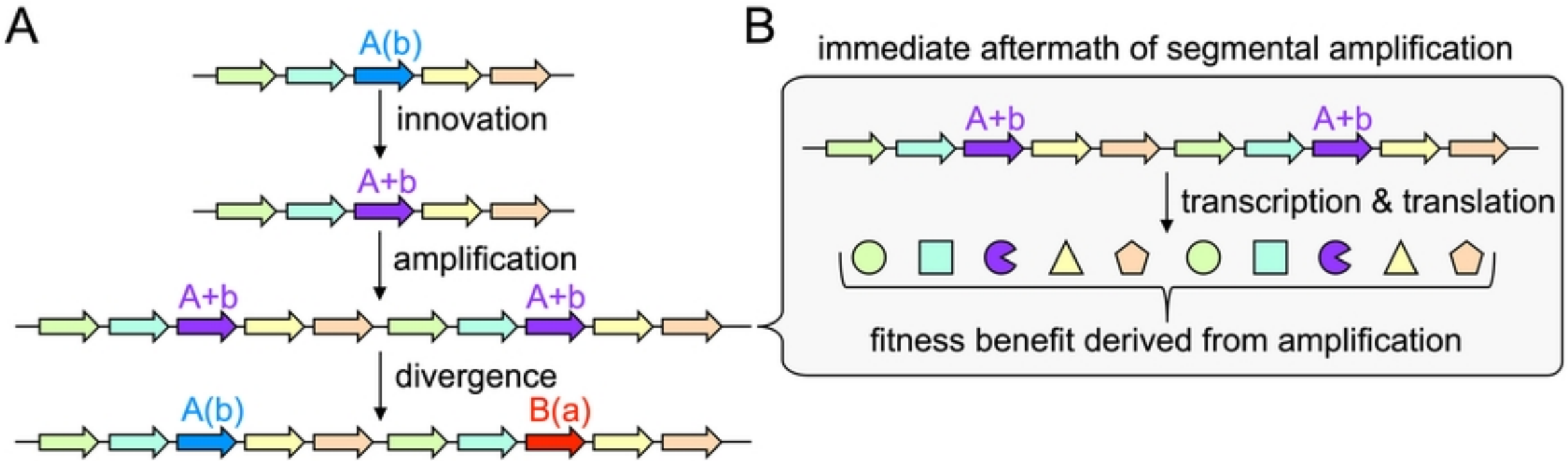
Name	Sequence (5' → 3')[a]	Use
RF35-fwd	CATTACGGATTCACTGGCCG	amplifying *tetR-ccdB-cat* cassette
RF36-rvs	CGGTAAATAGCTTGCCTGCTC	amplifying *tetR-ccdB-cat* cassette
RF61-fwd	CGATCGACTCTAGCTAGAGGATC	amplifying *sacB-gtm*^{R} cassette
RF62-rvs	GAAACGGATGAAGGCACGAAC	amplifying *sacB-gtm*^{R} cassette
RF71-fwd	GGCGTCATTTTGCGTGTTTCAG	amplifying region upstream of *proBA*
R72-rvs	cctctagctagagtcgatcgCAACCATCTGCGAGCAAAGC	amplifying region upstream of *proBA*
RF73-fwd	ttcgtgccttcatccgtttcGGCATGGACGAGCTGTACAAG	amplifying region downstream of *proBA*
RF74-rvs	CCACTCAGCGTTAATTACCTACACG	amplifying region downstream of *proBA*
RF81-rvs	ttgtacagctcgtccatgccCAACCATCTGCGAGCAAAGC	amplifying region upstream of *proBA* to replace *sacB-gtm*^{R} with markerless deletion of *proBA*
RF82-fwd	gctttgctcgcagatggttgGGCATGGACGAGCTGTACAAG	amplifying region upstream of *proBA* to replace *sacB-gtm*^{R} with markerless deletion of *proBA*
RF88-fwd	GCAACAAAACGCCATGCTTT	amplifying *proBA** with mutant M2 promoter
RF89-rvs	TGGCCTTGTGAATTATAAACGC	amplifying *proBA** with mutant M2 promoter
RF91-fwd	GTCGGCTTTATTATTCTTGCGCC	amplifying region upstream of the target site between *rpoS* and *ygbN*
RF97-rvs	cggccagtgaatccgtaatgGGCTGGCCTTTTCTGTGCAC	amplifying region upstream of the target site between *rpoS* and *ygbN*
RF94-fwd	GAAGATGCGGAATTTGATGAGAACG	amplifying region downstream of the target site between *rpoS* and *ygbN*
RF98-rvs	gagcaggcaagctatttaccgTCGCTTGAGACTGGCCTTTCTG	amplifying region downstream of the target site between *rpoS* and *ygbN*
RF99-rvs	aaagcatggcgttttgttgcGGCTGGCCTTTTCTGTGCAC	amplifying region upstream of the target site between *rpoS* and *ygbN* to replace *tet*^{R}-*ccdB-cat* with M2-*proBA**
RF100-	cgtttataattcacaaggccaTCGCTTGAGAC	amplifying region downstream

