

1       **Mutations that improve the efficiency of a weak-link enzyme are rare**  
2                   **compared to adaptive mutations elsewhere in the genome**

3

4       Andrew B. Morgenthaler<sup>1,2</sup>, Wallis R. Kinney<sup>1,2</sup>, Christopher C. Ebmeier<sup>1</sup>, Corinne M. Walsh<sup>2,3</sup>,  
5       Daniel J. Snyder<sup>4</sup>, Vaughn S. Cooper<sup>4</sup>, William M. Old<sup>1</sup>, Shelley D. Copley<sup>1,2</sup>

6

7       <sup>1</sup>Department of Molecular, Cellular, and Developmental Biology, University of Colorado

8       Boulder, Boulder, United States

9       <sup>2</sup>Cooperative Institute for Research in Environmental Sciences, University of Colorado Boulder,  
10      Boulder, United States

11      <sup>3</sup>Department of Ecology and Evolutionary Biology, University of Colorado Boulder, Boulder,  
12      United States

13      <sup>4</sup>Department of Microbiology and Molecular Genetics, and Microbial Genomic Sequencing  
14      Center, University of Pittsburgh, Pittsburgh, United States

15

16      **Abstract**

17           New enzymes often evolve by amplification and divergence of genes encoding enzymes  
18      with a weak ability to provide a new function. Experimental studies to date have followed the  
19      evolutionary trajectory of an amplified gene, but have not addressed other mutations in the  
20      genome when fitness is limited by an evolving gene. We have adapted *Escherichia coli* in which  
21      an enzyme's weak secondary activity has been recruited to serve an essential function. While the  
22      gene encoding the "weak-link" enzyme amplified in all eight populations, mutations improving  
23      the new activity occurred in only one. This beneficial allele quickly swept the amplified array,

24 displacing the parental allele. Most adaptive mutations, however, occurred elsewhere in the  
25 genome. We have identified the mechanisms by which three of the classes of mutations increase  
26 fitness. These mutations may be detrimental once a new enzyme has evolved, and require  
27 reversion or compensation, leading to permanent changes in the genome.

28

## 29 **Introduction**

30 The expansion of huge superfamilies of enzymes, transcriptional regulators, transporters,  
31 and signaling molecules from single ancestral genes has been a dominant process in the  
32 evolution of life (Bergthorsson et al., 2007; Chothia et al., 2003; Glasner et al., 2006; A. L.  
33 Hughes, 1994; Ohno, 1970; Todd et al., 2001). The emergence of new protein family members  
34 has enabled organisms to access new nutrients, sense new stimuli, and respond to changing  
35 conditions with ever more sophistication (Bouquin & Khila, 2017; Conant & Wolfe, 2008; Nei &  
36 Rooney, 2005; Reams & Neidle, 2004; Starr et al., 2017; Storz, 2016).

37 The Innovation-Amplification-Divergence (IAD) model (Figure 1) posits that evolution  
38 of new genes by duplication and divergence begins when a physiologically irrelevant side  
39 activity of a protein, known as a promiscuous activity (Copley, 2003, 2017), becomes important  
40 for fitness due to a mutation or environmental change (Bergthorsson et al., 2007; Francino, 2005;  
41 A. L. Hughes, 1994; Näsvall et al., 2012). Often this newly useful activity is inefficient, making  
42 the enzyme the “weak-link” in metabolism. Gene duplication/amplification provides a ready  
43 mechanism to improve fitness by increasing the abundance of a weak-link enzyme. If further  
44 mutations lead to evolution of an enzyme capable of efficiently carrying out the newly needed  
45 function, selective pressure to maintain a high copy number will be removed, allowing extra

46 copies to be lost and leaving behind two paralogs (or just one gene encoding a new enzyme if the  
47 original function is no longer needed).

48 While the IAD model provides a satisfying theoretical framework for the process of gene  
49 duplication and divergence, our understanding of the process is far from perfect. Although the  
50 signatures of gene duplication and divergence are obvious in extant genomes, we have little  
51 information about the genome contexts and environments in which new enzymes arose.

52 Laboratory evolution offers the possibility of tracking this process in real time. In a landmark  
53 study, Näsvall et al. used laboratory evolution to demonstrate that a gene encoding an enzyme  
54 with two inefficient activities required for synthesis of histidine and tryptophan amplified and  
55 diverged to alleles encoding two specialists within 2000 generations (Näsvall et al., 2012).

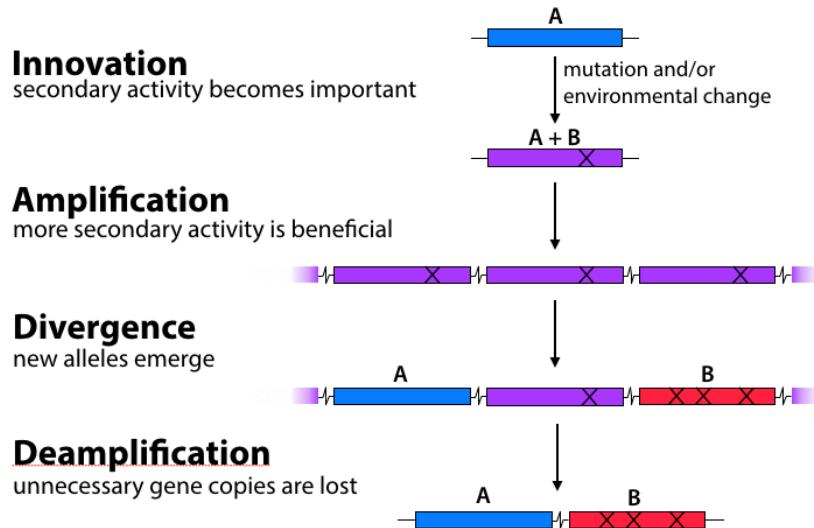
56 However, this study followed only mutations in the diverging gene. In general, when an  
57 organism is exposed to a novel selection pressure that requires evolution of a new enzyme, any  
58 mutation – either in the gene encoding the weak-link enzyme or elsewhere in the genome – that  
59 improves fitness will provide a selective advantage.

60 We have generated a model system in *E. coli* for exploring the relative importance of  
61 mutations in a gene encoding a weak-link enzyme and elsewhere in the genome. ProA ( $\gamma$ -  
62 glutamyl phosphate reductase, Figure 2) is essential for proline synthesis in *E. coli*. ArgC (*N*-  
63 acetylglutamyl phosphate reductase) catalyzes a similar reaction in the arginine synthesis  
64 pathway, although the two enzymes are not homologous (Goto et al., 2003; Ludovice et al.,  
65 1992; Page et al., 2003). ProA can reduce *N*-acetylglutamyl phosphate, but its activity is too  
66 inefficient to support growth of a  $\Delta argC$  strain of *E. coli* in glucose. However, a point mutation  
67 that changes Glu383 to Ala allows slow growth of the  $\Delta argC$  strain in glucose. Enzymatic assays  
68 show that E383A ProA (ProA\*) has severely reduced activity with  $\gamma$ -glutamyl semialdehyde

69 (GSA), but substantially improved activity with *N*-acetylglutamyl semialdehyde (NAGSA)  
70 (Khanal et al., 2015; McLoughlin & Copley, 2008). (It is necessary to assay kinetic parameters  
71 in the reverse direction because the substrates for the forward reaction are too unstable to prepare  
72 and purify.) Glu383 is in the active site of the enzyme; the change to Ala may create extra room  
73 in the active site to accommodate the larger substrate for ArgC, but at a cost to the ability to bind  
74 and orient the native substrate. The poor efficiency of the weak-link ProA\* creates strong  
75 selective pressure for improvement of both proline and arginine synthesis during growth of  
76  $\Delta$ argC *E. coli* on glucose as a sole carbon source.

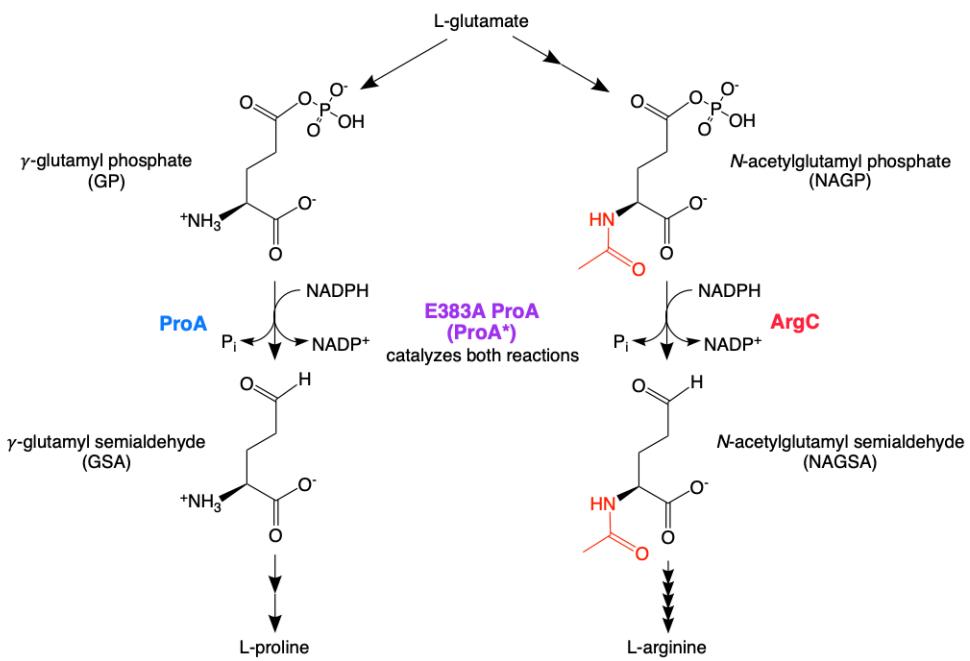
77 We adapted eight replicate populations of  $\Delta$ argC *proA\** *E. coli* in minimal medium  
78 supplemented with glucose and proline for up to 1000 generations to identify mechanisms by  
79 which the impairment in arginine synthesis could be alleviated. Our expectation that  
80 amplification of *proA\** would be beneficial was borne out in all populations. However, whole-  
81 genome sequencing of the adapted populations showed that a mutation in *proA\** occurred in only  
82 one population. Indeed, most of the adaptive mutations occurred outside of *proA\**. We have  
83 identified the mechanisms by which three common classes of mutations allow a relatively rapid  
84 increase in fitness while the more difficult process of improving the weak-link enzyme  
85 progresses. However, they increase fitness at a cost to presumably well-evolved functions.

86 Our results demonstrate that mutations elsewhere in the genome play an important role  
87 during the process of gene amplification and divergence when the inefficient activity of a weak-  
88 link enzyme limits fitness. Thus, the process of evolution of a new enzyme by gene duplication  
89 and divergence is inextricably intertwined with mutations elsewhere in the genome that improve  
90 fitness by other mechanisms.



91

**Figure 1.** The Innovation-Amplification-Divergence (IAD) model of gene evolution. A physiologically irrelevant side activity B of an enzyme may become physiologically relevant due to a mutation or environmental change. Gene amplification increases the abundance of the weak-link enzyme. Acquisition of mutations can improve the efficiency of the newly important activity B. Once sufficient B activity is achieved, selection is relaxed and extra gene copies are lost, leaving behind two paralogs.



92 **Figure 2.** Reactions catalyzed by ProA ( $\gamma$ -glutamyl phosphate reductase) in proline synthesis and ArgC ( $N$ -acetylglutamyl phosphate reductase) in arginine synthesis. E383A ProA (ProA\*) can catalyze both reactions, albeit poorly.

93 **Results**

94 **Growth rate of  $\Delta argC proA^*$  *E. coli* increased 3-fold within a few hundred generations of  
95 adaptation in M9/glucose/proline**

96 We generated a progenitor strain for laboratory evolution by replacing *argC* with the *kan*<sup>r</sup>  
97 antibiotic resistance gene, modifying *proA* to encode ProA\*, and introducing a mutation in the -  
98 10 region of the promoter of the *proBA* operon that was previously shown to occur frequently  
99 during adaptation of the  $\Delta argC$  strain (Kershner et al., 2016). We also introduced *yfp*  
100 downstream of *proA*\* and deleted several genes (*fimAICDFGH* and *csgBAC*, which are required  
101 for the formation of fimbriae and curli, respectively (Barnhart & Chapman, 2006; Proft & Baker,  
102 2009)) to minimize the occurrence of biofilms. We evolved eight parallel lineages of this strain  
103 (AM187, Table 1) in M9 minimal medium supplemented with 0.2% glucose, 0.4 mM proline,  
104 and 20 µg/mL kanamycin in a turbidostat to identify mutations that improve synthesis of  
105 arginine. We used a turbidostat rather than a serial transfer protocol because turbidostats can  
106 maintain cultures in exponential phase and thereby avoid selection for mutations that simply  
107 decrease lag phase or improve survival in stationary phase. Turbidostats also avoid population  
108 bottlenecks during serial passaging that can result in loss of genetic diversity.

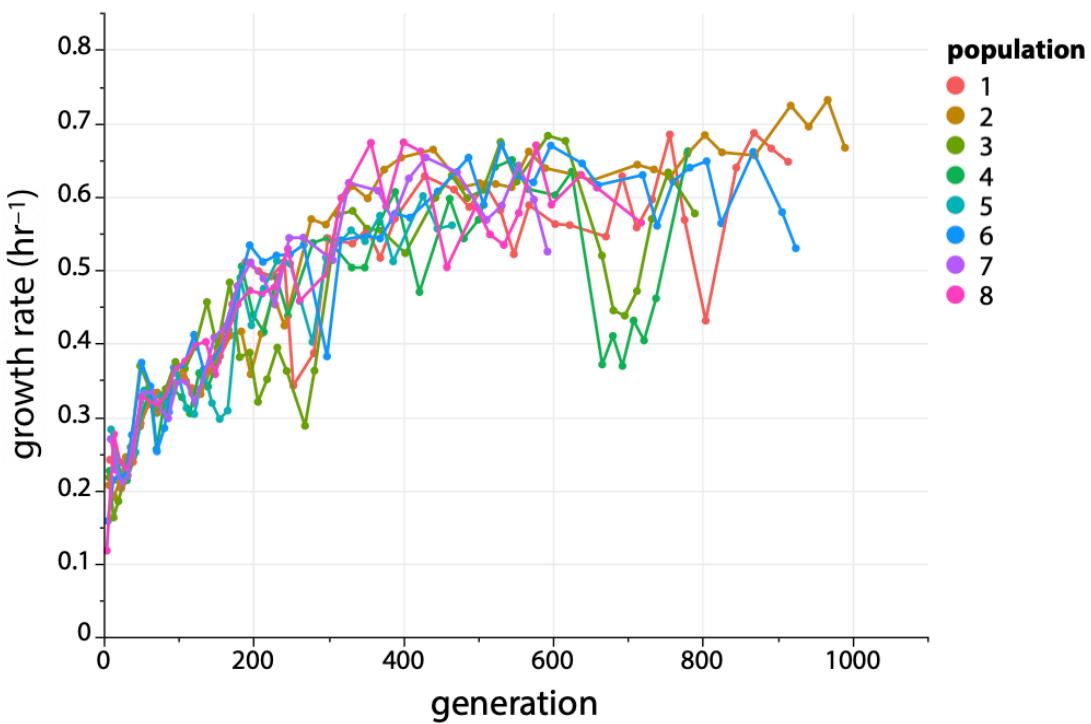
109 Growth rate in each culture tube was averaged over each 24-hour period and was used to  
110 calculate the number of generations each day. Each culture was maintained until a biofilm  
111 formed (33-57 days, corresponding to 470-1000 generations). While it is possible to restart  
112 cultures from individual clones after biofilm formation, this practice introduces a severe  
113 population bottleneck. Thus, we decided to stop the adaptation for each population when a  
114 biofilm formed. For this reason, every population was adapted for a different number of  
115 generations.

**Table 1.** Strains used in this work.

strain	genotype	notes
<b>AM008</b>	<i>Escherichia coli</i> BW25113; <i>argC::kan</i> <sup>r</sup>	Keio strain (Baba et al., 2006)
<b>AM187</b>	AM008 + -45 C→T [M2 promoter mutation from (Kershner et al., 2016)] upstream of <i>proB</i> + A1148→C (changes Glu383 to Ala) in <i>proA</i> + <i>yfp</i> construct downstream of <i>proBA</i> consisting of (in order from 5' to 3') BBa_B0015 terminator, P3 promoter, synthetic RBS, <i>yfp</i> (see Materials and Methods); Δ <i>fimAICDFGH</i> ; Δ <i>csgBAC</i>	parental strain for adaptation, GenBank accession number CP037857
<b>AM209</b>	<i>Escherichia coli</i> BL21(DE3); <i>argC::kan</i> <sup>r</sup> ; <i>proA::cat</i>	expression and purification of wild-type and mutant ProAs
<b>AM239</b>	AM187 + 58 bp deletion upstream of <i>argB</i> (pos. 4145856-4145913) <sup>a</sup>	
<b>AM241</b>	AM187 + C4145901→G (24 bp upstream of <i>argB</i> start codon)	
<b>AM242</b>	AM187 + C4145903→A (22 bp upstream of <i>argB</i> start codon)	
<b>AM243</b>	AM187 + C4145903→T (22 bp upstream of <i>argB</i> start codon)	
<b>AM244</b>	AM187 + C4145907→A (18 bp upstream of <i>argB</i> start codon)	
<b>AM245</b>	AM187 + 38 bp duplication upstream of <i>argB</i> (pos. 4145912-4145949)	
<b>AM267</b>	<i>Escherichia coli</i> BL21; <i>carAB::kan</i> <sup>r</sup>	expression and purification of wild-type and mutant carbamoyl phosphate synthetases
<b>AM279</b>	AM187 + C1169→T in <i>ygcB</i> (changes Ala390 to Val in Cas3)	
<b>AM320</b>	AM187 + T1116→G in <i>proA</i> * (changes Phe372 to Leu)	
<b>AM327</b>	AM187 + 82 bp deletion upstream of <i>pyrE</i> (pos. 3808881-3808962)	
<b>AM329</b>	AM187 + 82 bp deletion upstream of <i>pyrE</i> (pos. 3808881-3808962); C1169→T <i>ygcB</i> (changes Ala390 to Val in Cas3)	

116     <sup>a</sup> Genome positions refer to the sequence of strain AM187 (GenBank accession number CP037857), which was  
 117     modified from the *E. coli* BW25113 sequence (GenBank accession number CP009273 (Grenier et al., 2014)) based  
 118     on the mutations that had been introduced.

119 Over the course of the experiment, growth rate increased 2.5-3.5-fold for all eight  
120 populations (Figure 3). This improvement corresponds to a change in doubling time from ~3.3  
121 hours to ~1 hour. As expected, the rate at which growth rate increased slowed over the course of  
122 adaptation. Occasional dips in growth rate occurred during the adaptation (Figure 3). These dips  
123 are artifacts arising from temporary aberrations in selective conditions due to turbidostat  
124 malfunctions that prevented introduction of fresh medium, causing the cultures to enter  
125 stationary phase. Occasionally cultures were saved as frozen stocks until the turbidostat was  
126 fixed (see Materials and Methods). Restarting cultures from frozen stocks may have caused a  
127 temporary drop in growth rate.

