Main Manuscript for

Title: The transcriptomic and proteomic ramifications of segmental amplification

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

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

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

[1]To whom correspondence may be addressed. Email: shelley.copley@colorado.edu.

Competing interests statement: The authors declare no competing interests.

Classification: Biological Sciences (major); Evolution (minor)

Keywords: gene duplication, segmental amplification, microbial evolution, ProA, Lon protease

This PDF includes:

Main Text

Figures 1 to 8

Abstract

Gene amplification can drive adaptation by rapidly increasing the cellular dosage of critical gene products. Segmental amplifications often encompass large genomic regions surrounding the gene(s) under selection for higher dosage. Overexpression of co-amplified neighboring genes imposes a substantial 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 modulate 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 arose in all populations. Amplicons ranged in size (9 to 125 kb) and copy number (2 to 12 copies). We performed RNA-seq and label-free proteomics to quantify expression of amplified genes present at 2, 6, and 12 copies. mRNA expression generally scales with gene copy number, but protein expression scales less well with both gene copy number and mRNA expression. We characterize the molecular mechanisms underlying discrepancies between gene copy number and expression for several cases. We also show that segmental amplifications can have system-wide consequences by indirectly altering expression of non-amplified genes. Our findings indicate that the fitness benefit derived from segmental amplification depends on the combined effects of amplification size, gene content, and copy number as well as collateral effects on non-amplified genes.

Significance Statement

Gene amplification frequently drives rapid adaptation when organisms are challenged by harsh conditions. However, gene amplification rarely amplifies only the gene(s) under selection for higher dosage. Segmental amplifications often include many co-amplified neighbor genes. Because segmental amplifications can reach very high copy number, overexpression of co-amplified genes can impose an energetic burden and perturb physiology, yet this aspect of gene amplification has received little attention. Here, we performed transcriptomic and proteomic analysis of laboratory-evolved *Escherichia coli* strains containing large, high-copy-number amplifications. We found that mRNA levels, but often not protein

50 levels, scale well with gene copy number. We identified amplified genes that exhibit discrepancies
51 between gene copy number and mRNA/protein expression and examine the molecular mechanisms
52 underlying these discrepancies.
53

Main Text

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 (1–10) and eukaryotes (11–18) and is implicated in the development and proliferation of cancer cells (19–23).

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, 24) (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 newly advantageous but weak 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 (25–29).

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 reach above 10 copies and/or span over 100 kb during adaptation to stresses such as starvation or antimicrobial exposure (4–10, 14, 18, 30, 31). Despite the evolutionary importance of gene amplification, little attention has been given to the consequences of co-amplifying neighboring genes (Fig. 1B). Co-amplification of neighboring genes would be expected to inappropriately increase levels of the encoded proteins and cause physiological perturbations and/or energetic burdens. These collateral effects could have important repercussions on population dynamics in communities of organisms that are competing to adapt in the face of an environmental challenge or opportunity.

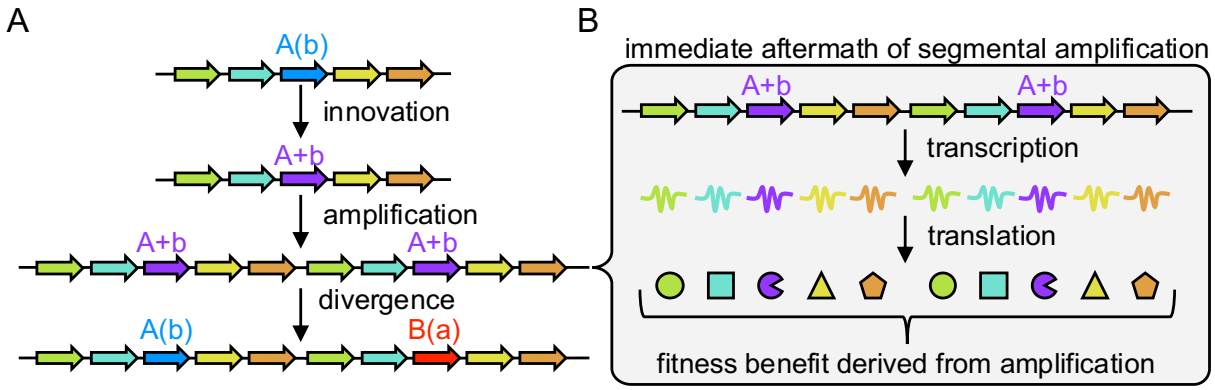


Fig. 1. The expression and function of co-amplified neighboring genes can affect the fitness benefit derived from segmental amplification. (A) The Innovation-Amplification-Divergence (IAD) model. Capital letters, major function; lower-case letters, inefficient function; parentheses, promiscuous function. (B) The transcription and translation of all amplified genes determines the overall fitness benefit gained in the immediate aftermath of segmental amplification.

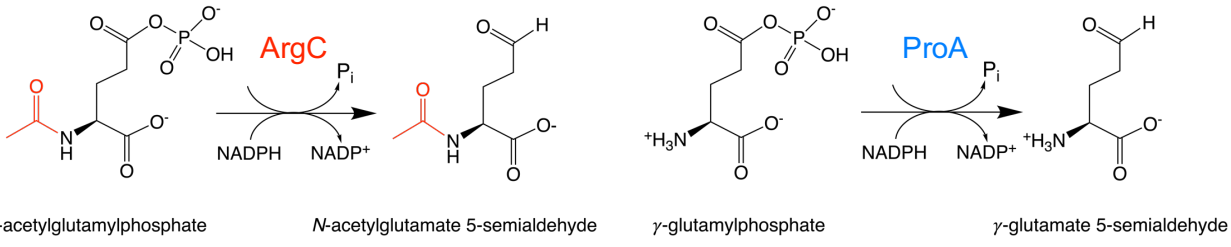
Previous investigations of the levels of mRNAs and proteins produced from duplicated/amplified genes in aneuploid yeast strains with one or two extra copies of whole chromosomes (32–37) and in cancer cells (38, 39) have shown that mRNA and, to a lesser extent, protein levels often, but not always, scale with gene copy number. Alterations in expression of co-amplified genes in these settings may be due to either intrinsic regulatory mechanisms that counteract inappropriate gene expression or to compensatory mutations that arose over many generations. A large-scale examination of hundreds of natural aneuploid yeast isolates found that most proteins expressed from aneuploid chromosomes exhibited attenuated levels (37), a phenomenon termed dosage compensation. Because these stable aneuploidies present in natural yeasts are likely long-standing adaptations, there has been time for multiple compensatory mutations to arise to counteract the most detrimental consequences of aneuploidy and alter the proteome of these isolates. Another recent study assessing the expression of individually duplicated yeast genes similarly found that protein levels were attenuated in some cases, although the precise molecular mechanisms that dampened protein expression were usually not clear (40). The expression of high-copy-number co-amplified genes in the immediate aftermath of segmental amplification is largely unknown. Consequently, we lack a broad understanding of the intrinsic capacity of

104 regulatory mechanisms to buffer against the collateral effects of segmental amplification. These
105 regulatory mechanisms can operate at the transcriptional, posttranscriptional and posttranslational levels
106 to attenuate expression. The level at which such regulation occurs affects the energetic burden incurred
107 following segmental amplification. Repressing the transcription of a high-copy-number operon is obviously
108 less energetically costly than degrading the encoded proteins following translation. However, the findings
109 of several studies indicate that transcript levels generally correspond to gene copy number and that most
110 dosage compensation happens at the posttranscriptional and posttranslational levels (17, 33–37, 41)

111 We have developed a model system in *Escherichia coli* to characterize the immediate
112 ramifications of segmental amplification on the transcriptome and proteome. The *E. coli* enzymes ArgC
113 and ProA are nonhomologous reductases that catalyze similar reactions in arginine and proline
114 biosynthesis, respectively (Fig. 2). We previously deleted *argC* from the *E. coli* genome and selected for
115 recruitment of another enzyme to replace the missing ArgC activity during growth on minimal medium
116 (42). A strain with a point mutation in *proA*, which causes substitution of Glu383 to Ala in the active site of
117 ProA, was able to restore arginine production. E383A ProA (ProA*) catalyzes both its native reaction in
118 proline synthesis and the weak promiscuous activity required for arginine synthesis (Fig. 2). The poor
119 activity of ProA* compromises growth in minimal media in which synthesis of proline and arginine is
120 required for growth, leading to strong selective pressure for amplification of *proA*.* We previously
121 observed segmental amplifications encompassing *proA** ranging in size from 5-165 kb and reaching copy
122 numbers between 3-20 copies in experimentally evolved *E. coli* populations (4, 43).

123

124



125

126 **Fig. 2. Reactions catalyzed by the reductases ArgC and ProA in *E. coli*.** The acetyl group of *N*-
127 acetylglutamylphosphate (red) is the only difference between the two substrates.

128

129 Many of the genes neighboring the *proBA** operon are pseudogenes or uncharacterized cryptic
130 prophage genes. Thus, amplification of *proA** in its native locus does not provide an ideal system for
131 assessing the effects of co-amplification of genes near the gene under selection for higher copy number.
132 For this work, we relocated the *proBA** operon to a new site neighboring a gene that we hypothesized
133 might cause physiological perturbations if overexpressed following segmental amplification. We evolved
134 eight replicate populations of this *E. coli* strain for 145-200 generations in a turbidostat. Whole genome
135 sequencing revealed that all populations contained segmental amplifications of differing sizes and copy
136 numbers. We examined the transcriptomes and proteomes of two populations, those with the highest
137 copy number and largest amplifications, to determine the collateral effects of segmental amplification. The
138 expression of many amplified genes scales with gene copy number at the mRNA and protein levels.
139 However, in several cases, mRNA and/or protein levels either exceed or fall short of the level expected
140 based solely on gene copy number. We have examined the molecular mechanisms for several cases in
141 which such discrepancies exist. We also found significant changes in levels of hundreds of non-amplified
142 genes, demonstrating the profound and global impact of segmental amplification on cellular physiology.
143 Through this survey of the system-wide consequences of segmental amplification, we identified a
144 compensatory mutation that appears to prevent overexpression of unnecessary proteins, which may
145 represent a broad class of adaptations to proteotoxic stress following gene amplification.
146

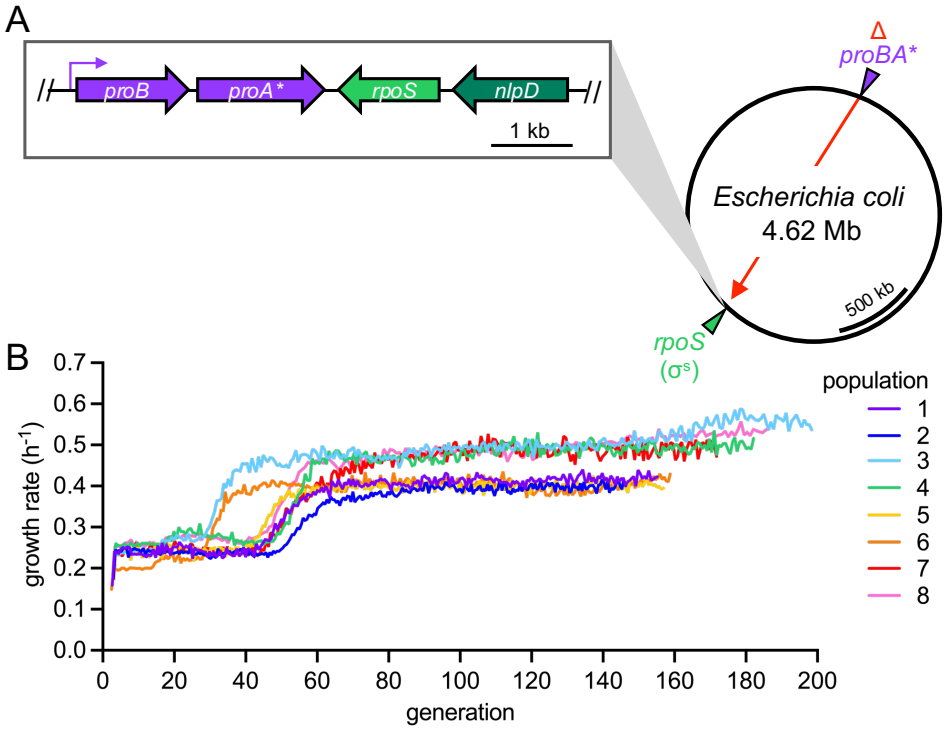
147 ## Results and Discussion

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

150 We deleted the *proBA** operon from its native locus and relocated it next to *rpoS,* which encodes
151 the alternative sigma factor RpoS/σ^S (also known as σ^{38}) (Fig. 3A). We included the upstream region
152 containing a C to T promoter mutation 45 bp upstream of *proB* that increases transcription of *proBA** (44).
153 The *proBA** operon was oriented convergently to *rpoS* so that transcriptional read-through from *proBA**
154 would not increase *rpoS* expression. A dual *rrnB* T1/T7 TE transcriptional terminator was placed
155 immediately after *proA**. The alternative sigma factor RpoS serves as a global regulator by interacting
156 with RNA polymerase to activate transcription of numerous genes important for redirecting cellular

157 resources from growth toward survival (45, 46). Consequently, expression of RpoS is typically associated
158 with a slowing of growth during stressful conditions or limited nutrient availability. Elevated expression of
159 genes in the RpoS regulon confers improved resistance to starvation, high temperatures, low pH,
160 oxidative stress and osmotic shock (45, 46). We hypothesized that overexpression of RpoS following co-
161 amplification with *proA** might perturb cellular physiology because it would shift cells toward a
162 transcriptional profile intended as a stress response to halt growth while the cells are growing
163 exponentially with adequate nutrients.

