Proteostasis modulates gene dosage evolution in antibiotic-resistant bacteria

Chinmaya Jena, Saillesh Chinnaraj, Soham Deolankar and Nishad Matange[#]

Department of Biology, Indian Institute of Science Education and Research, Pashan, Pune 411008 India

[#]nishad@iiserpune.ac.in

Abstract

Evolution of gene expression frequently drives antibiotic resistance in bacteria. We had previously (Patel and Matange, *eLife*, 2021) shown that in *Escherichia coli*, mutations at the *mgrB* locus were beneficial in trimethoprim and led to overexpression of dihydrofolate reductase (DHFR), encoded by the *folA* gene. Here, we show that DHFR levels are further enhanced by spontaneous duplication of a genomic segment encompassing *folA* and spanning hundreds of kilobases. This duplication was rare in wild type *E. coli*. However, its frequency was elevated in a *lon*-knockout strain, altering the mutational landscape early during trimethoprim adaptation. We then exploit this system to investigate the relationship between trimethoprim pressure and *folA* copy number. During long-term evolution, *folA* duplications were frequently reversed. Reversal was slower under antibiotic pressure, first requiring the acquisition of point mutations in DHFR or its promoter. Unexpectedly, despite resistance-conferring point mutations, some populations under high trimethoprim pressure maintained *folA* duplication to compensate for low abundance DHFR mutants. We find that evolution of gene dosage depends on expression demand, which is generated by antibiotic and exacerbated by proteolysis of drug-resistant mutants of DHFR. We propose a novel role for proteostasis as a determinant of copy number evolution in antibiotic-resistant bacteria.

Introduction

The dosage of a gene, i.e., copy number relative to the genome, directly impacts its level of expression and can alter organismal phenotype and fitness [1-3]. Mutations that increase copy numbers, termed as gene duplications and amplifications (GDAs), are implicated in diverse biological phenomena including cancer progression [4-6], speciation [7-9], reproductive isolation [10-12] and antimicrobial resistance (AMR) [2, 13]. GDAs are important for rapid adaptation since their frequencies of occurrence are higher by several orders of magnitude (~1000-fold) than point mutations [2, 13-16]. Following the initial GDA mutation, however, amplified genes could be lost over time, pseudogenized, sub-functionalized or neo-functionalized [17-19]. Which of these possibilities is realized depends on the nature and strength of selection imposed by the environment, as well as the fitness cost associated with maintaining multiple gene copies [17-19].

In bacteria, GDA mutations are frequently encountered during adaptation to antibiotics [13] or heavy metals [20, 21]. They occur by RecA-dependent nonequal recombination as well as RecA-independent processes and often involve repeat sequences such as in IS-elements [22-25]. Amplified genes in drug-resistant bacteria code for effectors of resistance such as drug inactivating enzymes, drug target modifiers or "by-pass" mechanisms [26-28], but may also compensate for resistance-associated fitness costs [29]. More recently, the view that GDAs are less "hard-wired" than point mutations has emerged. For instance, transient alterations in gene dosage are a common mechanism of heteroresistance, i.e. resistant sub-populations within a drug-sensitive bacterial isolate [20, 30, 31]. There is also evidence to suggest that copy number variations are more common in bacteria than appreciated. It is estimated that as many as 20% of cells in a bacterial population may harbour a duplication of some genomic region [32]. These copy number polymorphisms create within-population heterogeneities in gene expression, serving as a bet-hedging strategy with adaptive value in fluctuating environments [33].

Dihydrofolate reductase (DHFR) is a ubiquitous enzyme crucial for growth and division. It catalyzes the formation of tetrahydrofolate, an important cofactor in nucleotide and amino acid biosynthesis. DHFR is coded by an essential gene in most organisms and its deletion or pharmacological inhibition results in cell death or growth stasis [34]. Not surprisingly, anti-folates are effective therapeutic agents for diverse pathologies [35, 36]. For instance,

methotrexate, used in the treatment of cancers and autoimmune conditions [37], pyrimethamine, used as an anti-malarial [38], and trimethoprim, used as an anti-bacterial [39], all of which are competitive inhibitors of DHFR. Interestingly, copy number amplification and consequent overexpression of DHFR confers resistance in all three systems [40-43]. Thus, the gene dosage of DHFR appears to be tightly regulated across evolutionarily divergent organisms, making it a useful system to understand the principles underlying selection and evolutionary fate of GDA mutations.

In *Escherichia coli* DHFR is coded by the *folA* gene. Several mutation hot spots for trimethoprim resistance have been identified in this enzyme [44-47]. Additionally, mechanisms that transcriptionally up-regulate *folA* expression are also known in trimethoprim resistant *E. coli*. These include cis-regulatory mutations in the *folA* promoter and trans-acting loss of function mutations in *mgrB*, which upregulate *folA* transcription in a PhoQP-dependent manner [44-46, 48]. Earlier work from our group identified Lon, a conserved quality control protease in bacteria, as a post-translational regulator of trimethoprim resistance. Lon degraded mutant DHFRs and reduced their half-lives *in vivo*, directly affecting genotype-phenotype relationships of resistance-conferring mutations [49, 50]. In this study, we first compare early mutational events in wild type and Lon-deficient *E. coli* challenged with trimethoprim. We show that a large genomic duplication encompassing *folA* confers trimethoprim resistance in both genetic backgrounds, though its frequency of occurrence is elevated in *lon*-knockout *E. coli*. We then use this system to ask what factors determine the fate of duplicated *folA* gene copies during evolution at different trimethoprim pressures.

Results

A large IS-flanked duplication encompassing *folA* mediates rapid adaptation to trimethoprim in *E. coli*

We had previously evolved 5 replicate populations of wild type *E. coli* K-12 MG1655 in sub-MIC trimethoprim (300 ng/mL) for 25 generations. We then isolated and sequenced the genome of a dominant trimethoprim-resistant clone from each population (designated as WTMPR1-5). These sequencing results have been reported earlier, in Patel and Matange (2021) [46]. Briefly, all isolates harboured loss-of-function mutations in the *mgrB* gene (Figure 1 A) that resulted in PhoQP-dependent up-regulation of DHFR expression [46]. Additionally, isolates WTMPR1, 3 and 5 had mutations at the active site of DHFR, while WTMPR4 harboured a large genomic duplication of ~0.6 Mb encompassing the *folA* gene (Figure 1 A, B). We performed similar evolution experiments starting with an isogenic *E. coli Δlon* strain to ask if early adaptation to trimethoprim was altered by Lon-deficiency. Among trimethoprim-resistant isolates derived from *E. coli Δlon*, designated as LTMPR1-5, four harboured mutations at the *mgrB* locus (Figure 1 A). However, none of them had mutations in DHFR or its promoter. Instead, LTMPR2-4 had evolved GDA mutations similar to WTMPR4, while LTMPR1 had a duplication in an overlapping but shorter genomic stretch (Figure 1 A, B). In all trimethoprim-resistant isolates, duplicated genomic regions were flanked by IS-elements (Figure 1 B). Isolates with and without genomic duplications showed comparable trimethoprim IC50 values (Figure 1 A), indicating that GDA mutations were an alternate genetic path to evolving resistance and occurred more frequently in Lon-deficient *E. coli*.

Seeking to explain the higher frequency of GDA mutations in the *lon*-knockout, we asked whether IS-transposases were sensitive to proteolysis by Lon. Indeed, heterologously expressed InsB and InsL had greater stability in the *lon*-knockout (Figure S1), suggesting that their accumulation may contribute to higher GDA frequency in *E. coli* Δ*lon*. If true it would follow that a higher proportion of GDA mutants would be isolated from *E. coli Δlon* during adaptation to other environments as well. In agreement with this prediction, GDA mutations were detected after evolution of *E. coli Δlon* in 2 of the 5 additional growth conditions tested, namely antibiotic-free and spectinomycin-supplemented media (Figure 1 C, S2). In all cases

120 duplications were flanked by IS elements (Figure S2). Further, the regions of the genome that
121 were duplicated were similar among isolates derived from the same environment, but
122 different across environments (Figure 1 A, S2). On the other hand, wild type *E. coli* did not
123 show GDA mutations in any of the other tested conditions.
124
125 The duplicated genomic region in trimethoprim-resistant isolates included genes previously
126 associated with resistance to trimethoprim, namely, *folA* which codes for DHFR, and *acrA*
127 and *B*, which code for components of the AcrAB-TolC efflux pump (Figure 1 B). In one of
128 the isolates, i.e. LTMPR1, the amplified stretch encompassed *folA* but not *acrAB* without
129 affecting trimethoprim IC50 values (Figure 1 A, B). We therefore reasoned that duplication of
130 *folA* was the main effector of adaptation to trimethoprim. Immunoblotting showed that DHFR
131 protein was indeed overproduced in isolates LTMPR1 and WTMPR4 over and above the
132 increase mediated by inactivation of *mgrB* alone (Figure 2 A).
133
134 Though DHFR overproduction may contribute to the selection of genomic amplifications in
135 trimethoprim-resistant bacteria, ~173 (LTMPR1) and ~531 (WTMPR4) other genes were also
136 duplicated in these isolates. To understand the global transcriptional burden imposed by this
137 amplification, we analysed the gene expression profiles of WTMPR4 and LTMPR1 using
138 RNA-sequencing (Supplementary File 1). Since both isolates had loss-of-function mutations
139 in *mgrB*, which cause large changes in global gene expression [51], we also analysed the
140 transcriptomes of *E. coli ΔmgrB* and *E. coli ΔlonΔmgrB* strains for comparison
141 (Supplementary File 1). The level of *folA* mRNA was higher in LTMPR1 (2.3-fold) and
142 WTMPR4 (5.7-fold) over respective controls (Supplementary File 1), in agreement with
143 immunoblotting data. Other duplicated genes were mildly overexpressed by ~1.7-fold on an
144 average in LTMPR1 (over *E. coli ΔlonΔmgrB*) and ~1.3-fold in WTMPR4 (over *E. coli*
145 *ΔmgrB*) (Figure 2 B, Supplementary File 1). However, there was large variation in expression
146 level among duplicated genes in these isolates (Figure 2 B), indicating that increase in gene
147 dosage did not necessarily result in enhanced expression levels. Instead, global gene
148 expression profiles of WTMPR4 and LTMPR1 were correlated with *E. coli ΔmgrB* and *E.*
149 *coli ΔlonΔmgrB* respectively (Figure 2 C, D) and most of the top 50 up/down-regulated genes
150 in WTMPR4 and LTMPR1 could be attributed to loss of *mgrB* alone (Figure S3). In line with
151 this result, strains with and without genomic amplifications had comparable fitness costs (~20
152 %) in drug-free media, which could also be attributed to inactivation of *mgrB* (Figure 2 E).