fwd	TGGCCTTTCTG	of the target site between *rpoS* and *ygbN* to replace *tet*^{R}-*ccdB-cat* with M2-*proBA**
RF133-rvs	CCGCGTCTCATCTTTATCGGTG	checking presence of 54/88 kb amplification in evolved clones from population 2 (binds near upstream junction)
RF134-rvs	GGCGCTATCAGTAATCGCAGC	checking presence of 39/76 kb amplification in evolved clones from population 3 (binds near upstream junction)
RF135-fwd	TCGTATGAAGTGATGCATGGGG	checking presence of 88 kb or 76 kb amplification in evolved clones from populations 2 or 3, respectively (binds near downstream junction)
RF136-fwd	CGTCACGCGAATACCTTTGATTTG	checking presence of 54 kb amplification in evolved clones from population 2 (binds near downstream junction)
RF137-fwd	CACCAGGTTGCGTATGTTGAGAAG	checking presence of 39 kb amplification in evolved clones from population 3 (binds near downstream junction)
RF138-fwd	GCAAACGTGGTAGAGAAGGCAC	amplifying *prlF-yhaV*
RF139-rvs	GTCAGTCGAGCCATCAGTGGAAG	amplifying *prlF-yhaV*
RF140-fwd	GGCTGTAAAAGGACAGTGAATCATG	sequencing *prlF*
RF141-rvs	GCTGAATACGGTATAGGCATCTG	sequencing *prlF*
1357-fwd	CGTGCAGGCGATTGATAA	quantifying *proA** copy number by qPCR
1358-rvs	CTGTTCACGGCACAGTTT	quantifying *proA** copy number by qPCR
1359-fwd	CGTAGATCTGACGGTGAATTT	quantifying *gyrB* copy number by qPCR
1360-rvs	CGTTGGTGTTTCGGTAGTA	quantifying *gyrB* copy number by qPCR
1361-fwd	CCCGTGGCTGAAAGTTAAA	quantifying *icd* copy number by qPCR
1362-rvs	CAGGTTCATACAGGCGATAAC	quantifying *icd* copy number by qPCR