164 We used a single founder colony of the Δ*argC proBA** relocation strain to start eight replicate
165 populations in a turbidostat. The cultures were grown in M9 minimal medium with glucose as the sole
166 carbon source to impose strong selection for mutations that increase ProA* dosage. After 30-60
167 generations, we observed improvements in growth rate for all populations (Fig. 3B). Populations 3 and 8
168 showed an additional slight increase in growth rate around generation 180. We ended the experiment
169 after 145-200 generations to limit the opportunity for compensatory mutations that might alleviate harmful
170 effects of segmental amplification to arise. Similarly, we wanted to limit the chance for amplicon
171 remodeling to remove unnecessary copies of amplified genes (43).
172



173

174 **Fig. 3. Evolution of an *E. coli* strain with the *proBA** operon relocated next to *rpoS.*** (A) The *proBA**
175 operon with a mutant constitutive promoter was deleted from its native locus and relocated directly next to
176 the *rpoS* gene. (B) Growth rate increased during laboratory evolution of the *proBA**-*rpoS* relocation strain
177 in minimal medium in a turbidostat.

178

179 We sequenced genomic DNA (gDNA) isolated from each population at the end of the evolution
180 experiment. All populations had segmental amplifications encompassing *proBA** and *rpoS.* The amplicons
181 ranged in size from 9-125 kb and in copy number from 2-12 copies (Fig. 4A and B). Copy numbers were
182 confirmed by quantitative PCR (qPCR) using primers targeting *proA**. Most of the segmental
183 amplifications were mediated by the transposable insertion sequence elements IS186 and IS1, leading to
184 similar amplification junctions across populations. IS186 transposition into *serV* served as an upstream
185 junction point in 7 of the 8 populations and IS186 transposition into the intergenic region between *ygcE*
186 and *queE* served as a downstream junction in 6 of the 8 populations (Fig. 4A and *SI Appendix,* Table S1).
187 IS1 transposition was detected at amplification junctions in populations 3 and 7 (*SI Appendix,* Table S1).

188 As neither IS element was present at these sites in the ancestral strain, it appears that these sites may be
189 hotspots for IS element insertion.
190 We detected two amplicons at frequencies ≥5% in 6 of the 8 populations based on the genome
191 sequencing data (Fig. 4A). The presence of two amplicons could be due to two competing clades with
192 distinct amplicons in the population or to an amplicon nested inside a larger amplicon in a single clade.
193 Populations 2 and 3 stood out as being the most likely to contain a nested amplicon due to the detection
194 of multiple (≥3) amplification junctions and the presence of amplified regions with distinct copy numbers
195 based on the read coverage data (*SI Appendix,* Fig. S1A and S2A). Thus, we sought to experimentally
196 discern whether there were nested amplicons or competing clades with different amplicons in populations
197 2 and 3. To distinguish between these possibilities, we carried out PCR reactions designed to amplify
198 products that span putative amplification junctions (*SI Appendix,* Fig. S1 and S2). Population 2 appears to
199 have two distinct clades; almost all colonies (19 out of 20) had either the ~54 kb (4/19) or the ~88 kb
200 amplicon (15/19) (*SI Appendix,* Fig. S1). Only one colony (1/20) had both amplicons, suggesting that if a
201 clade with a nested amplicon exists in population 2, it is present at a low frequency. In contrast, the
202 majority of colonies isolated from population 3 (13 out of 20) showed PCR products for both the ~40 and
203 ~76 kb amplicons, indicating that most cells contained a smaller approximately 12-copy amplification
204 nested inside a larger duplication (Fig. 4C and *SI Appendix,* Fig. S2). Of the remaining isolates tested
205 from population 3, four had only the ~40 kb amplicon, while three had only the ~76 kb amplicon.
206 We did not perform diagnostic PCRs for the remaining populations with clear high-frequency and
207 low-frequency amplicons based on the read coverage data. In populations 1 and 7, we suspect that low-
208 frequency clades with ~88 kb amplicons are being outcompeted by clades with smaller amplicons (~56
209 and ~33 kb). In populations 4 and 8, low-frequency clades with ~9 and ~67 kb amplicons, respectively,
210 might have arisen more recently than the clades with larger amplicons (~88 and 125 kb) and thus might
211 still represent a small fraction of the population. Alternatively, the ~9 and ~67 kb amplicons might have
212 arisen by remodeling of the larger amplicons in clades that are now in the process of overtaking the
213 original clades.
214

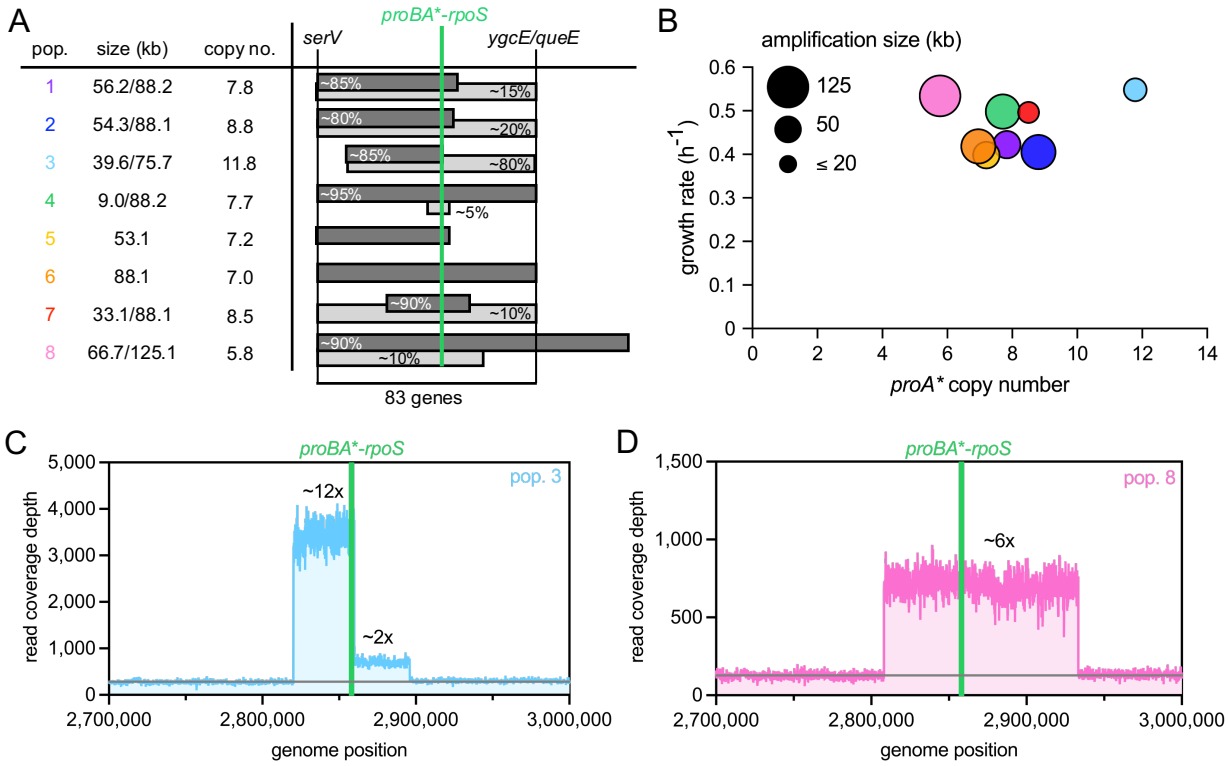


Fig. 4. Segmental amplifications occurred during evolution of all *proBA*-rpoS* relocation populations. (A) Sizes and copy numbers of amplicons detected in the eight populations. In populations with two amplicons, dark gray bars represent the higher-frequency amplicon and light gray bars represent the lower-frequency amplicon. Only amplicons estimated to be present at population frequencies of ≥5% are shown. Population-level amplicon frequencies were estimated based on diagnostic PCR for populations 2 and 3 and read coverage data for the remaining six populations. (B) Growth rate, *proA** copy number, and amplification size for each population at the end of the evolution experiment. Only the most abundant amplicon is shown for each population for simplicity. (C) The nested amplicons in population 3. (D) The 125 kb amplicon in population 8.

We focused on populations 3 and 8 for transcriptomic and proteomic analysis because they possessed the highest-copy-number and largest amplifications, respectively (Fig. 4C and D). Most cells in population 3 appear to have a ~40 kb amplification present at 12 copies nested inside a ~76 kb duplication (Fig. 4C), while population 8 primarily consists of a clade with a 6-fold amplification of a 125 kb region, the largest amplicon we observed (Fig. 4D). We detected a putative compensatory mutation

231 that was fixed in population 8 at the end of the evolution experiment, which will be discussed in-depth
232 later. Because our goal was to characterize the immediate effects of segmental amplification on the
233 transcriptome and proteome in the absence of compensatory mutations, we sequenced population 8
234 samples at earlier timepoints and determined that this mutation was not present in any isolates screened
235 at generation 125 (47, 48).
236
237 **Expression of most amplified genes is elevated in the aftermath of segmental amplification**
238 We performed RNA-seq and label-free quantitative proteomics on replicate cultures started from
239 single colonies isolated from population 3 at generation 199 and from population 8 at generation 125. We
240 verified that the population 3 isolate contained the ~40 kb amplification nested inside a larger ~76 kb
241 duplication by PCR using primers near the amplification junctions. Transcriptomic and proteomic analyses
242 of this isolate allowed us to characterize how mRNA and protein levels changed at two different
243 amplification levels, 2 and 12 copies, within the same genome. The ancestral *rpoS-proBA** relocation
244 strain was used as the reference strain for determining expression fold-changes.
245 Levels of mRNAs and small regulatory RNAs (sRNAs) expressed from amplified genes generally
246 scale linearly with gene copy number (Fig. 5A and *SI Appendix,* Fig. S3, R^2 = 0.63). On average, RNA
247 levels increased 2.1 ± 0.1-, 5.2 ± 0.2-, and 10.0 ± 0.6-fold (mean ± SEM) for genes present at 2, 6 and 12
248 copies, respectively (Fig. 5A). Protein levels increased on average by 2.3 ± 0.1-, 5.0 ± 0.2-, and 10.5 ±
249 1.3-fold for genes present at 2, 6, and 12 copies, respectively (Fig. 5B). However, protein fold-change
250 values varied widely, especially for 6- and 12-copy genes. Levels of proteins expressed from amplified
251 genes correlate relatively weakly with both gene copy number (R^2 = 0.47) and mRNA level (R^2 = 0.35) (*SI*
252 *Appendix,* Fig. S3). Similar results have been observed in aneuploid yeast (37, 47). To some extent the
253 poor correlation may be because estimation of protein abundance by mass spectrometry is less precise
254 than estimation of RNA levels by RNA-seq. However, in many cases, posttranscriptional and
255 posttranslational regulatory mechanisms likely modulate RNA stability, translation and protein degradation
256 in ways that attenuate the impact of increased gene copy number.
257