153 Thus, we concluded that large GDA mutations encompassing *folA* led to DHFR
154 overexpression and represented an alternate mechanism for trimethoprim-resistance in *E. coli*
155 with undetectable additional cost.
156
157 **Maintenance of *folA* duplication is linked to antibiotic-generated demand for DHFR**
158 **expression**
159
160 We next investigated the evolutionary fate of duplicate *folA* gene copies by evolving the
161 LTMPR1 isolate, which harboured the shortest duplication, in the absence of antibiotic. We
162 also evolved this isolate at sustained drug pressure, i.e. trimethoprim at the same
163 concentration at which it had originally emerged (300 ng/mL). Using three replicate
164 populations in each condition, we asked how DHFR expression and trimethoprim IC50
165 values changed over ~252 generations of evolution and used genome sequencing to identify
166 the underlying mutational basis. As controls, we performed similar evolution experiments
167 with WTMPR4 (which harboured a GDA and a functional Lon protease) and WTMPR5
168 (which harboured a *folA* point mutation and a functional Lon protease).
169
170 *Evolution in trimethoprim-free media*
171
172 In drug-free conditions, trimethoprim IC50 of LTMPR1 rapidly declined and a concomitant
173 loss of colony formation on trimethoprim-supplemented plates occurred (Figure 3 A, B).
174 Evolved WTMPR4, but not WTMPR5, showed similar loss of colony formation on
175 trimethoprim indicating that resistance mediated by GDA mutations was far less stable than
176 point mutations regardless of the presence of Lon (Figure 3 B). For LTMPR1, these trends
177 were explained by a decline in DHFR expression levels over evolution and reversion to a
178 single *folA* copy (Figure 3 C, D, S4). Mutation in *mgrB* was, however, consistently
179 maintained till the end of the experiment (Supplementary File 2). In line with these data,
180 trimethoprim IC50s of evolving populations stabilized at a value similar to *E. coli*
181 *ΔlonΔmgrB* (Figure 3 A). Interestingly, a second GDA mutation ("GDA-2", ~2.2 Mb to 3.3
182 Mb) emerged in these populations (Figure 3 D, S4). The GDA-2 mutation was apparently lost
183 and regained multiple times during the experiment (Figure 3 D, S4), suggesting a dynamic
184 maintenance in the population rather than fixation. Further, the region of the genome
185 amplified in GDA-2 was similar to what we had observed when *E. coli Δlon* was evolved in

186 drug-free media (Figure S2, S4), indicating that GDA-2 was likely a fitness enhancing
187 mutation and unrelated to trimethoprim adaptation.
188
189 To further investigate the emergence of GDA-2 during evolution of LTMPR1 in drug-free
190 media we sequenced 6 randomly picked colonies from each of the lineages at the final time
191 point (Figure 3 E). All the sequenced isolates had a single copy of *folA* and a majority of
192 isolates showed the presence of GDA-2, corroborating population sequencing results (Figure
193 3 E, Supplementary File 2). Interestingly, some of the isolates that did not harbour GDA-2
194 had evolved mutations in the *rpoS* gene (Figure 3 E, Supplementary File 2). Loss-of-function
195 mutations in RpoS are fixed at late time points during adaptation to trimethoprim and
196 enhance the fitness of *E. coli* without conferring resistance [46, 51]. We also sequenced a few
197 randomly picked isolates from WTMPR4 and WTMPR5 evolved lines (2 from each of the 3
198 replicate lineages in each case) to check for the presence of GDA-2 and *rpoS* mutations. As
199 expected, all WTMPR4 isolates had reverted to a single copy of *folA*, while WTMPR5
200 isolates retained their ancestrally inherited *folA* mutation. None of them had evolved the
201 GDA-2 mutation, while a few of the isolates from WTMPR4 and WTMPR5 lines had
202 acquired *rpoS* mutations (Figure 3 F, G).
203
204 Consistent with earlier studies, and the role of mutations in *rpoS* in general fitness
205 enhancement in *E. coli*, we found that an LTMPR1-derived isolate with mutant *rpoS*
206 recovered fitness in drug-free media (Figure 4 A). Interestingly, an isolate with the GDA-2
207 mutation had also recovered fitness (Figure 4 A). Coupled with the fact that the GDA-2
208 mutation occurred only in a *lon*-deficient background, and GDA-2 and *rpoS* mutations were
209 mutually exclusive we wondered whether the fitness enhancing effects of GDA-2 were also
210 mediated by reduced RpoS activity. Lon is known to reduce the levels of RpoS during the
211 exponential growth phase of *E. coli* [52]. Therefore, in a *lon*-knockout we reasoned that at
212 least some targets of RpoS would be overexpressed. To test this prediction, we analysed the
213 transcriptome of *E. coli Δlon* and specifically looked at the expression levels of 168 known
214 targets of RpoS (Supplementary File 3). Of these, 41 genes were overexpressed in the *lon*-
215 knockout by at least 1.5-fold over wild type, including genes induced in the stationary phase
216 and by various stresses such as *poxB*, *aidB* and *cbpA* (Figure 4 B, C). Levels of *rpoS* mRNA
217 were however unaffected in this strain. Next, we compared the expression levels of these 41
218 RpoS-targets in LTMPR1 and *E. coli Δlon ΔmgrB* with *E. coli Δlon* as the reference and

219 found that most had similar expression levels in these three strains (Figure 4 C). Indeed a few
220 of these genes showed higher expression levels in LTMPR1 and *E. coli Δlon ΔmgrB*, which
221 was not unexpected since loss of *mgrB* is known stabilize RpoS [53]. Finally, we checked
222 their expression in isolates derived from LTMPR1 which harboured either GDA-2 or *rpoS*
223 mutations. Indeed, in both isolates RpoS target genes were down-regulated indicating a re-
224 setting of RpoS activity (Figure 4 C). Further, in the isolate harbouring GDA-2, *rpoS* mRNA
225 levels were themselves lowered to 0.49x of *E. coli Δlon* (Supplementary File 3). Finally,
226 deletion of *rpoS* from *E. coli Δlon* enhanced the fitness of the strain (Figure 4 D). Thus,
227 evolution of LTMPR1 in antibiotic-free media was characterized by rapid reversal of GDA-
228 mediated resistance and acquisition of mutations that enhanced fitness by reducing the levels
229 of RpoS-transcribed genes (Figure 4 D).
230
231 *Evolution at constant trimethoprim pressure*
232
233 In the presence of trimethoprim (300 ng/mL), antibiotic resistance level and DHFR
234 expression remained high over the course of evolution of LTMPR1 (Figure 5 A, B, C).
235 WTMPR4 and WTMPR5 too remained resistant, and in the case of WTMPR4 an increased
236 colony forming efficiency was observed at a late time point (Figure 5 B). These data
237 suggested that under drug pressure high *folA* copy number was likely to be maintained.
238 Indeed, population sequencing of LTMPR1 lineages showed that the ancestral GDA mutation
239 was maintained until 196 generations of evolution in all three replicates (Figure 5 D, S5). By
240 252 generations, however, a majority of bacteria in Lines A and C had reverted to a single
241 *folA* copy (Figure 5 D, S5) and had acquired a point mutation either in the *folA* gene (Line C)
242 or its promoter (Lines A) (Figure 5 D, Supplementary File 4). Line B, on the other hand,
243 retained high *folA* copy number, but with a reduction in the length of the genomic duplication
244 (Figure 5 D, S5). Line B had also acquired a mutation in the *folA* promoter, however its
245 frequency in the population was low (Figure 5 D, Supplementary File 4).
246
247 In agreement with population sequencing, none of the resistant isolates derived from
248 LTMPR1 Line A at the final time point showed the ancestral GDA. Instead, they all
249 harboured a $C_{-35}T$ mutation in the *folA* promoter that is known to up-regulate DHFR
250 expression [51] (Figure 5 E, Supplementary File 4). Among Line B isolates, many had
251 retained the ancestral duplication, without acquiring any additional mutations in *folA*. One of