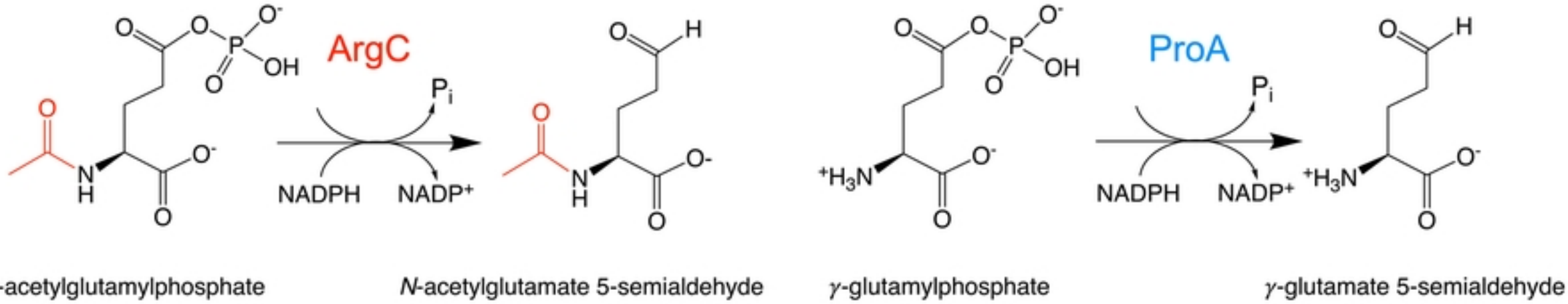
1031 [a]lowercase letters indicate non-annealing sequence overlaps used for splicing by
1032 overlap extension PCR

References

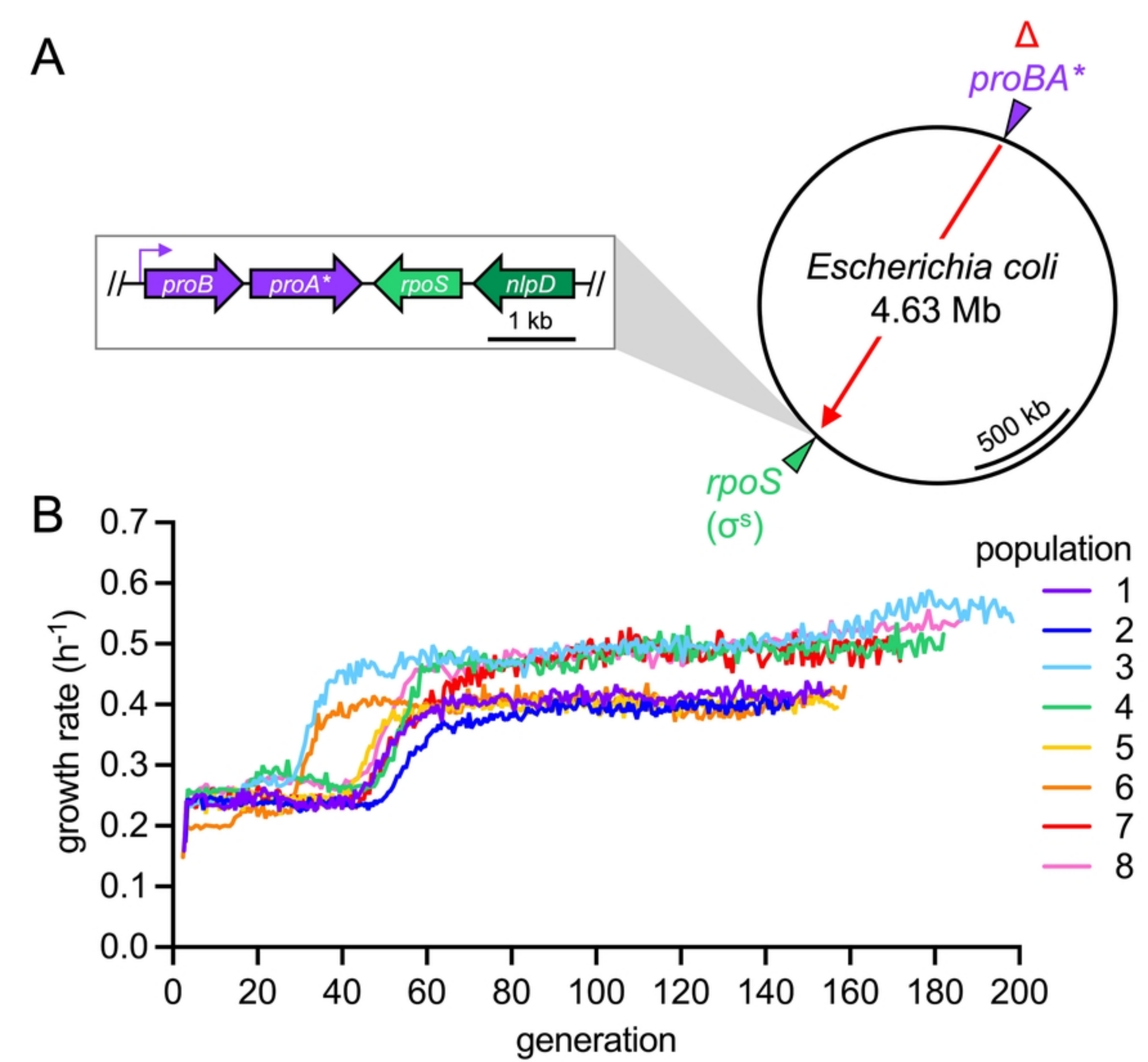
1. Baba T, Ara T, Hasegawa M, Takai Y, Okumura Y, Baba M, et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. Mol Syst Biol. 2006;2: 2006.0008. doi:10.1038/msb4100050
2. Morgenthaler AB, Kinney WR, Ebmeier CC, Walsh CM, Snyder DJ, Cooper VS, et al. Mutations that improve efficiency of a weak-link enzyme are rare compared to adaptive mutations elsewhere in the genome. eLife. 2019;8: e53535. doi:10.7554/eLife.53535
3. Morgenthaler AB, Fritts RK, Copley SD. Amplicon remodeling and genomic mutations drive population dynamics after segmental amplification. Mol Biol Evol. 2022;39: msab289. doi:10.1093/molbev/msab289
4. Kershner JP, Yu McLoughlin S, Kim J, Morgenthaler A, Ebmeier CC, Old WM, et al. A synonymous mutation upstream of the gene encoding a weak-link enzyme causes an ultrasensitive response in growth rate. J Bacteriol. 2016;198:2853–2863. doi:10.1128/JB.00262-16
5. Quandt J, Hynes MF. Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. Gene. 1993;127:15–21. doi:https://doi.org/10.1016/0378-1119(93)90611-6
6. Bossi L, Ratel M, Laurent C, Kerboriou P, Camilli A, Eveno E, et al. NusG prevents transcriptional invasion of H-NS-silenced genes. PLoS Genet. 2019;15: e1008425. doi:https://doi.org/10.1371/journal.pgen.1008425
7. Datta S, Costantino N, Court DL. A set of recombineering plasmids for gram-negative bacteria. Gene. 2006;379: 109–115. doi:https://doi.org/10.1016/j.gene.2006.04.018