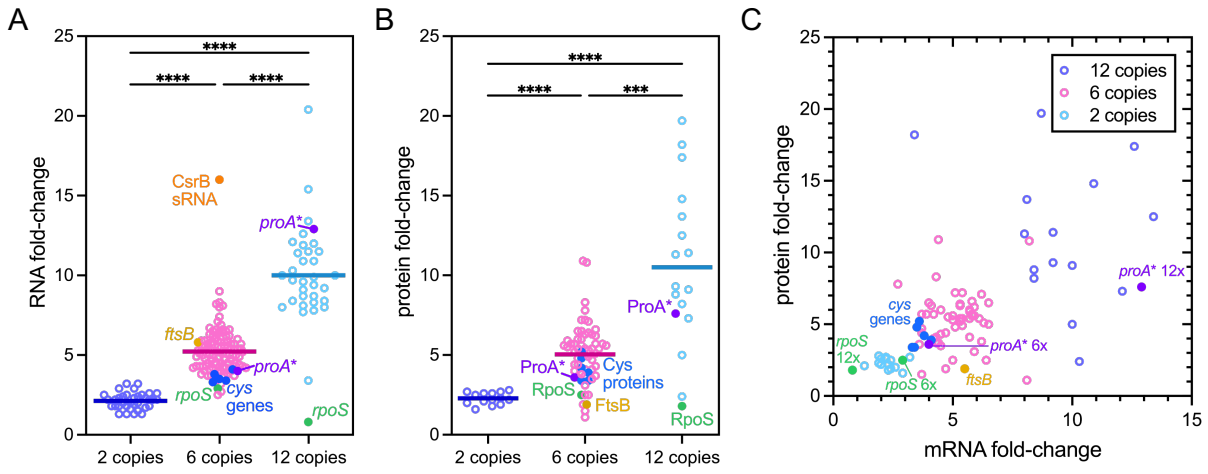


Fig. 5. Effect of segmental amplification on expression of RNA and protein from amplified genes. (A) Fold-change of RNAs expressed from amplified genes. (B) Fold-change of proteins expressed from amplified genes. (C) Protein fold-change as a function of mRNA fold-change for amplified genes. Asterisks indicate statistically significant differences between groups based on a non-parametric Kruskal-Wallis test with Dunn's multiple comparisons test (*** p <0.001; **** p <0.0001). Amplified genes that are discussed in the main text are indicated by different colors.

Expression of *rpoS* is attenuated in both populations 3 and 8

The most glaring discrepancy we observed between gene copy number and mRNA and protein levels is the minimal change in *rpoS* mRNA and RpoS protein when *rpoS* is present at 12 copies in population 3 (Fig. 5A and B). The junction between the smaller amplicon and the larger duplicated region in population 3 (Fig. 4C) lies between *rpoS* and *nlpD*; the nested amplification separates *rpoS* from six of seven potential promoters by moving them hundreds of kb upstream of the *rpoS* start codon (Fig. 6A). Thus, population 3 avoids problematic overexpression of *rpoS* even when the gene copy number is very high. Expression of *rpoS* is also not proportional to the 6 copies of the gene in population 8 (Fig. 5C). RpoS levels are tightly regulated by a combination of multiple transcriptional regulators, sRNAs, and an adaptor-mediated proteolysis pathway (45). The combined activity of these regulatory mechanisms may keep RpoS levels in check in the absence of the canonical stresses to which it normally responds.

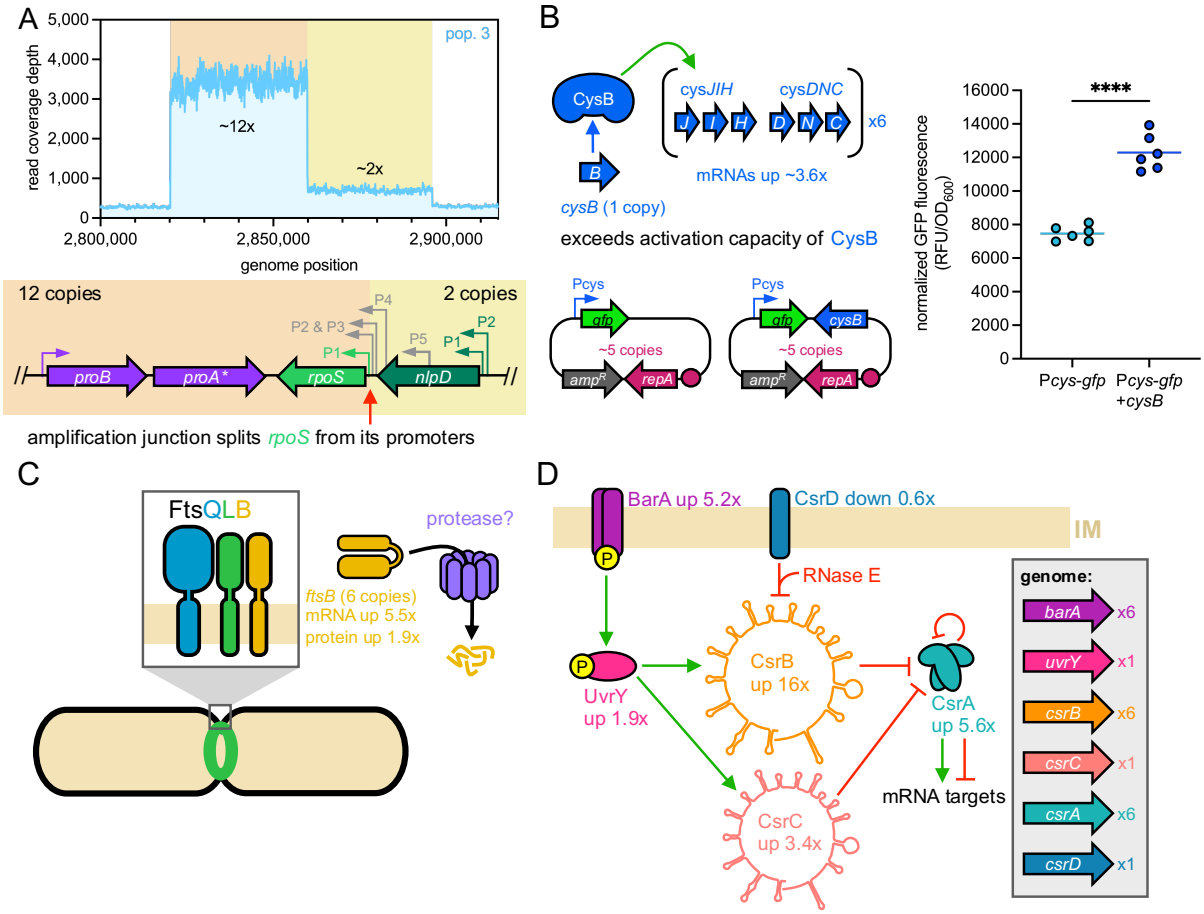


Fig. 6. Mechanisms that lead to discrepancies between gene copy number and transcript or protein levels. (A) The nested amplification junction present in population 3 separates *rpoS* from 6 of 7 potential promoters. (B) Normalized GFP fluorescence for cells carrying a plasmid with a copy number of 5 encoding *gfp* under control of a CysB-activated promoter with and without *cysB* on the same plasmid. Asterisks indicate statistically significant differences between groups based on an unpaired *t* test (**** $p < 0.0001$); n=6 biological replicates. (C) Lower-than-expected FtsB protein levels result from decreased FtsB stability and thus more degradation in the absence of enough of its interaction partners FtsQ and FtsL. (D) Disproportionate expression of genes involved in the Csr system results from both amplification of some but not all genes and complex interactions between RNA and protein components of the circuit.

Amplification of *cysJIH-DNC* outstrips the availability of the transcriptional activator CysB

The *cysJIH* and *cysDNC* operons, which encode proteins involved in sulfate assimilation, were amplified to 6 copies in population 8, but levels of the encoded mRNAs increased by an average of only

292 3.6-fold (Fig. 5C and Dataset S1). Expression of these operons requires the transcriptional activator
293 CysB, which is encoded outside the amplified region and is thus present at only a single copy (Fig. 6B).
294 We hypothesized that the relatively modest transcriptional response of the *cysJIH* and *cysDNC* operons
295 following amplification was due to insufficient CysB to initiate transcription at all of the extra copies of the
296 *cysDNC* and *cysJIH* operons. To test this hypothesis, we constructed a plasmid-based transcriptional
297 reporter system with GFP expression under the control of a CysB-activated promoter (the promoter of the
298 *cysDNC* operon) (Fig. 6B). We constructed two versions of this low-copy-number reporter plasmid (~5
299 copies per cell), one lacking *cysB* and another encoding *cysB.* Cells harboring the reporter plasmid
300 lacking *cysB* rely on the single chromosomal copy of *cysB* to activate *gfp* transcription. Conversely, cells
301 harboring the plasmid containing *cysB* have the chromosomal copy of *cysB* along with ~5 extra copies of
302 *cysB* from the plasmid*,* a scenario in which the copy number of *cysB* is commensurate with the copy
303 number of the *cysDNC* and *cysJIH* operons. We found that cells with the reporter plasmid containing *cysB*
304 had 1.6-fold higher GFP fluorescence than cells lacking *cysB* on the reporter plasmid (Fig. 6B). This
305 finding suggests that the dampened *cysDNC* and *cysJIH* transcript levels after amplification are due to
306 outstripping of the activation capacity of CysB.
307
308 **Dosage imbalance leads to degradation of a protein produced at a higher level than its interaction**
309 **partners**
310 FtsB is an essential cell division protein that forms a 1:1:1 complex with FtsL and FtsQ (48). *ftsB*
311 is located within the 6-fold amplified region in population 8, but *ftsL* and *ftsQ* are over 2.7 Mb away from
312 *ftsB* and not amplified. *ftsB* expression was increased by 5.5-fold at the transcript level, but only 1.9-fold
313 at the protein level (Fig. 5C and Dataset S1). Proteins that form complexes are often unstable in the
314 absence of their interaction partners due to improper folding and subsequent degradation by proteases
315 (49, 50). It has been shown that FtsL and FtsQ can both stabilize FtsB, preventing its degradation (51).
316 Depletion of either FtsL or FtsQ in *E. coli* resulted in reduced levels of FtsB and simultaneous appearance
317 of a smaller FtsB degradation product (51). Thus, the imbalance between the cellular levels of FtsB and
318 FtsLQ we observe after co-amplification of *ftsB* likely leads to decreased stability and increased