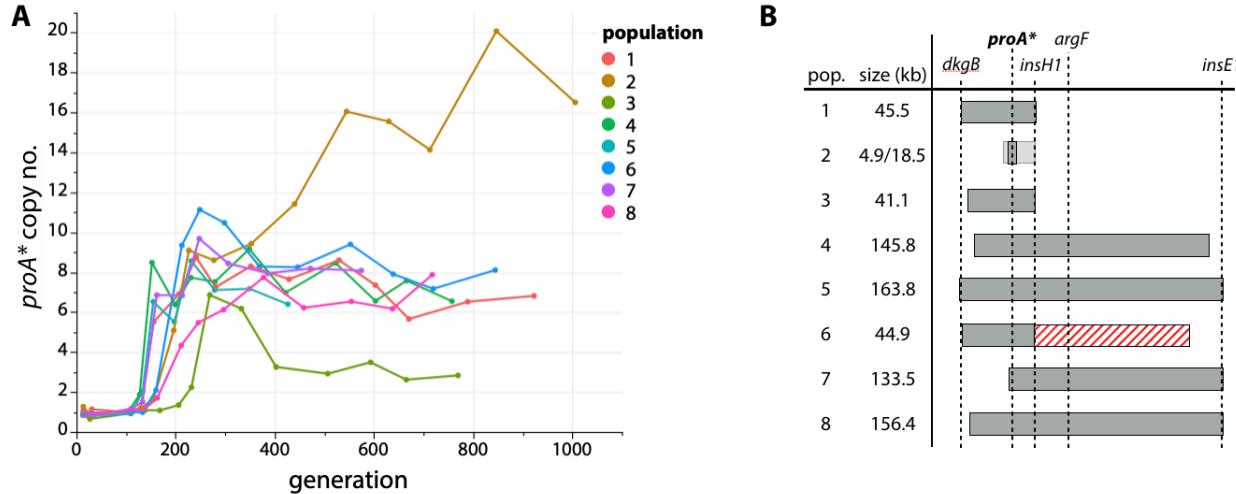


**Figure 3.** Growth rate increases ~3-fold during adaptation of  $\Delta argC$  M2-*proA*\* *E. coli* in M9 minimal medium containing 0.2% glucose, 0.4 mM proline and 20  $\mu$ g/mL kanamycin.

128     **The copy number and extent of amplification around *proA*\* varied among replicate  
129     populations**

130           We monitored *proA*\* copy number during the adaptation experiment using qPCR of  
131     population genomic DNA (Figure 4A). *proA*\* was present in at least 6 copies by generation 300  
132     in all eight populations. Six of the populations maintained 6-9 copies for the remainder of the  
133     adaptation. In contrast, *proA*\* copy number in population 2 increased to as many as 20 copies. In  
134     population 3, *proA*\* copy number dropped to three by generation 400.

135           We identified the boundaries of the amplified regions in all eight populations by  
136     sequencing population genomic DNA (Figure 4B, Supplementary File 1). The amplified region  
137     in population 2 was unusually small, spanning only 4.9 kb and resulting in co-amplification of  
138     only two other genes besides *proBA*\*. Population 2 also appeared to have a second region of  
139     amplification of 18.5 kb. (Whether these two distinct amplification regions coexisted in the same  
140     clone or as two separate clades within the population could not be determined from population  
141     genome sequencing.) In contrast, the amplified regions in the other seven populations ranged  
142     from 41.1 to 163.8 kb, encompassing between 55 and 177 genes. We attribute the variation in  
143     *proA*\* copy number to these differences in the size of the amplified region on the genome. The  
144     population with the smallest amplified region (4.9 kb, population 2) carries fewer multicopy  
145     genes and thus should incur a lower fitness cost, allowing *proA*\* to reach a higher copy number  
146     (Adler et al., 2014; Kugelberg et al., 2006; Pettersson et al., 2009; Reams et al., 2010).



**Figure 4.** *proA\** is amplified during adaptation. (A) *proA\** copy number in each adapted population as measured by qPCR. (B) Regions of genomic amplification in each adapted population based population-wide, whole-genome sequencing. Population 2 had two overlapping regions of amplification, both of which included *proA\** (shown as differently shaded bars). Population 6 had a 95.1 kb deletion (shown as a red striped bar) immediately downstream of the amplified region.

147

148 **A mutation in *proA\** led to deamplification in population 3**

149 The decrease in *proA\** copy number in population 3 was particularly noteworthy since it  
 150 might have been an indication that a mutation had improved the neo-ArgC activity of ProA\*,  
 151 resulting in a decreased need for multiple copies. In fact, a mutation in *proA\** that changes  
 152 Phe372 to Leu (Figure 5A) was observed in population 3. E383A F372L ProA will be designated  
 153 ProA\*\* hereafter. Introduction of this mutation into the parental strain (which carried *proA\**)  
 154 increased growth rate by 69% (Figure 5B), confirming that the mutation is adaptive and not a  
 155 random hitchhiker. In contrast, no mutations in *proA\** were identified in any of the other  
 156 populations.

157 The neo-ArgC and native ProA activities of wild-type, ProA\*, and ProA\*\* were assayed  
 158 (in the reverse direction) with NAGSA and GSA, respectively (Table 2). The  $k_{cat}/K_M, \text{NAGSA}$  for

159 ProA\*\* is 3.6-fold higher than that of ProA\* and nearly 80-fold higher than that for ProA. In  
160 contrast, there is no difference between  $k_{cat}/K_{M,GSA}$  for ProA\* and ProA\*\*.

161 To determine when the mutation that changes Phe372 to Leu in ProA\* occurred, we  
162 performed sequencing of population genomic DNA at generations 270, 440, 630 and at the end  
163 of the adaptation (Figure 5C, 130x, 122x, 70x and 81x sequencing depth, respectively). *proA*\*\*  
164 was present in 9% of the sequencing reads of the population by generation 270. By the time  
165 deamplification of *proA*\* had occurred at generation 440, the frequency of *proA*\*\* had risen to  
166 21% of sequencing reads. By the end of the adaptation, *proA*\*\* was fixed in the population, yet 3  
167 copies remained in the genome, suggesting that ProA\*\* does not have sufficient neo-ArgC  
168 activity to be present at a single copy in the genome.

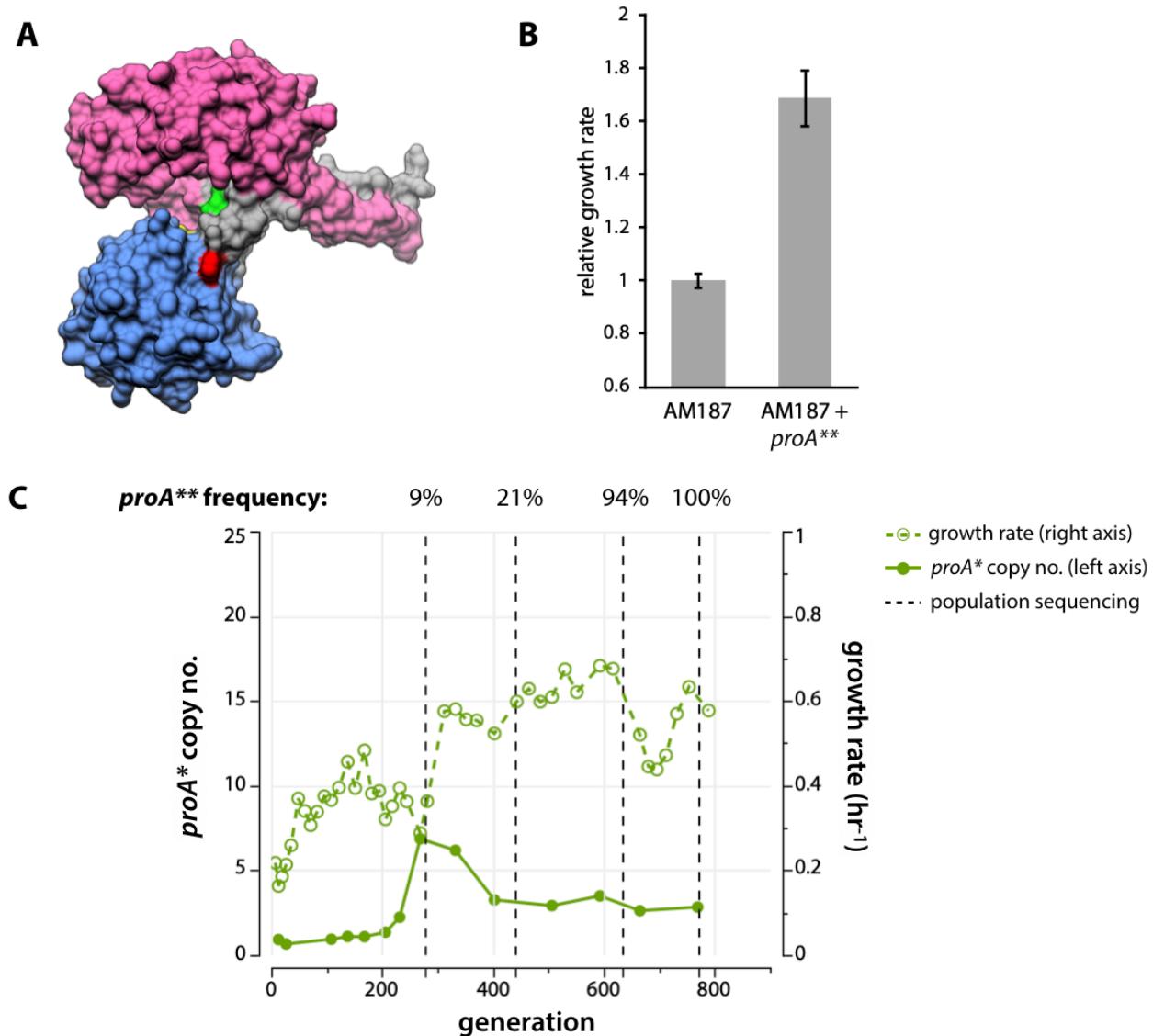
169 The fact that a mutation that improved the neo-ArgC activity of ProA\* occurred in only  
170 one population was surprising considering that ProA\* is the weak-link enzyme limiting growth  
171 rate. Because the growth rates of all 8 populations improved substantially (Figure 3), mutations  
172 outside of the *proBA*\* operon must also be contributing to fitness.

173

**Table 2.** Kinetic parameters for GSA and NAGSA dehydrogenase activities of ProA, ProA\*, and ProA\*\*.

	GSA activity (ProA)			NAGSA activity (neo-ArgC)		
	$k_{cat}$ ( $s^{-1}$ )	$K_M$ (mM)	$k_{cat}/K_{M,GSA}$ ( $M^{-1} s^{-1}$ )	$k_{cat}$ ( $s^{-1}$ )	$K_M$ (mM)	$k_{cat}/K_{M,NAGSA}$ ( $M^{-1} s^{-1}$ )
WT	$16 \pm 0.3$	$0.22 \pm 0.01$	$72000 \pm 2000$	$0.0083 \pm 0.0009$	$0.30 \pm 0.09$	$28 \pm 9$
ProA* (E383A)	$0.0076 \pm 0.0008$	$0.20 \pm 0.04$	$37 \pm 8$	$0.046 \pm 0.002$	$0.076 \pm 0.009$	$610 \pm 74$
ProA** (E383A F372L)	$0.023 \pm 0.005$	$0.42 \pm 0.14$	$55 \pm 22$	$0.21 \pm 0.01$	$0.095 \pm 0.011$	$2200 \pm 260$

174 <sup>a</sup> Values reported were calculated from a nonlinear least squares regression of three replicates at each substrate  
175 concentration  $\pm$  standard error.



**Figure 5.** *proA\** acquired a beneficial mutation in population 3. (A) Crystal structure of *Thermotoga maritima* ProA (PDB 1O20) (Page et al., 2003). Yellow, catalytic cysteine; green, equivalent of *E. coli* ProA Glu383; red, equivalent of *E. coli* ProA Phe372; magenta, NADPH-binding domain; blue, catalytic domain; gray, oligomerization domain. (B) Change in growth rate when the mutation changing Phe372 to Leu is introduced into the genome of the parental strain (AM187). Error bars represent one standard error from the mean,  $N = 4$ . (C) Growth rate (right axis, dotted lines) and *proA\** copy number (left axis, solid lines) for population 3. Vertical dotted lines indicate when population genomic DNA was sequenced. The frequency of the *proA\*\** allele at each time point is noted above the plot.

176     **Mutations outside of *proA\** improved fitness**

177              Population genome sequencing at the end of the experiment and at several intermediate  
178              timepoints from populations with longer adaptations (Figure S1) revealed that the final  
179              populations contained 13-178 mutations present at frequencies  $\geq 5\%$ , 3-5 mutations at  
180              frequencies  $\geq 30\%$ , and 1-4 mutations (not including amplification of *proA\**) that were fixed  
181              (100%). We found several mutations in the same genes in different populations, suggesting that  
182              these mutations confer a fitness advantage (Supplementary File 1).

183              The first mutation to appear in all populations was either an 82 bp deletion in the *rph*  
184              pseudogene directly upstream of *pyrE* or a C→T mutation in the intergenic region between *rph*  
185              and *pyrE*. PyrE is required for *de novo* synthesis of pyrimidine nucleotides. Both of these  
186              mutations have arisen in other *E. coli* evolution experiments, and have been shown to restore a  
187              known PyrE deficiency in the BW25113 *E. coli* strain (Blank et al., 2014; Bonekamp et al.,  
188              1984; Conrad et al., 2009; Jensen, 1993; Knöppel et al., 2018). Thus, these mutations are general  
189              adaptations to growth in minimal medium and do not pertain to the selective pressures caused by  
190              the weak-link enzyme ProA\*.

191              A second early mutation in four of the populations occurred in *ygcB*. This mutation  
192              changes Ala390 to Val in Cas3, a nuclease/helicase in the Type I CRISPR/Cas system in *E. coli*  
193              (Howard et al., 2011). We introduced this mutation into the genome of the parent AM187 and  
194              compared the growth rates of the mutant and AM187 (Figure S2). Surprisingly, we saw no  
195              significant change in growth rate. Since this mutation appeared about the same time as the  
196              mutations upstream of *pyrE*, we wondered whether the *ygcB* mutation might only improve  
197              growth rate in the context of restored *pyrE* expression. Thus, we also tested the growth rate of a  
198              strain with the Cas3 mutation and the 82 bp deletion upstream of *pyrE*. Again, we saw no

199 significant change in relative growth rate (Figure S2). Thus, the *ygcB* mutation is most likely a  
200 neutral hitchhiker. The most likely explanation for its prevalence is that it was present in a clade  
201 of the parental population that later rose to a high frequency when an additional beneficial  
202 mutation was acquired by one of its members.

203

204 **Mutations upstream of *argB* increase ArgB abundance**

205 All eight final populations contained mutations in the intergenic region upstream of *argB*  
206 and downstream of *kan<sup>r</sup>*. These mutations were fixed in two populations, and present at  
207 frequencies of 9-82% in the other populations (Figure 6A).

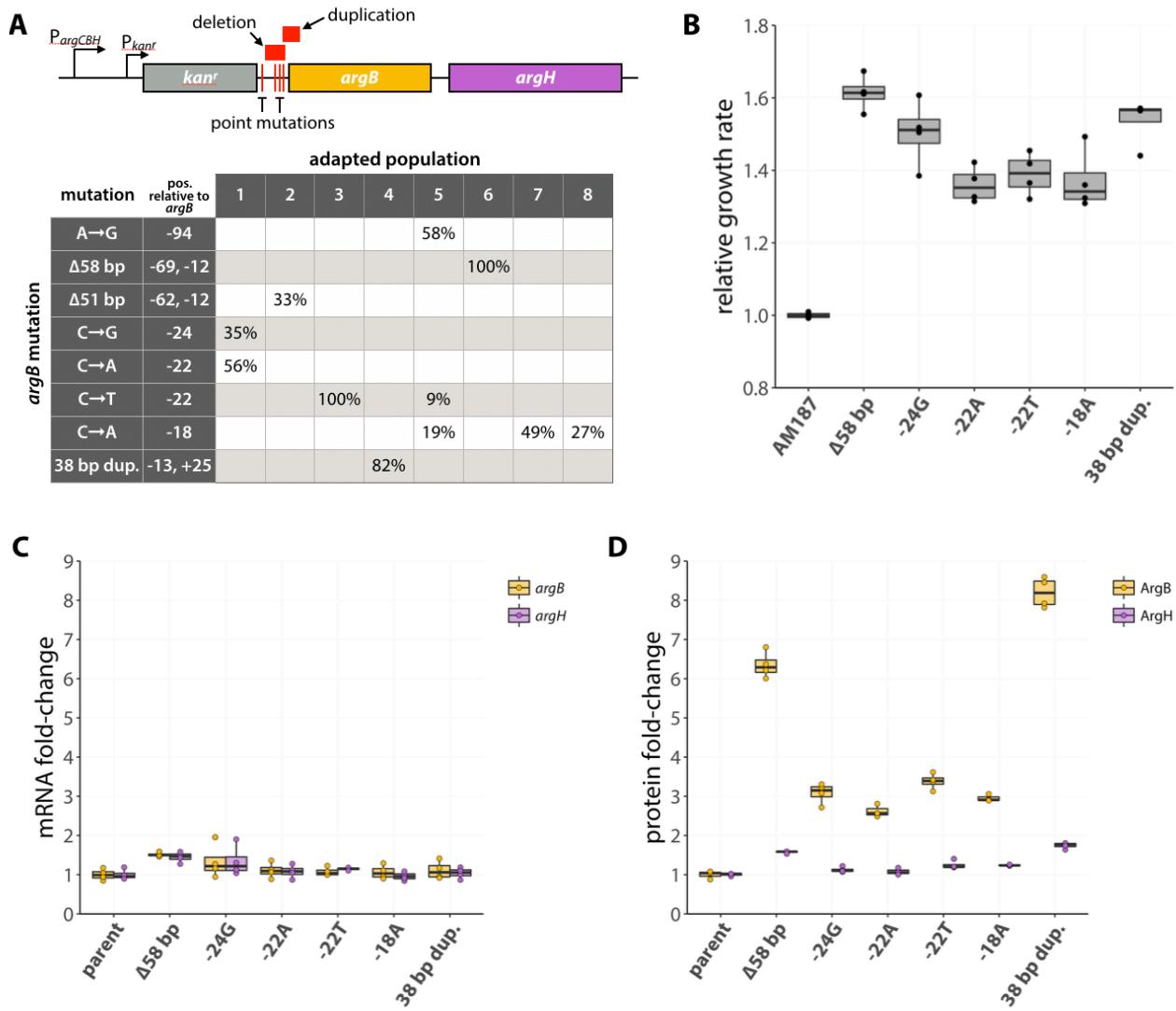
208 We reintroduced six of the mutations upstream of *argB* into the parental strain AM187.  
209 The mutations increased growth rate by 36-61% (Figure 6B). Levels of mRNAs for *argB* and  
210 *argH*, which is immediately downstream of *argB*, were little affected by the mutations (Figure  
211 6C). However, levels of ArgB protein measured by label-free high resolution Orbitrap mass  
212 spectrometry increased 2.6-8.2-fold (Figure 6D). In contrast, ArgH levels increased only  
213 modestly. These data suggest that the mutations upstream of *argB* increase translational  
214 efficiency of *argB* mRNA.