252 the isolates from this Line that had reverted to a single *folA* copy had gained the $A_{-64}T$ *folA*
253 promoter mutation (Figure 5 E, Supplementary File 4). In Line C too we found that isolates
254 that had lost the ancestral GDA had evolved a point mutation in the *folA* promoter ($+TG_{-30}$) or
255 a $Trp_{30}Arg$ mutation in DHFR (Figure 5 E, Supplementary File 4). We found similar results in
256 sequenced isolates from WTMPR4 lineages, i.e. loss of GDA and gain of a point mutation
257 either in *folA* or its promoter (Figure 5 F). WTMPR5-derived isolates did not accumulate
258 additional mutations in *folA* but did acquire mutations in *rpoS* (Figure 5 G), showing that at
259 this drug pressure additional mutations in *folA* were not required. Interestingly, two isolates
260 from LTMPR1 Lines B and C had evolved new GDA mutations encompassing *folA*. In one of
261 the Line B isolates (Isolate 2) the ancestral GDA was replaced by 5x amplification of a
262 shorter genomic region (~12 kb) encompassing *folA* (Figure S6). Both ends of this newly
263 amplified region were flanked by IS1A (Figure S6). On the other hand, one of the Line C
264 isolates (Isolate 5) had an expanded duplication beyond what was originally found in
265 LTMPR1 (Figure S6). Also, both copies of *folA* in this isolate coded for DHFR-$Trp_{30}Arg$
266 (Figure 5 E, S6), suggesting that the original GDA was lost and an extended GDA mutation
267 was gained after fixation of the mutant *folA* allele. Thus, even under drug pressure, in most
268 bacteria GDA mutations were reversed. Notably, reversion to a single *folA* copy was slower
269 than in drug-free media as it first required the fixation of point mutations at the *folA* locus.
270 However, in a smaller sub-population, more than one *folA* gene copy was maintained. This
271 maintenance was dynamic and likely involved repeated loss and gain of GDA mutations.
272
273 To further investigate the replacement of GDA by point mutants, we checked to see if under
274 drug pressure point mutants could invade a population harbouring GDA mutations. We
275 allowed LTMPR1-derived isolates with point mutations in *folA* (marked with Δ*lacZ::Kan*) to
276 compete with a GDA-harbouring isolate from the same time point in evolution (to nullify the
277 effects of other media adaptations) and asked how rapidly the former could invade a
278 population of the latter. Since for this assay we used point mutant strains genetically marked
279 with Δ*lacZ::Kan* we verified the neutrality of the Δ*lacZ::Kan* marker by checking if the
280 marked isolate could invade its unmarked parent. Indeed, in the presence of trimethoprim
281 (300 ng/mL) point mutants were enriched over the GDA mutant in co-culture by about 100-
282 fold (Figure 6 A) recreating the results of our evolution experiment. A possible explanation
283 for this finding may be higher resistance level conferred by *folA* point mutations than GDA.
284 However, most isolates with *folA* point mutations or harbouring GDAs had similar IC50

285 values as the ancestral LTMPR1 (Figure 6 B). These isolates also had comparable doubling
286 times in the presence of 300 ng/mL trimethoprim (Figure 6 B). Thus, though there were no
287 large observable differences in growth rate or drug IC50 values between strains, point
288 mutants could indeed invade and outcompete GDA mutations, reflecting either small
289 differences in fitness between them or their relative stabilities in culture.
290
291 **Maintenance of high *folA* gene copy number under drug pressure is driven by unstable**
292 **mutants of DHFR**
293
294 Though sustained drug pressure did extend the lifetime of *folA* gene duplication, acquisition
295 of trimethoprim-resistant point mutations facilitated reversion to a single gene copy. We
296 therefore asked whether *folA* duplication could be maintained under conditions of increasing
297 antibiotic pressure, i.e., trimethoprim concentrations starting at 1 μg/mL and ending at 10
298 μg/mL over 252 generations of evolution (Figure 7 A). Under these conditions, trimethoprim
299 IC50 values increased sharply over the course of evolution eventually reaching ~40 to ~160-
300 fold higher than ancestral LTMPR1 (Figure 7 A). Colony formation at high trimethoprim
301 concentrations also increased at later generations of evolution (Figure 7 B), reflecting the
302 high drug pressure. Similarly evolved WTMPR4 and WTMPR5 also showed enhanced
303 colony formation at high trimethoprim concentrations (Figure 7 B).
304
305 Curiously, we observed that the 3 replicate lineages of LTMPR1 showed different trends in
306 DHFR expression. Line A maintained a high level of DHFR protein for the duration of the
307 experiment. On the other hand, Lines B and C showed high DHFR expression until 196
308 generations, after which it sharply declined (Figure 7 C). Genome sequencing revealed that
309 all 3 lines convergently evolved the $Trp_{30}Arg$ mutation in DHFR early during evolution and
310 had fixed this mutation by 196 generations. All 3 lines also retained *folA* duplication until this
311 time point (Figure 7 D, S7). Subsequently, Line A evolved an additional $C_{-35}T$ *folA* promoter
312 mutation and had partially lost the ancestral GDA (Figure 7 D, S7, Supplementary File 5). In
313 Lines B and C, a second missense mutation in DHFR, i.e. $Tyr_{151}Asp$, appeared between 196
314 and 252 generations (Figure 7 D, Supplementary File 5). Line B also showed an $Ile_{94}Leu$
315 mutation in DHFR at low frequency at generation 252 (Figure 7 D, Supplementary File 5).
316 Unexpectedly, despite acquiring these additional mutational hits and lower DHFR expression

317 level, bacteria in LTMPR1 Lines B and C retained two copies of *folA* in contrast to Line A
318 (Figure 7 D, S7, Supplementary File 5).
319
320 Individual resistant clones from each of the lineages provided greater insight into the genetic
321 trajectories of evolution. In LTMPR1 Line A, 2 of the 6 sequenced isolates had lost gene
322 duplication and acquired the $Trp_{30}Arg$ coding and $C_{-35}T$ promoter mutations in *folA* (Figure 7
323 E, Supplementary File 5). Interestingly, the other 4 sequenced isolates from Line A retained
324 the ancestral GDA and harboured the same combination of *folA* point mutations, but only in 1
325 of the 2 gene copies (Figure 7 E, Supplementary File 5). Ten of the twelve isolates from
326 LTMPR1 Lines B and C, on the other hand, retained *folA* duplication. In all ten isolates, both
327 copies of *folA* had acquired $Trp_{30}Arg$ and $Tyr_{151}Asp$ mutations (Figure 7 E, Supplementary
328 File 5). Unlike LTMPR1, all sequenced isolates of WTMPR4 at 252 generations of evolution
329 had lost their GDA mutation and gained 2 *folA* mutations (Figure 7 F). WTMPR5 too had
330 gained an additional *folA* mutation (Figure 7 G). Notably, none of the isolates from WTMPR4
331 and WTMPR5 lineages had the $Trp_{30}Arg$ and $Tyr_{151}Asp$ combination of mutations that
332 occurred repeatedly in LTMPR1 lineages.
333
334 We next sought to understand why some LTMPR1 bacteria had retained high *folA* copy
335 number despite acquiring multiple point mutations in the drug target. To address this
336 question, we first asked what the advantage of *folA* duplication was in the two different
337 mutational trajectories that were observed for LTMPR1. Isolates from Line A had similar
338 IC50 values for trimethoprim regardless of whether they had one or two copies of *folA*
339 (Figure 8 A). On the contrary, isolates from Line B that harboured only one copy of mutant
340 *folA* (coding for DHFR-$Trp_{30}Arg$-$Tyr_{151}Asp$) displayed ~3.5-fold lower IC50 than those that
341 harboured two copies (Figure 8 A). Immunoblotting showed that isolates from Line A had
342 similar expression level of DHFR (Figure 8 B). However, DHFR protein level was
343 significantly compromised by the $Trp_{30}Arg$-$Tyr_{151}Asp$ mutation combination and double the
344 number of *folA* gene copies restored this defect (Figure 8 B). Thus, higher gene dosage of
345 DHFR was dispensable once a promoter up-regulatory mutation was fixed. On the other
346 hand, bacteria with alleles of DHFR harbouring multiple missense mutations continued to
347 benefit from overexpression due to *folA* duplication.
348
349 Mutations at position Trp_{30} in DHFR are accompanied by lower *in vivo* proteolytic stability
350 due to perturbation of the hydrophobic core of the protein [54]. This, in turn, affects the

351 expression level of DHFR [49, 54]. Interestingly, residue Tyr_{151} lies in the same hydrophobic
352 pocket as Trp_{30} (Figure 8 C). To test the effect of combination of mutations at these sites on
353 trimethoprim resistance and DHFR expression, we first heterologously expressed wild type
354 DHFR, Trp_{30}Arg or Tyr_{151}Asp single mutants and the Trp_{30}Arg-Tyr_{151}Asp double mutant in
355 wild type *E. coli*. As expected, DHFR Trp_{30}Arg expression increased the IC50 of
356 trimethoprim. Surprisingly, neither DHFR Tyr_{151}Asp, nor the Trp_{30}Arg-Tyr_{151}Asp double
357 mutant conferred resistance to the antibiotic. Indeed, both mutants conferred lower IC50
358 values than wild type DHFR. This effect was due to loss of hydrophobicity at the Tyr_{151}
359 position, as DHFR-Tyr_{151}Leu and DHFR-Tyr_{151}Phe conferred similar IC50 values as wild
360 type DHFR (Figure 8 D). Since the Trp_{30}Arg-Tyr_{151}Asp mutation combination had evolved in
361 a *lon*-deficient background, we next expressed these mutants in *E. coli Δlon.* Here too, DHFR
362 Trp_{30}Arg conferred trimethoprim resistance, while the Tyr_{151}Asp mutant did not. Strikingly,
363 the combination of Trp_{30}Arg and Tyr_{151}Asp conferred very high-level trimethoprim resistance
364 (Figure 8 D) demonstrating synergistic epistasis that was contingent on the absence of Lon
365 protease. The levels of expression of DHFR mutants in wild type and *lon*-deficient bacteria
366 mirrored trimethoprim resistance levels indicating that clearance by Lon altered their
367 genotype-phenotype relationship (Figure 8 E). Based on these data, we concluded that the
368 action of proteostatic machinery generated additional demand for expression by clearing
369 mutant-DHFR proteins. This additional demand drove the maintenance of *folA* duplications
370 in bacteria evolving at very high drug pressures.
371
372