319 degradation of FtsB protein without affecting *ftsB* transcript levels, which scale with gene copy number
320 (Fig. 6C).
321
322 **The structure of the Csr circuit seems to buffer it against dosage imbalances caused by**
323 **segmental amplification**
324 In population 8, *csrA*, *csrB* and *barA* are present within the 6-copy amplicon. The levels of CsrA
325 and BarA proteins are increased by 5.6- and 5.2-fold, respectively. However, the level of the sRNA CsrB is
326 disproportionally increased by 16-fold (Fig. 6D and Dataset S1). These genes are components of the
327 carbon storage regulator (Csr) system, a complex regulatory circuit consisting of both proteins and sRNAs
328 that posttranscriptionally coordinates expression of genes involved in diverse processes such as
329 glycolysis, glycogen synthesis and flagellar motility (52, 53). *csrC*, *csrD* and *uvrY*, which are also part of
330 the Csr system, were not amplified, but the levels of the sRNA CsrC and the protein UvrY are increased
331 by 3.4- and 1.9-fold respectively, while the level of CsrD is modestly decreased by ~40% (Fig. 6D and
332 Datasets S2 and S3). Thus, amplification of some Csr system genes appears to lead to perturbations in
333 the levels other components.
334 The striking and concurrent increases in the levels of CsrB and CsrC sRNAs in population 8 are
335 likely due to a combination of effects mediated by the intricate network of regulatory interactions in the
336 Csr circuit (Fig. 6D). CsrB and CsrC are functionally similar, as both sRNAs can bind and sequester CsrA
337 homodimers in an inactive state (52, 53). Despite *csrB* being within the amplified region, while *csrC* is
338 located >1.1 Mb away and not amplified, both sRNAs are largely subject to the same forms of regulation
339 in the Csr system. The 5.2-fold increase in the level of the histidine kinase BarA and the 1.9-fold increase
340 in the level of the cognate response regulation UvrY would be expected to increase transcription of *csrB*
341 and *csrC.* In addition, the decreased expression of CsrD, which recruits RNase E to degrade CsrB and
342 CsrC sRNAs, should also contribute to increased levels of CsrB and CsrC.
343 To our surprise, the substantial perturbations to sRNA and protein levels in this elaborate
344 regulatory circuit seem to have negligible effects on the stability and translation of mRNAs regulated by
345 CsrA. CsrA blocks translation of *glgCAP* mRNAs, which encode glycogen synthesis enzymes (52, 53).
346 However, we see little change in GlgCAP protein levels (Dataset S2). Conversely, CsrA protects *flhDC*

347 mRNAs, which encode the master regulator of flagellar synthesis, from degradation by RNAse E, thereby
348 promoting translation (52, 53). Although *flhDC* transcript levels are slightly elevated, we did not detect
349 FlhDC proteins in our proteomics data (Dataset S2 and S3). We observed a similar modest increase in
350 *flhDC* transcript levels in population 3, in which neither *csrA* nor *csrB* is amplified, suggesting that CsrA
351 plays little role in altering *flhDC* mRNA levels in these strains. The effect of the elevated CsrA expression
352 is likely blunted by the increased levels of both CsrB and CsrC, which sequester CsrA in an inactive form.
353 Thus, the net result of moderately elevated CsrA levels may be insignificant in the context of other
354 regulators that modulate transcription of *glgCAP* and *flhDC* (54). Overall, the intricate wiring of the Csr
355 regulatory circuit, replete with different forms of activation and inhibition, seems to buffer it against
356 stoichiometric imbalances between interacting sRNAs and proteins. This intrinsic buffering capacity
357 should be useful for acclimating to environmental changes while avoiding dysregulation, but it also seems
358 to have the incidental effect of limiting severe physiological perturbations following segmental
359 amplification.
360
361 **Segmental amplification has system-wide effects**
362 The segmental amplifications in populations 3 and 8 resulted in significant changes hundreds of
363 transcripts (Fig. 7), including mRNAs and sRNAs encoded by genes that were not present in the amplified
364 regions. Expression of two classes of non-amplified genes was clearly increased in both populations 3
365 and 8: ribosomal RNA (rRNA) and ribosomal protein genes; and colanic acid capsular polysaccharide
366 synthesis and export genes (Fig. 7 and Dataset S3).
367

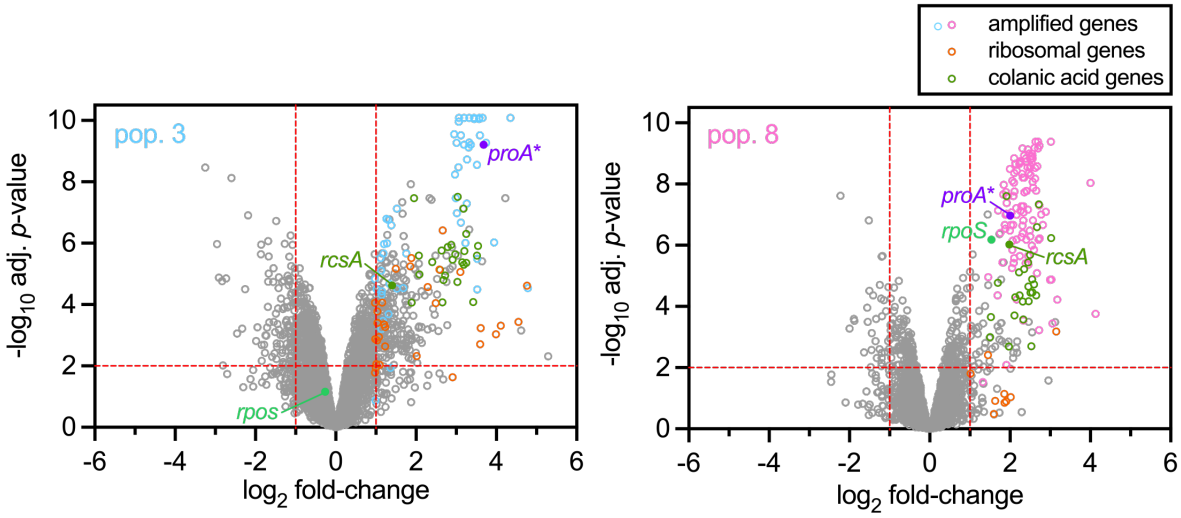
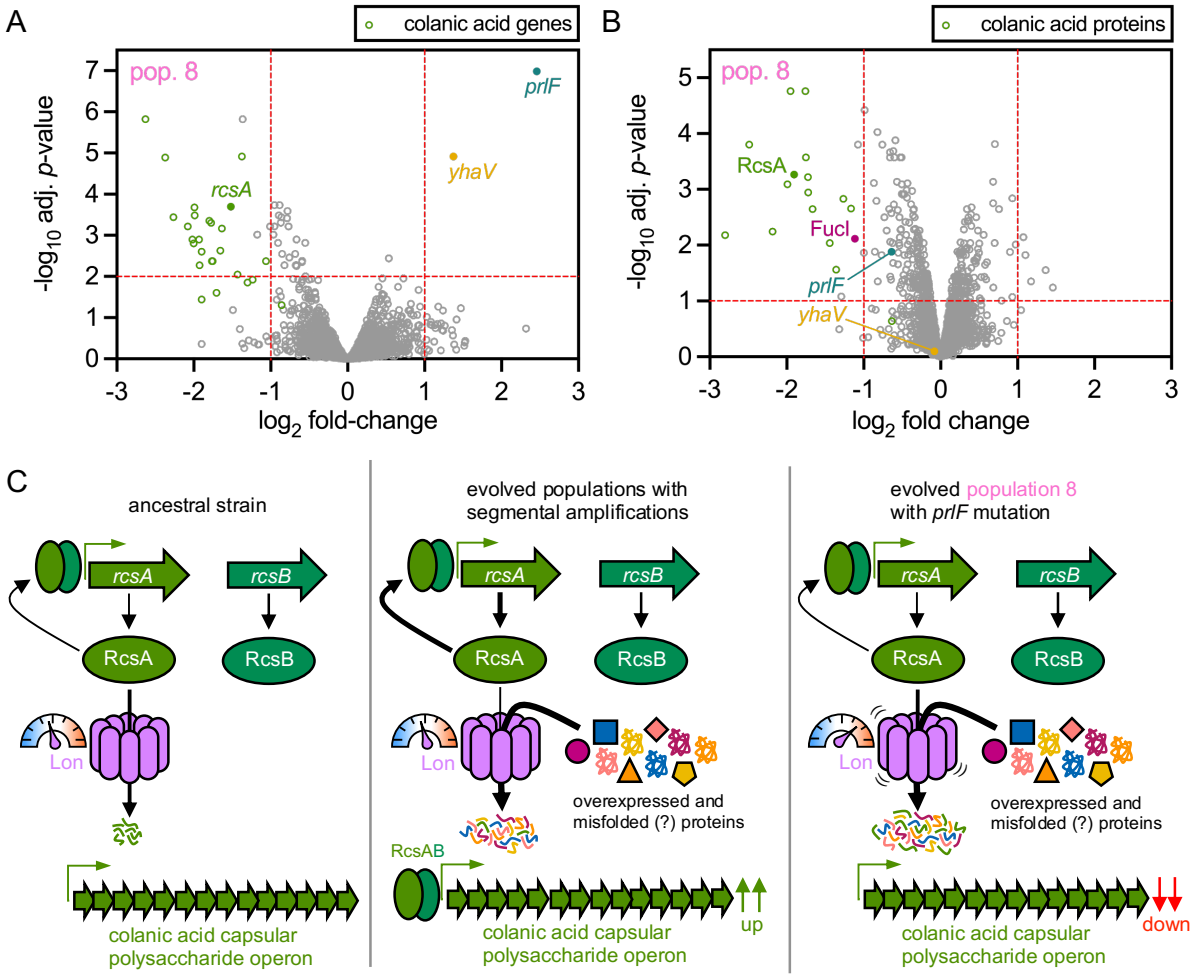


Fig. 7. The global transcriptomic consequences of segmental amplification in populations 3 and 8. Amplified genes that are differentially expressed in populations 3 and 8 compared to the ancestral strain are indicated by blue and pink points, respectively. Amplified genes *proA** and *rpoS* and classes of non-amplified genes are indicated by different colors.

The increased expression of rRNA and ribosomal protein genes is likely an indirect effect of the amplification of *proA**, which alleviates the shortage of proline and arginine in the ancestral strain and should dampen the stringent response that diminishes transcription of ribosomal genes (55). The increased expression of about 25 colanic acid capsular polysaccharide synthesis genes appears to be caused by overexpression of the transcriptional activator RcsA. Although *rcsA* is not among the co-amplified genes, *rcsA* transcript levels are increased by 2.6- and 4-fold and protein levels are increased by 8.2- and 4.6-fold in populations 3 and 8, respectively (Fig. 7 and Datasets S2 and S3). RcsA forms a heterodimer with RscB that stimulates transcription of *rcsA* itself in addition to colanic acid synthesis genes (56–58). RcsA is an unstable protein that is rapidly degraded by Lon and HslUV proteases, which typically prevents excess RcsA from accumulating and activating its own transcription(56–60). Because proteolysis appears to be the primary mechanism by which RcsA expression is kept low, a physiological perturbation that impairs global protein homeostasis might be expected to attenuate the capacity of Lon and HslUV proteases to degrade RcsA. We hypothesize that inappropriate overexpression of co-amplified genes following segmental amplification might divert Lon and other proteases away from degrading their

388 usual client proteins, such as RcsA, and instead shift their activity toward degrading overexpressed and
389 potentially misfolded proteins. The resulting accumulation of RcsA would, in turn, increase transcription of
390 *rcsA* itself as well as colanic acid synthesis genes.
391
392 **A compensatory mutation elevates Lon protease activity and alleviates adverse effects of**
393 **segmental amplification**
394 As mentioned previously, we identified a putative compensatory mutation that was fixed in
395 population 8 at the end of the evolution experiment. This mutation is a small 7 bp duplication in *prlF*,
396 which encodes the antitoxin of the PrlF-YhaV toxin-antitoxin system (61). This exact mutation has been
397 reported to alleviate toxic overexpression of a fusion protein by hyperactivating the Lon protease (61, 62).
398 While the molecular mechanism linking this toxin-antitoxin system to the activity of Lon protease is not
399 clear, this *prlF* mutation may help ameliorate general proteotoxic stress, as Lon plays an important role in
400 global protein homeostasis by degrading misfolded proteins in addition to particular client proteins (63,
401 64). This *prlF* mutation was not detected in colonies isolated from population 8 at generation 125 yet rose
402 to fixation in approximately 60 generations by generation 186. The highly beneficial nature of this *prlF*
403 mutation is consistent with the notion that increasing protein degradation by elevating Lon activity
404 alleviates the some of the fitness cost associated with segmental amplification.
405 Our initial transcriptomic and proteomic analysis of population 8 used a clonal isolate from
406 generation 125, which lacked the *prlF* mutation, in order to characterize the immediate collateral
407 consequences of segmental amplification in the absence of compensatory mutations. To determine the
408 global effect of the Lon-hyperactivating *prlF* mutation, we also isolated a clone from population 8 at
409 generation 186 with the *prlF* mutation and performed RNA-seq and label-free quantitative proteomics in
410 parallel with our other analyses. The population 8 isolates from generations 125 and 186 have the same
411 ~125 kb amplicon at approximately 6 copies (*SI Appendix* Fig. S4), and the only difference between the
412 two isolates is the *prlF* mutation. The *prlF* mutation causes a significant decrease in the mRNA levels of
413 many colanic acid synthesis and export genes, including *rcsA* (Fig. 8A). Because the *prlF* mutation
414 increases the activity of Lon protease and Lon is known to readily degrade RcsA (56–58), lower RcsA
415 levels are expected to result in less transcription of colanic acid-related genes.