215 Increased translational efficiency can be due to increased accessibility of the region  
216 surrounding the Shine-Dalgarno site and the start codon, which might be achieved in three ways:  
217 1) a decrease in the stability of the lowest-free-energy secondary structure in this region (Bentele  
218 et al., 2013; Espah Borujeni et al., 2014; Goodman et al., 2013); 2) a change in the ensemble of  
219 mRNA structures such that the Shine-Dalgarno sequence is more likely to be single-stranded in  
220 multiple accessible structures (Espah Borujeni et al., 2014; Salis et al., 2009); or 3) a decrease in  
221 the folding rate of the mRNA such that a ribosome can bind to the unfolded mRNA emerging

222 behind a preceding ribosome before the mRNA folds and obscures the Shine-Dalgarno sequence  
223 (a process called ribosomal drafting, (Espah Borujeni & Salis, 2016)). The *argB* mRNA, like  
224 16% of  $\gamma$ -proteobacterial mRNAs (Scharff et al., 2011), lacks a canonical Shine-Dalgarno  
225 sequence, but the ribosome is expected to bind to a region encompassing the start codon and at  
226 least the upstream 8-10 nucleotides. We calculated the minimum free energy secondary  
227 structures of the 140-nucleotide RNA sequences encompassing the intergenic region affected by  
228 the various mutations through 33 bp downstream of the *argB* start codon (Figure S3). Direct  
229 comparisons of the predicted free energies of these structures are not particularly informative,  
230 especially for mutants with large deletions or duplications. However, for 5 of the 8 of the mutant  
231 structures, the probability that the 5'-UTR from about 20 nt upstream of the start codon is  
232 sequestered in a stem-loop in the lowest free-energy structure is decreased relative to the parent  
233 (Figure S3B); the increased accessibility of this region should increase translation efficiency.  
234 However, for three cases (-94 A→G, -22 C→A, and -18 C→A), this region is equally or more  
235 likely to be sequestered in a stem-loop. The thermodynamic stability of this region is clearly not  
236 the only factor responsible for the effects of the mutations upstream of *argB*.

237 We also considered the possibility that mutations upstream of *argB* might increase  
238 expression by increasing ribosome drafting. Figure S4 shows the predicted folding times of RNA  
239 sequences encompassing 30 nucleotides downstream to 30 nucleotides upstream of the *argB* start  
240 codon (totaling 63 nucleotides) for each mutant except the -94 A→G mutant. Folding is  
241 predicted to be significantly slower for three of the seven mutant RNAs, which should increase  
242 the efficiency of translation. For four of the mutants, however, folding rate is similar to or even  
243 faster than that of the parental mRNA. For three of these mutants (the 58 bp deletion, -22 C→T,  
244 and the 38 bp duplication), the secondary structure prediction shown in Figure S3 suggests that

245 the translation initiation region is less likely to be sequestered in a hairpin. Thus, the effects of 6  
246 of the 8 mutations upstream of *argB* can be explained by a decrease in secondary structure  
247 stability around the ribosome binding site or a decrease in the folding rate of the mRNA in this  
248 region, or both. The effects of the -94 A→G and -22 C→A mutations, however, cannot be  
249 explained by either of these mechanisms.



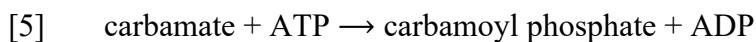
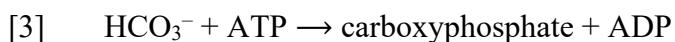
**Figure 6.** Several adaptive mutations occurred upstream of *argB*. (A) Locations of adaptive mutations (red) found upstream of *argB* (yellow) and *argH* (purple). *argC* was replaced with *kan<sup>r</sup>* in the parental strain AM187, giving this operon two promoters, one native to the operon (*P<sub>argCBH</sub>*), and the other introduced with the *kan<sup>r</sup>* gene (*P<sub>kanr</sub>*). The table shows the percentages of each adapted population that contained a given *argB* mutation at the final time point. Six of the *argB* mutations were introduced into the genome of the ancestral strain and changes in growth rate (B), gene expression (C), and protein abundance (D) were determined.

251     **Mutations in *carB* either increase activity or impact allosteric regulation**

252         We found eight different mutations in *carB* in six of the evolved populations: four  
253         missense mutations, three deletions ( $\geq 12$  bp), and one 21 bp duplication (Figure 7A). CarB, the  
254         large subunit of carbamoyl phosphate synthetase (CPS), forms a complex with CarA to catalyze  
255         production of carbamoyl phosphate from glutamine, bicarbonate, and two molecules of ATP (Eq.  
256         [1]) (Rubino et al., 1986, 1987).



257         Synthesis of carbamoyl phosphate involves four reactions that take place in three separate active  
258         sites connected by a molecular tunnel of  $\sim 100$  Å in length (Thoden et al., 2002). CarA catalyzes  
259         hydrolysis of glutamine to glutamate and ammonia (Eq. [2]), while CarB phosphorylates  
260         bicarbonate to form carboxyphosphate in its first active site (Eq. [3]). Ammonia from the CarA  
261         active site is channeled to CarB, where it reacts with carboxyphosphate to form carbamate (Eq.  
262         [4]). Carbamate migrates to a second active site within CarB where it reacts with ATP to form  
263         carbamoyl phosphate and ADP (Eq. [5]).



264         Carbamoyl phosphate feeds into both the pyrimidine and arginine synthesis pathways and its  
265         production is regulated in response to intermediates or products of both pathways (Figure 7B), as  
266         well as by IMP (a purine) (Pierrat & Raushel, 2002). It is inhibited by UMP (a pyrimidine) and  
267         moderately activated by IMP. UMP and IMP compete to bind the same region of CarB (Eroglu  
268         & Powers-Lee, 2002). The allosteric effects of UMP and IMP are dominated, however, by  
269         activation by ornithine. Ornithine, an intermediate in arginine synthesis that reacts with

270 carbamoyl phosphate, binds to and activates CarB even when UMP is bound (Figure 7C)  
271 (Braxton et al., 1999; Eroglu & Powers-Lee, 2002).

272 Seven of the eight mutations found in *carB* affect residues in the allosteric domain of  
273 CarB. The lone mutation outside of the allosteric domain changes Gly369, which is immediately  
274 adjacent to the allosteric region in the 3D structure, to Val (Figure 7C).

275 The kinetic parameters for CPS activity [determined as the glutamine- and bicarbonate-  
276 dependent ATPase activity (Eq. [1])] of all eight CPS mutants are shown in Table 3. All  
277 mutations decreased  $k_{cat}/K_{m,ATP}$  by 34-63%, with the exception of the mutation that changes  
278 Lys966 to Glu, which nearly doubles  $k_{cat}/K_{m,ATP}$ . None of the mutations affected the enzyme's  
279 ability to couple ATP hydrolysis with carbamoyl phosphate production (Figure S5).

280

**Table 3.** Kinetic parameters<sup>a</sup> for the glutamine-dependent ATPase reaction of wild-type and variant carbamoyl phosphate synthetases.

Enzyme	$K_{M,ATP}$ (mM)	$k_{cat}$ (s <sup>-1</sup> )	$k_{cat}/K_{M,ATP}$ (M <sup>-1</sup> s <sup>-1</sup> )	UMP $K_d$ (μM)	UMP <i>a</i>	ornithine $K_d$ (μM)	ornithine <i>a</i>
WT	1.05 ± 0.08	13.5 ± 0.3	12.9 (± 1.0) ×10 <sup>3</sup>	0.81 ± 0.04	0.27 ± 0.01	130 ± 7	3.28 ± 0.03
G369V	3.31 ± 0.25	21.5 ± 0.7	6.51 (± 0.54) ×10 <sup>3</sup>	1.53 ± 0.29	0.48 ± 0.02	372 ± 20	11.8 ± 0.13
L960P	1.12 ± 0.05	9.10 ± 0.15	8.12 (± 0.41) ×10 <sup>3</sup>	na	na	na	na
L964Q	1.25 ± 0.08	8.04 ± 0.17	6.41 (± 0.42) ×10 <sup>3</sup>	na	na	na	na
K966E	0.97 ± 0.06	20.6 ± 0.4	21.2 (± 1.4) ×10 <sup>3</sup>	0.61 ± 0.04	0.21 ± 0.01	181 ± 34	3.23 ± 0.08
Δ12 bp (at nt 2906) <sup>b</sup>	1.09 ± 0.06	4.80 ± 0.09	4.39 (± 0.25) ×10 <sup>3</sup>	na	na	na	na
Δ132 bp (at nt 2986)	1.16 ± 0.06	6.47 ± 0.11	5.57 (± 0.31) ×10 <sup>3</sup>	na	na	na	na
Δ12 bp (at nt 3108)	1.30 ± 0.10	5.86 ± 0.16	4.51 (± 0.37) ×10 <sup>3</sup>	na	na	na	na
21 bp dup. (at nt 3130)	1.40 ± 0.12	9.70 ± 0.30	6.94 (± 0.64) ×10 <sup>3</sup>	597 ± 133	0.51 ± 0.03	na	na

281 <sup>a</sup> Values reported were calculated from a nonlinear least squares regression of data for three replicates at each  
282 substrate concentration ± standard error. Values for  $K_d$  and *a* for UMP and ornithine were determined by fitting the  
283 data to the following equation:  $v/v_0 = (aL + K_d)/(L + K_d)$ , where L is the ligand concentration, v is the initial reaction

284 rate,  $v_0$  is the initial reaction rate in the absence of ligand,  $a$  is  $v/v_0$  at infinite  $L$ , and  $K_d$  is the apparent dissociation  
285 constant. No  $K_d$  or  $a$  values are given (indicated by *na*) when the inhibition by the allosteric ligand was too weak to  
286 measure.

287 <sup>b</sup>Nucleotide (nt) numbers refer to the position of deletions or duplications in *carB*.

288

289 We also measured the effect of mutations on UMP inhibition and ornithine activation of  
290 CPS (Table 3, Figure 7D-E). Regulation of the K966E variant, the enzyme for which  $k_{cat}/K_{m,ATP}$   
291 was nearly doubled, was minimally affected. Five of the variants showed a complete loss of  
292 allosteric regulation by both UMP and ornithine. The variant with the 21 bp duplication retained  
293 modest inhibition by UMP (~50% reduction of turnover at high UMP concentrations), but only  
294 at very high concentrations of UMP; the apparent  $K_{d,UMP}$  was increased by 740-fold. Similarly,  
295 G369V CPS retained partial inhibition by UMP (~50% reduction of turnover at high UMP  
296 concentrations). The apparent  $K_{d,UMP}$  of the G369V enzyme was only doubled, but this variant  
297 showed a 3.5-fold increase in activation at high ornithine concentrations.

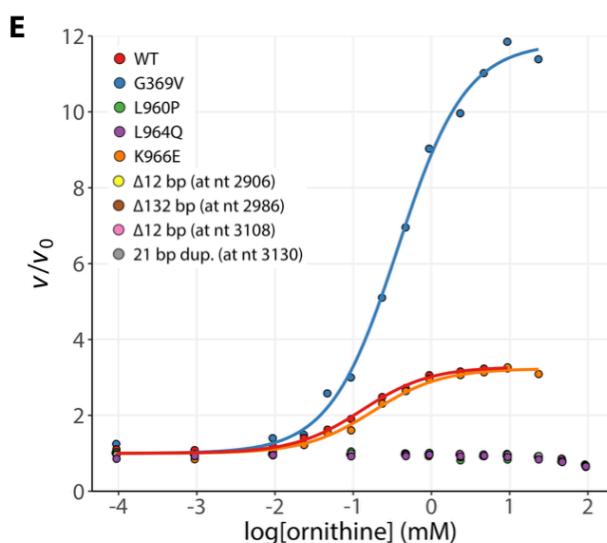
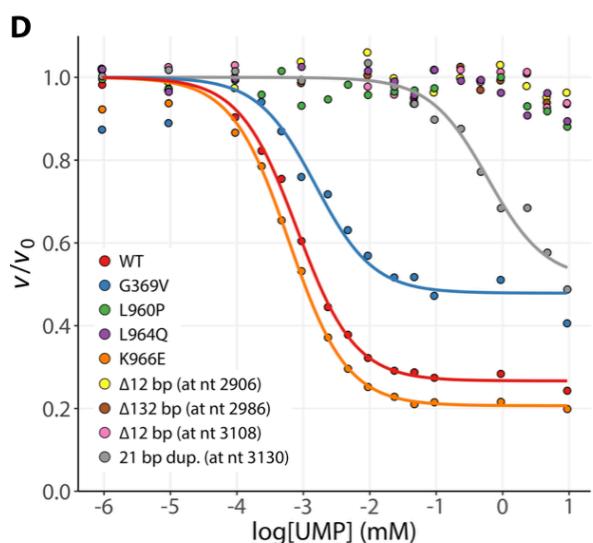
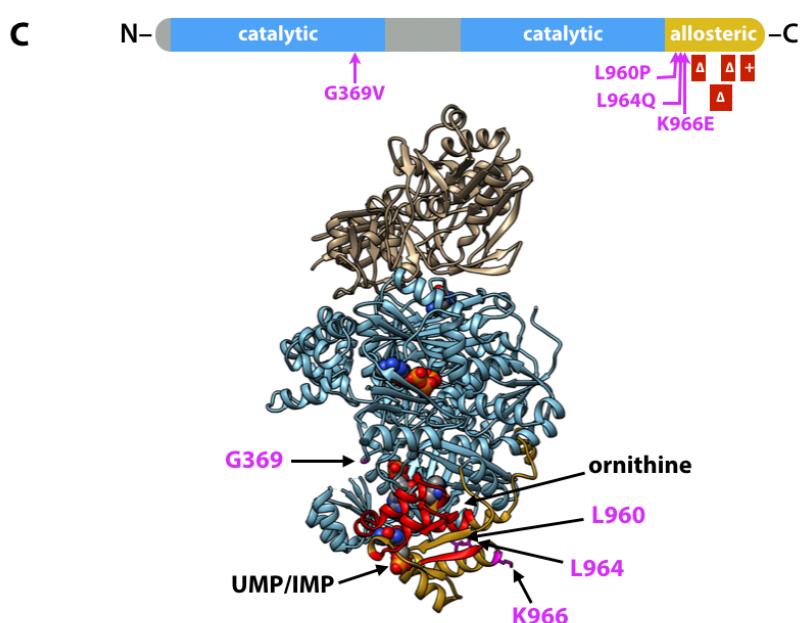
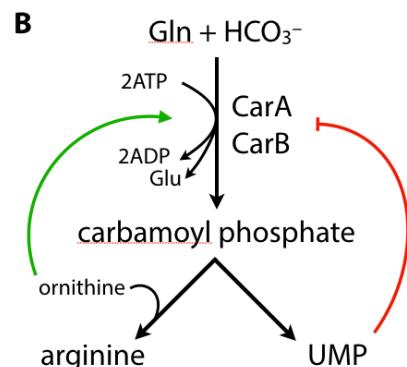
<i>carB</i> mutation	$k_{cat}/K_M$	UMP inhibition	L-Orn activation	adapted population							
				1	2	3	4	5	6	7	8
G369V	↓	↓	↑		6%						
L960P	↓	X	X							12%	
L964Q	↓	X	X						37%		
K966E	↑	n.c.	n.c.								78%
Δ12 bp (2906-2917 nt)	↓	X	X								9%
Δ132 bp (2986-3117 nt)	↓	X	X	44%							
Δ12 bp (3108-3119 nt)	↓	X	X				5%				
21 bp dup. (3150 nt)	↓	↓	X		6%						

Nucleotide (nt) numbers below mutant descriptions indicate where in the 3222 nt *carB* gene deletions or duplications occurred

↑ and ↓ indicate that a kinetic parameter was either increased or decreased, respectively, in the variant enzyme relative to the wild-type enzyme

X indicates loss of activity

n.c. indicates no change in the kinetic parameter relative to the wild-type enzyme



**Figure 7.** Several adaptive mutations occurred in *carB*. (A) Maximum percentage of each *carB* mutation found in the population at any time during the adaptation. (B) Allosteric regulation of carbamoyl phosphate synthetase. CarA and CarB are the small and large subunits of carbamoyl phosphate synthetase, respectively. (C) CarB functional domains (top) and crystal structure of *E. coli* CarAB (PDB 1CE8, bottom) (Thoden et al., 1999). Beige, CarA; blue, CarB; yellow, allosteric domain of CarB; red, residues that are deleted or duplicated in the adapted strains; magenta, point mutations that occur in the adapted strains. IMP and ornithine bound to the allosteric domain are shown as spheres. One of the two bound ATP molecules can be seen as spheres in the center of CarB. (D-E) Influence of UMP or L-ornithine concentration on the glutamine-dependent ATPase activity of CarAB mutants. Reaction rates are shown relative to  $v_0$ , the reaction rate in the absence of ligand.

298

299 **Discussion**

300 Recruitment of promiscuous enzymes to serve new functions followed by mutations that  
301 improve the promiscuous activity has been a dominant force in the diversification of metabolic  
302 networks (Copley, 2017; Glasner et al., 2006; Khersonsky & Tawfik, 2010; O'Brien &  
303 Herschlag, 1999; Rauwerdink et al., 2016). New enzymes may be important for fitness or even  
304 survival when an organism is exposed to a novel toxin or source of carbon or energy, or when  
305 synthesis of a natural product can enable manipulation of competing organisms. This process  
306 also contributes to non-orthologous gene replacement, which can occur when a gene is lost  
307 during a time in which it is not required, but its function later becomes important again and is  
308 replaced by recruitment of a non-orthologous promiscuous enzyme (Albalat & Cañestro, 2016;  
309 Ferla et al., 2017; Juárez-Vázquez et al., 2017; Olson, 1999).

310 We have modeled a situation in which a new enzyme is required by deleting *argC*, which  
311 is essential for synthesis of arginine in *E. coli*. Previous work showed that the most readily  
312 available source of neo-ArgC activity that enables  $\Delta argC$  *E. coli* to grow on glucose as a sole  
313 carbon source is a promiscuous activity of ProA. However, a point mutation that changes Glu383  
314 to Ala is required to elevate the promiscuous activity to a physiologically useful level. This

315 mutation substantially damages the native function of the enzyme, creating an inefficient  
316 bifunctional enzyme whose poor catalytic abilities limit growth rate on glucose. It is important to  
317 note that the decrease in the efficiency of the native reaction may be a critical factor in the  
318 recruitment of ProA to serve as a neo-ArgC because it will diminish inhibition of the newly  
319 important reaction by the native substrate (Khanal et al., 2015; McLoughlin & Copley, 2008).