373 **Discussion**

374

375 Ohno's hypothesis, proposed in 1970, provided a conceptual framework to think about the
376 evolution and eventual fate of duplicated genes [55]. Several theoretical and experimental
377 studies have built upon the original idea to delineate the evolutionary outcomes of gene
378 duplication [18, 56, 57]. More recently, experimental evolution of microbes has been used to
379 test some of the suggestions made by Ohno and others. This approach has the inherent
380 advantage of greater control over the selection pressure applied and available genetic tools to
381 interrogate molecular mechanisms. In most studies of this kind, an artificial system was
382 constructed which was then evolved under defined selection pressures [33, 58-61]. Our study
383 began with serendipitous isolation of spontaneous trimethoprim-resistant mutants of *E. coli*
384 that overproduced DHFR because of a *folA* gene duplication. We exploited this system to
385 understand the relationship between selection pressure and gene duplication in the context of
386 drug-resistance evolution. The selection pressure for higher gene expression, i.e. "expression
387 demand", can be understood as the amount of a gene product that is required to maintain
388 fitness in a particular environment. Antibiotics directly increase expression demand by
389 inhibiting the activity of their targets. Our study shows that increasing gene dosage is a rapid,
390 but transient evolutionary response to ensure that expression demand is met. On the other
391 hand, point mutations are relatively stable routes to mitigating expression demand. Over short
392 time scales, changes in gene dosage can counter the effects of antibiotic. However, their
393 inherent instability results in almost inevitable replacement with phenotypically equivalent
394 point mutations. Thus, continued antibiotic exposure delayed but could not prevent reversion
395 to low gene copy number in most of our experiments. Further, we discover a novel
396 association between proteostasis and gene dosage evolution. In *E. coli* evolving at high
397 trimethoprim pressure, bacterial protein quality control exacerbated expression demand by
398 clearing mutant DHFR-s. As a result, drug-resistant mutants of DHFR with low proteolytic
399 stability drove the maintenance of *folA* gene duplication (Figure 9). Based on these findings
400 we propose that a combination of factors, which include environmental and physiological,
401 alter the relationship between gene copy number and fitness, and drive gene dosage
402 evolution.

403

404 A key finding from our study was the two-fold influence that proteostatic machinery had on
405 gene copy number. Firstly, Lon protease-deficiency promoted GDA mutations, possibly by
406 allowing accumulation of IS-tranposases in *E. coli*. This observation resonates with earlier

407 work by Nicoloff and co-workers who found that large IS-element flanked genomic
408 duplications in tetracycline-resistant *E. coli* were commonly preceded by spontaneous
409 mutations that inactivated *lon* [62, 63]. Further, the role of Lon protease in regulating IS-
410 element mobility by a similar mechanism as proposed by us has been previously
411 demonstrated for IS1 and IS903, strengthening our argument [64, 65]. We note however that
412 other mechanisms by which Lon may modulate the frequency of IS-mediated GDAs in *E.*
413 *coli* cannot be ruled out. For instance, Lon protease homologs from several organisms
414 directly bind DNA [66-70], though the role of this binding in DNA transactions is unclear.
415 Similarly, Lon is a regulator of the recombination protein, RecA, by its action on LexA [71]
416 and in *Salmonella* RecA stimulated the rate of gene duplications mediated by IS3 [23]. The
417 second impact of protein quality control on GDA was the action of proteases on
418 trimethoprim-resistant mutants of DHFR, which promoted gene duplication by generating
419 additional demand for expression (Figure 9). Importantly, which of these two effects
420 dominated was determined by a combination of environment, i.e. drug concentrations, and
421 other mutations acquired at the *folA* locus during evolution. In particular, mutations in the
422 *folA* promoter could alleviate expression demand and hence reduce selection for high gene
423 copy number.
424
425 The interplay between GDA and point mutations during evolution agrees in many respects
426 with a recent study by Tomanek and Guet [61]. They use an elegantly designed artificial
427 system [33, 61] in which expression of the *galK* gene in *E. coli* was driven by a very low
428 activity promoter. This strain was then allowed to evolve in different concentrations of
429 galactose. Like the present study, they found that the activity of IS-elements biased
430 evolutionary trajectories in favor of duplications. They also showed that co-occurrence of
431 duplications and point mutations was selected only at high expression demand, in agreement
432 with our findings. Finally, they argued that gene duplication events may constrain evolution
433 by point mutations. Once again, our study corroborates these results at the short term in that
434 the *lon*-knockout which showed a higher propensity for GDA mutations did not evolve point
435 mutations in *folA* at early time points of evolution. However, given the relative low stability
436 of large GDAs, these were inevitably lost over longer duration of evolution and replaced with
437 point mutations. Thus, it is likely that the different mutational profiles of IS-biased
438 populations may be relevant only at short durations of evolution. It is important to note that
439 though we have focused on the effects of mutations at the *folA* locus, which are directly under
440 trimethoprim pressure, many of the isolates from our evolution experiments also harboured

additional mutations at other loci. For instance, mutations in *rpoS* were seen throughout our evolution lines and at least in drug-free media were significant contributors to bacterial fitness. It is not yet clear what the role of other mutations are likely to be under antibiotic selection, though it may be envisaged that these additional mutations play the role of media adaptations. We did not take their effects into in this study account since we expected them to be minor contributors to fitness especially during evolution under trimethoprim selection. However, their role in modulating the stability and selection of GDA mutations in drug-free/low drug environments may be more significant and remains to be investigated in the future.

In summary, the present study uncovers the complex relationship between gene copy number, point mutations and environment, and identifies proteostasis as a regulator of gene dosage evolution in bacteria. These findings open the door to alternative explanations for why gene duplications may be maintained over long periods of evolution and provides insight into how gene expression demand is mitigated by multiple mechanisms in bacteria.

Materials and Methods

Strains, plasmids and culture conditions

Escherichia coli K-12 MG1655 was used as wild type for all experiments. *E. coli* DH10B was used for plasmid construction and manipulation. *E. coli* strains were cultured in Luria-Bertani (LB) broth with shaking at 180-200 rpm or on LB agar plates at 37 °C. Ampicillin was used at a concentration of 100 µg/mL for selection of plasmid-transformed strains. Trimethoprim was added to growth media at required concentrations as needed. The strains and plasmids used in this study are listed in Table 1.

Table 1. List of strains and plasmids used in the study

Sr. No.	**Strain**	**Description**	**Source**
1.	*Escherichia coli* K-12 MG1655	Wild type *E. coli*; used as ancestor for evolution experiments and as reference strains for phenotypic and genotypic comparisons	Laboratory stocks
Gene knockouts generated in E. coli K-12 MG1655			
1.	*E. coli ΔmgrB*	Isogenic knockout of *mgrB* gene (*ΔmgrB::Kan*)	[46]
2.	*E. coli ΔlacZ*	Isogenic knockout of *lacZ* gene (*ΔlacZ::Cat*)	[46]
3.	*E. coli Δlon*	Isogenic knockout of *lon* gene (*Δlon::Kan*)	[49]
4.	*E. coli ΔlonΔlacZ*	Double knockout of *lon* (*Δlon::Kan*) and *lacZ* (*ΔlacZ::Cat*) genes	[46]
5.	*E. coli ΔmgrBΔlon*	Double knockout of *lon* (*Δlon::Cat*) and *mgrB* (*ΔmgrB::Kan*) genes	[46]
6.	*E. coli ΔlonΔrpoS*	Double knockout of *lon* (*Δlon::Cat*) and *rpoS* (*ΔrpoS::Kan*) genes	[46]
Strains/Isolates derived from short-term adaptive laboratory evolution			
1.	WTMPR1, WTMPR2, WTMPR3, WTMPR4 and	Trimethoprim-resistant derivatives of *E. coli* wild-type obtained by laboratory	[46] Note: These

	WTMPR5	evolution at 300 ng/ml of trimethoprim for ~25 generation, followed by selection on agar plates supplemented with 1 μg/mL of trimethoprim	strains are labelled as TMPR1-5 in the original publication. They have been renamed WTMPR1-5 to indicate that they were derived after evolution of wild type.
2.	LTMPR1, LTMPR2, LTMPR3, LTMPR4 and LTMPR5	Trimethoprim-resistant derivatives of *E. coli Δlon* obtained by laboratory evolution at 300 ng/ml of trimethoprim for ~25 generation, followed by selection on agar plates supplemented with 1 μg/mL of trimethoprim	This study
Plasmids			
1.	pPRO-*insB*	*InsB* gene cloned into the pPROEx-HtC expression plasmid in frame-with an N-terminal hexa-histidine tag.	This study
2.	pPRO-*insL*	*InsL* gene cloned into the pPROEx-HtA expression plasmid in frame-with an N-terminal hexa-histidine tag.	This study
3.	pPRO-*folA*	*FolA* gene cloned into the pPROEx-HtB expression plasmid in frame-with an N-terminal hexa-histidine tag.	[54]
4.	pPRO-*folA* $Trp_{30}Arg$	*FolA* gene cloned into the pPROEx-HtB expression plasmid, harbouring the $Trp_{30}Arg$ mutation in frame-with an N-terminal hexa-histidine tag.	[54]
5.	pPRO-*folA* $Tyr_{151}Asp$	*FolA* gene cloned into the pPROEx-HtB expression plasmid, harbouring the	This study

		Tyr_{151}Asp mutation in frame-with an N-terminal hexa-histidine tag.	
6.	pPRO-*folA* Trp_{30}*Arg* Tyr_{151}*Asp*	*FolA* gene cloned into the pPROEx-HtB expression plasmid, harbouring the Trp_{30}Arg and Tyr_{151}Asp mutations in frame-with an N-terminal hexa-histidine tag.	This study
7.	pPRO-*folA* Tyr_{151}*Phe*	*FolA* gene cloned into the pPROEx-HtB expression plasmid, harbouring the Tyr_{151}Phe mutation in frame-with an N-terminal hexa-histidine tag.	This study
8.	pPRO-*folA* Tyr_{151}*Leu*	*FolA* gene cloned into the pPROEx-HtB expression plasmid, harbouring the Tyr_{151}Leu mutation in frame-with an N-terminal hexa-histidine tag.	This study

Laboratory evolution of antibiotic resistance

Short-term evolution of antibiotic resistance

For short term laboratory evolution of antibiotic resistance, wild type and *E. coli Δlon* populations were cultured in 96-well plates in LB medium incorporating sub-inhibitory concentrations of different antibiotics. Following concentrations of antibiotics were used: trimethoprim (300 ng/mL; MIC/3); spectinomycin (5 μg/mL; MIC/10); nalidixic acid (4.25 μg/mL; MIC/3); erythromycin (10 μg/mL; MIC/10); rifampicin (4.25 μg/mL; MIC/3). Control populations were also set up in antibiotic-free LB. Five replicates were included for each strain in each antibiotic. Total volume of culture in each well was 150 μL. 10 % of the population was passaged to fresh growth media every 24 hours, for 7 days, resulting in ~25 generations of evolution. At the end of 7 days, the evolved populations were frozen as glycerol stocks.