A
A(b)
innovation
A+b
amplification
A+b
A+b
divergence
A(b)
B(a)
B
immediate aftermath of segmental amplification
A+b
A+b
transcription & translation
fitness benefit derived from amplification

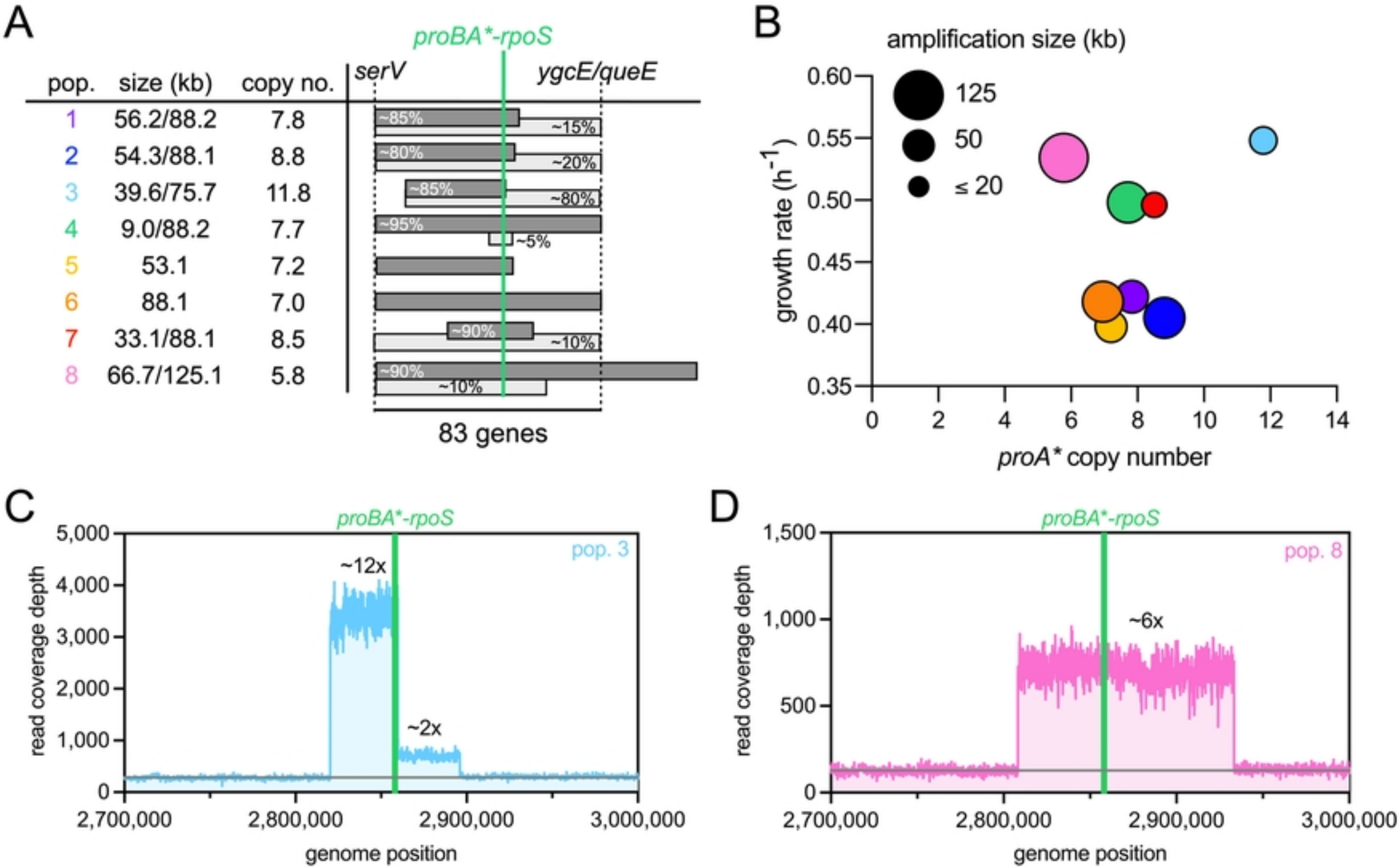
Figure


ArgC
Pi
NADPH
NADP+
ProA
Pi
NADPH
NADP+
N-acetylglutamylphosphate
N-acetylglutamate 5-semialdehyde
γ-glutamylphosphate
γ-glutamate 5-semialdehyde