320 When we previously carried out short-term adaptation of  $\Delta argC\ proA^*$  *E. coli* in glucose,  
321 we observed four mutations that increased fitness by increasing expression of *proA*\*: a G→A  
322 substitution in the -35 promoter region (M1) and a C→T substitution in the -10 promoter region  
323 (M2) upstream of the *proBA*\* operon, a synonymous mutation in *proB* that strengthens a cryptic  
324 promoter upstream of *proA*\* (M3), and amplification of a region of the genome that includes the  
325 *proBA*\* operon (Kershner et al., 2016). The point mutations occurred independently, never  
326 together. Amplification occurred even after acquisition of M1, M2, or M3. The amplification of  
327 *proA*\* suggests that it has the potential to diverge to restore efficient ProA function and evolve  
328 an efficient neo-ArgC.

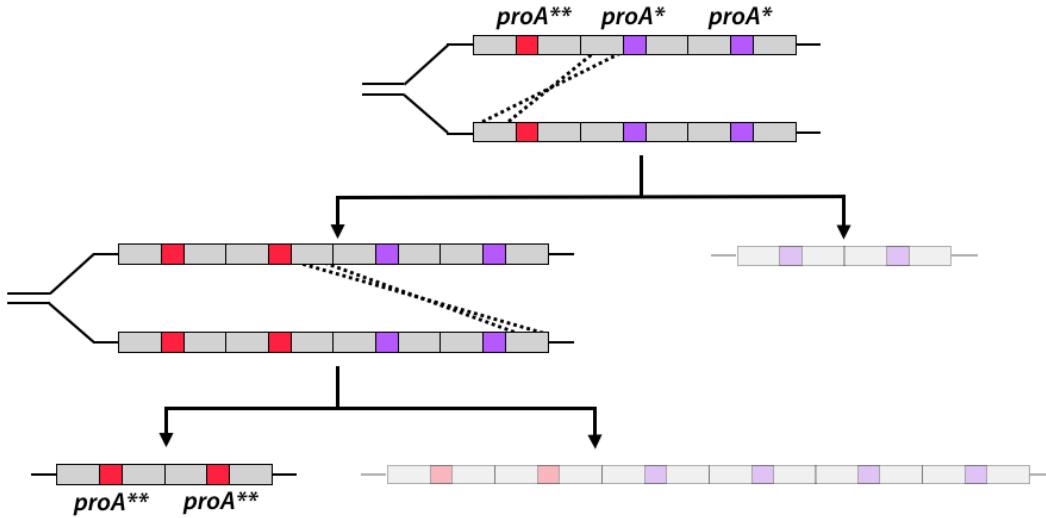
329 We chose to carry out longer-term evolution of a  $\Delta argC\ proA^*$  strain in glucose in the  
330 presence of proline to specifically address the evolution of an efficient neo-ArgC. We introduced  
331 the M2 promoter mutation in the parent strain to ensure that all populations had the same  
332 promoter mutation during adaptation. After 470-1000 generations of adaptation, growth rate had  
333 increased ~3-fold in all eight replicate cultures. We have focused on five types of genetic  
334 changes that clearly increase fitness: (1) mutations upstream of *pyrE*; (2) amplification of a  
335 variable region of the genome surrounding the *proBA*\* operon; (3) a mutation in *proA*\* that  
336 changes Phe372 to Leu; (4) mutations upstream of *argB*; and (5) mutations in *carB*. (Each of the  
337 final populations contains additional mutations that may also contribute to fitness, but these

338 mutations were typically found in low abundance and/or in only one population.) The *pyrE*  
339 mutations have previously been shown to be a general adaptation of *E. coli* BW25113 to growth  
340 in minimal medium (Blank et al., 2014; Conrad et al., 2009; Jensen, 1993; Knöppel et al., 2018).  
341 The other four types of mutations are specific adaptations to the bottleneck in arginine synthesis  
342 caused by substitution of the weak-link enzyme ProA\* for ArgC. Interestingly, only two of these  
343 – gene amplification and the mutation in *proA*\* – directly involved the weak-link enzyme ProA\*.

344 Surprisingly, we saw evolution of *proA*\* towards a more efficient neo-ArgC in only one  
345 population (Table 2). In this population, *proA* copy number dropped from ~7 to ~3 within 100  
346 generations (Figure 5C). This pattern is consistent with the IAD model; copy number is expected  
347 to decrease as mutations increase the efficiency of the weak-link activity. However, the fact that  
348 copy number did not return to one implies that the neo-ArgC activity of ProA\*\* is not sufficient  
349 to justify a single copy of the gene.

350 Because ~3 copies of *proA*\*\* remained in the population (Figure 5C) and the progenitor  
351 *proA*\* was not detectable, all copies in the amplified array have clearly acquired the mutation  
352 that changes Phe372 to Leu – i.e. the more beneficial *proA*\*\* allele has “swept” the amplified  
353 array. This observation has important implications for the IAD model. In the original conception  
354 of the IAD model, it was proposed that amplification of a gene increases the opportunity for  
355 different beneficial mutations to occur in different alleles, and then for recombination to shuffle  
356 these mutations (Bergthorsson et al., 2007; Francino, 2005). Both phenomena would increase the  
357 rate at which sequence space can be searched and thereby increase the rate at which a new  
358 enzyme evolves. In order for this to occur, however, it would be necessary for individual alleles  
359 to acquire different beneficial mutations before recombination occurred. This scenario is  
360 inconsistent with the relative frequencies of point mutations and recombination between large

361 homologous regions in an amplified array (Anderson & Roth, 1981; Reams et al., 2010). Point  
362 mutations occur at a frequency between  $10^{-9}$  and  $10^{-10}$  per nucleotide per cell division depending  
363 on the genomic location (Jee et al., 2016), and thus between  $10^{-6}$  and  $10^{-7}$  per gene per cell  
364 division for a gene the size of *proA*. If 10 copies of an evolving gene were present, then the  
365 frequency of mutation in a single allele would be between  $10^{-5}$  and  $10^{-6}$  per cell division.  
366 Homologous recombination after an initial duplication event is orders of magnitude more  
367 frequent, occurring in ~1 of every 100 cell divisions (Reams et al., 2010). Thus, homologous  
368 recombination between replicating chromosomes in a cell could result in a selective allelic sweep  
369 long before a second beneficial mutation occurs in a different allele in the amplified array. Figure  
370 8 depicts how this could occur after acquisition of an advantageous mutation in one allele in an  
371 amplified array. This is indeed the result that we observed; heterozygosity among *proA\** alleles  
372 was lost within 500 generations (Figure 5C). More recent papers depict selective amplification of  
373 beneficial alleles before acquisition of additional mutations (Andersson et al., 2015; Näsvall et  
374 al., 2012); our results support this revision of the original IAD model. It is possible that alleles  
375 encoding enzymes that are diverging toward two specialists might recombine to explore  
376 combinations of mutations. However, such recombination might not accelerate evolution, as  
377 mutations that lead toward one specialist enzyme would likely be incompatible with those that  
378 lead toward the other specialist enzyme.



**Figure 8.** Homologous recombination of an amplified *proA\** array with one *proA\*\** allele can rapidly lead to a daughter cell with only *proA\*\** alleles. Each arrow represents one cell duplication event and shows the genotypes of the two resulting daughter cells. The less-fit daughter cell from each recombination event is grayed out.

379

380        While growth rate improved substantially in all populations, a beneficial mutation in  
381        *proA\** arose in only one, suggesting that either mutations that improve the neo-ArgC activity are  
382        uncommon, or that their fitness effects are smaller than those caused by mutations elsewhere in  
383        the genome that also improve arginine synthesis. We identified two primary mechanisms that  
384        apparently improve arginine synthesis without affecting the efficiency of the weak-link enzyme  
385        ProA\* itself.

386        ArgB (*N*-acetylglutamate kinase) catalyzes the second step in arginine synthesis by  
387        phosphorylating *N*-acetylglutamate to form NAGP, the native substrate for ArgC and a  
388        secondary substrate for ProA\* (Figure 2). We identified eight mutations upstream of *argB*; the  
389        six we tested improved growth rate by 36-61% and increased the abundance of ArgB by 2.6-8.2-  
390        fold. Notably, ArgB levels were increased even though the levels of *argB* mRNA were  
391        unchanged (Figure 6). The increase in protein levels without a concomitant increase in mRNA

392 levels suggests that these mutations impact the efficiency of translation. Secondary structure  
393 around the translation initiation site plays a key role because this region must be unfolded in  
394 order to bind to the small subunit of the ribosome (Hall et al., 1982; Scharff et al., 2011). Indeed,  
395 a study of the predicted secondary structures of 5000 genes from bacteria, mitochondria and  
396 plastids, many of which lack canonical Shine-Dalgarno sequences (as does *argB*), showed that  
397 secondary structure around the start codon is markedly less stable than up- or down-stream  
398 regions (Bentele et al., 2013; Espah Borujeni et al., 2014; Goodman et al., 2013; Scharff et al.,  
399 2011). Our computational studies of the effect of mutations on the predicted lowest free energy  
400 secondary structures of the region surrounding the start codon of *argB* suggest that the  
401 thermodynamic stability of this region plays a role in the beneficial effects of most of the observe  
402 mutations (Figure S3). In addition, some of the mutations slow the predicted rate of mRNA  
403 folding around the start codon, which would increase the probability of ribosomal drafting  
404 (Figure S4). Both effects would lead to an increase in ArgB abundance. An increase in the level  
405 of ArgB should increase the concentration of the NAGP substrate for the weak-link ProA\*,  
406 thereby pushing material through this bottleneck in the arginine synthesis pathway.

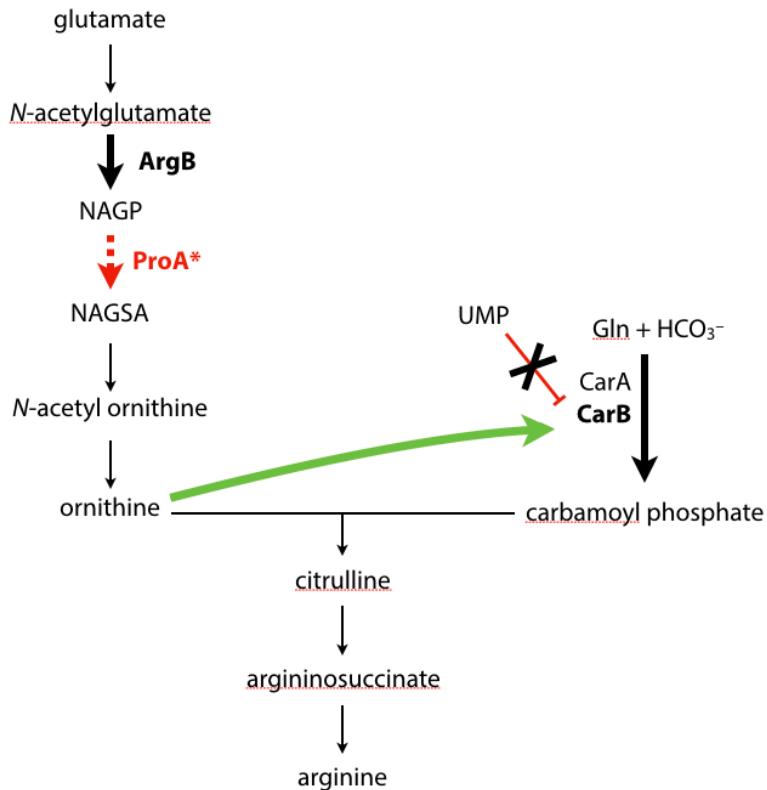
407 The adaptive mutations in *carB* increase catalytic turnover, decrease inhibition by UMP,  
408 or increase activation by ornithine of CPS. All of these effects should increase the level of CPS  
409 activity in the cell and consequently the level of carbamoyl phosphate. Why would this be  
410 advantageous? Carbamoyl phosphate reacts with ornithine, which will be in short supply due to  
411 the upstream ProA\* bottleneck, to generate citrulline (Legrain & Stalon, 1976) (Figure 9). If  
412 ornithine transcarbamoylase is not saturated with respect to carbamoyl phosphate, then  
413 increasing carbamoyl phosphate levels should increase citrulline production and thereby increase  
414 flux into the lower part of the arginine synthesis pathway. Although we do not know whether

415 ornithine transcarbamoylase is saturated with respect to carbamoyl phosphate *in vivo* (the  $K_M$  for  
416 carbamoyl phosphate is 360  $\mu\text{M}$  (Baur et al., 1990), but the intracellular concentration of  
417 carbamoyl phosphate is unknown), the occurrence of so many mutations that increase carbamoyl  
418 phosphate synthase activity supports the notion that they lead to an increase in carbamoyl  
419 phosphate that potentiates flux through the arginine synthesis pathway.

420 The majority of adaptive mutations we observed in *carB* cause loss of the exquisite  
421 allosteric regulation that controls flux through this important step in pyrimidine and arginine  
422 synthesis. This tight regulation likely evolved due to the energetically costly reaction catalyzed  
423 by CPS, which consumes two ATP molecules. While a constitutively active CPS is beneficial in  
424 the short term to improve arginine synthesis, it will likely be detrimental once arginine  
425 production no longer limits growth. We term mutations such as those observed in *carB*  
426 “expedient” mutations because they provide a quick fix when cells are under strong selective  
427 pressure, but at a cost to a previously well-evolved function. The damage caused by expedient  
428 mutations may be repairable later by reversion, compensatory mutations or horizontal gene  
429 transfer. Interestingly, the latter two repair processes may contribute to sequence divergence  
430 between organisms that has typically been attributed to neutral drift, but rather may be due to  
431 scars left from previous selective processes.

432 A particularly striking conclusion from this work is that most of the mutations that  
433 improved fitness under these strong selective conditions did not impact the gene encoding the  
434 weak-link enzyme, but rather adjusted fluxes in the metabolic network to compensate for the  
435 bottleneck in metabolism by other mechanisms. Not surprisingly, the process of evolution of a  
436 new enzyme by gene duplication and divergence does not take place in isolation, but is  
437 inextricably intertwined with mutations in the rest of the genome. The ultimate winner in a

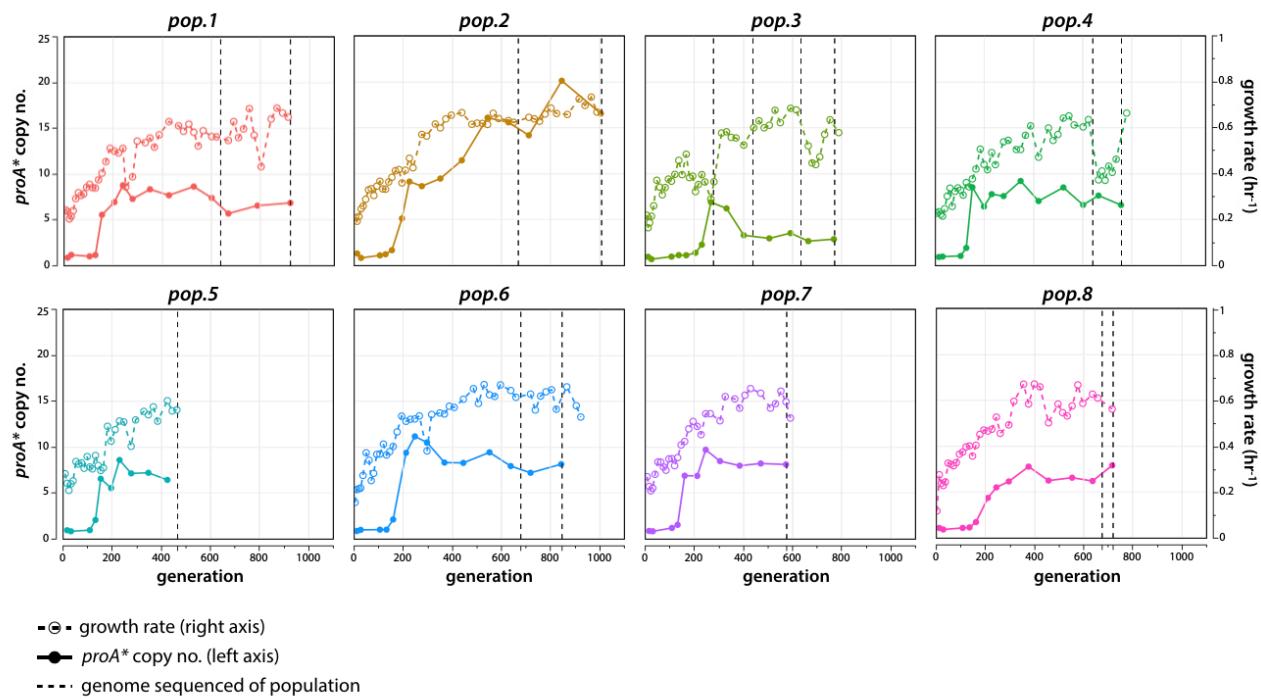
438 microbial population exposed to a novel selective pressure that requires evolution of a new  
439 enzyme may be the clone that succeeds in evolving an efficient enzyme that no longer limits  
440 fitness while accumulating the least damaging, or at least the most easily repaired, expedient  
441 mutations.



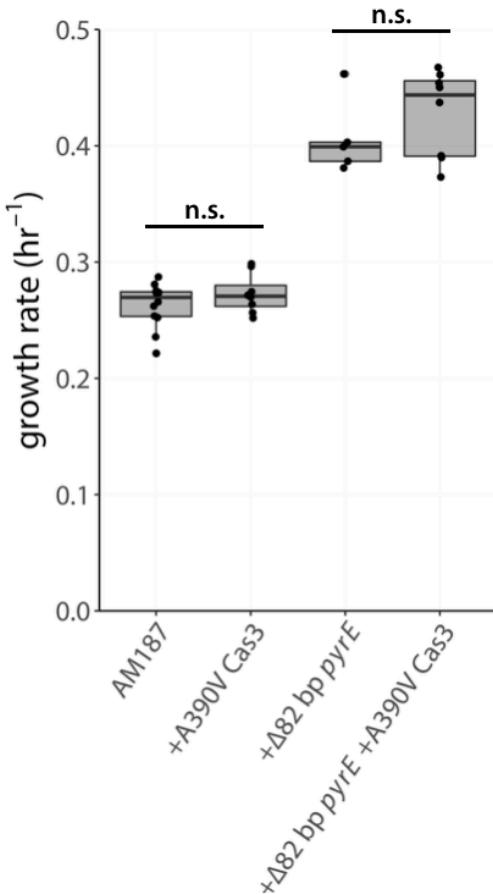
**Figure 9.** Adaptive mutations are predicted to increase flux through the arginine synthesis pathway. The pathway bottleneck and weak-link enzyme ProA\* is shown in red. Steps in the arginine synthesis pathway that are affected by adaptive mutations are highlighted in bold type.