To check for the evolution of resistance and isolation of resistant clones for further analyses, frozen populations were revived in drug-free LB broth and grown overnight (14-16 hours). Saturated cultures were serially diluted and spotted on LB agar plates supplemented with the appropriate antibiotic at MIC. Colonies were randomly picked, cultured overnight in drug-free LB and frozen in 50% glycerol at -80 °C until further use.

490

491 *Long-term evolution at different trimethoprim pressures*

492 For long-term evolution, trimethoprim-resistant isolates LTMPR1, WTMPR4 and WTMPR5
493 were first revived from frozen stocks in drug-free LB overnight. Triplicate evolving lineages
494 were then established by sub-culturing 1% of the saturated culture of each ancestor into 2 mL
495 of appropriate medium. For relaxed selection drug-free LB was used. For constant antibiotic
496 pressure, trimethoprim was added at a concentration of 300 ng/mL. For increasing antibiotic
497 pressure, trimethoprim was increased from 1 μg/mL to 10 μg/mL over course of the
498 experiment as shown in Figure 4 A. Cultures were grown for 10-14 hours (~7 generations)
499 and then passaged (1%) into fresh media. This was continued for 36 passages, i.e. ~252
500 generations. Aliquots of cultures were frozen periodically for further analyses.

501 For isolating individual clones, populations were revived from frozen stocks at 252
502 generations of evolution in drug-free LB, serially diluted and then spotted onto LB agar
503 plates supplemented with trimethoprim (1, 5 or 10 μg/mL). Plates were incubated for 18-20
504 hours at 37 °C and resistant colonies were picked randomly for further analyses.

505

506 ***Measuring trimethoprim resistance***

507

508 *Trimethoprim IC50*

509 Trimethoprim Inhibitory Concentration - 50 (IC50), i.e. the antibiotic concentration needed
510 for 50 % growth inhibition, was calculated using a broth dilution assay. Frozen glycerol
511 stocks of appropriate bacterial strains were revived in 1-3 mL LB for 6-8 hours. In a 96-well
512 plate, 150 μL of LB was added to individual wells. Trimethoprim stocks were added, and the
513 antibiotic was serially diluted. No antibiotic was added to a few wells, which served as
514 controls. Next, 1.5 μL of the appropriate bacterial culture was added to each well. The
515 peripheral wells of the plate were filled with water to prevent dehydration. The 96-well plate
516 was then incubated at 37°C for 15-18 hrs. Optical density at 600 nm was measured for each
517 well using a plate reader (Ensight, Perkin Elmer). Obtained data were first normalised to
518 growth in drug-free wells (set to 1) and then fitted to a sigmoidal log(inhibitor)-response
519 curve using Graphpad Prism software. The equation for this model is as follows:

520 $Y= Min + (Max-Min)/(1+10^{(X-Log(IC50))})$

521 where,

522 Min: Minimum value of normalised OD and Max: maximum value of normalised OD

523 Values of IC50 were estimated from the above curve fit and averaged across 3 independent
524 measurements for every strain.
525
526 *Colony forming efficiency*
527 To calculate the colony forming efficiency of evolving populations at different trimethoprim
528 concentrations, frozen glycerol stocks were first revived in 1 mL LB for 14-15 hours.
529 Saturated cultures were serially diluted (10-fold). Diluted cultures were then spotted (10 mL)
530 on LB agar plates containing 0, 1, 5 and 10 μg/mL of trimethoprim. Spots were allowed to
531 dry completely and the plates were incubated at 37 °C for ~18 hours. Colonies were counted
532 from the least dense spot and multiplied by the appropriate dilution factor to determine
533 colony forming units (CFUs)/mL.
534
535 *Growth curves*
536 Appropriate frozen stocks were revived overnight in drug-free LB (1-3 mL) and 1μl of the
537 saturated culture was added to 200 μL of media supplemented with trimethoprim (300 ng/ml)
538 in the wells of a sterile uncoated 96-well plate. Optical density at 600 nm was measured
539 during growth at 37 °C at 15-minute intervals over a 24-hour period on a Perkin-Elmer
540 Ensight Multimode Plate Reader. Doubling time (DT) (in mins) was estimated from log
541 phase, using the following formula:

$$DT = \ln(2) / (\frac{\ln\left(\frac{OD2}{OD1}\right)}{T2 - T1})$$

542
543 Where OD1 and OD2 represent optical densities at the onset and end of log phase
544 respectively, and T1 and T2 represent time elapsed (in mins), corresponding to OD1 and
545 OD2.
546
547 **Genome sequencing, identification of mutations and detection of GDA mutations**
548
549 Genomic DNA was extracted from late log-phase cultures of appropriate strain/resistant
550 isolates/evolving populations using phenol:chloroform:isoamyl alcohol extraction as
551 previously described [72]. Extracted genomic DNA was then cleaned-up using HiGenoMB
552 DNA isolation and purification kit (Himedia, India) as per manufacturer's instructions and

quantitated spectrophotometrically. The integrity was checked by running purified DNA on an agarose gel before further analyses.

For whole-genome resequencing, paired-end whole-genome next-generation sequencing (NGS) was performed on a MiSeq system (Illumina, USA) with read lengths of 150-250 bps. Library preparation and sequencing services were provided by Eurofins (Bangalore, India) and Genotypic (Bangalore, India). Processed reads were aligned to the reference genome *E. coli* K-12 MG1655 (U00096) using bowtie2. Variant calling and prediction of new junctions in the genome to identify large structural mutations was done using Breseq [73] using default settings in population or clonal modes as needed. All variants and new junctions that were already present in the ancestor were excluded. Variants unique to the test sample were classified as "mutations".

For detection of GDA mutations, 'coverage' vs 'coordinate in reference genome' plot generated by the bam.cov command in Breseq was used [73]. Contiguous stretches greater than 10 kb with coverage depth of 1.5x or more, compared to the rest of the genome, were classified as GDA mutations. The presence of IS-elements or repeat sequences was manually checked using the coordinates of the GDA.

Calculating relative fitness of E. coli strains

Relative fitness of gene knockouts or evolved clones was estimated using competition assays. Saturated cultures of test strains and reference strains (*E. coli ΔlacZ* or *E. coli ΔlacZΔlon*) were mixed 1:1 to a total of 1% in 3 mL LB. Strains were allowed to compete for ~7 generations (24 hours incubation). Mixed cultures were diluted and plated on LB agar supplemented with IPTG (1 mM) and X-gal (50 µg/mL) at the start and end of the competition experiment. Plates were incubated for 18-24 hrs, and blue and white colonies were counted to determine the CFUs/mL at initial and final time points.

Relative fitness (w) was calculated using the following formula:

$$w = \ln\left(\frac{Tf}{Ti}\right) \Big/ \ln\left(\frac{Rf}{Ri}\right)$$

where T_f and T_i are the final and initial CFUs/mL of the test strain and R_f and R_i are the final and initial CFUs/mL of the reference strain, respectively.

583 ***Estimation of DHFR expression by immunoblotting***

584

585 Immunoblotting was used to determine the expression level of DHFR in various
586 strains/mutants of *E. coli* as described in Patel and Matange (2021) [46]. Briefly, equal lysate
587 protein (5 μg) was subjected to SDS-PAGE on a 15% gel and electroblotted onto PVDF
588 membranes. The membrane was blocked with 5% BSA solution and cut into 2 halves at the
589 25 kDa molecular weight band. The lower part of the membrane was probed with anti-DHFR
590 polyclonal IgG (100 ng/mL) and the upper part was probed with anti-FtsZ serum (1:50,000).
591 HRP-linked anti-rabbit IgG (1:10,000) was used as the secondary antibody. Hybridised
592 antibodies were detected using chemiluminescence. Band intensities were quantitated in
593 ImageJ and normalised to appropriate control (set to 1). Fold change in DHFR was calculated
594 as follows:

595 $$Fold\ change_{DHFR} = {}^{Normalised\ intensity_{DHFR}}/_{Normalised\ intensity_{FtsZ}}$$

596

597 ***Gene expression analysis using RNA-sequencing***

598

599 Transcriptome analyses were performed using RNA sequencing. For RNA-seq, saturated
600 overnight culture of appropriate strains/isolates was inoculated (1 %) into 3 mL LB in
601 triplicate and incubated at 37 ºC for 3 hours with shaking at 180 rpm. The cultures were
602 centrifuged and bacterial pellets were resuspended in 1 mL of RNAlater (Invitrogen) for
603 storage. RNA extraction, sequencing and preliminary data analyses were performed by
604 Eurofins (Bangalore, India). Total RNA was extracted using RNAeasy spin columns and
605 quantitated using Qubit and Bioanalyzer. Bacterial rRNA was depleted using Ribozero kit.
606 Bacterial mRNA was then reverse transcribed, library prepared and quality control performed
607 using Tapestation platform. Paired end sequencing was performed on Illumina platform with
608 read lengths of 150 bp read length. Processed reads were aligned to the reference genome
609 (U00096.3) using HISTAT2(version 2.1.0). Abundance estimation was done using
610 featureCounts(version1.34.0). Differential gene expression (DGE) analysis was done by
611 comparing the expression of individual genes in test and reference samples. The output of the
612 DGE analyses were $\log_2$(fold change) values for each gene and P-values for statistical
613 significance.