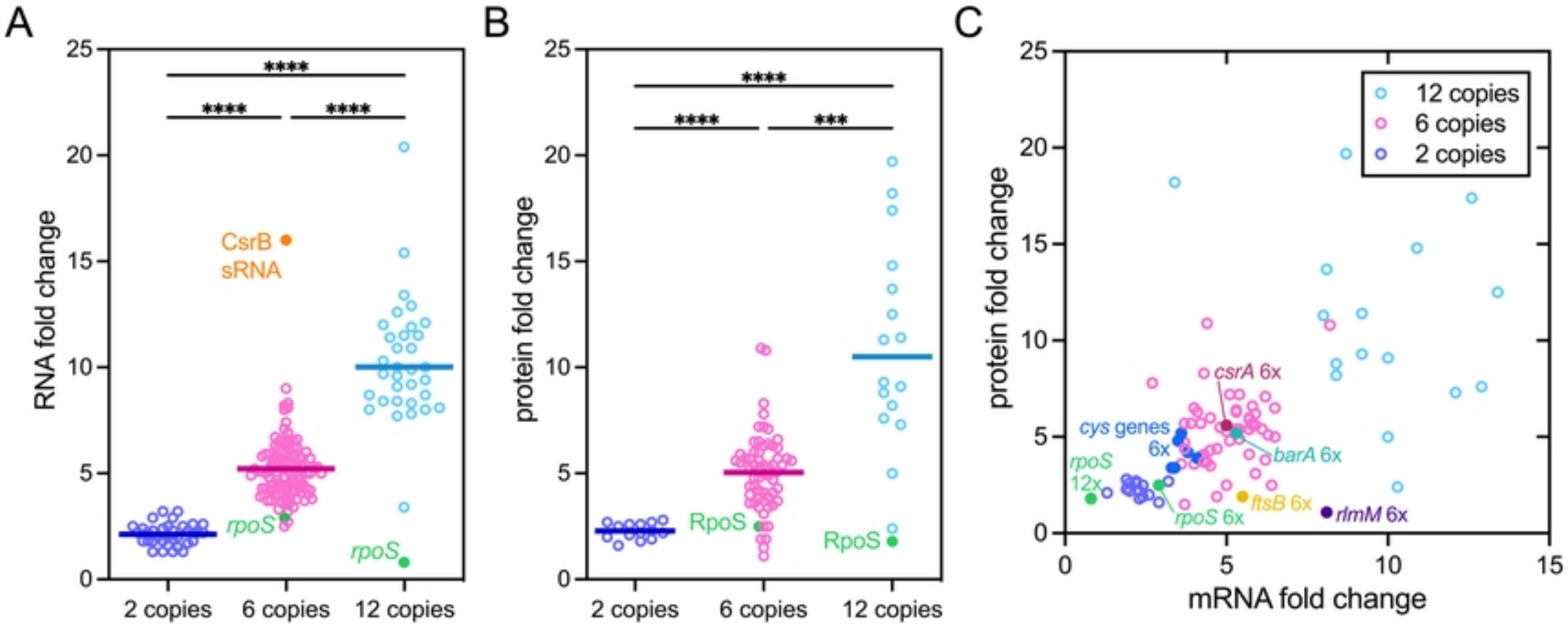
Figure


A
proB
proA*
rpoS
nlpD
1 kb
Δ
proBA*
Escherichia coli
4.63 Mb
500 kb
rpoS
(σs)
B
growth rate (h-1)
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0
20
40
60
80
100
120
140
160
180
200
generation
population
1
2
3
4
5
6
7
8