442

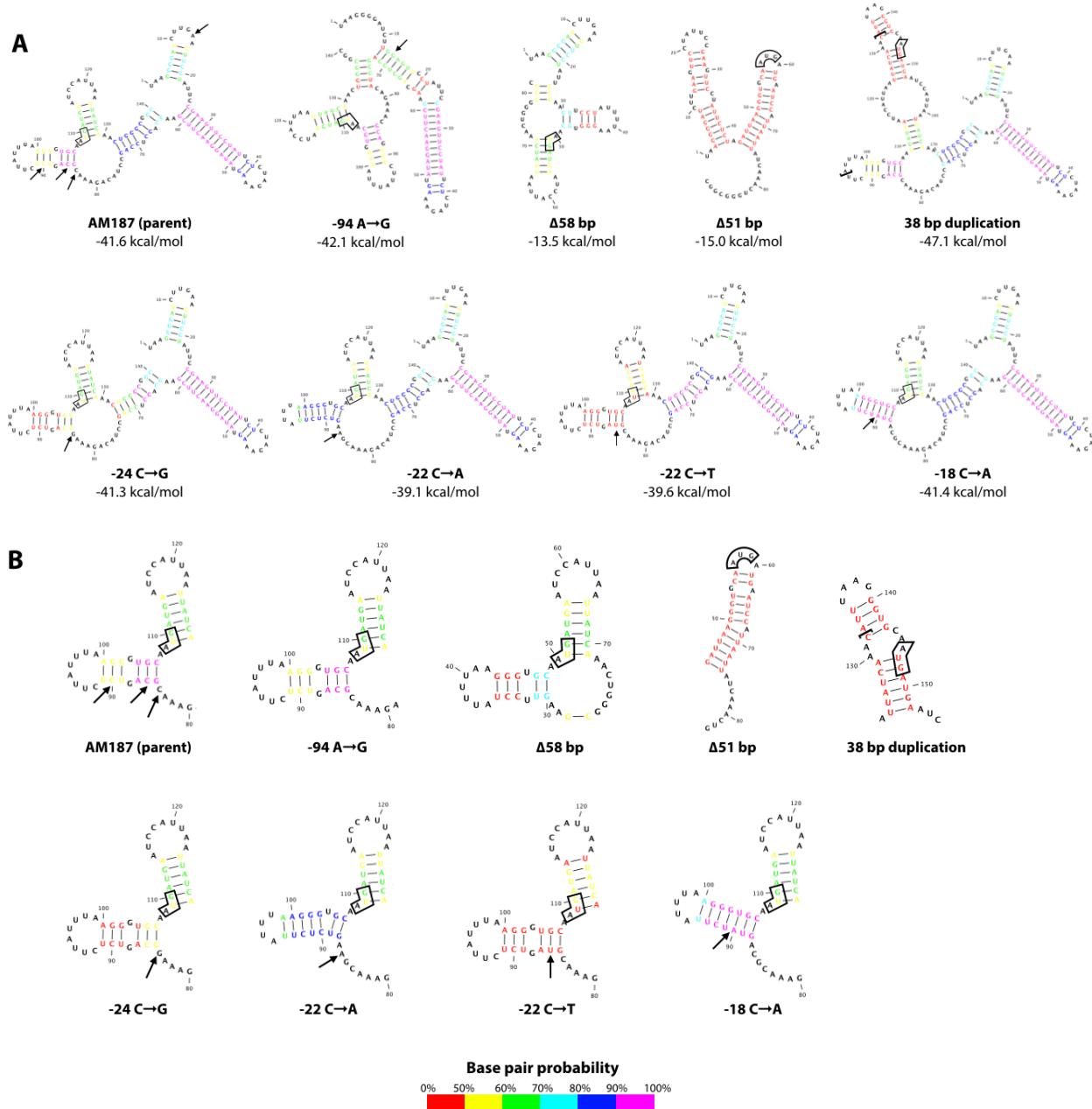
443 **Supplementary Figures**



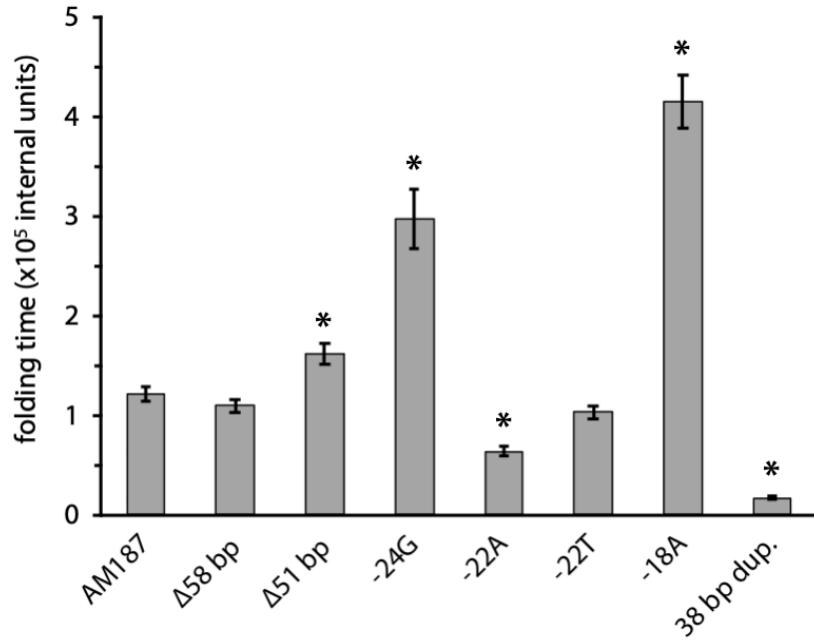
**Figure S1.** Growth rate (right axis, dotted lines) and *proA\** copy number (left axis, solid lines) for each adapted population. Vertical dotted lines indicate when population genomic DNA was sequenced.



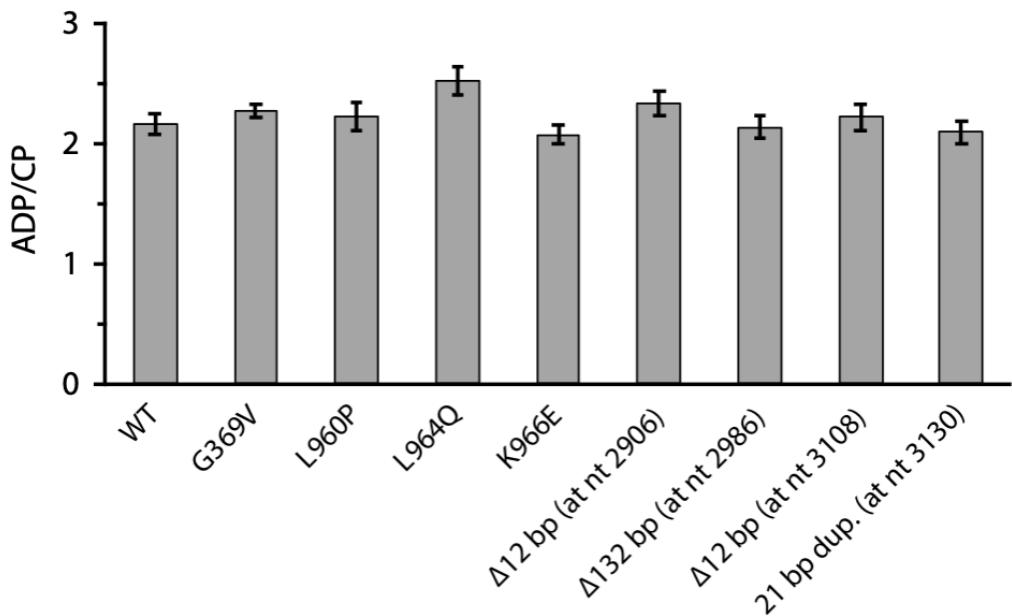
**Figure S2.** Effects of the A390V Cas3 and *rph-pyrE* mutations on growth rate of the parental strain AM187. (All  $p$  values are  $>0.1$  by a one-tailed Student's  $t$ -test).



**Figure S3.** *argB* mRNA secondary structure comparison. (A) The entire intergenic region between *kan<sup>r</sup>* and *argB* plus 33 nucleotides of the *argB* coding region was used for RNA structure prediction. The start codon for *argB* is boxed in each structure. Arrows in the AM187 structure indicate where an adaptive point mutation occurred. Arrows in mutant structures indicate which nucleotide was changed. In the 38 bp duplication mutant structure, the brackets contain the 38 inserted nucleotides. (B) Same as in (A), but showing only the region surrounding the *argB* start codon.



**Figure S4.** Folding time of a region encompassing 30 nucleotides upstream of the *argB* start codon plus 33 nucleotides of the *argB* coding region based on Kinfold simulations (Wolfinger et al., 2004). Error bars represent one standard error from the mean, N = 500.



**Figure S5.** Ratio of ADP to carbamoyl phosphate (CP) produced by wild-type and mutant carbamoyl phosphate synthetases.

444 **Materials and Methods**

445 **Materials**

446 Common chemicals were purchased from Sigma-Aldrich (St. Louis, MO) and Fisher  
447 Scientific (Fair Lawn, NJ).

448 NAGSA was synthesized enzymatically from *N*-acetylornithine using *N*-acetylornithine  
449 aminotransferase (ArgD) in a 25 mL reaction as described previously by Khanal et al. 2015 and  
450 stored at -70 °C. NAGSA concentrations were determined using the *o*-aminobenzaldehyde assay  
451 as described previously (Albrecht et al., 1962; Mezl & Knox, 1976).

452 GSA was synthesized enzymatically from L-ornithine using *N*-acetylornithine  
453 aminotransferase (ArgD) as described previously by Khanal et al. 2015 and stored at -70 °C.  
454 GSA concentrations were determined using the *o*-aminobenzaldehyde assay as described  
455 previously (Albrecht et al., 1962; Mezl & Knox, 1976).

456 Plasmids and primers used in this study are listed in Tables S1 and S2, respectively.

457

458 **Strains and culture conditions**

459 Strains used in this study are listed in Table 1. *E. coli* cultures were routinely grown in  
460 LB medium at 37 °C with 20 µg/mL kanamycin, 100 µg/mL ampicillin, 20 µg/mL  
461 chloramphenicol, or 10 µg/mL tetracycline, as required. Adaptation of strain AM187 was  
462 performed at 37 °C in M9 minimal medium containing 0.2% glucose, 0.4 mM proline, and 20  
463 µg/mL kanamycin (Adaptation Medium).

464

465 **Strain construction**

466 The parental strain for the adaptation experiment (AM187) was constructed from the  
467 Keio collection *argC::kan<sup>r</sup>* *E. coli* BW25113 strain (Baba et al., 2006). The *fimAICDFGH* and  
468 *csgBAC* operons were deleted (to slow biofilm formation), and the M2 *proBA* promoter mutation  
469 (Kershner et al., 2016) and the point mutation in *proA* that changes Glu383 to Ala (McLoughlin  
470 & Copley, 2008) were inserted into the genome using the scarless genome editing technique  
471 described in Kim et al. 2014. We initially hoped to measure *proA\** copy number during  
472 adaptation using fluorescence, although ultimately qPCR proved to be a better approach. Thus,  
473 we inserted *yfp* downstream of *proA\** under control of the P3 promoter (Mutalik et al. 2013) and  
474 with a synthetically designed ribosome binding site (Espah Borujeni et al., 2014; Salis et al.,  
475 2009). A double transcription terminator (BioBrick Part: BBa\_B0015) was inserted immediately  
476 downstream of *proBA\** to prevent read-through transcription of *yfp*. We also inserted a NotI cut  
477 site immediately downstream of *proA\** to enable cloning of individual *proA\** alleles after  
478 amplification if necessary. A Fis binding site located 32 bp downstream of *proA* was preserved  
479 because it might impact *proA* transcription. The NotI-2xTerm-*yfp* cassette was inserted  
480 downstream of *proA\** using the scarless genome editing technique described in Kim et al. 2014.  
481 The genome of the resulting strain AM187 was sequenced to confirm that there were no  
482 unintended mutations and deposited to NCBI GenBank under accession number CP037857.

483 Most mutations observed during the adaptation experiment were introduced into the  
484 parental AM187 strain using the scarless genome editing protocol described in Kim et al. 2014.  
485 This protocol is preferable to Cas9 genome editing for introduction of point mutations and small  
486 indels because it does not require introduction of synonymous PAM mutations that have the  
487 potential to affect RNA structure. The 58 bp deletion upstream of *argB* and the 82 bp deletion in

488 *rph* were introduced using Cas9-induced DNA cleavage and  $\lambda$  Red recombinase-mediated  
489 homology-directed repair with a linear DNA fragment. Sequences of the protospacers and  
490 mutations cassettes used for Cas9 genome editing procedures are listed in Tables S3 and S4.  
491 Details of the genome editing procedures are provided in the supplementary material.

492

493 **Laboratory adaptation**

494 Adaptation of strain AM187 in Adaptation Medium was carried out in eight replicate  
495 tubes in a custom turbidostat constructed as described by Takahashi et al. 2015. To start the  
496 adaptation, strain AM187 was grown to exponential phase ( $OD_{600} = 0.7$ ) in LB/kanamycin at 37  
497 °C. Cells were centrifuged at 4,000 x g for 10 min at room temperature and resuspended in an  
498 equal volume of PBS. The suspended cells were washed twice more with PBS and resuspended  
499 in PBS. This suspension was used to inoculate all eight turbidostat chambers to give an initial  
500  $OD_{600}$  of 0.01 in 14 mL of Adaptation Medium. The turbidostat was set to maintain an  $OD_{650}$  of  
501 0.4 by diluting individual cultures with an appropriate amount of fresh medium every 60  
502 seconds.

503 A 3 mL portion of each population was collected every 2-3 days; 800  $\mu$ l was used to  
504 make a 10% glycerol stock, which was then stored at -70 °C. The remaining sample was pelleted  
505 for purification of genomic DNA using the Invitrogen PureLink Genomic DNA Mini Kit  
506 according to the manufacturer's protocol.

507 The turbidostat occasionally malfunctioned, requiring the adaptation to be paused. When  
508 this occurred, the populations were subjected to centrifugation at 4,000 x g for 10 min at room  
509 temperature and the pelleted cells were resuspended in 1.6 mL of Adaptation Medium. Half of  
510 the resuspension was used to make a -70 °C 10% glycerol stock, and the other half to purify

511 genomic DNA. When the turbidostat was fixed, the frozen stock was thawed and the cells were  
512 collected by centrifugation at 16,000 x g for 1 min at room temperature. The pelleted cells were  
513 resuspended in 1 mL of PBS, washed, and resuspended in 500 µL of Adaptation Medium. The  
514 entire resuspension was used to inoculate the appropriate chamber of the turbidostat. Sometimes  
515 the adaptation had to be restarted from a frozen stock of a normal sample (as opposed to the  
516 entire population as just described), resulting in a more significant population bottleneck. In this  
517 case, the entire frozen stock was thawed and only 700 µL washed as described above to be used  
518 for the inoculation. The remaining 300 µL of the glycerol stock were re-stored at -70 °C in case  
519 the frozen stock was needed for downstream analysis.

520

## 521 **Calculation of growth rate and generations during adaptation**

522 The turbidostat takes an OD<sub>650</sub> reading every ~3 sec and dilutes the cultures every 60  
523 seconds. Thus, readings between dilutions can be used to calculate an average growth rate each  
524 day based on the following equation:

$$[6] \quad \bar{\mu} = \frac{\sum_i^n \frac{\ln(N_{t_{1,i}}/N_{t_{0,i}})}{t_{1,i}-t_{0,i}}}{n}$$

525 where  $\bar{\mu}$  is the average growth rate in hr<sup>-1</sup>, n is the number of independent growth rate  
526 calculations within a given 24 hr period,  $N_{t_2}$  is the OD<sub>650</sub> reading right before the dilution,  $N_{t_1}$  is  
527 the OD<sub>650</sub> reading right after the dilution, and  $t_1$  and  $t_0$  are the times at which the OD<sub>650</sub> was  
528 measured. The number of generations per day (g) was then calculated from  $\bar{\mu}$  using Eq. [7].

$$[7] \quad g = \frac{24 h}{\ln(2)/\bar{\mu}}$$

529

530 **Measurement of *proA\** copy number**

531 The copy number of *proA\** was determined by qPCR of purified population genomic  
532 DNA. *gyrB* and *icd*, which remained at a single copy in the genome throughout the adaptation  
533 experiment, were used as internal reference genes. Details of the experimental procedure and  
534 data analysis are provided in the supplementary material.

535

536 **Whole-genome sequencing**

537 Libraries were prepared from purified population genomic DNA using a modified  
538 Illumina Nextera protocol and multiplexed onto a single run on an Illumina NextSeq500 to  
539 produce 151-bp paired-end reads (Baym et al., 2015), giving a 60-130-fold coverage of the strain  
540 AM187 genome. Reads were trimmed using BBtools v35.82 (DOE Joint Genome Institute) and  
541 mapped using *breseq* v0.32.1 using the polymorphism (mixed population) option (Deatherage &  
542 Barrick, 2014).

543

544 **Growth rate measurements**

545 Growth rates for each strain were calculated from growth curves measured in  
546 quadruplicate. Overnight cultures were grown in LB/kanamycin at 37 °C from glycerol stocks.  
547 Ten µL of each overnight culture was used to inoculate 4 mL of LB/kanamycin and the cultures  
548 were allowed to grow to mid-exponential phase ( $OD_{600}$  0.3-0.6) at 37 °C with shaking. The  
549 cultures were subjected to centrifugation at 4,000 x g for 10 min at room temperature and the  
550 pellets resuspended in an equal volume of PBS. The pellets were washed once more in PBS. The  
551 cells were diluted to an OD of 0.001 in Adaptation Medium and a 100 µL aliquot was loaded  
552 into each well of a 96-well plate. The plates were incubated in a Varioskan (Thermo Scientific)

553 plate reader at 37 °C with shaking every 5 minutes for 1 minute. The absorbance at 600 nm was  
554 measured every 20 minutes for up to 500 hours. The baseline absorbance for each well (the  
555 average over several smoothed data points before growth) was subtracted from each point of the  
556 growth curve. Growth parameters (maximum specific growth,  $\mu_{\max}$ ; lag time,  $\lambda$ ; maximum  
557 growth,  $A_{\max}$ ) were estimated by non-linear regression using the modified Gompertz equation  
558 (Zwietering et al., 1990). Non-linear least-squares regression was performed in Excel using the  
559 Solver feature.