614

615 ***Cloning and site-directed mutagenesis***

616

617 *Cloning of InsB and InsL transposases*

618 *InsB* was amplified from *E. coli* K-12 MG1655 genomic DNA by PCR using primers
619 insB1_fwd_*Nco*I (5'-GGCGCCATGGATATGCCGGGCAACAGCCCG-3')
620 insB1_rev_*Eco*RI (5'- CCTTTGAATTCTTATTGATAGTGTTTTATG-3'). For amplification
621 of *insL*, primers insL_*Eco*RI_fwd (5'- ATGCCGAATTCATGAATTACTCTCACGATAAC-
622 3') and insL_*Hin*dIII_rev (5'- ATGCCAAGCTTTTAGTTCTTCTTTTCGGATCC-3') were
623 used. Amplified *insB* gene was digested with *Nco*I and *Eco*RI restriction endonuclease and
624 ligated with similarly digested pPROEx-HtC plasmid. Amplified *insL* was digested with
625 *Eco*RI and *Hin*dIII restriction endonucleases and ligated with similarly digested pPROEx-
626 HtA plasmid. Positive clones were screening using restriction digestion. Clones were
627 transformed into *E. coli* wild type or *Δlon* for further analyses.

628

629 *Site-directed mutagenesis for DHFR*

630 Site-directed mutagenesis was performed to introduce mutations into the coding sequence of
631 *folA* in pPRO-folA as described in [54]. Briefly, PCR was performed using pPRO-folA or
632 pPRO-folA $Trp_{30}Arg$ as template and the appropriate forward and reverse mutagenic primers
633 given below:

634 folA_$Tyr_{151}Asp$_fwd_*Bgl*II:
635 5'-CTCTCACAGCGATTGCTTTGAGATCTTGGAGCGGCGGGGCC-3'
636 folA_$Tyr_{151}Asp$_rev_*Bgl*II:
637 5'-CGCCGCTCCAAGATCTCAAAGCAATCGCTGTGAGAGTTCTGCG-3'
638 folA_$Tyr_{151}Leu$_fwd_*Bgl*II:
639 5'-TCTCACAGCTTATGCTTTGAGATCTTGGAGCGGCGGGGCC-3'
640 folA_$Tyr_{151}Leu$_rev_*Bgl*II:
641 5'- CGCCGCTCCAAGATCTCAAAGCATAAGCTGTGAGAGTTCTGCG-3'
642 folA_$Tyr_{151}Phe$_fwd_*Bgl*II:
643 5'- CTCACAGCTTTTGCTTTGAGATCTTGGAGCGGCGGGGCC-3'
644 folA_$Tyr_{151}Phe$_rev_*Bgl*II:
645 5'- GCCGCTCCAAGATCTCAAAGCAAAAGCTGTGAGAGTTCTGCG-3'

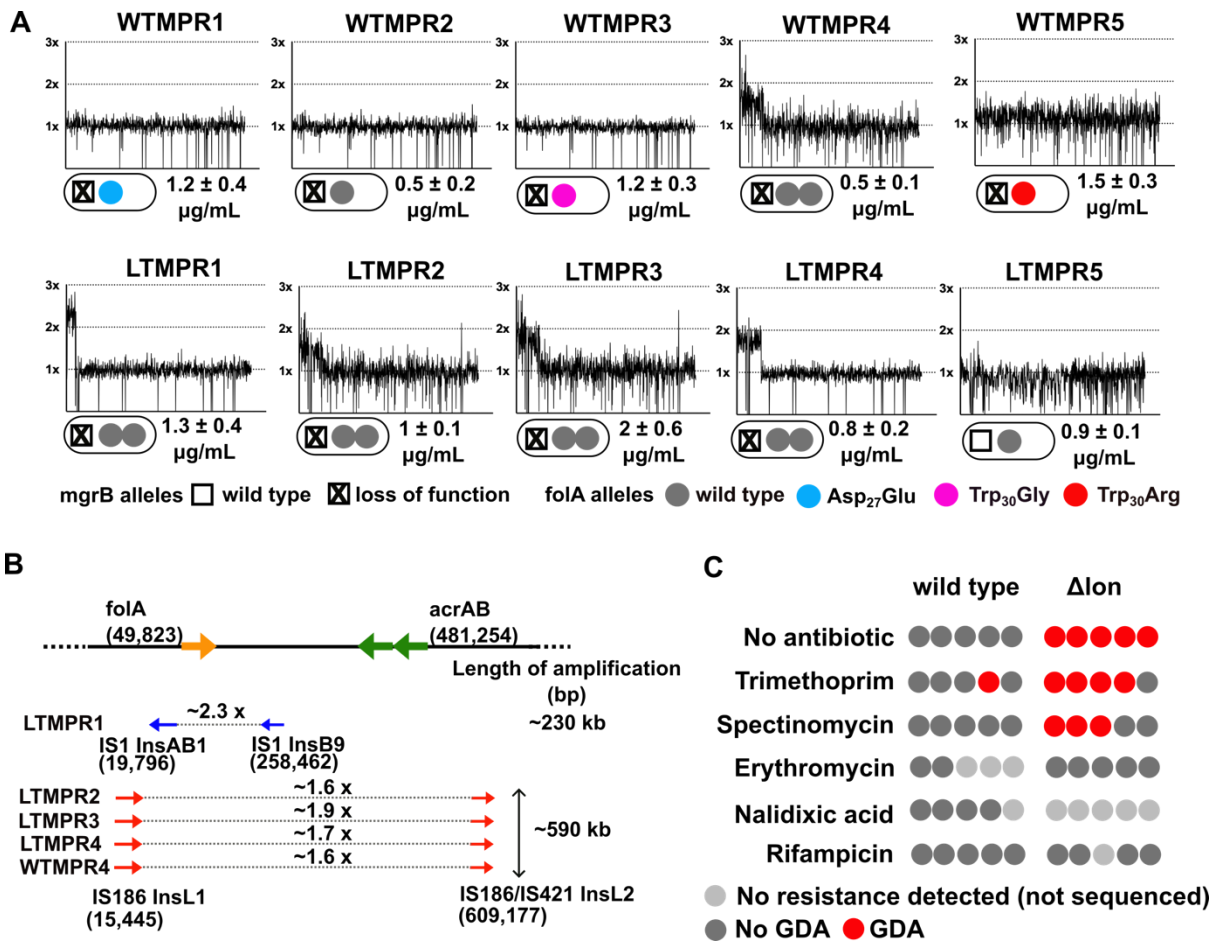
646 PCR products were digested with *Dpn*I overnight, transformed into chemically competent *E.*
647 *coli* DH10B and selected on LB agar plates containing ampicillin. Plasmids habouring the

648 desired mutations were screened by restriction digestion and confirmed by sequencing.
649 Sequenced clones were transformed into *E. coli* wild type or *Δlon* for further analyses.
650
651 **In-cell stability of InsB and InsL**
652
653 A chloramphenicol-chase assay was used to assess the in-cell stability of heterologously
654 expressed InsB and InsL transposases as described in Matange (2020) [49] with a few
655 modifications. Overnight cultures of *E. coli* wild type or *Δlon* harbouring plasmids pPRO-
656 insB or pPRO-insL were diluted 1% in 5 mL LB supplemented with ampicillin and IPTG
657 (0.1mM). After 4 hours of growth at 37 °C, 1 mL of culture was pelleted. The cell pellet was
658 lysed in 100 µL of 1x Laemmli sample loading buffer by heating at 95°C for 10-20 mins.
659 Chloramphenicol was added to the remaining culture at a final concentration of 25 µg/mL to
660 stop protein synthesis and the culture was placed back in the incubator. After 15, 45, and 90
661 minutes 1 mL aliquots of the culture were pelleted and lysed. Lysates were subjected to SDS-
662 PAGE followed by immunoblotting to detect the levels of InsB and InsL. Anti-Hexahis
663 monoclonal antibody (Invitrogen) was used as the primary antibody for InsL. Anti-Hexahis
664 polyclonal antibody (Invitrogen) was used as the primary antibody for InsB. Band intensities
665 were calculated in ImageJ and normalised to control (set to 1) to calculate fraction of protein
666 remaining after chloramphenicol treatment.
667
668 **Acknowledgments**
669 We acknowledge Dr. Manjula Reddy for providing us with the anti-FtsZ polyclonal antibody.
670 We acknowledge Bruhad Dave for his valuable inputs during discussions at early stages of
671 this project and Prof. Satyajit Rath, Prof. Sandhya Visweswariah and Dr. Girish Deshpande
672 for their comments on the manuscript. Funding for this work was provided by
673 DBT/Wellcome Trust India Alliance. NM is a recipient of the India Alliance Intermediate
674 Fellowship. CJ is a recipient of a Senior Research Fellowship from the CSIR, Govt. of India.
675
676 **Supplementary Files.**
677
678 **Supplementary File 1.** Gene expression changes in LTMPR1 and WTMPR4 isolates
679 **Supplementary File 2.** Genomic changes associated with evolution of LTMPR1 in drug-free
680 media