416



417

418 **Fig. 8. A compensatory mutation that increases activity of Lon protease decreases expression of**
419 **colonic acid synthesis and export genes.** (A) The effect of the *prlF* mutation on the transcriptome of
420 population 8. (B) The effect of the *prlF* mutation on the proteome of population 8. Gene expression was
421 compared between a generation 125 isolate lacking the *prlF* mutation and a generation 185 isolate with
422 the *prlF* mutation. Colanic acid gene/proteins are indicated in green. Other genes/proteins discussed in
423 the text are indicated with different colors. (C) Postulated effect of the compensatory *prlF* mutation that
424 hyperactivates Lon protease, leading to increased degradation of RcsA and therefore decreased
425 expression of RcsAB-activated colanic acid capsular polysaccharide biosynthesis genes.

426

427 Only two genes, *prlF* itself and *yhaV*, which together form an operon and are thus transcribed as
428 a single polycistronic mRNA (61), show dramatic upregulation due to the *prlF* mutation. The PrlF antitoxin

429 and YhaV toxin together form a transcriptional repressor that represses transcription of their own operon
430 (61). The *prlF* mutation is a 7 bp duplication that causes a frameshift, which truncates the C-terminus of
431 PrlF and impairs the repressor function of PrlF-YhaV (61). Thus, diminished repression of the *prlF-yhaV*
432 operon due to the *prlF* mutation increases transcript levels for these genes.

433 At the proteome level, we observe a clear decrease in the expression of colanic acid synthesis
434 and export proteins, matching their dampened expression at the transcript level (Fig. 8A and B). Our data
435 support the hypothesis that the *prlF* mutation causes a massive attenuation of many colanic acid-related
436 genes at the transcript and protein levels by increasing Lon-mediated proteolysis of the transcriptional
437 activator RcsA. Increasing degradation of RcsA alone likely prevents the unnecessary expression of over
438 20 genes involved in colanic acid synthesis and export. We also wondered whether increasing Lon
439 activity via the *prlF* mutation might decrease the inappropriate overexpression of any co-amplified genes.
440 The level of FucI, a fucose isomerase, is increased by 7.2-fold in the generation 125 isolate with 6 copies
441 of *fucI* compared to the ancestral strain (Dataset S1). In the generation 186 isolate with the *prlF* mutation,
442 FucI levels are reduced by >50% despite *fucI* still being present at 6 copies (Fig. 8B). Because fucose is
443 not present in the growth medium, FucI expression is not needed for growth. The reduction in FucI levels
444 suggests that the hyperactivated Lon in the generation 186 isolate contributes to FucI degradation.
445 However, other proteins encoded in the amplified *fucAO* and *fucPIKUR* operons do not exhibit similar
446 decreases in expression following hyperactivation of Lon (Dataset S1).

447 While the transcription of the *prlF-yhaV* operon is highly upregulated due to the *prlF* mutation,
448 their expression at the protein level is not (Fig. 8 A and B). This result is not surprising as the 7 bp
449 duplication in *prlF* introduces a premature stop codon that has been shown to have polar effects on *yhaV*,
450 possibly by decreasing translation of YhaV from polycistronic *prlF-yhaV* mRNAs (61). The truncated PrlF
451 that results from this frameshift mutation has a shorter and more basic C-terminus. The altered C-
452 terminus of PrlF and/or diminished interaction with YhaV may reduce PrlF stability and therefore its level
453 (Fig. 8B).

454 The effect of the *prlF* mutation on the transcriptome and proteome of cells with large segmental
455 amplifications indicates that it is a compensatory mutation. We hypothesize that by elevating the activity
456 of Lon protease, the *prlF* mutation helps alleviate the fitness costs associated with inadvertant

457 overexpression of amplified and non-amplified genes (Fig. 8C). In population 8, the *prlF* mutation, which
458 increases Lon protease activity (62), arises and fixes in about 60 generations, indicating it is highly
459 beneficial. When comparing isolates with and without the *prlF* mutation, our data show that
460 hyperactivating Lon leads to more degradation of client protein RcsA and thus less transcription and lower
461 protein levels of colanic acid synthesis genes (Fig. 8). Therefore, we view the *prlF* mutation as
462 representing a class of compensatory mutations that are broadly beneficial for counteracting proteotoxic
463 stress by increasing proteolysis, which could result from segmental amplification or other forms of stress
464 that disrupt protein homeostasis.
465
466 **Summary**
467 The strong selection for increased ProA* activity in our evolving populations resulted in rapid
468 segmental amplification encompassing *proA** and many neighboring genes. Transcription of most
469 amplified genes correlates well with copy number. However, there are notable exceptions. Expression of
470 *rpoS* is attenuated in population 3 by loss of several of its promoters. Excessive RpoS would be expected
471 to be problematic because RpoS promotes a transcriptional program that slows growth and alters
472 resource allocation when cells are starving or stressed. This response would be inappropriate when
473 glucose is abundant and proline and arginine synthesis have been improved due to amplification of *proA*.*
474 It is tempting to speculate that the loss of *rpoS* promoters allowed the small amplicon nested within the
475 originally duplicated region to rise to 12 copies.
476 We also observed other mechanisms for attenuation of mRNA expression that do not rely on a
477 fortuitous amplification junction. Transcription of amplified genes in two operons encoding sulfate
478 assimilation genes is attenuated due to inadequate levels of a transcriptional activator encoded by a gene
479 that was not amplified. Such a mismatch could occur in many situations when only some genes in a
480 regulon are amplified. Regulons often include operons that are scattered across the genome and are
481 controlled by transcriptional activators that may not be in close proximity to the operons they control.
482 While attenuation in this case appears to be due to outstripping the capacity of a transcriptional activator,
483 we expect that the opposite situation could also arise; i.e. expression of amplified genes could be
484 exacerbated if the capacity of a transcriptional repressor is outstripped.

485 Post-transcriptional mechanisms also appear to contribute to attenuation of protein levels
486 produced from amplified genes. For example, it is well accepted that subunits of protein complexes that
487 are produced in excess are often degraded, probably because they misfold due to exposure of surfaces
488 that are normally involved in protein-protein interactions (49, 50). Indeed, excess FtsB has been shown to
489 be degraded due to insufficient levels of its interacting partner proteins FtsQ and FtsL (51).
490 The complex regulation of many genes in *E. coli* may contribute to attenuation of expression of
491 amplified genes. *rpoS* expression is regulated by multiple transcription factors, sRNAs, and adaptor-
492 mediated proteolysis, which allows integration of multiple signals in the decision to activate the stress
493 response while preventing inopportune overexpression. Likewise, the Csr system integrates multiple
494 inputs in a complex regulatory circuit in order to appropriately fine-tune translation of specific mRNAs.
495 While these systems evolved to enable *E. coli* to respond appropriately to changing environments, a
496 propitious side effect is the ability to buffer the effects of segmental amplification.
497 Segmental amplification can have global consequences. Approximately 86% and 49% of genes
498 whose expression is increased by ≥2-fold at the transcript level in populations 3 and 8, respectively, are
499 not amplified. While co-amplification of genes neighboring a gene under selection for increased copy
500 number has previously been recognized to contribute to fitness, our work adds a previously
501 unappreciated nuance; segmental amplification can also alter expression of mRNAs and proteins from
502 hundreds of unamplified genes. Thus, the immediate fitness effects of segmental amplification are due to
503 the combination of the desirable effect of amplifying the gene under selection for increase dosage and the
504 collateral effects (both detrimental and beneficial) of altered expression of co-amplified genes within the
505 amplicon as well as genes elsewhere in the genome. Finally, the fitness costs associated with segmental
506 amplification can be mitigated, at least partially, by compensatory mutations that alleviate the energetic
507 burden or physiological perturbations that accompany a particular amplification.
508
509 **Materials and Methods**
510 **Strains and growth conditions**
511 *E. coli* strains, plasmids, and primers used in this work are listed in *SI Appendix*, Tables S2, S3,
512 and S4. *E. coli* strains were routinely cultured in lysogeny broth (LB) or on LB agar plates containing 1.5%

513 agar at 37 °C. M9 salts minimal medium (12.8 g/L Na_2HPO_2•$7H_2O$, 3 g/L KH_2PO_4, 0.5 g/L NaCl, 1 g/L
514 NH_4Cl) was supplemented with glucose at a final concentration of 0.2% (v/v). Antibiotics were used at the
515 following final concentrations unless otherwise indicated: ampicillin, 100-150 µg/mL; kanamycin, 20
516 µg/mL; chloroamphenicol, 20 µg/mL; and gentamicin, 15 µg/mL.
517
518 **Strain construction**
519 All strains used in this work are derived from *E. coli* BW25113. The Δ*argC*::*kan*^{R} M2-*proBA**-*yfp*
520 Δ*fimAICDFGH* Δ*csgBAC* Δ82 bp *rph-pyrE E. coli* strain AM327 was previously constructed from the Keio
521 collection Δ*argC*::*kan*^{R} *E. coli* BW25113 strain as described in (4). The Δ*fimAICDFGH* Δ*csgBAC*
522 deletions help prevent biofilm formation during laboratory evolution in a turbidostat. The 82 bp deletion in
523 *rph* and upstream of *pyrE* fixes a defect in pyrimidine synthesis due to a frameshift in *rph* in *E. coli*
524 BW25113 and other K-12 strains (65, 66). The M2 promotor mutation is a C > T point mutation 45 bp
525 upstream of *proB* that increases transcription of the *proBA* operon as described in (44). The AM187 strain
526 (GenBank accession number CP037857.2) and its derivative AM327 contain a 1.8-Mb inversion in the
527 genome that occurred between two IS3 elements (43).
528 For this study, we deleted the M2-*proBA**-*yfp* region in the *E. coli* AM327 genome using λ-Red
529 recombineering (67). We transformed the temperature-sensitive plasmid pSIM6 (68) containing the λ-Red
530 recombinase genes (*exo*, *beta* and *gam*) under control of a heat-inducible promoter into AM327 by
531 electroporation. After recovery in super optimal broth with catabolite repression medium (SOC) (BD
532 Difco™, Franklin Lakes, NJ) for ~3 h at 30 °C, cells were plated on LB ampicillin and incubated at 30 °C
533 to select for transformants. We replaced *proBA** and *yfp* in AM327 with a *sacB-gtm*^{R} selection-
534 counterselection cassette amplified from the plasmid pJQ200SK (69). We amplified ~1 kb fragments from
535 the regions upstream and downstream of the *proBA* operon using primers that introduced ~20 bp ends
536 complementary to the ends of the *sacB-gtm*^{R} cassette. We assembled a linear *proBA*-yfp* replacement
537 fragment from these three fragments using splicing by overlap extension (SOE) PCR using Q5® High-
538 Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA). The AM327 strain harboring pSIM6 was
539 inoculated into LB ampicillin, grown to an optical density at 600 nm (OD_{600}) of 0.2-0.6 at 30 °C and then
540 incubated at 42 °C for 15 min to induce expression of the λ Red genes. Cells were then centrifuged at