560

561 **Measurement of *argB* and *argH* gene expression by RT-qPCR**

562         Overnight cultures were grown in LB/kanamycin at 37 °C from glycerol stocks. Ten µL  
563 of each overnight culture was used to inoculate 4 mL of LB/kanamycin and the cultures were  
564 grown to mid-exponential phase (OD<sub>600</sub> 0.3-0.6) at 37 °C with shaking. Exponential phase  
565 cultures were centrifuged at 4,000 x g for 10 min and pellets resuspended in equal volume PBS.  
566 Pellets were washed once more in PBS. The cells were diluted to an OD of 0.001 in 4 mL of  
567 Adaptation Medium and grown to an OD<sub>600</sub> of 0.2-0.3. Four 2-mL aliquots of culture were  
568 thoroughly mixed with 4 mL of RNAProtect Bacteria Reagent (Qiagen) and incubated at room  
569 temperature for 5 min before centrifugation at 4,000 x g for 12 min at room temperature. Pellets  
570 were frozen in liquid N<sub>2</sub> and stored at -70 °C. Procedures for purification of RNA, reverse  
571 transcription, qPCR, and data analysis are described in the supplementary material.

572

573 **Measurement of ArgB and ArgH levels by mass spectrometry**

574         Freezer stocks were used to streak each strain on LB/kanamycin. Four parallel 2-mL  
575 aliquots of LB/kanamycin were inoculated with individual colonies and the cultures were grown

576 to mid-exponential phase at 37 °C with shaking. One mL of each culture was subjected to  
577 centrifugation at 16,000 x g for 1 min at room temperature. The cell pellets were resuspended in  
578 1 mL PBS and washed twice more in PBS before resuspension and dilution to an OD of 0.001 in  
579 5 mL of Adaptation Medium. Cultures were grown to an OD<sub>600</sub> of 0.2-0.3 at 37 °C with shaking  
580 and then chilled on ice for 10 min before pelleting by centrifugation at 4,000 x g at 4 °C. Cell  
581 pellets were frozen in liquid N<sub>2</sub> and stored at -70 °C.

582 Frozen cell pellets were thawed and lysed in 60 µL 50 mM Tris-HCl, pH 8.5, containing  
583 4% (w/v) SDS, 10 mM TCEP and 40 mM chloroacetamide in a Bioruptor Pico sonication device  
584 (Diagenode) using 10 cycles of 30 seconds on, 30 seconds off, followed by boiling for 10 min,  
585 and then another 10 cycles in the Bioruptor. The lysates were subjected to centrifugation at  
586 15,000 x g for 10 minutes at 20 °C and protein concentrations in the supernatant were  
587 determined by tryptophan fluorescence (Wiśniewski & Gaugaz, 2015). Ten µL of each sample  
588 (3-6 µg protein/µL) was digested using the SP3 method (C. S. Hughes et al., 2014). Carboxylate-  
589 functionalized speedbeads (GE Life Sciences) were added to the lysates. Addition of acetonitrile  
590 to 80% (v/v) caused the proteins to bind to the beads. The beads were washed twice with 70%  
591 (v/v) ethanol and once with 100% acetonitrile. Protein was digested and eluted from the beads  
592 with 15 µL of 50 mM Tris buffer, pH 8.5 with 1 µg endoproteinase Lys-C (Wako) for 2 hours  
593 with shaking at 600 rpm at 37 °C in a thermomixer (Eppendorf). One µg of trypsin (Pierce) was  
594 then added to the solution and incubated at 37 °C overnight with shaking at 600 rpm. Beads were  
595 collected by centrifugation and then placed on a magnet to more reliably remove the elution  
596 buffer containing the digested peptides. The peptides were then desalted using an Oasis HLB  
597 cartridge (Waters) according to the manufacturer's instructions and dried in a speedvac.

598           Samples were suspended in 12 µL of 3% (v/v) acetonitrile/0.1% (v/v) trifluoroacetic acid  
599       and 500 ng of peptides were directly injected onto a C18 1.7 µm, 130 Å, 75 µm X 250 mm M-  
600       class column (Waters), using a Waters M-class UPLC. Peptides were eluted at 300 nL/minute  
601       using a gradient from 3% to 20% acetonitrile over 100 minutes into an Orbitrap Fusion mass  
602       spectrometer (Thermo Scientific). Precursor mass spectra (MS1) were acquired at a resolution of  
603       120,000 from 380-1500 m/z with an AGC target of  $2.0 \times 10^5$  and a maximum injection time of  
604       50 ms. Dynamic exclusion was set for 20 seconds with a mass tolerance of  $\pm 10$  ppm.  
605       Precursor peptide ion isolation width for MS2 fragment scans was 1.6 Da using the quadrupole,  
606       and the most intense ions were sequenced using Top Speed with a 3-second cycle time. All MS2  
607       sequencing was performed using higher energy collision dissociation (HCD) at 35% collision  
608       energy and scanning in the linear ion trap. An AGC target of  $1.0 \times 10^4$  and 35-second maximum  
609       injection time was used. Rawfiles were searched against the Uniprot *Escherichia coli* database  
610       using Maxquant version 1.6.1.0 with cysteine carbamidomethylation as a fixed modification.  
611       Methionine oxidation and protein N-terminal acetylation were searched as variable  
612       modifications. All peptides and proteins were thresholded at a 1% false discovery rate (FDR).  
613

614       **Enzyme overexpression plasmids**

615       *proA* and *proA\** were amplified from the genomes of *E. coli* BW25113 and AM187,  
616       respectively, and inserted into a linearized pET-46 vector backbone by Gibson assembly (Gibson  
617       et al., 2009). A sequence encoding a 6xHis-tag followed by a 2xVal-linker was incorporated at  
618       the N-terminus of each protein. The *proA\*\** expression plasmid was constructed using the Q5  
619       Site-Directed Mutagenesis Kit (NEB).

620        *argC* was cloned into a pTrcHisB vector backbone. The final plasmid included an N-  
621        terminal 6xHis-tag followed by a Gly-Met-Ala-Ser linker and with Met1 of ArgC removed.

622        The *argD* and *argI* expression plasmids from the ASKA collection (Kitagawa et al.,  
623        2005) were used for expression of *N*-acetyltornithine aminotransferase and ornithine  
624        transcarbamoylase, respectively. In each case, the expression plasmid included a sequence  
625        encoding an N-terminal 6xHis-tag followed by a Thr-Asp-Pro-Ala-Leu-Arg-Ala linker.

626        Wild-type *carAB* was amplified from the genome of AM187 and inserted into a  
627        linearized pCA24N vector backbone by Gibson assembly (Gibson et al., 2009). The final  
628        construct included an N-terminal 6xHis-tag on CarA followed by a Thr-Asp-Pro-Ala-Leu-Arg-  
629        Ala linker. CPS mutant plasmids were constructed using the Q5 Site-Directed Mutagenesis Kit  
630        (NEB).

631        The correct sequences for all constructs were confirmed by Sanger sequencing.

632

### 633        **Protein purification**

634        Wild-type and variant ProAs were expressed in strain AM209 (BL21(DE3) *argC::kan'*  
635        *proA::cat*) to enable purification in the absence of wild-type ProA and ArgC. Carbamoyl  
636        phosphate synthetase (CPS) consists of a stable complex between CarA and CarB. Thus, in order  
637        to purify CPS, both *carA* and *carB* were co-expressed on the same plasmid with a His-tag on  
638        CarA. Wild-type and variant CPSs were expressed in strain AM267 (BL21 *carAB::kan'*) to  
639        enable purification in the absence of wild-type CPS.

640        Enzymes were expressed and purified using the following protocol with minor variations  
641        between the proteins. A small scraping from the glycerol stock of each expression strain was  
642        used to inoculate LB plus appropriate antibiotics. The cultures were grown overnight with

643 shaking at 37 °C. Overnight cultures were diluted 1:100 into 500 mL-2 L of LB plus appropriate  
644 antibiotic and grown with shaking at 37 °C. Once the OD<sub>600</sub> reached 0.5-0.9, IPTG was added to  
645 a final concentration of 0.5 mM. Growth was continued at 30 °C for 5 hrs while shaking. Cells  
646 were harvested by centrifugation at 5,000 x g for 20 min at 4 °C. Cell pellets were stored at -70  
647 °C until protein purification.

648 Frozen cell pellets were resuspended in 5x the cell pellet weight of ice-cold 20 mM  
649 sodium phosphate, pH 7.4, containing 300 mM NaCl and 10 mM imidazole. Fifty µL of protease  
650 inhibitor cocktail (Sigma-Aldrich, P8849) was added for each gram of cell pellet. Lysozyme was  
651 added to a final concentration of 0.2 mg/mL and the cells were lysed by probe sonication (20 sec  
652 of sonication followed by 30 sec on ice, repeated 3 times). Cell debris was removed by  
653 centrifugation at 18,000 x g for 20 min at 4 °C. The soluble fraction was then loaded onto 1 mL  
654 or 3 mL HisPur Ni-NTA Spin Columns (Thermo Scientific) and His-tagged protein was purified  
655 according to the manufacturer's protocol. If the soluble fraction was larger than the column  
656 volume, the initial binding step was repeated before washing to accommodate all the soluble  
657 fraction. Bound protein was eluted with one column volume of 20 mM sodium phosphate, pH  
658 7.4, containing 300 mM NaCl and increasing amounts of imidazole (100 mM, 250 mM, and  
659 finally 500 mM). Two separate elutions were performed with 500 mM imidazole. Fractions  
660 containing the protein of interest were pooled and dialyzed overnight against 6-12 L of exchange  
661 buffer at 4 °C. (ProA and ArgC were dialyzed against 20 mM potassium phosphate, 20 mM  
662 DTT, pH 7.5. N-acetylornithine aminotransferase was dialyzed against 20 mM potassium  
663 phosphate, pH 7.5. CPS was dialyzed against 100 mM potassium phosphate, pH 7.6. Ornithine  
664 transcarbamoylase was dialyzed against 20 mM Tris-acetate, pH 7.5.) Protein purity was  
665 assessed by SDS-PAGE and concentration measured using the Qubit protein assay kit with a

666 Qubit 3.0 fluorometer (Invitrogen). Purified protein was stored at 4 °C for short-term storage,  
667 and frozen in liquid nitrogen and stored at -70 °C for long-term storage.

668

### 669 **GSA and NAGSA dehydrogenase assays**

670 The native and neo-ArgC activities of ProA were assayed in the reverse direction  
671 (dehydrogenase reaction) because the lability of the forward substrates  $\gamma$ -glutamyl phosphate and  
672 *N*-acetylglutamyl phosphate make them difficult to purify. The change in the dehydrogenase  
673 activity due to a mutation is proportional to the change in the reductase activity according to the  
674 Haldane relationship (Haldane, 1930; McLoughlin & Copley, 2008).

675 Assaying ProA's dehydrogenase activity using  $\gamma$ -glutamyl semialdehyde (GSA) and *N*-  
676 acetylglutamyl semialdehyde (NAGSA) as substrates is complicated by the equilibrium of GSA  
677 and NAGSA with their hydrated forms, as well as GSA's intramolecular cyclization to form  
678 pyrroline-5-carboxylate (P5C) (Bearne & Wolfenden, 1995; Mezl & Knox, 1976). In order to  
679 measure the concentration of the free aldehyde form of these substrates, we mixed 15  $\mu$ M ProA  
680 or ArgC with 2 mM "GSA" (including the hydrate and P5C) or 2 mM "NAGSA" (including the  
681 hydrate), respectively, in a solution containing 100 mM potassium phosphate, pH 7.6, and 1 mM  
682 NADP<sup>+</sup> and measured the burst in NADPH production (Khanal et al., 2015). The concentrations  
683 of GSA+P5C+hydrate or NAGSA+hydrate were determined using the *o*-aminobenzaldehyde  
684 assay (Albrecht et al., 1962; Mezl & Knox, 1976). The absorbance at 340 nm due to formation of  
685 NADPH exhibited a burst followed by a linear phase when measured for 60 seconds. We assume  
686 that the burst corresponds to reduction of the free aldehyde form of GSA or NAGSA and the rate  
687 of the linear phase is determined by the conversion of the hydrate (and P5C in the case of GSA)

688 to the free aldehyde. We calculated the magnitude of the burst by fitting either all of the data or  
689 the linear portion of the data to one of the following equations.

$$[8] \quad f(x) = mx + b$$

$$[9] \quad f(x) = mx + b(1 - e^{-x})$$

690 where  $x$  is time in seconds,  $m$  is the slope of the linear phase, and  $b$  is the magnitude of the burst  
691 and thus proportional to the starting concentration of the free aldehyde form of the substrate. In  
692 the case of the linear fit, only the linear portion of the  $A_{340}$  data was used. Eq. [8] was used to  
693 calculate NAGSA free aldehyde concentration because the exponential equation did not fit the  
694 data well. Eq. [9] was used to calculate GSA free aldehyde concentration. We repeated the  
695 assay three times and averaged the magnitude of the burst to calculate free aldehyde  
696 concentrations for solutions of GSA and NAGSA (under these buffer and temperature  
697 conditions) of 4.5% and 4.2% of the total concentration of free aldehyde + hydrate (+ P5C for  
698 GSA), respectively.

699 GSA and NAGSA dehydrogenase activities were measured by monitoring the appearance  
700 of NADPH at 340 nm in reaction mixtures containing 100 mM potassium phosphate, pH 7.6, 1  
701 mM NADP<sup>+</sup>, varying concentrations of NAGSA or GSA, and catalytic amounts of ProA, ProA\*,  
702 and ProA\*\*. All kinetic measurements were done at 25 °C. Values for  $K_M$  refer to the  
703 concentration of the free aldehyde form of the substrate.

704

## 705 Assays for carbamoyl phosphate synthetase activity and allosteric regulation

706 Kinetic assays for carbamoyl phosphate synthetase (CPS) were carried out with minor  
707 modifications of the methods described in Pierrat and Raushel 2002. The rate of ATP hydrolysis  
708 was measured at 37 °C by coupling production of ADP to oxidation of NADH using pyruvate  
709 kinase, which converts ADP and PEP to ATP and pyruvate, and lactate dehydrogenase, which

710 reduces pyruvate to lactate. Loss of NADH was monitored at 340 nm. Reaction mixtures  
711 consisted of 50 mM HEPES, pH 7.5, containing 10 mM MgCl<sub>2</sub>, 100 mM KCl, 20 mM potassium  
712 bicarbonate, 10 mM L-glutamine, 1 mM PEP, 0.2 mM NADH, saturating amounts of pyruvate  
713 kinase and lactate dehydrogenase (Sigma-Aldrich, P0294), and varying amounts of ATP (0.01 to  
714 8 mM). Reactions were initiated by adding 0.2 μM CPS to a final concentration of 0.2 μM. The  
715 effects of UMP and ornithine were measured under the same reaction conditions but with a fixed  
716 ATP concentration of 0.2 mM and varying concentrations of either UMP or ornithine.

717 Carbamoyl phosphate production was measured with minor modifications of previously  
718 described procedures (Snodgrass & Parry, 1969; Stapleton et al., 1996). Formation of carbamoyl  
719 phosphate by CPS was coupled with formation of citrulline by ornithine transcarbamoylase;  
720 citrulline forms a yellow complex ( $\epsilon_{464} = 37800 \text{ M}^{-1} \text{ cm}^{-1}$  (Snodgrass & Parry, 1969)) when  
721 mixed with diacetyl monoxime and antipyrine. Reaction mixtures consisted of 50 mM HEPES,  
722 pH 7.5, 10 mM MgCl<sub>2</sub>, 100 mM KCl, 20 mM potassium bicarbonate, 10 mM L-glutamine, 4  
723 mM ATP, 10 mM L-ornithine, and 0.7 μM ornithine transcarbamoylase. Reactions (0.25 mL)  
724 were initiated by adding CPS at a final concentration of 0.2 μM. After incubation for 2.5 min at  
725 37 °C, reactions were quenched by addition of 1 mL of a solution consisting of 25%  
726 concentrated H<sub>2</sub>SO<sub>4</sub>, 25% H<sub>3</sub>PO<sub>4</sub> (85%), 0.25% (w/v) ferric ammonium sulfate, and 0.37% (w/v)  
727 antipyrine, followed by addition of 0.5 mL of 0.4% (w/v) diacetyl monoxime/7.5% (w/v) NaCl.  
728 The quenched reaction mixtures were placed in a boiling water bath for 15 min before  
729 measurement of OD<sub>464</sub>. Control reactions contained all components except CPS.  
730

731 **RNA structure prediction**

732 RNA secondary structures for *argB* mRNAs were predicted using CLC Main Workbench  
733 8.1. The entire intergenic region between *kan<sup>r</sup>* and *argB* plus the first 33 nucleotides of *argB*  
734 were included in the structure prediction. The first 33 nucleotides were included because an  
735 mRNA-bound ribosome prevents another ribosome from binding to the mRNA until it has  
736 moved past the first 33 nucleotides (Steitz, 1969). Thus, at least the first 33 nucleotides are  
737 available for folding with the upstream region when a mRNA is being translated.

738

739 **Calculation of RNA folding times**

740 Folding times for a 63-nucleotide region (30 nt downstream and 30 nt upstream of the  
741 *argB* start codon) surrounding the start codon of *argB* mRNAs were calculated using the Kinfold  
742 program (v1.3) from the ViennaRNA v2.4.11 package (Wolfinger et al., 2004). Kinfold utilizes a  
743 Monte Carlo algorithm to calculate the folding time of each RNA sequence to the lowest free  
744 energy structure. We simulated 500 folding trajectories for each structure.

745

746 **Acknowledgements**

747 This work was supported by the NASA Astrobiology Institute (NNA15BB04A). We  
748 thank Craig Joy (University of Colorado Boulder, Physics Department) and Chris Takahashi  
749 (University of Washington) for help building the turbidostat.

750

751 **Funding**

Funder	Grant reference number	Author
National Aeronautics and Space Administration	NNA15BB04A	Shelley D. Copley

(NASA) Astrobiology  
Institute

Department of Defense  
(DOD)/Defense Advanced  
Research Projects Agency  
(DARPA)

13-34-RTA-FP-007

William M. Old

752 The funders had no role in study design, data collection and interpretation, or the decision to submit the work for  
753 publication.

754

## 755 Additional files

756 **Supplementary file 1.** Mutations found during the adaptation experiment. Sheet 1: mutations  
757 that were present at frequencies greater than 30% or that appeared in different populations. Sheet  
758 2: genome positions of the amplified regions surrounding *proA*\*. Sheet 3: other amplified or  
759 deleted genomic regions in the adapted populations.

760 **Supplementary file 2.** Supplementary Materials and Methods.

761

## 762 Data availability

763 The genome sequence of *E. coli* strain AM187 used in this study has been deposited to NCBI  
764 GenBank under accession number CP037857.