681 **Supplementary File 3.** Gene expression changes in isolates derived from LTMPR1 evolution
682 in drug-free media containing mutation in *rpoS* or GDA-2
683 **Supplementary File 4.** Genomic changes associated with evolution of LTMPR1 in
684 trimethoprim (300 ng/mL)
685 **Supplementary File 5.** Genomic changes associated with evolution of LTMPR1 in
686 increasing trimethoprim pressure
687 **Supplementary File 6.** Genome sequencing of isolates derived from WTMPR4 evolution at
688 different trimethoprim pressures
689 **Supplementary File 7.** Genome sequencing of isolates derived from WTMPR5 evolution at
690 different trimethoprim pressures
691
692

693 **Figures**

694



695

696

697 **Figure 1. Adaptation to antibiotics by large genomic duplications is favoured in *lon*-**
698 **deficient *E. coli*. A.** Coverage depth plots and mutations at the *folA* and *mgrB* loci of 5
699 independently evolved trimethoprim-resistant isolates derived from wild type (WTMPR1-5)
700 or *Δlon E. coli* (LTMPR1-5). Resistant clones were isolated after ~25 generations of
701 evolution at 300 ng/mL of trimethoprim. Coverage depth plots show the number of reads
702 from Illumina short-read sequencing (y-axis) mapped to each genomic coordinate of the
703 reference genome (x-axis). Coverage of 1x, 2x and 3x are marked by dotted lines for
704 reference. Mutations in *mgrB* and *folA,* and trimethoprim IC50 values (mean ± S.D. from 3
705 independent measurements) are provided for each isolate below the appropriate coverage
706 depth plot. **B.** Cartoon (not to scale) showing the duplicated genomic stretch and flanking IS-
707 elements in trimethoprim-resistant isolates. Blue arrows represent IS1, while red arrows
708 represent IS186. The positions of *folA* and *acrAB* genes, implicated in trimethoprim
709 resistance in earlier studies, relative to the duplicated stretch are shown. Fold increase in the

number of reads corresponding to the *folA* gene are indicated for each isolate. **C.** Summary of the prevalence of GDA mutations detected in wild type or *E. coli Δlon* after ~25 generations of evolution in control media (no antibiotic) or in media supplemented with sub-MIC antibiotics. Each circle represents one of the 5 replicates in the evolution experiment. Light grey circles represent replicates in which no resistance was detected and hence were not sequenced. Dark grey circles indicate that the whole population or a resistant isolate was sequenced but no GDA mutation was detected. Red circles indicate that a GDA mutation was identified.

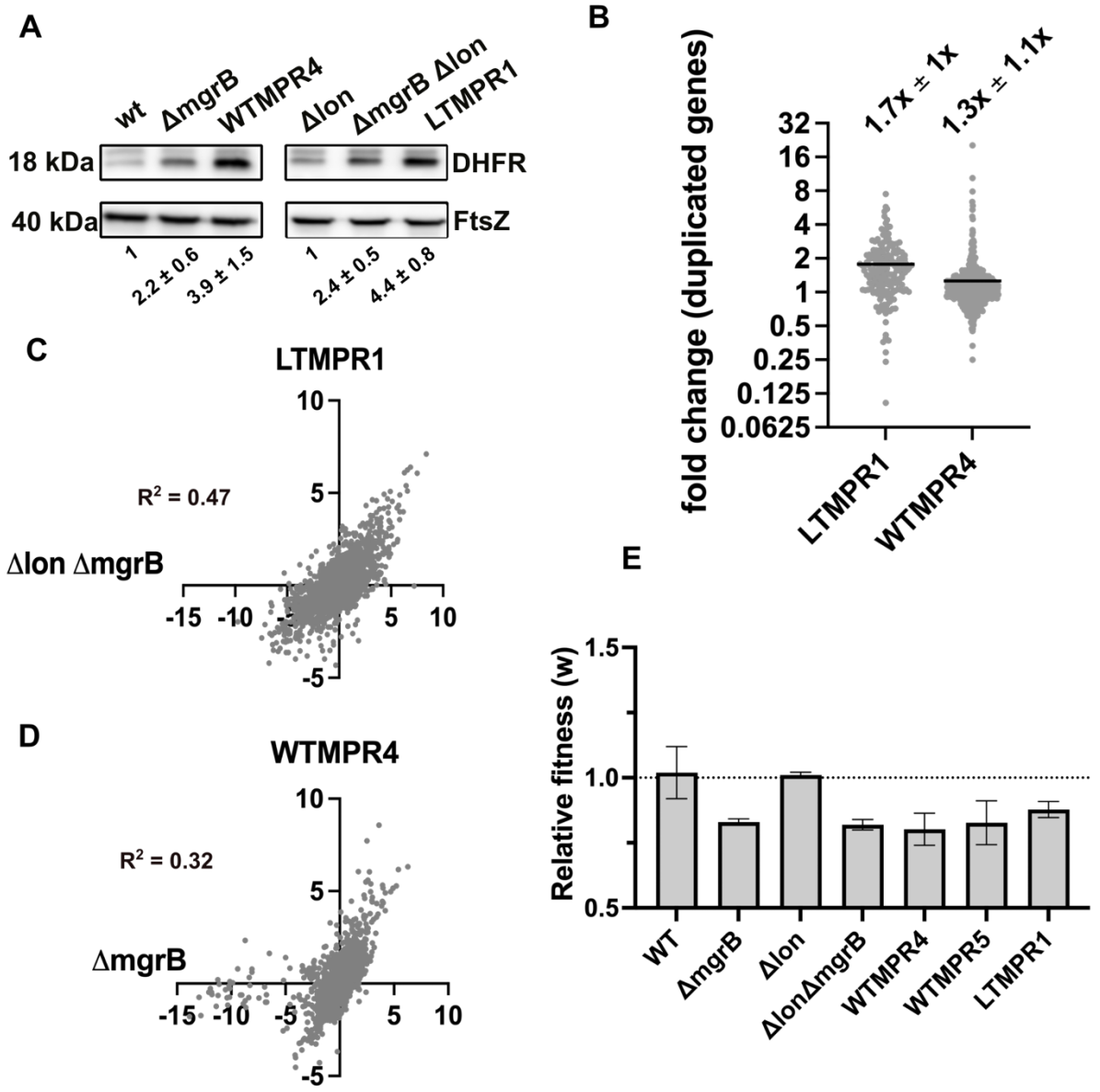


Figure 2. Overexpression of DHFR in trimethoprim-resistant isolates with genomic duplication. A. DHFR protein expression in trimethoprim-resistant isolates WTMPR4 and LTMPR1, compared with respective ancestors and *mgrB*-knockout strains. DHFR protein

723 was detected by immunoblotting using an anti-DHFR polyclonal antibody. FtsZ was used as
724 loading control. A representative immunoblot from 3 biological replicates is shown.
725 Quantitation was performed by calculating band intensities using image analysis. Fold change
726 values of DHFR levels over controls (wild type or *E. coli Δlon*, set to 1) is shown as mean ±
727 S.D. from triplicate measurements. **B.** Expression level of duplicated genes in LTMPR1 and
728 WTMPR4, determined by RNA-sequencing, expressed as fold over *E. coli ΔlonΔmgrB* and
729 *E. coli ΔmgrB* respectively. Each point represents fold change value for a single gene. Mean
730 of the scatter is shown as a black line, and its value is provided along with standard deviation
731 above the plot. **C, D.** Correlation between expression levels of genes in LTMPR1 (C) and
732 WTMPR4 (D) with *E. coli ΔlonΔmgrB* and *E. coli ΔmgrB* respectively. Gene expression
733 levels were estimated using RNA-sequencing and fold changes were calculated using wild
734 type *E. coli* as reference for WTMPR4 and *E. coli ΔmgrB*, and *E. coli Δlon* as reference for
735 LTMPR1 and *E. coli ΔlonΔmgrB*. Each point represents $\log_2$(fold change) value for a single
736 gene. The resulting scatter was fit to a simple linear regression and the obtained R^2 values are
737 provided. **E.** Relative fitness (w) of indicated *E. coli* strains calculated using competition
738 experiments with *E. coli ΔlacZ* as reference. For strains derived from *E. coli Δlon, E. coli*
739 *ΔlonΔlacZ* was used as reference. Neutrality of *ΔlacZ* was established by competition with
740 the unmarked ancestral strains, i.e. wild type (WT) and *E. coli Δlon*. No change in relative
741 fitness compared to the reference strain (w = 1) is indicated by a dotted line. Mean ± S.D.
742 from three independent measurements is plotted.