541 4,500 x g at 4 °C, washed with ice-cold deionized water and immediately electroporated with ~0.5-2.0 µg
542 of the *proBA** replacement linear fragment. Following electroporation, cells were inoculated into fresh
543 SOC (BD Difco™) and allowed to recover at 30 °C for ~3 h before being spread onto LB gentamicin.
544 Gentamicin-resistant colonies were screened for integration of *sacB-gtm*^{R} into the *proBA* locus by colony
545 PCR using One*Taq*® DNA Polymerase (New England Biolabs).
546 We constructed a fragment to delete the *sacB-gtm*^{R} cassette by amplifying ~1 kb regions
547 upstream and downstream of the *proBA* operon with overlapping ends. The two PCR products were
548 assembled together using SOE PCR. Cells containing pSIM6 and the *sacB-gtm*^{R} selection-
549 counterselection marker in place of *proBA** were inoculated into LB ampicillin, grown to an OD_{600} of 0.2-
550 0.6 at 30 °C and subjected to a second round of λ-Red recombineering as described above. Cells were
551 electroporated with ~0.5-2.0 µg of the *sacB-gtm*^{R} replacement fragment. Following electroporation, cells
552 were inoculated into fresh SOC and allowed to recover at 30 °C for ~3 h before being spread onto salt-
553 free (i.e. no NaCl) LB plates containing 10% (w/v) sucrose and incubated at 30 °C for counterselection.
554 Colonies that grew on salt-free LB sucrose plates were patched onto LB gentamicin plates to confirm
555 gentamicin sensitivity due to loss of the *sacB-gtm*^{R} cassette and screened for markerless deletion of
556 *proBA** by colony PCR and Sanger sequencing.
557 To relocate the M2-*proBA** operon next to *rpoS* in the *E. coli* genome, we first introduced a *tetR-*
558 *ccdB-cat* selection-counterselection cassette at the target site between *rpoS* and *ygbN.* We amplified the
559 *tetR-ccdB-cat* cassette from pDLM3 (70). We amplified ~1 kb fragments upstream and downstream of the
560 target site using primers that introduced ~20 bp ends complementary to the ends of the *tetR-ccdB-cat*
561 cassette. These three fragments were assembled together using SOE PCR. The Δ*proBA** strain RF72
562 harboring pSIM6 was inoculated into LB ampicillin, grown to an OD_{600} of 0.2-0.6 at 30 °C and subjected to
563 λ-Red recombineering as described above. Cells were electroporated with ~0.5-2.0 µg of the *tetR-ccdB-*
564 *cat* integration fragment. Following electroporation, cells were inoculated into fresh SOC and allowed to
565 recover at 30 °C for ~3 h or at room temperature overnight before being spread onto LB chloramphenicol
566 plates. The plates were incubated at 30 °C for 1-2 days. Chloramphenicol-resistant colonies were
567 screened for integration of *tetR-ccdB-cat* between *ygbN* and *rpoS* by colony PCR. The *tetR-ccdB-cat*
568 cassette was then replaced with M2-*proBA*.* We amplified the M2-*proBA** operon (without *yfp*) from

569 AM187 gDNA and ~1 kb fragments upstream and downstream of the intergenic site between *rpoS* and
570 *ygbN* from RF72 gDNA by PCR. These three fragments were assembled using SOE-PCR to generate the
571 *tetR-ccdB-cat* replacement cassette. The *ygbN-tetR-ccdB-cat-rpoS* strain harboring pSIM6 was
572 inoculated into LB ampicillin, grown to an OD_{600} of 0.2-0.6 at 30 °C and subjected to λ-Red
573 recombineering as described above. Cells were electroporated with ~0.5-2.0 µg of the *tetR-ccdB-cat*
574 replacement cassette. Following electroporation, cells were inoculated into fresh SOC and allowed to
575 recover at 30 °C for ~3 h or at room temperature overnight before being spread onto LB
576 anhydrotetracycline (2-6 ug/mL) plates for counterselection. The plates were incubated at 30 °C
577 overnight. Colonies that grew on LB anhydrotetracycline plates were patched onto LB chloramphenicol
578 plates to confirm chloramphenicol sensitivity and screened for integration of M2-*proBA** into the intergenic
579 site between *ygbN* and *rpoS.* The resulting M2-*proBA**-*rpoS* strain was streaked onto LB
580 anhydrotetracycline plates and incubated at 40 °C overnight to cure the temperature-sensitive pSIM6
581 plasmid. Colonies were patched onto fresh LB anhydrotetracycline and LB ampicillin plates to confirm
582 ampicillin sensitivity due to loss of pSIM6. Genomic integration of M2-*proBA** next to *rpoS* was confirmed
583 by Sanger sequencing.
584
585 **Laboratory Evolution**
586 To initiate laboratory evolution of the *proBA**-*rpoS* relocation strain (RF90), cells from a frozen
587 glycerol stock were streaked onto LB kanamycin plates and the plates were incubated at 37 °C overnight.
588 A single founder colony was inoculated into LB kanamycin and incubated at 37 °C overnight. The
589 overnight culture was centrifuged at 8,000 x g at room temperature and the cells were washed and
590 resuspended in M9 medium without glucose. A 125 µL aliquot of cells was inoculated into each of eight
591 glass vials containing 25 mL (1:200 dilution) of M9 minimal medium containing 0.2% glucose and
592 kanamycin (20 µg/mL). The vials were placed into the vial sleeves of an eVOLVER turbidostat (FynchBio).
593 Each vial contained a micro-flea magnetic stir bar (10 mm x 3 mm). The eVOLVER was set in 'turbidostat'
594 operation mode at 37 °C with a stir rate of 8, a lower OD_{600} threshold of 0.15, an upper OD_{600} threshold of
595 0.25 and a vial volume of 25 mL. When cultures reached the upper OD_{600} threshold of 0.25, ~7-10 mL of
596 fresh M9 + glucose medium was pumped into vials and an equivalent volume of culture was removed so

597 that cultures were diluted to an OD_{600} of ~0.15. Cultures were evolved in this manner for 145-200
598 generations (~14 days). Approximately every 40-60 h, ~7-10 mL samples of cultures expelled from the
599 vials during dilution were collected and centrifuged at 4,500 x g at room temperature. The cell pellets
600 were resuspended in 1 mL of M9 medium without glucose. Aliquots (0.5 mL) of concentrated cells from
601 each population were mixed 1:1 with 50% glycerol and stored at -70 °C. The remaining 0.5 mL cell
602 suspensions were centrifuged at 8,000 x g at room temperature, the supernatants were removed and the
603 cell pellets were stored at -20 °C for extraction of genomic DNA (gDNA).
604
605 **Whole Genome Sequencing and Analysis**
606 Cell pellets were thawed and gDNA was extracted using the PureLink™ Genomic DNA Mini Kit
607 (Invitrogen) according to the manufacturer's protocol for Gram-negative bacteria. Purified gDNA samples
608 were stored at -20 °C. Whole genome sequencing was performed at SeqCenter (formerly the Microbial
609 Genome Sequencing Center; Pittsburgh, PA). Sample libraries were prepared with the Illumina DNA Prep
610 kit and IDT 10 bp UDI indices and sequenced on an Illumina NextSeq 2000, producing 2x151 bp paired-
611 end reads. Demultiplexing, quality control, and adapter trimming was performed with Illumina BCL
612 Convert v3.9.1. The resulting fastq files were further processed using fastp v0.20.1 (71) to trim and
613 remove low quality reads with the following settings: -q 15 -u 20 -t 1. Mutations in evolved populations
614 were identified using breseq v0.37.1 (72, 73) in polymorphism mode with bowtie 2.5.0 (74) and R 4.2.2 by
615 mapping processed fastq reads to the genome of the *E. coli proBA*-rpoS* strain. To visualize read
616 coverage depth of segmental amplifications encompassing *proBA**, read coverage depth for every 50 bp
617 for a 300 kb genome window was exported using the breseq BAM2COV subcommand: -a seq
618 id:2700000-3000000 -t -1 --resolution 6000.
619
620 **Quantitative PCR for measurement of *proA** copy number**
621 *proA** copy number was measured using qPCR of purified gDNA as described in (4) with slight
622 modifications. Briefly, PowerSYBR Green PCR master mix (Applied Biosytems by Thermo Scientific) was
623 combined with primers and gDNA in 48-well plates (Thermo Scientific) and samples were analyzed using
624 a StepOne Real-Time PCR System (Applied Biosystems). The genes *gyrB* and *icd*, which remained at

625 one copy in evolved populations, were used as internal reference genes for quantification of *proA** copy
626 number. Threshold cycle (C_T) measurements for duplicate technical replicates of each gDNA sample were
627 averaged. The ΔC_T values for *proA** and single copy reference genes were calculated by subtracting the
628 C_T for gDNA from the ancestral *proBA*-rpoS* strain from the C_T for gDNA from the evolved populations.
629 The ΔC_T values of reference genes *gyrB* and *icd* were averaged for each sample and $\Delta\Delta C_T$ values were
630 then calculated by subtracting the mean reference ΔC_T from *proA** ΔC_T for each sample. *proA** copy
631 number was calculated as $2^{-\Delta\Delta CT}$ for each sample. Primer sets used for qPCR are listed in *SI Appendix*
632 Table S4.
633
634 **RNA extraction**
635 Glycerol stocks of ancestral *proBA*-rpoS E. coli* and evolved populations 3 (at generation ~199)
636 and 8 (at generation ~125) were streaked onto M9 plates containing 0.4% glucose. Single colonies of
637 each strain were inoculated into 4 mL M9 containing 0.2% glucose and the cultures were grown at 37 °C
638 overnight with shaking at 200 rpm. The overnight cultures were inoculated into 25 mL of fresh M9
639 containing 0.2% glucose and 20 µg/mL kanamycin in 125 mL Erlenmeyer flasks at a 1:50 dilution for the
640 ancestral strain and a 1:100 dilution for the evolved populations to account for differences in growth rate.
641 Four replicate cultures of each strain were grown at 37 ºC with shaking at 200 rpm for ~7 h to an OD_{600} of
642 ~0.2-0.5. Cultures were centrifuged at 4,500 x g for 10 min at 4 °C. Cell pellets were resuspended in 5 mL
643 M9 salts without glucose. Four mL of each cell suspension was flash frozen in a dry ice-ethanol (95%)
644 bath and stored at -70 °C prior to RNA extraction. The remaining 1 mL of each cell suspension was
645 centrifuged at 8,000 x g for 2 min. The supernatant was removed and the cell pellets were flash frozen in
646 a dry ice-ethanol (95%) bath and stored at -20 °C prior to sample preparation for proteomics.
647 Total RNA was extracted from cells using RNAprotect Bacteria Reagent and RNeasy Mini Kit
648 (Qiagen, Hilden, Germany). Frozen 4-mL cell suspensions were thawed on ice, mixed 2:1 with
649 RNAprotect Bacteria Reagent, incubated for 5 min at room temperature and centrifuged at 4,500 x g for
650 10 min at room temperature. Cell pellets were resuspended in TE buffer with lysozyme (1 mg/mL) to lyse
651 the cells. RNA was purified using RNeasy Mini Kits according to the manufacturer's instructions except
652 that optional on-column DNase digestion was performed by mixing 10 µL TURBO™ DNase (Invitrogen,