765

## 766 **Supplementary file 2: Supplementary Materials and Methods**

**Table S1.** Plasmids used in this study.

Plasmid	Use/description	Source/Ref
pSLTS	scarless genome editing/arabinose-inducible λ Red recombinase, anhydrotetracycline-inducible meganuclease I-SceI, temperature-sensitive origin of replication, ampicillin resistance	(Kim et al., 2014)
pSIM5	introduction of linear DNA fragments into the genome/heat-inducible λ Red recombinase, temperature sensitive origin of replication, chloramphenicol resistance	(Datta et al., 2006)

pSIM27	introduction of linear DNA fragments into the genome/heat-inducible λ Red recombinase, temperature-sensitive origin of replication, tetracycline resistance	(Datta et al., 2006)
pAM053	Cas9-mediated genome editing/constitutively-expressed <i>cas9</i> , heat-inducible λ Red recombinase, temperature-sensitive origin of replication, chloramphenicol resistance	This study
pET-46	backbone for IPTG-inducible expression of wild-type and mutant ProAs/T7 promoter upstream of 6xHis-tag, ampicillin resistance	Novagen
pCA24N	backbone for IPTG-inducible expression of wild-type and mutant carbamoyl phosphate synthetases/chloramphenicol resistance	(Kitagawa et al., 2005)
pCA24N- <i>argI</i>	IPTG-inducible expression of ornithine transcarbamoylase/chloramphenicol resistance	(Kitagawa et al., 2005)
pCA24N- <i>argD</i>	IPTG-inducible expression of <i>N</i> -acetylornithine aminotransferase/chloramphenicol resistance	(Kitagawa et al., 2005)
pAM068	expresses the Cas9 guide RNA that targets the upstream region of <i>argB</i> for introduction of the 58 bp deletion upstream of <i>argB</i> /temperature-sensitive origin of replication, ampicillin resistance	This study
pAM100	expresses the Cas9 guide RNA that targets <i>rph</i> for introduction of the 82 bp deletion upstream of <i>pyrE</i> /temperature sensitive origin of replication, ampicillin resistance	This study

767

**Table S2.** Primers used in this study.

name	sequence (5'→3')	description
1357	CGTGCAGGCGATTGATAA	Forward primer for <i>proA</i> qPCR
1358	CTGTTCACGGCACAGTTT	Reverse primer for <i>proA</i> qPCR
1359	CGTAGATCTGACGGTGAAATT	Forward primer for <i>gyrB</i> qPCR
1360	CGTTGGTGTTCGGTAGTA	Reverse primer for <i>gyrB</i> qPCR
1361	CCCGTGGCTGAAAGTTAAA	Forward primer for <i>icd</i> qPCR
1362	CAGGTTCATACAGGCGATAAC	Reverse primer for <i>icd</i> qPCR
hcaT_F	CGTGGTGGCGGAAGTCATTATC	Forward primer for <i>hcaT</i> RT-qPCR
hcaT_R	CGCCGAGATCAACAGCATATCG	Reverse primer for <i>hcaT</i> RT-qPCR
cysG_F	GGTGGCGAAGAGCTGGAAA	Forward primer for <i>cysG</i> RT-qPCR
cysG_R	GAATACCCGAATAGGCAGAGCAA	Reverse primer for <i>cysG</i> RT-qPCR
argB_F	GGCGGGAACGGCAAATAAAA	Forward primer for <i>argB</i> RT-qPCR
argB_R	CCGTCACCGAGAAACAAACC	Reverse primer for <i>argB</i> RT-qPCR
argH_F	TGGTGGAAACCGCACAAAC	Forward primer for <i>argH</i> RT-qPCR
argH_R	GCGCCAGCATCTAACATAG	Reverse primer for <i>argH</i> RT-qPCR

rAM209	GACATAGCGTGGCTACCCG	Forward primer for amplifying linear mutation fragment for the 58 bp deletion upstream of <i>argB</i>
rAM192	CAGCCCTTCATCAGCTCATCC	Reverse primer for amplifying linear mutation fragment for the 58 bp deletion upstream of <i>argB</i>
rAM307	AGCGGTTGCGATCTGGAAT	Forward primer for amplifying linear mutation fragment for the 82 bp deletion in <i>rph</i> upstream of <i>pyrE</i>
rAM308	ATGGTTTCATGCCTTCGCTC	Reverse primer for amplifying linear mutation fragment for the 82 bp deletion in <i>rph</i> upstream of <i>pyrE</i>

768

**Table S3.** Protospacers used for Cas9-mediated scarless genome editing.

genome modification	gRNA plasmid	protospacer (5'→3')
58 bp deletion upstream of <i>argB</i>	pAM068	GAGACTGCGTTCTGTAGGC
82 bp deletion in <i>rph</i> upstream of <i>pyrE</i>	pAM100	GCTACTCATCTTGTGGCTC

769

**Table S4.** Mutation cassettes used for Cas9-mediated scarless genome editing. Brackets represent the primer annealing regions for amplifying the mutation cassettes.

genome modification	sequence (5'→3')
58 bp deletion upstream of <i>argB</i>	[GACATAGCGTGGCTACCCG]TGATATTGCTGAAGAGCTTGGCGGC GAATGGGCTGACCGCTCCTCGTCTTACGGTATGCCGCTCCGA TTCGCAGCGCATCGCCTCTATGCCCTCTTGACGAGTTCTTCAAT AAGGGGATCTTGAAGTTCTATTCCGAAGTTCTATTAAAGGGTGC AATGATGAATCCATTAATTATCAAACGGCGCGTACTGCTGGAT AGTGAAGAGGCCTGGAACGTCTGTTAGCGCACTGGTGAATTATC GTGAGTCACATCAGCGTCCGCTGGTATTGTGCACGGCGGGTTG CGTGGT[GGATGAGCTGATGAAAGGGCTG]
82 bp deletion in <i>rph</i> upstream of <i>pyrE</i>	[AGCGGTTGCGATCTGGAAT]ACGTTGAAGACTCTGCCGCAGAGAC CGACATGAACGTAGTGATGACCGAAGACGGCGCATATTGAAGT GCAGGGGACGGCAGAAGGCCTGGCAAAGTGTAGTTAAGGC GACTGATGAGTCGCCTTTGTCTGTAGAAAAGTAAGATGAG[G AGCGAAGGCATGAAACCAT]

770

## 771 Genome editing

772 Strain AM187 was first transformed with a helper plasmid (pAM053, Table S1) encoding

773 *cas9* under the control a weak constitutive promoter (pro1 from Davis, Rubin, and Sauer 2011),

774     λ Red recombinase genes (*exo*, *gam*, and *bet*) under the control of a heat-inducible promoter, and  
775     a temperature-sensitive origin of replication (Datta et al., 2006). These cells were grown to an  
776     OD<sub>600</sub> of 0.2-0.4 at 30 °C and then incubated at 42 °C with shaking for 15 min to induce  
777     expression of the λ Red recombinase genes. The cells were immediately subjected to  
778     electroporation with 100 ng of either pAM068 or pAM100, which encode guide RNAs targeting  
779     20-nucleotide sequences upstream of *argB* or in *rph*, respectively (Table S3), and 450 ng of a  
780     linear homology repair template that will introduce the desired deletion into the genome (Table  
781     S4). (Linear homology repair templates were amplified from genomic DNA of clones isolated  
782     during the adaptation experiment that contained the desired deletions and the PCR fragments  
783     were gel-purified. Primers used to generate the linear DNA mutation fragments are listed in  
784     Table .) The cells were allowed to recover at 30 °C for 2-3 hours before being spread onto  
785     LB/ampicillin plates. Surviving colonies contained the guide RNA plasmid and the desired  
786     deletion, as confirmed by Sanger sequencing. pAM053 and the guide RNA plasmids, both of  
787     which have temperature-sensitive origins of replication, were cured by growth of individual  
788     colonies at 37 °C.

789                 Strain AM209 was constructed from *E. coli* BL21(DE3) for expression of wild-type and  
790     mutant ProAs. We deleted *argC* and *proA* to ensure that any activity measured during *in vitro*  
791     assays was not due to trace amounts of ArgC or wild-type ProA. To accomplish these deletions,  
792     we amplified and gel-purified DNA fragments containing antibiotic resistance genes (kanamycin  
793     and chloramphenicol for deletion of *argC* and *proA*, respectively) flanked by 200-400 bp of  
794     sequences homologous to the upstream and downstream regions of either *argC* or *proA*. *E. coli*  
795     BL21(DE3) cells containing pSIM27 (Datta et al., 2006) – a vector containing heat-inducible λ  
796     Red recombinase genes – were grown in LB/tetracycline to an OD of 0.2-0.4 and then incubated

797 in a 42 °C shaking water bath for 15 min to induce expression of  $\lambda$  Red recombinase genes. The  
798 cells were then immediately subjected to electroporation with 100 ng of the appropriate linear  
799 DNA mutation cassette. Successful transformants were selected on either LB/kanamycin or  
800 LB/chloramphenicol plates.

801 Strain AM267 was constructed from *E. coli* BL21(DE3) for expression of wild-type and  
802 mutant carbamoyl phosphate synthetases (CPS). We deleted *carAB* to ensure that any activity  
803 measured during *in vitro* assays was not due to trace amounts of wild-type CPS. To accomplish  
804 the deletion, we amplified and gel-purified a DNA fragment containing the kanamycin resistance  
805 gene flanked by 40 bp of sequence homologous to the upstream and downstream regions of  
806 either *argC* or *proA*. *E. coli* BL21 cells containing pSIM5 (Datta et al., 2006) – a vector carrying  
807 heat-inducible  $\lambda$  Red recombinase genes – were grown in LB/chloramphenicol to an OD of 0.2-  
808 0.4 and then incubated in a 42 °C shaking water bath for 15 min to induce expression of  $\lambda$  Red  
809 recombinase genes. The cells were then immediately subjected to electroporation with 100 ng of  
810 the appropriate linear DNA mutation cassette. Successful transformants were selected on either  
811 LB/kanamycin plates.

812

813 **Measurement of *proA\** copy number**

814 The primer sets used for each gene are listed in Table S2. PowerSYBR Green PCR  
815 master mix (Thermo Scientific) was used according to the manufacturer's protocol. A standard  
816 curve using variable amounts of AM187 genomic DNA was run on every plate to calculate  
817 efficiencies for each primer set. Primer efficiencies were calculated with the following equation:

$$[10] \quad E_x = 10^{-\left(\frac{1}{m}\right)}$$

818 where  $E$  is the efficiency of primer set  $x$ , and  $m$  is the slope of the plot of  $C_t$  vs. starting quantity  
819 for the standard curve.  $proA^*$  copy number was then calculated with the following equation  
820 (Hellemans et al., 2007):

$$[11] \quad n = \frac{E_{proA}^{\Delta C_{t,proA}}}{\sqrt{E_{gyrB}^{\Delta C_{t,gyrB}} \times E_{icd}^{\Delta C_{t,icd}}}}$$

821 where  $n$  is the  $proA^*$  copy number, and  $\Delta C_{t,x}$  is the difference in  $C_t$ 's measured during  
822 amplification of AM187 and sample genomic DNA with primer set  $x$ .

823

#### 824 **Measurement of *argB* and *argH* gene expression by RT-qPCR**

825 RNA was purified using the Invitrogen PureLink RNA Mini Kit according to the  
826 manufacturer's protocol. The cell lysate produced during the PureLink protocol was  
827 homogenized using the QIAshredder column (Qiagen) prior to RNA purification. After RNA  
828 purification, each sample was treated with TURBO DNase (Invitrogen) according to the  
829 manufacturer's protocol. Reverse transcription (RT) was performed with 250-600 ng of RNA  
830 using SuperScript IV VILO (Invitrogen) master mix according to the manufacturer's protocol.

831 qPCR of cDNA was performed to measure the fold-change in expression of *argB* and  
832 *argH* in mutant strains compared to that in AM187. *hcaT* and *cysG* were used as reference genes  
833 (Zhou et al., 2011). The primer sets used for each gene are listed in Table S2. A standard curve  
834 using variable amounts of *E. coli* BW25113 genomic DNA was run to calculate the primer  
835 efficiencies for each primer set. Fold-changes in expression of *argB* and *argH* were calculated as  
836 described above for calculations of  $proA^*$  copy number.

837

838

839 **References**

- 840 Adler, M., Anjum, M., Berg, O. G., Andersson, D. I., & Sandegren, L. (2014). High fitness costs  
841 and instability of gene duplications reduce rates of evolution of new genes by duplication-  
842 divergence mechanisms. *Molecular Biology and Evolution*, 31(6), 1526–1535.  
843 <https://doi.org/10.1093/molbev/msu111>
- 844 Albalat, R., & Cañestro, C. (2016). Evolution by gene loss. *Nature Reviews Genetics*, 17(7),  
845 379–391. <https://doi.org/10.1038/nrg.2016.39>
- 846 Albrecht, A. M., Scher, W. I., & Vogel, H. J. (1962). Determination of aliphatic aldehydes by  
847 spectrophotometry. *Analytical Chemistry*, 34(3), 398–400.  
848 <https://doi.org/10.1021/ac60183a028>
- 849 Anderson, P., & Roth, J. (1981). Spontaneous tandem genetic duplications in *Salmonella*  
850 *typhimurium* arise by unequal recombination between rRNA (rrn) cistrons. *Proceedings of*  
851 *the National Academy of Sciences*, 78(5), 3113–3117.  
852 <https://doi.org/10.1073/pnas.78.5.3113>
- 853 Andersson, D. I., Jerlström-Hultqvist, J., & Näsvall, J. (2015). Evolution of New Functions De  
854 Novo and from Preexisting Genes. *Cold Spring Harbor Perspectives in Biology*, 7(6).  
855 <https://doi.org/10.1101/cshperspect.a017996>
- 856 Baba, T., Ara, T., Hasegawa, M., Takai, Y., Okumura, Y., Baba, M., ... Mori, H. (2006).  
857 Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio  
858 collection. *Molecular Systems Biology*, 2, 2006.0008. <https://doi.org/10.1038/msb4100050>
- 859 Barnhart, M. M., & Chapman, M. R. (2006). Curli biogenesis and function. *Annual Review of*  
860 *Microbiology*, 60(1), 131–147. <https://doi.org/10.1146/annurev.micro.60.080805.142106>
- 861 Baur, H., Tricot, C., Stalon, V., & Haas, D. (1990). Converting catabolic ornithine

- 862 carbamoyltransferase to an anabolic enzyme. *Journal of Biological Chemistry*, 265(25),  
863 14728–14731.
- 864 Baym, M., Kryazhimskiy, S., Lieberman, T. D., Chung, H., Desai, M. M., & Kishony, R. K.  
865 (2015). Inexpensive multiplexed library preparation for megabase-sized genomes. *PLoS  
866 ONE*, 10(5), 1–15. <https://doi.org/10.1371/journal.pone.0128036>
- 867 Bearne, S. L., & Wolfenden, R. (1995). Glutamate gamma-semialdehyde as a natural transition  
868 state analogue inhibitor of *Escherichia coli* glucosamine-6-phosphate synthase.  
869 *Biochemistry*, 34(36), 11515–11520.
- 870 Bentele, K., Saffert, P., Rauscher, R., Ignatova, Z., & Blüthgen, N. (2013). Efficient translation  
871 initiation dictates codon usage at gene start. *Molecular Systems Biology*, 9(675), 1–10.  
872 <https://doi.org/10.1038/msb.2013.32>
- 873 Bergthorsson, U., Andersson, D. I., & Roth, J. R. (2007). Ohno’s dilemma: evolution of new  
874 genes under continuous selection. *Proceedings of the National Academy of Sciences of the  
875 United States of America*, 104(43), 17004–17009. <https://doi.org/10.1073/pnas.0707158104>
- 876 Blank, D., Wolf, L., Ackermann, M., & Silander, O. K. (2014). The predictability of molecular  
877 evolution during functional innovation. *Proceedings of the National Academy of Sciences*,  
878 111(8), 3044–3049. <https://doi.org/10.1073/pnas.1318797111>
- 879 Bonekamp, F., Clemmesen, K., Karlström, O., & Jensen, K. F. (1984). Mechanism of UTP-  
880 modulated attenuation at the *pyrE* gene of *Escherichia coli*: an example of operon polarity  
881 control through the coupling of translation to transcription. *The EMBO Journal*, 3(12),  
882 2857–2861. <https://doi.org/10.1002/j.1460-2075.1984.tb02220.x>
- 883 Bouquin, A. Le, & Khila, A. (2017). *Taxon-restricted genes at the origin of a novel trait  
884 allowing access to a new environment*. 390(October), 386–390.

- 885 https://doi.org/10.1126/science.aan2748
- 886 Braxton, B. L., Mullins, L. S., Raushel, F. M., & Reinhart, G. D. (1999). Allosteric dominance in  
887 carbamoyl phosphate synthetase. *Biochemistry*, 38(5), 1394–1401.
- 888 https://doi.org/10.1021/bi982097w
- 889 Chothia, C., Gough, J., Vogel, C., & Teichmann, S. A. (2003). Evolution of the protein  
890 repertoire. *Science*, 300(5626), 1701–1703. https://doi.org/10.1126/science.1085371
- 891 Conant, G. C., & Wolfe, K. H. (2008). Turning a hobby into a job: how duplicated genes find  
892 new functions. *Nature Reviews. Genetics*, 9(12), 938–950. https://doi.org/10.1038/nrg2482
- 893 Conrad, T. M., Joyce, A. R., Applebee, M. K., Barrett, C. L., Xie, B., Gao, Y., & Palsson, B. T.  
894 (2009). Whole-genome resequencing of *Escherichia coli* K-12 MG1655 undergoing short-  
895 term laboratory evolution in lactate minimal media reveals flexible selection of adaptive  
896 mutations. *Genome Biology*, 10(10), 1–12. https://doi.org/10.1186/gb-2009-10-10-r118
- 897 Copley, S. D. (2003). Enzymes with extra talents: Moonlighting functions and catalytic  
898 promiscuity. *Current Opinion in Chemical Biology*, 7(2), 265–272.  
899 https://doi.org/10.1016/S1367-5931(03)00032-2
- 900 Copley, S. D. (2017). Shining a light on enzyme promiscuity. *Current Opinion in Structural  
901 Biology*, 47, 167–175. https://doi.org/10.1016/j.sbi.2017.11.001
- 902 Datta, S., Costantino, N., & Court, D. L. (2006). A set of recombineering plasmids for gram-  
903 negative bacteria. *Gene*, 379(1–2), 109–115. https://doi.org/10.1016/j.gene.2006.04.018
- 904 Davis, J. H., Rubin, A. J., & Sauer, R. T. (2011). Design, construction and characterization of a  
905 set of insulated bacterial promoters. *Nucleic Acids Research*, 39(3), 1131–1141.  
906 https://doi.org/10.1093/nar/gkq810
- 907 Deatherage, D. E., & Barrick, J. E. (2014). Identification of mutations in laboratory-evolved