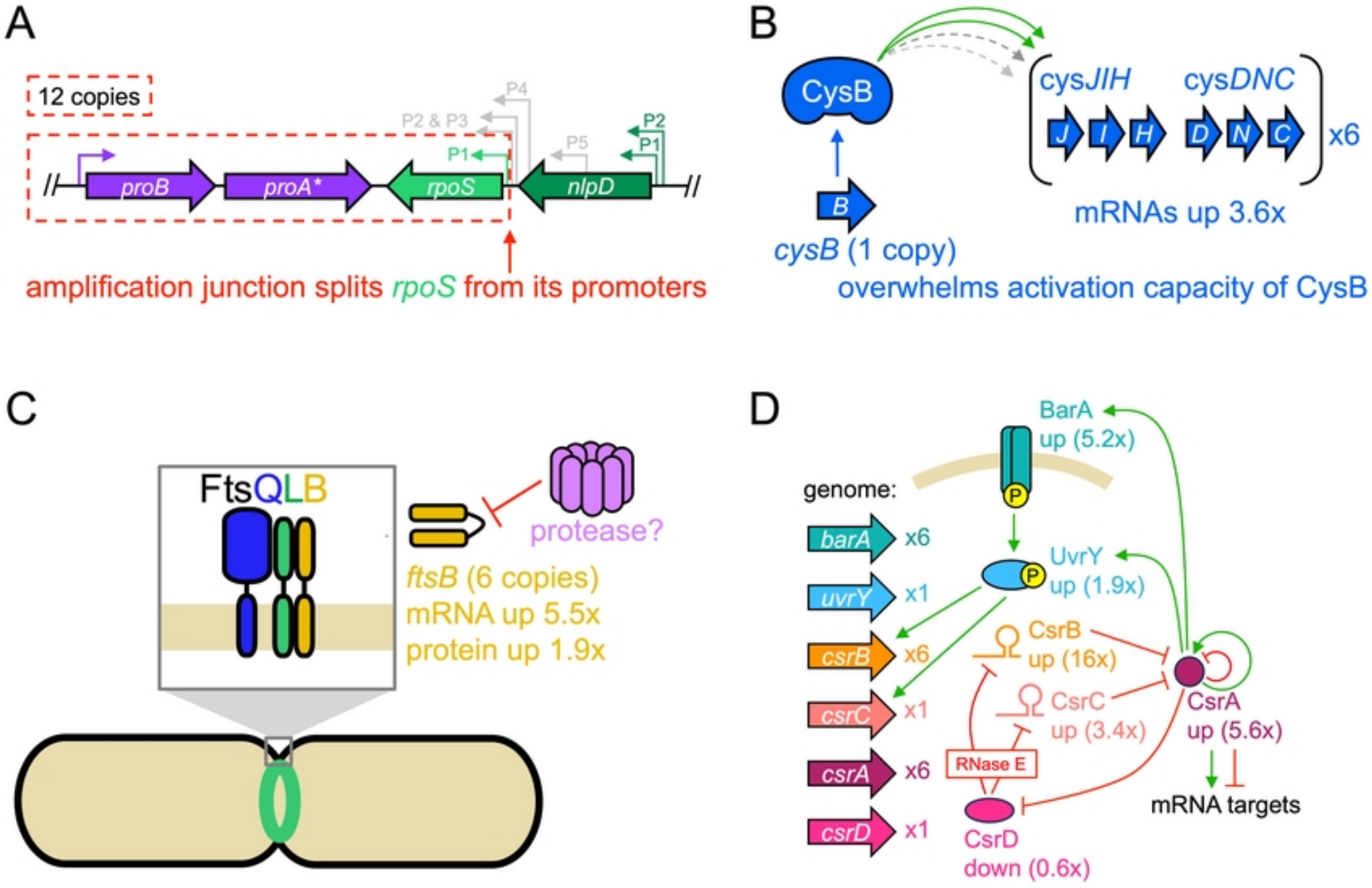
Figure


A
proBA*-rpoS
serV
ygcE/queE
pop.
size (kb)
copy no.
1 56.2/88.2 7.8
2 54.3/88.1 8.8
3 39.6/75.7 11.8
4 9.0/88.2 7.7
5 53.1 7.2
6 88.1 7.0
7 33.1/88.1 8.5
8 66.7/125.1 5.8
~85%
~15%
~80%
~20%
~85%
~80%
~95%
~5%
~90%
~10%
~90%
~10%
83 genes
B
amplification size (kb)
125
50
≤ 20
growth rate (h-1)
0.60
0.55
0.50
0.45
0.40
0.35
0
2
4
6
8
10
12
14
proA* copy number
C
proBA*-rpoS
pop. 3
~12x
~2x
read coverage depth
5,000
4,000
3,000
2,000
1,000
0
2,700,000
2,800,000
2,900,000
3,000,000
genome position
D
proBA*-rpoS
pop. 8
~6x
read coverage depth
1,500
1,000
500
0
2,700,000
2,800,000
2,900,000
3,000,000
genome position