653 Waltham, MA) with 70 µL buffer RDD (Qiagen), adding 80 µL of DNase solution directly to the spin
654 column membrane of each sample and incubating at room temperature for 15 min. RNA samples were
655 stored at -20 °C.
656
657 **RNA sequencing and differential expression analysis**
658 RNA sequencing was performed at SeqCenter (formerly the Microbial Genome Sequencing
659 Center; Pittsburgh, PA). Three replicate RNA samples were sequenced for each strain. RNA samples
660 were treated with DNase (Invitrogen). Sample libraries were prepared with the Illumina Stranded Total
661 RNA Prep Ligation with Ribo-Zerp Plus kit and 10 bp IDT for Illumina indices and sequenced on a
662 NextSeq 2000, producing 2x51 bp paired-end reads. Demultiplexing, quality control, and adapter
663 trimming was performed with Illumina BCL Convert v3.9.1.
664 To map reads to the *E. coli* genome to generate read counts for each gene, fastq files were
665 processed with fastp v0.20.1 (71) to trim and remove low quality reads with the following settings: -q 15 -u
666 20. Processed reads were aligned to the *E. coli* reference genome using bowtie2 (74). Read alignment
667 files were converted into sorted.bam files using samtools (75, 76). Aligned reads were mapped to *E. coli*
668 genes using HTSeq (77) to generate read counts with the settings: htseq-count -f bam –r pos -s no -t
669 gene -i Name -m intersection-nonempty --nonunique all.
670 Differential expression analysis was performed in R v4.3.1/RStudio (Posit team; Boston, MA) with
671 the following Bioconductor/R packages: edgeR (78, 79), limma (80), glimma, gplots, RColorBrewer and
672 NFM. Read counts generated from HTSeq were filtered to remove lowly expressed genes and normalized
673 using trimmed mean of M values with edgeR. Read counts were transformed into log counts per million
674 (CPM) with the voom function of limma for each sample. Differential expression analysis of transformed
675 logCPM data was performed with the contrasts.fit and eBayes functions in limma. Differential expression
676 data were exported and plotted in GraphPad Prism v10.0.2 (GraphPad Software; Boston, MA)
677
678 **Label-free proteomics and differential expression analysis**
679 Frozen cell pellets were lysed in 50 mM Tris-HCl, pH 8.5 containing 5% (w/v) sodium dodecyl
680 sulfate, 10 mM tris(2-carboxyethylphosphine) and 40 mM chloroacetamide by boiling for 10 minutes at 95

681 °C followed by probe sonication (1 sec on, 1 sec off) for 1 minute. Each sample was digested using the
682 SP3 method described in (81). Carboxylate-functionalized SpeedBeads (GE Life Sciences) were added
683 to the lysates. Addition of acetonitrile to 80% (v/v) caused proteins to bind to the beads. The beads were
684 washed twice with 80% (v/v) ethanol and twice with 100% acetonitrile. Proteins were digested with Lys-
685 C/Trypsin (Promega) at 37 °C overnight with rotation. SpeedBeads were collected by centrifugation and
686 placed on a magnet to hold the SpeedBead pellet together so that the digested peptides in the
687 supernatant could be extracted. The peptides were then desalted using an Oasis HLB cartridge (Waters)
688 according to the manufacturer's instructions and dried in a SpeedVac.
689 The dried peptide samples were suspended in 3% (v/v) acetonitrile/0.1% (v/v) trifluoroacetic acid
690 and 1 µg was directly injected onto a C18 1.7 µm, 130 Å, 75 µm X 250 mm M-class column (Waters)
691 using a Thermo Ultimate 3000 RSLCnano UPLC. Peptides were eluted at 300 nL/minute using a gradient
692 from 2% to 20% acetonitrile over 100 min into a Q-Exactive HF-X mass spectrometer (Thermo Scientific).
693 Precursor mass spectra (MS1) were acquired at a resolution of 120,000 from 350 to 1550 m/z with an
694 automatic gain control (AGC) target of 3E6 and a maximum injection time of 50 milliseconds. Precursor
695 peptide ion isolation width for MS2 fragment scans was 1.4 m/z, and the top 12 most intense ions were
696 sequenced. All MS2 spectra were acquired at a resolution of 15,000 with higher energy collision
697 dissociation (HCD) at 27% normalized collision energy. An AGC target of 1E5 and 100 milliseconds
698 maximum injection time was used. Raw files were searched against the Uniprot *Escherichia coli*
699 databases UP000002032 and UP000000625 using Maxquant version 2.0.3.0 with cysteine
700 carbamidomethylation as a fixed modification. Methionine oxidation and protein N-terminal acetylation
701 were searched as variable modifications. All peptides and proteins were thresholded at a 1% false
702 discovery rate (FDR). Statistical analysis was performed on $\log_2$-transformed iBAQ and LFQ intensities
703 using limma (80).
704
705 **Plasmid-based P*cys-gfp* reporter system**
706 To construct P*cys-gfp* reporter plasmids, the backbone of the low copy number (~5 copies)
707 plasmid pTHSSe_43 (82), including the pSC101 origin of replication with *repA* variant (E93V) gene and

708 ampicillin resistance cassette, was linearized and amplified by PCR. A 312 bp fragment including the
709 CysB-activated promoter of the *cysDNC* operon (P*cys*) through the start codon was amplified by PCR
710 from gDNA from the ancestral *proBA*-rpoS* relocation strain. A 777 bp fragment of the monomeric
711 superfolder *gfp* gene lacking a start codon was amplified from pAA002. For the plasmid containing *cysB*,
712 a 1,402 bp fragment including *cysB* with ~200 bp upstream and downstream was amplified by PCR from
713 gDNA from the ancestral strain. All primers used to amplify PCR fragments had ~20 bp overlapping ends
714 for Gibson assembly (83). Fragments were mixed together at an insert to vector ratio of approximately 4:1
715 with Gibson Assembly Master Mix (New England Biolabs) and incubated at 50 °C for 1 h. Gibson
716 Assembly reactions were diluted 4-fold with sterile deionized water and the product was transformed into
717 chemically competent *E. coli* DH5α by heat shock at 42 °C for 30 sec according to manufacturer's
718 protocol. Following heat shock, cells were inoculated into SOC (BD Difco™) and allowed to recover at 37
719 °C with shaking at 200 rpm for 1-2 h, then spread onto LB ampicillin plates. Ampicillin-resistant
720 transformant colonies were screened for proper plasmid assembly by colony PCR. P*cys-gfp* reporter
721 plasmids were purified using the Monarch® Plasmid Miniprep Kit (New England Biolabs) according to the
722 manufacturer's protocol. The ancestral *proBA*-rpoS* relocation *E. coli* strain was grown at 37 °C in LB to
723 an OD_{600} of ~0.3-0.5, washed with room temperature deionized water and immediately electroporated
724 with 20-40 ng of purified plasmid. Following electroporation, cells were inoculated into SOC and allowed
725 to recover at 37 °C with shaking at 200 rpm for ~1 h, then spread onto LB ampicillin plates. Ampicillin-
726 resistant and GFP-positive colonies were screened by colony PCR to confirm presence of the proper
727 P*cys-gfp* reporter plasmid. All primers used to construct strains and plasmids are listed in *SI Appendix*
728 Table S4.

729 Starter cultures were prepared by inoculating 5 µL of thawed frozen stock of each strain into 4 mL
730 M9 supplemented with 0.2% glucose, 0.5 mM L-arginine, 0.4 mM L-proline, 20 µg/mL kanamycin and 150
731 µg/mL ampicillin and incubating cultures overnight at 37 °C with shaking at 200 rpm. The next day, 40 µL
732 of each overnight culture was used to inoculate 4 mL of fresh M9 medium (1:100 dilution) supplemented
733 as described above but with 100 µg/mL ampicillin. Cultures were grown at 37 °C with shaking at 200 rpm
734 for ~6 hours to an OD_{600} of ~0.4-0.5. One mL of each culture was harvested by centrifugation at 8,000 x g
735 for 3 min at room temperature. The cell pellets were resuspended in 1 mL of fresh M9 medium lacking

736 carbon sources and antibiotics. A 100 µL aliquot of samples from both strains (six biological replicates and
737 two technical replicates) was transferred to the wells of a 96-well Corning® Costar black/clear bottom
738 microplate. Sterile M9 medium was added to wells of the same plate as blank controls. Cell density
739 (OD_{600}) and GFP fluorescence (Excitation: 479 nm; Emission: 520 nm; Optics Position: Top) of each
740 sample was measured using a BioTek Synergy H1 microplate reader (Agilent Technologies, Santa Clara,
741 CA). The OD_{600} and GFP fluorescence of blank medium only controls were averaged and subtracted from
742 all sample measurements as background. GFP fluorescence measurements for each sample were
743 normalized by dividing by OD_{600} values. The normalized GFP fluorescence measurements for each set of
744 technical replicates were averaged.
745

746 **References**

747 1. W. Paulander, D. I. Andersson, S. Maisnier-Patin, Amplification of the Gene for Isoleucyl–
748 tRNA Synthetase Facilitates Adaptation to the Fitness Cost of Mupirocin Resistance in
749 Salmonella enterica. *Genetics* **185**, 305–312 (2010).
750 2. J. Näsvall, L. Sun, J. R. Roth, D. I. Andersson, Real-Time Evolution of New Genes by
751 Innovation, Amplification, and Divergence. *Science (1979)* **338**, 384 LP – 387 (2012).
752 3. S. C. Seaton, *et al.*, Genome-wide selection for increased copy number in Acinetobacter
753 baylyi ADP1: locus and context-dependent variation in gene amplification. *Mol Microbiol*
754 **83**, 520–535 (2012).
755 4. A. B. Morgenthaler, *et al.*, Mutations that improve efficiency of a weak-link enzyme are
756 rare compared to adaptive mutations elsewhere in the genome. *Elife* **8**, e53535 (2019).
757 5. H. S. Girgis, *et al.*, Single-molecule nanopore sequencing reveals extreme target copy
758 number heterogeneity in arylomycin-resistant mutants. *Proceedings of the National*
759 *Academy of Sciences* **118**, e2021958118 (2021).
760 6. M. Steinrueck, C. C. Guet, Complex chromosomal neighborhood effects determine the
761 adaptive potential of a gene under selection. *Elife* **6**, e25100 (2017).
762 7. K. P. T. Silva, G. Sundar, A. Khare, Efflux pump gene amplifications bypass necessity of
763 multiple target mutations for resistance against dual-targeting antibiotic. *Nat Commun* **14**,
764 3402 (2023).
765 8. M. Saathoff, *et al.*, Gene amplifications cause high-level resistance against albicidin in
766 gram-negative bacteria. *PLoS Biol* **21**, e3002186 (2023).
767 9. L. Wang, *et al.*, Multiple genetic paths including massive gene amplification allow
768 Mycobacterium tuberculosis to overcome loss of ESX-3 secretion system substrates.
769 *Proceedings of the National Academy of Sciences* **119**, e2112608119 (2022).
770 10. M. Maybin, *et al.*, IS1-mediated chromosomal amplification of the arn operon leads to
771 polymyxin B resistance in Escherichia coli B strains. *mBio* **0**, e00634-24 (2024).
772 11. C. J. Brown, K. M. Todd, R. F. Rosenzweig, Multiple duplications of yeast hexose
773 transport genes in response to selection in a glucose-limited environment. *Mol Biol Evol*
774 **15**, 931–942 (1998).
775 12. D. Gresham, *et al.*, The Repertoire and Dynamics of Evolutionary Adaptations to
776 Controlled Nutrient-Limited Environments in Yeast. *PLoS Genet* **4**, e1000303 (2008).