- 908        microbes from next-generation sequencing data using *breseq*. In L. Sun & W. Shou (Eds.),  
909        *Engineering and Analyzing Multicellular Systems: Methods and Protocols* (pp. 165–188).  
910        [https://doi.org/10.1007/978-1-4939-0554-6\\_12](https://doi.org/10.1007/978-1-4939-0554-6_12)
- 911        Eroglu, B., & Powers-Lee, S. G. (2002). Unmasking a functional allosteric domain in an  
912        allosterically nonresponsive carbamoyl-phosphate synthetase. *Journal of Biological  
913        Chemistry*, 277(47), 45466–45472. <https://doi.org/10.1074/jbc.M208185200>
- 914        Espah Borujeni, A., Channarasappa, A. S., & Salis, H. M. (2014). Translation rate is controlled  
915        by coupled trade-offs between site accessibility, selective RNA unfolding and sliding at  
916        upstream standby sites. *Nucleic Acids Research*, 42(4), 2646–2659.  
917        <https://doi.org/10.1093/nar/gkt1139>
- 918        Espah Borujeni, A., & Salis, H. M. (2016). Translation initiation is controlled by RNA folding  
919        kinetics via a ribosome drafting mechanism. *Journal of the American Chemical Society*,  
920        138(22), 7016–7023. <https://doi.org/10.1021/jacs.6b01453>
- 921        Ferla, M. P., Brewster, J. L., Hall, K. R., Evans, G. B., & Patrick, W. M. (2017). Primordial-like  
922        enzymes from bacteria with reduced genomes. *Molecular Microbiology*, 105(4), 508–524.  
923        <https://doi.org/10.1111/mmi.13737>
- 924        Francino, M. P. (2005). An adaptive radiation model for the origin of new gene functions. *Nature  
925        Genetics*, 37(6), 573–577. <https://doi.org/10.1038/ng1579>
- 926        Gibson, D. G., Young, L., Chuang, R., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009).  
927        Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*,  
928        6(5), 343–345. <https://doi.org/10.1038/nmeth.1318>
- 929        Glasner, M., Gerlt, J., & Babbitt, P. (2006). Evolution of enzyme superfamilies. *Current Opinion  
930        in Chemical Biology*, 10(5), 492–497. <https://doi.org/10.1016/j.cbpa.2006.08.012>

- 931 Goodman, D. B., Church, G. M., & Kosuri, S. (2013). Causes and effects of N-terminal codon  
932 bias in bacterial genes. *Science*, 342(6157), 475–479.  
933 <https://doi.org/10.1126/science.1241934>
- 934 Goto, M., Agari, Y., Omi, R., Miyahara, I., & Hirotsu, K. (2003). Expression, purification and  
935 preliminary X-ray characterization of N-acetyl- $\gamma$ -glutamyl-phosphate reductase from  
936 *Thermus thermophilus* HB8. *Acta Crystallographica - Section D Biological  
937 Crystallography*, 59(2), 356–358. <https://doi.org/10.1107/S0907444902020802>
- 938 Grenier, F., Matteau, D., Baby, V., & Rodrigue, S. (2014). Complete genome sequence of  
939 *Escherichia coli* BW25113. *Genome Announcements*, 2(5), 90005.  
940 <https://doi.org/10.1128/genomeA.01038-14>
- 941 Haldane, J. B. S. (1930). *Enzymes*. London, New York, Longmans, Green.
- 942 Hall, M. N., Gabay, J., Débarbouillé, M., & Schwartz, M. (1982). A role for mRNA secondary  
943 structure in the control of translation initiation. *Nature*, 295(5850), 616–618.  
944 <https://doi.org/10.1038/295616a0>
- 945 Hellmanns, J., Mortier, G., De Paepe, A., Speleman, F., & Vandesompele, J. (2007). qBase  
946 relative quantification framework and software for management and automated analysis of  
947 real-time quantitative PCR data. *Genome Biology*, 8(2), R19. <https://doi.org/10.1186/gb-2007-8-2-r19>
- 949 Howard, J. A. L., Delmas, S., Ivančić-Baće, I., & Bolt, E. L. (2011). Helicase dissociation and  
950 annealing of RNA-DNA hybrids by *Escherichia coli* Cas3 protein. *The Biochemical  
951 Journal*, 439(1), 85–95. <https://doi.org/10.1042/BJ20110901>
- 952 Hughes, A. L. (1994). The evolution of functionally novel proteins after gene duplication.  
953 *Proceedings of the Royal Society of London. Series B: Biological Sciences*, 256(1346), 119–

- 954 124. <https://doi.org/10.1098/rspb.1994.0058>
- 955 Hughes, C. S., Foehr, S., Garfield, D. A., Furlong, E. E., Steinmetz, L. M., & Krijgsveld, J.
- 956 (2014). Ultrasensitive proteome analysis using paramagnetic bead technology. *Molecular*
- 957 *Systems Biology*, 10(10), 757–757. <https://doi.org/10.15252/msb.20145625>
- 958 Jee, J., Rasouly, A., Shamovsky, I., Akivis, Y., Steinman, S. R., Mishra, B., & Nudler, E. (2016).
- 959 Rates and mechanisms of bacterial mutagenesis from maximum-depth sequencing. *Nature*,
- 960 534(7609), 693–696. <https://doi.org/10.1038/nature18313>
- 961 Jensen, K. F. (1993). The *Escherichia coli* K-12 “wild types” W3110 and MG1655 have an *rph*
- 962 frameshift mutation that leads to pyrimidine starvation due to low *pyrE* expression levels.
- 963 *Journal of Bacteriology*, 175(11), 3401–3407. <https://doi.org/10.1128/jb.175.11.3401-3407.1993>
- 965 Juárez-Vázquez, A. L., Edirisinghe, J. N., Verduzco-Castro, E. A., Michalska, K., Wu, C., Noda-
- 966 García, L., ... Barona-Gómez, F. (2017). Evolution of substrate specificity in a retained
- 967 enzyme driven by gene loss. *eLife*, 6, 1–21. <https://doi.org/10.7554/eLife.22679>
- 968 Kershner, J. P., Yu McLoughlin, S., Kim, J., Morgenthaler, A., Ebmeier, C. C., Old, W. M., &
- 969 Copley, S. D. (2016). A synonymous mutation upstream of the gene encoding a weak-link
- 970 enzyme causes an ultrasensitive response in growth rate. *Journal of Bacteriology*, 198(20),
- 971 2853–2863. <https://doi.org/10.1128/JB.00262-16>
- 972 Khanal, A., Yu McLoughlin, S., Kershner, J. P., & Copley, S. D. (2015). Differential effects of a
- 973 mutation on the normal and promiscuous activities of orthologs: Implications for natural
- 974 and directed evolution. *Molecular Biology and Evolution*, 32(1), 100–108.
- 975 <https://doi.org/10.1093/molbev/msu271>
- 976 Khersonsky, O., & Tawfik, D. S. (2010). Enzyme promiscuity: A mechanistic and evolutionary

- 977 perspective. *Annual Review of Biochemistry*, 79(1), 471–505.
- 978 <https://doi.org/10.1146/annurev-biochem-030409-143718>
- 979 Kim, J., Webb, A. M., Kershner, J. P., Blaskowski, S., & Copley, S. D. (2014). A versatile and
- 980 highly efficient method for scarless genome editing in *Escherichia coli* and *Salmonella*
- 981 *enterica*. *BMC Biotechnology*, 14(1), 84. <https://doi.org/10.1186/1472-6750-14-84>
- 982 Kitagawa, M., Ara, T., Arifuzzaman, M., Ioka-Nakamichi, T., Inamoto, E., Toyonaga, H., &
- 983 Mori, H. (2005). Complete set of ORF clones of *Escherichia coli* ASKA library (A
- 984 complete set of *E. coli* K-12 ORF archive): unique resources for biological research. *DNA*
- 985 *Research*, 12(5), 291–299. <https://doi.org/10.1093/dnarese/dsi012>
- 986 Knöppel, A., Knopp, M., Albrecht, L. M., Lundin, E., Lustig, U., Näsvall, J., & Andersson, D. I.
- 987 (2018). Genetic adaptation to growth under laboratory conditions in *Escherichia coli* and
- 988 *Salmonella enterica*. *Frontiers in Microbiology*, 9(April), 756.
- 989 <https://doi.org/10.3389/fmicb.2018.00756>
- 990 Kugelberg, E., Kofoid, E., Reams, A. B., Andersson, D. I., & Roth, J. R. (2006). Multiple
- 991 pathways of selected gene amplification during adaptive mutation. *Proceedings of the*
- 992 *National Academy of Sciences of the United States of America*, 103(46), 17319–17324.
- 993 <https://doi.org/10.1073/pnas.0608309103>
- 994 Legrain, C., & Stalon, V. (1976). Ornithine carbamoyltransferase from *Escherichia coli* W.
- 995 Purification, structure and steady-state kinetic analysis. *European Journal of Biochemistry*,
- 996 63(1), 289–301. <https://doi.org/10.1111/j.1432-1033.1976.tb10230.x>
- 997 Ludovice, M., Martin, J. F., Carrachas, P., & Liras, P. (1992). Characterization of the
- 998 *Streptomyces clavuligerus* argC gene encoding N-acetylglutamyl-phosphate reductase:
- 999 expression in *Streptomyces lividans* and effect on clavulanic acid production. *Journal of*

- 1000        *Bacteriology*, 174(14), 4606–4613. <https://doi.org/10.1128/jb.174.14.4606-4613.1992>
- 1001        McLoughlin, S. Y., & Copley, S. D. (2008). A compromise required by gene sharing enables  
1002        survival: Implications for evolution of new enzyme activities. *Proceedings of the National  
1003        Academy of Sciences of the United States of America*, 105(36), 13497–13502.  
1004        <https://doi.org/10.1073/pnas.0804804105>
- 1005        Mezl, V. A., & Knox, W. E. (1976). Properties and analysis of a stable derivative of pyrroline-5-  
1006        carboxylic acid for use in metabolic studies. *Analytical Biochemistry*, 74(2), 430–440.  
1007        [https://doi.org/10.1016/0003-2697\(76\)90223-2](https://doi.org/10.1016/0003-2697(76)90223-2)
- 1008        Mutualik, V. K., Guimaraes, J. C., Cambray, G., Lam, C., Christoffersen, M. J., Mai, Q.-A., ...  
1009        Endy, D. (2013). Precise and reliable gene expression via standard transcription and  
1010        translation initiation elements. *Nature Methods*, 10(4), 354–360.  
1011        <https://doi.org/10.1038/nmeth.2404>
- 1012        Näsvall, J., Sun, L., Roth, J. R., & Andersson, D. I. (2012). Real-time evolution of new genes by  
1013        innovation, amplification, and divergence. *Science*, 338(6105), 384–387.  
1014        <https://doi.org/10.1126/science.1226521>
- 1015        Nei, M., & Rooney, A. P. (2005). Concerted and birth-and-death evolution of multigene families.  
1016        *Annual Review of Genetics*, 39(1), 121–152.  
1017        <https://doi.org/10.1146/annurev.genet.39.073003.112240>
- 1018        O'Brien, P. J., & Herschlag, D. (1999). Catalytic promiscuity and the evolution of new  
1019        enzymatic activities. *Chemistry & Biology*, 6(4), R91–R105. [https://doi.org/10.1016/S1074-5521\(99\)80033-7](https://doi.org/10.1016/S1074-5521(99)80033-7)
- 1021        Ohno, S. (1970). *Evolution by Gene Duplication*. <https://doi.org/10.1007/978-3-642-86659-3>
- 1022        Olson, M. V. (1999). When less is more: Gene loss as an engine of evolutionary change. *The*

- 1023        *American Journal of Human Genetics*, 64(1), 18–23. <https://doi.org/10.1086/302219>
- 1024        Page, R., Nelson, M. S., von Delft, F., Elsliger, M.-A., Canaves, J. M., Brinen, L. S., ... Wilson,
- 1025        I. A. (2003). Crystal structure of  $\gamma$ -glutamyl phosphate reductase (TM0293) from
- 1026        *Thermotoga maritima* at 2.0 Å resolution. *Proteins: Structure, Function, and*
- 1027        *Bioinformatics*, 54(1), 157–161. <https://doi.org/10.1002/prot.10562>
- 1028        Pettersson, M. E., Sun, S., Andersson, D. I., & Berg, O. G. (2009). Evolution of new gene
- 1029        functions: simulation and analysis of the amplification model. *Genetica*, 135(3), 309–324.
- 1030        <https://doi.org/10.1007/s10709-008-9289-z>
- 1031        Pierrat, O. A., & Raushel, F. M. (2002). A functional analysis of the allosteric nucleotide
- 1032        monophosphate binding site of carbamoyl phosphate synthetase. *Archives of Biochemistry*
- 1033        and *Biophysics*, 400(1), 34–42. <https://doi.org/10.1006/abbi.2002.2767>
- 1034        Proft, T., & Baker, E. N. (2009). Pili in Gram-negative and Gram-positive bacteria — structure,
- 1035        assembly and their role in disease. *Cellular and Molecular Life Sciences*, 66(4), 613–635.
- 1036        <https://doi.org/10.1007/s00018-008-8477-4>
- 1037        Rauwerdink, A., Lunzer, M., Devamani, T., Jones, B., Mooney, J., Zhang, Z.-J., ... Dean, A. M.
- 1038        (2016). Evolution of a catalytic mechanism. *Molecular Biology and Evolution*, 33(4), 971–
- 1039        979. <https://doi.org/10.1093/molbev/msv338>
- 1040        Reams, A. B., Kofoid, E., Savageau, M., & Roth, J. R. (2010). Duplication frequency in a
- 1041        population of *Salmonella enterica* rapidly approaches steady state with or without
- 1042        recombination. *Genetics*, 184(4), 1077–1094. <https://doi.org/10.1534/genetics.109.111963>
- 1043        Reams, A. B., & Neidle, E. L. (2004). Selection for gene clustering by tandem duplication.
- 1044        *Annual Review of Microbiology*, 58(1), 119–142.
- 1045        <https://doi.org/10.1146/annurev.micro.58.030603.123806>

- 1046 Rodriguez-Beltran, J., Hernandez-Beltran, J. C. R., DelaFuente, J., Escudero, J. A., Fuentes-
- 1047 Hernandez, A., MacLean, R. C., ... San Millan, A. (2018). Multicopy plasmids allow
- 1048 bacteria to escape from fitness trade-offs during evolutionary innovation. *Nature Ecology &*
- 1049 *Evolution*, 2(5), 873–881. <https://doi.org/10.1038/s41559-018-0529-z>
- 1050 Rubino, S. D., Nyunoya, H., & Lusty, C. J. (1986). Catalytic domains of carbamyl phosphate
- 1051 synthetase. Glutamine-hydrolyzing site of *Escherichia coli* carbamyl phosphate synthetase.
- 1052 *The Journal of Biological Chemistry*, 261(24), 11320–11327.
- 1053 Rubino, S. D., Nyunoya, H., & Lusty, C. J. (1987). In vivo synthesis of carbamyl phosphate from
- 1054 NH<sub>3</sub> by the large subunit of *Escherichia coli* carbamyl phosphate synthetase. *Journal of*
- 1055 *Biological Chemistry*, 262(9), 4382–4386.
- 1056 Salis, H. M., Mirsky, E. A., & Voigt, C. A. (2009). Automated design of synthetic ribosome
- 1057 binding sites to control protein expression. *Nature Biotechnology*, 27(10), 946–950.
- 1058 <https://doi.org/10.1038/nbt.1568>
- 1059 Scharff, L. B., Childs, L., Walther, D., & Bock, R. (2011). Local absence of secondary structure
- 1060 permits translation of mRNAs that lack ribosome-binding sites. *PLoS Genetics*, 7(6),
- 1061 e1002155. <https://doi.org/10.1371/journal.pgen.1002155>
- 1062 Snodgrass, P. J., & Parry, D. J. (1969). The kinetics of serum ornithine carbamoyltransferase.
- 1063 *Journal of Laboratory and Clinical Medicine*, 73(6), 940–950.
- 1064 Stapleton, M. A., Javid-Majd, F., Harmon, M. F., Hanks, B. A., Grahmann, J. L., Mullins, L. S.,
- 1065 & Raushel, F. M. (1996). Role of conserved residues within the carboxy phosphate domain
- 1066 of carbamoyl phosphate synthetase. *Biochemistry*, 35(45), 14352–14361.
- 1067 <https://doi.org/10.1021/bi961183y>
- 1068 Starr, T. N., Picton, L. K., & Thornton, J. W. (2017). Alternative evolutionary histories in the

- 1069 sequence space of an ancient protein. *Nature*, 549(7672), 409–413.
- 1070 <https://doi.org/10.1038/nature23902>
- 1071 Steitz, J. A. (1969). Polypeptide chain initiation: Nucleotide sequences of the three ribosomal
- 1072 binding sites in bacteriophage R17 RNA. *Nature*, 224(5223), 957–964.
- 1073 <https://doi.org/10.1038/224957a0>
- 1074 Storz, J. F. (2016). Gene duplication and evolutionary innovations in hemoglobin-oxygen
- 1075 transport. *Physiology*, 31(3), 223–232. <https://doi.org/10.1152/physiol.00060.2015>
- 1076 Takahashi, C. N., Miller, A. W., Ekness, F., Dunham, M. J., & Klavins, E. (2015). A low cost,
- 1077 customizable turbidostat for use in synthetic circuit characterization. *ACS Synthetic Biology*,
- 1078 4(1), 32–38. <https://doi.org/10.1021/sb500165g>
- 1079 Thoden, J. B., Huang, X., Raushel, F. M., & Holden, H. M. (2002). Carbamoyl-phosphate
- 1080 synthetase: Creation of an escape route for ammonia. *Journal of Biological Chemistry*,
- 1081 277(42), 39722–39727. <https://doi.org/10.1074/jbc.M206915200>
- 1082 Thoden, J. B., Raushel, F. M., Wesenberg, G., & Holden, H. M. (1999). The binding of inosine
- 1083 monophosphate to *Escherichia coli* carbamoyl phosphate synthetase. *Journal of Biological*
- 1084 *Chemistry*, 274(32), 22502–22507. <https://doi.org/10.1074/jbc.274.32.22502>
- 1085 Todd, A. E., Orengo, C. A., & Thornton, J. M. (2001). Evolution of function in protein
- 1086 superfamilies, from a structural perspective. *Journal of Molecular Biology*, 307(4), 1113–
- 1087 1143. <https://doi.org/10.1006/jmbi.2001.4513>
- 1088 Wiśniewski, J. R., & Gaugaz, F. Z. (2015). Fast and sensitive total protein and peptide assays for
- 1089 proteomic analysis. *Analytical Chemistry*, 87(8), 4110–4116.
- 1090 <https://doi.org/10.1021/ac504689z>
- 1091 Wolfinger, M. T., Svrcek-Seiler, W. A., Flamm, C., Hofacker, I. L., & Stadler, P. F. (2004).

- 1092 Efficient computation of RNA folding dynamics. *Journal of Physics A: Mathematical and*  
1093 *General*, 37(17), 4731–4741. <https://doi.org/10.1088/0305-4470/37/17/005>
- 1094 Zhou, K., Zhou, L., Lim, Q., Zou, R., Stephanopoulos, G., & Too, H.-P. (2011). Novel reference  
1095 genes for quantifying transcriptional responses of *Escherichia coli* to protein overexpression  
1096 by quantitative PCR. *BMC Molecular Biology*, 12(1), 18. <https://doi.org/10.1186/1471->  
1097 2199-12-18
- 1098 Zwietering, M. H., Jongenburger, I., Rombouts, F. M., & van 't Riet, K. (1990). Modeling of the  
1099 bacterial growth curve. *Applied and Environmental Microbiology*, 56(6), 1875–1881.
- 1100