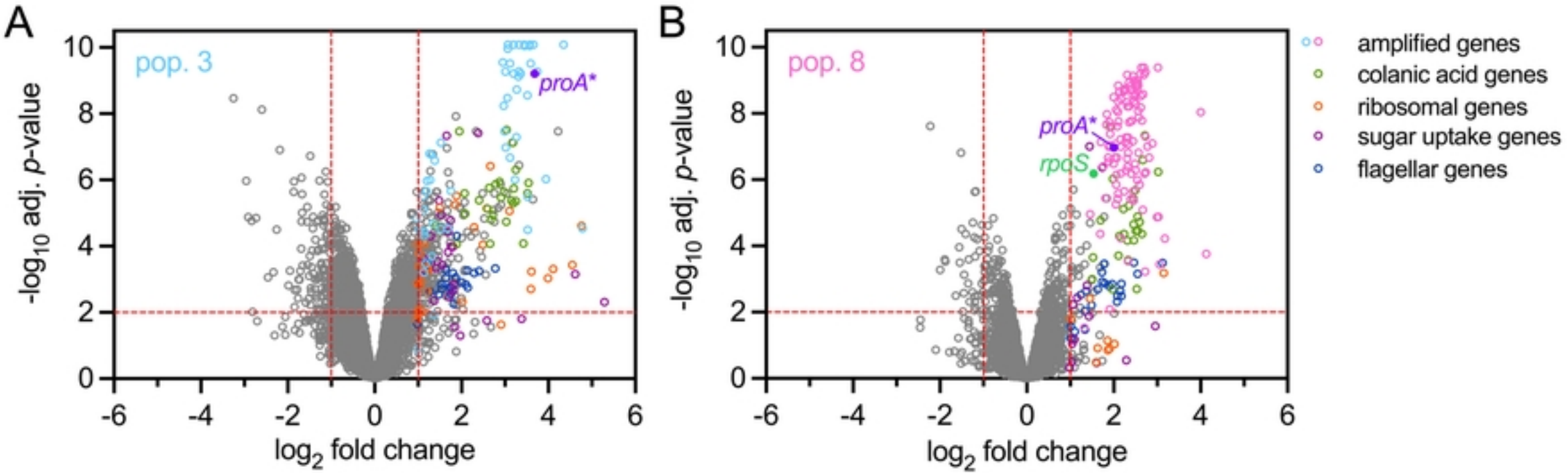
Figure


A
B
C
RNA fold change
protein fold change
mRNA fold change
2 copies
6 copies
12 copies
****
***
CsrB
sRNA
rpoS
RpoS
12 copies
6 copies
2 copies
csrA 6x
cys genes
6x
barA 6x
rpoS
12x
ftsB 6x
rlmM 6x
rpoS 6x

Figure


A
12 copies
P4
P2 & P3
P2
P1
P5
P1
proB
proA*
rpoS
nlpD
amplification junction splits rpoS from its promoters
B
CysB
cysJIH
cysDNC
J
I
H
D
N
C
x6
B
mRNAs up 3.6x
cysB (1 copy)
overwhelms activation capacity of CysB
C
FtsQLB
protease?
ftsB (6 copies)
mRNA up 5.5x
protein up 1.9x
D
BarA
up (5.2x)
genome:
barA
x6
uvrY
x1
csrB
x6
csrC
x1
csrA
x6
csrD
x1
UvrY
up (1.9x)
CsrB
up (16x)
CsrC
up (3.4x)
RNase E
CsrA
up (5.6x)
mRNA targets
CsrD
down (0.6x)

Figure


A
pop. 3
proA*
-log10 adj. p-value
log2 fold change
B
pop. 8
proA*
rpoS
amplified genes
colanic acid genes
ribosomal genes
sugar uptake genes
flagellar genes

Figure