13. Z. Gu, A. Cavalcanti, F.-C. Chen, P. Bouman, W.-H. Li, Extent of Gene Duplication in the Genomes of Drosophila, Nematode, and Yeast. *Mol Biol Evol* **19**, 256–262 (2002).
14. C. Payen, *et al.*, The Dynamics of Diverse Segmental Amplifications in Populations of Saccharomyces cerevisiae Adapting to Strong Selection. *G3: Genes|Genomes|Genetics* **4**, 399–409 (2014).
15. C. Bass, *et al.*, Gene amplification and microsatellite polymorphism underlie a recent insect host shift. *Proceedings of the National Academy of Sciences* **110**, 19460–19465 (2013).
16. P. Vande Zande, X. Zhou, A. Selmecki, The Dynamic Fungal Genome: Polyploidy, Aneuploidy and Copy Number Variation in Response to Stress. *Annu Rev Microbiol* (2023). https://doi.org/10.1146/annurev-micro-041320-112443.
17. P. Spealman, C. de Santana, T. De, D. Gresham, Multilevel gene expression changes in lineages containing adaptive copy number variants. *bioRxiv* 2023.10.20.563336 (2024). https://doi.org/10.1101/2023.10.20.563336.
18. S. A. Bergin, *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* **13**, e01777-22 (2022).
19. D. G. Albertson, Gene amplification in cancer. *Trends in Genetics* **22**, 447–455 (2006).
20. R. Beroukhim, *et al.*, The landscape of somatic copy-number alteration across human cancers. *Nature* **463**, 899–905 (2010).
21. T. Santarius, J. Shipley, D. Brewer, M. R. Stratton, C. S. Cooper, A census of amplified and overexpressed human cancer genes. *Nat Rev Cancer* **10**, 59–64 (2010).
22. U. Ben-David, A. Amon, Context is everything: aneuploidy in cancer. *Nat Rev Genet* **21**, 44–62 (2020).
23. C. D. Steele, *et al.*, Signatures of copy number alterations in human cancer. *Nature* **606**, 984–991 (2022).
24. U. Bergthorsson, D. I. Andersson, J. R. Roth, Ohno's dilemma: Evolution of new genes under continuous selection. *Proceedings of the National Academy of Sciences* **104**, 17004–17009 (2007).
25. S. D. Copley, Evolution of new enzymes by gene duplication and divergence. *FEBS J* **287**, 1262–1283 (2020).
26. B. L. Logeman, L. K. Wood, J. Lee, D. J. Thiele, Gene duplication and neo-functionalization in the evolutionary and functional divergence of the metazoan copper transporters Ctr1 and Ctr2. *Journal of Biological Chemistry* **292**, 11531–11546 (2017).
27. K. Voordeckers, *et al.*, Reconstruction of Ancestral Metabolic Enzymes Reveals Molecular Mechanisms Underlying Evolutionary Innovation through Gene Duplication. *PLoS Biol* **10**, e1001446 (2012).
28. S. A. Teichmann, M. M. Babu, Gene regulatory network growth by duplication. *Nat Genet* **36**, 492–496 (2004).
29. I. Nocedal, M. T. Laub, Ancestral reconstruction of duplicated signaling proteins reveals the evolution of signaling specificity. *Elife* **11**, e77346 (2022).
30. R. T. Todd, A. Selmecki, Expandable and reversible copy number amplification drives rapid adaptation to antifungal drugs. *Elife* **9**, e58349 (2020).
31. D. Belikova, A. Jochim, J. Power, M. T. G. Holden, S. Heilbronner, "Gene accordions" cause genotypic and phenotypic heterogeneity in clonal populations of Staphylococcus aureus. *Nat Commun* **11**, 3526 (2020).
32. N. Pavelka, *et al.*, Aneuploidy confers quantitative proteome changes and phenotypic variation in budding yeast. *Nature* **468**, 321–325 (2010).
33. J. Hose, *et al.*, Dosage compensation can buffer copy-number variation in wild yeast. *Elife* **4**, e05462 (2015).

34. N. Dephoure, *et al.*, Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *Elife* **3**, e03023 (2014).
35. E. M. Torres, *et al.*, Effects of Aneuploidy on Cellular Physiology and Cell Division in Haploid Yeast. *Science (1979)* **317**, 916 (2007).
36. E. M. Torres, M. Springer, A. Amon, No current evidence for widespread dosage compensation in S. cerevisiae. *Elife* **5**, e10996 (2016).
37. J. Muenzner, *et al.*, Natural proteome diversity links aneuploidy tolerance to protein turnover. *Nature* **630**, 149–157 (2024).
38. S. Myhre, *et al.*, Influence of DNA copy number and mRNA levels on the expression of breast cancer related proteins. *Mol Oncol* **7**, 704–718 (2013).
39. T. Geiger, J. Cox, M. Mann, Proteomic Changes Resulting from Gene Copy Number Variations in Cancer Cells. *PLoS Genet* **6**, e1001090 (2010).
40. D. Ascencio, *et al.*, Expression attenuation as a mechanism of robustness against gene duplication. *Proceedings of the National Academy of Sciences* **118**, e2014345118 (2021).
41. K. Ishikawa, K. Makanae, S. Iwasaki, N. T. Ingolia, H. Moriya, Post-Translational Dosage Compensation Buffers Genetic Perturbations to Stoichiometry of Protein Complexes. *PLoS Genet* **13**, e1006554- (2017).
42. S. Yu McLoughlin, S. D. Copley, A compromise required by gene sharing enables survival: Implications for evolution of new enzyme activities. *Proceedings of the National Academy of Sciences* **105**, 13497–13502 (2008).
43. A. B. Morgenthaler, R. K. Fritts, S. D. Copley, Amplicon Remodeling and Genomic Mutations Drive Population Dynamics after Segmental Amplification. *Mol Biol Evol* **39**, msab289 (2022).
44. J. P. Kershner, *et al.*, A Synonymous Mutation Upstream of the Gene Encoding a Weak-Link Enzyme Causes an Ultrasensitive Response in Growth Rate. *J Bacteriol* **198**, 2853 (2016).
45. A. Battesti, N. Majdalani, S. Gottesman, The RpoS-Mediated General Stress Response in Escherichia coli. *Annu Rev Microbiol* **65**, 189–213 (2011).
46. R. Hengge-Aronis, Signal Transduction and Regulatory Mechanisms Involved in Control of the σS (RpoS) Subunit of RNA Polymerase. *Microbiology and Molecular Biology Reviews* **66**, 373–395 (2002).
47. N. Dephoure, *et al.*, Quantitative proteomic analysis reveals posttranslational responses to aneuploidy in yeast. *Elife* **3**, e03023 (2014).
48. S. Du, J. Lutkenhaus, Assembly and activation of the Escherichia coli divisome. *Mol Microbiol* **105**, 177–187 (2017).
49. C. M. Brennan, *et al.*, Protein aggregation mediates stoichiometry of protein complexes in aneuploid cells. *Genes & Development* **33**, 1031–1047 (2019).
50. S. Mueller, *et al.*, Protein degradation corrects for imbalanced subunit stoichiometry in OST complex assembly. *Mol Biol Cell* **26**, 2596–2608 (2015).
51. M. D. Gonzalez, J. Beckwith, Divisome under Construction: Distinct Domains of the Small Membrane Protein FtsB Are Necessary for Interaction with Multiple Cell Division Proteins. *J Bacteriol* **191**, 2815–2825 (2009).
52. C. Pourciau, Y.-J. Lai, M. Gorelik, P. Babitzke, T. Romeo, Diverse Mechanisms and Circuitry for Global Regulation by the RNA-Binding Protein CsrA. *Front Microbiol* **11** (2020).
53. C. A. Vakulskas, A. H. Potts, P. Babitzke, B. M. M. Ahmer, T. Romeo, Regulation of Bacterial Virulence by Csr (Rsm) Systems. *Microbiology and Molecular Biology Reviews* **79**, 193–224 (2015).
54. D. M. Fitzgerald, R. P. Bonocora, J. T. Wade, Comprehensive Mapping of the Escherichia coli Flagellar Regulatory Network. *PLoS Genet* **10**, e1004649 (2014).

55. Z. D. Dalebroux, M. S. Swanson, ppGpp: magic beyond RNA polymerase. *Nat Rev Microbiol* **10**, 203–212 (2012).
56. W. Ebel, Trempy Janine E, Escherichia coli RcsA, a Positive Activator of Colanic Acid Capsular Polysaccharide Synthesis, Functions To Activate Its Own Expression. *J Bacteriol* **181**, 577–584 (1999).
57. V. Stout, A. Torres-Cabassa, M. R. Maurizi, D. Gutnick, S. Gottesman, RcsA, an unstable positive regulator of capsular polysaccharide synthesis. *J Bacteriol* **173**, 1738–1747 (1991).
58. A. S. Torres-Cabassa, S. Gottesman, Capsule synthesis in Escherichia coli K-12 is regulated by proteolysis. *J Bacteriol* **169**, 981–989 (1987).
59. C.-Y. Chang, H.-T. Hu, C.-H. Tsai, W.-F. Wu, The degradation of RcsA by ClpYQ(HslUV) protease in Escherichia coli. *Microbiol Res* **184**, 42–50 (2016).
60. E. Wall, N. Majdalani, S. Gottesman, The Complex Rcs Regulatory Cascade. *Annu Rev Microbiol* **72**, 111–139 (2018).
61. O. Schmidt, *et al.*, prlF and yhaV Encode a New Toxin–Antitoxin System in Escherichia coli. *J Mol Biol* **372**, 894–905 (2007).
62. W. B. Snyder, T. J. Silhavy, Enhanced export of beta-galactosidase fusion proteins in prlF mutants is Lon dependent. *J Bacteriol* **174**, 5661–5668 (1992).
63. E. Gur, R. T. Sauer, Recognition of misfolded proteins by Lon, a AAA+ protease. *Genes Dev* **22**, 2267–2277 (2008).
64. R. T. Sauer, T. A. Baker, AAA+ proteases: ATP-fueled machines of protein destruction. *Annu Rev Biochem* **80**, 587–612 (2011).
65. T. M. Conrad, *et al.*, Whole-genome resequencing of Escherichia coli K-12 MG1655 undergoing short-term laboratory evolution in lactate minimal media reveals flexible selection of adaptive mutations. *Genome Biol* **10**, R118 (2009).
66. K. F. Jensen, The Escherichia coli K-12 “wild types” W3110 and MG1655 have an rph frameshift mutation that leads to pyrimidine starvation due to low pyrE expression levels. *J Bacteriol* **175**, 3401–3407 (1993).
67. K. A. Datsenko, B. L. Wanner, One-step inactivation of chromosomal genes in Escherichia coli K-12 using PCR products. *Proceedings of the National Academy of Sciences* **97**, 6640–6645 (2000).
68. S. Datta, N. Costantino, D. L. Court, A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109–115 (2006).
69. J. Quandt, M. F. Hynes, Versatile suicide vectors which allow direct selection for gene replacement in Gram-negative bacteria. *Gene* **127**, 15–21 (1993).
70. L. Bossi, *et al.*, NusG prevents transcriptional invasion of H-NS-silenced genes. *PLoS Genet* **15**, e1008425 (2019).
71. S. Chen, Y. Zhou, Y. Chen, J. Gu, fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* **34**, i884–i890 (2018).
72. D. E. Deatherage, J. E. Barrick, “Identification of Mutations in Laboratory-Evolved Microbes from Next-Generation Sequencing Data Using breseq” in *Methods in Molecular Biology*, L. Sun, W. Shou, Eds. (Springer New York, 2014), pp. 165–188.
73. J. E. Barrick, *et al.*, Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics* **15**, 1039 (2014).
74. B. Langmead, S. L. Salzberg, Fast gapped-read alignment with Bowtie 2. *Nat Methods* **9**, 357–359 (2012).
75. H. Li, *et al.*, The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**, 2078–2079 (2009).
76. P. Danecek, *et al.*, Twelve years of SAMtools and BCFtools. *Gigascience* **10**, giab008 (2021).

77. G. H. Putri, S. Anders, P. T. Pyl, J. E. Pimanda, F. Zanini, Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics* **38**, 2943–2945 (2022).
78. M. D. Robinson, D. J. McCarthy, G. K. Smyth, edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
79. A. T. L. Lun, Y. Chen, G. K. Smyth, It's DE-licious: A Recipe for Differential Expression Analyses of RNA-seq Experiments Using Quasi-Likelihood Methods in edgeR. *Methods in molecular biology* **1418**, 391–416 (2016).
80. M. E. Ritchie, *et al.*, limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res* **43**, e47–e47 (2015).
81. C. S. Hughes, *et al.*, Ultrasensitive proteome analysis using paramagnetic bead technology. *Mol Syst Biol* **10**, 757 (2014).
82. T. H. Segall-Shapiro, E. D. Sontag, C. A. Voigt, Engineered promoters enable constant gene expression at any copy number in bacteria. *Nat Biotechnol* **36**, 352–358 (2018).
83. D. G. Gibson, *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343–345 (2009).