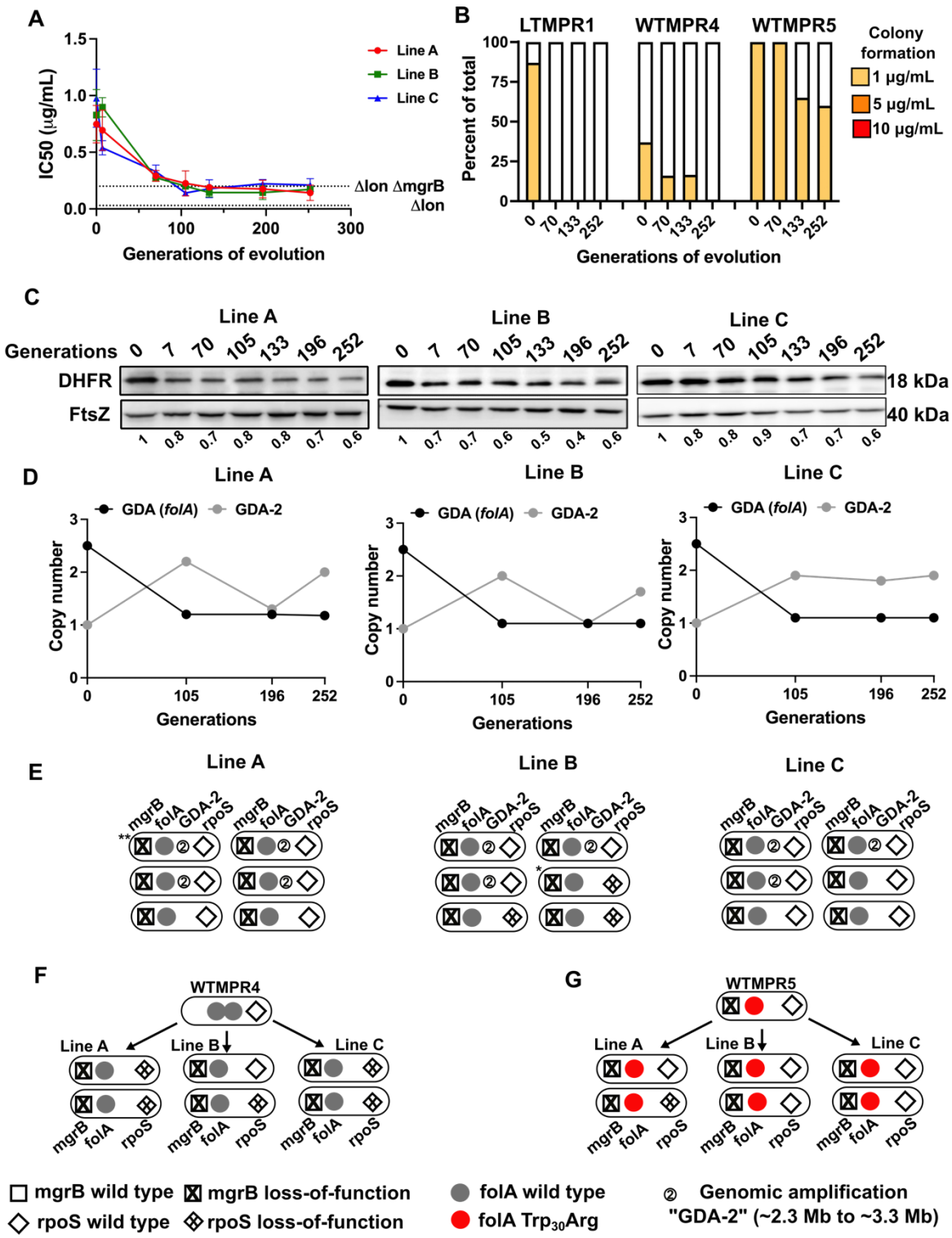


Figure 3. Evolutionary fate of *folA*-duplication in the absence of drug pressure. A. Trimethoprim IC50 values of 3 evolving lineages derived from LTMPR1 (A, B, C) over 252 generations of evolution in antibiotic-free medium. Mean ± S.D. from three measurements is plotted at each time point. IC50 values of *E. coli Δlon* and *ΔlonΔmgrB* are shown as dotted lines for reference. **B.** Colony formation of LTMPR1, WTMPR4 and WTMPR5 evolved in antibiotic-free medium at different generations. Fraction of the population capable of forming

750 colonies at 1, 5 and 10 μg/mL trimethoprim was calculated across 3 replicate lines. Mean
751 value from the three lines is plotted. The results of similar experiments performed on
752 ancestors (0 generations) are also provided for reference. **C.** DHFR expression during
753 evolution in the absence of trimethoprim in LTMPR1 lineages A, B and C, measured by
754 immunoblotting using anti-DHFR polyclonal antibody. FtsZ was used as a loading control.
755 Quantitation was performed by calculating band intensities using image analysis. DHFR
756 expression at each time point was normalized to the ancestor (i.e. 0 generations, set to 1).
757 Mean of three independent measurements is shown below each lane. **D.** Copy number of the
758 ancestral GDA encompassing *folA* (GDA(*folA*)) and GDA-2 at different time points of
759 evolution in the absence of trimethoprim. For GDA(*folA*), copy number was determined by
760 dividing number of reads from an Illumina sequencing experiment corresponding to *folA* by
761 the average number of reads mapping to the rest of the genome. **E.** Point mutations in *folA*,
762 *rpoS* and *mgrB*, and the "GDA-2" genomic duplication in 6 randomly picked colonies from
763 each of the LTMPR1 lines at 252 generations of evolution are shown schematically using the
764 appropriate symbols. Asterix (*/**) marks the genotypes that were carried forward for further
765 analyses. **F, G.** Point mutations in *folA*, *mgrB* and *rpoS* in isolates derived from 252
766 generations of evolution of WTMPR4 (F) and WTMPR5 (G) in no antibiotic. From each of
767 the evolving lines 2 random isolates were picked for genome sequencing. Various point
768 mutants at the 3 gene loci of interest are represented by appropriate symbols as shown.

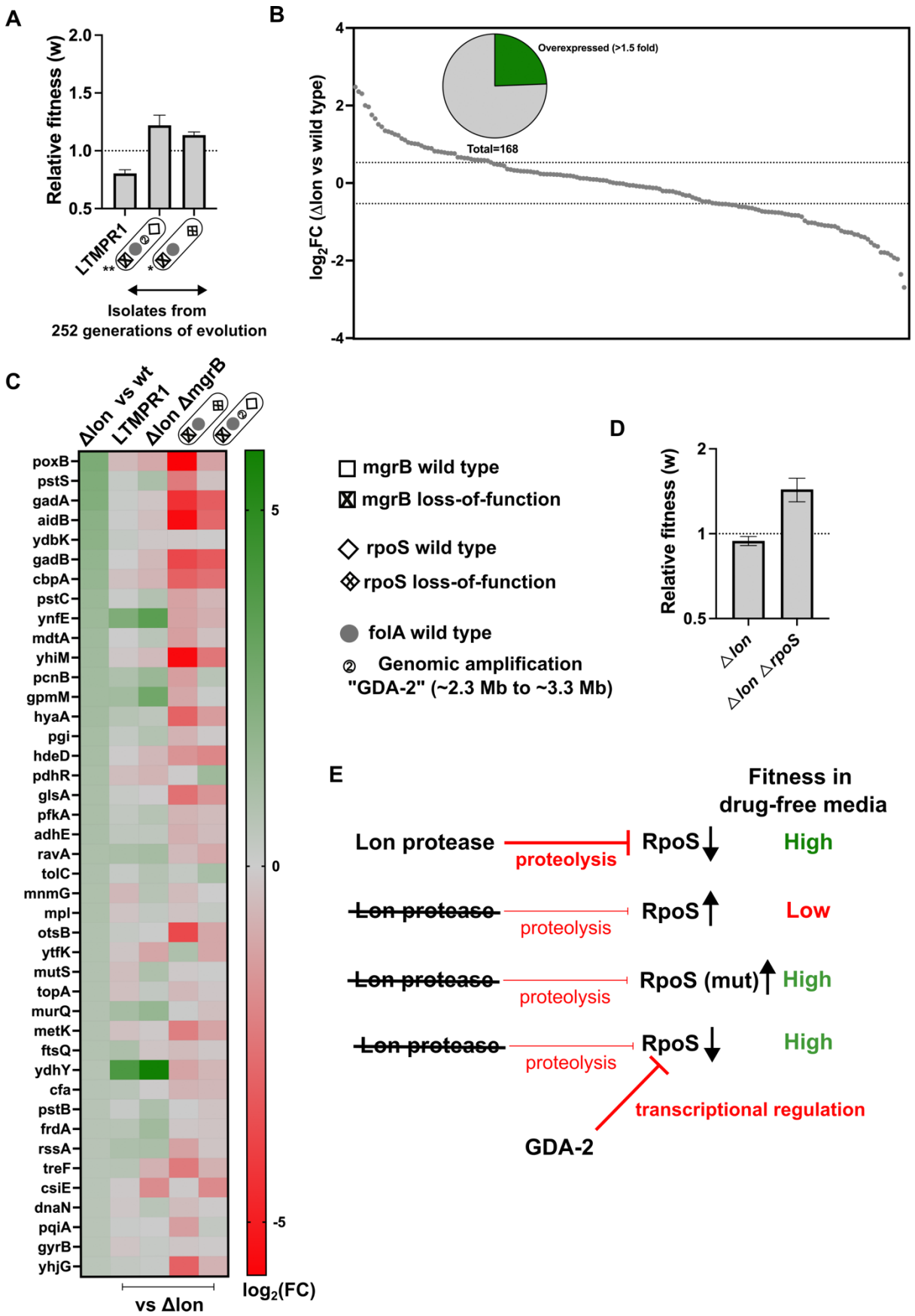


Figure 4. Fitness enhancement of LTMPR1 in drug-free media results from lower RpoS activity. A. Relative fitness (w) of isolates with the indicated genotypes derived from

772 LTMPR1 evolution in antibiotic-free media (see also Figure 3) calculated using a competitive
773 growth assay using *E. coli ΔlonΔlacZ* as the reference strain. Mean ± S.D. from three
774 independent measurements is plotted. No change in relative fitness compared to the reference
775 (w = 1) is shown as a dotted line for reference. **B.** Expression level of 168 known targets of
776 RpoS in *E. coli Δlon* compared to wild type expressed as $\log_2$(fold change) values determined
777 by RNA sequencing. Dotted lines at 1.5-fold higher and lower than wild type represent cut-
778 offs used to identify overexpressed and under-expressed genes. The pie chart shows the
779 fraction of RpoS targets that were overexpressed in *E. coli Δlon* by at least 1.5-fold in green.
780 **C.** Expression levels of 41 RpoS targets in *E. coli Δlon* compared to wild type are shown as a
781 heat map in the first vertical. The expression levels of these genes in LTMPR1, *E. coli*
782 *ΔlonΔmgrB* or LTMPR1-derived isolates compared to *E. coli Δlon* are provided in the
783 subsequent verticals. **D.** Relative fitness of *E. coli ΔlonΔrpoS* in antibiotic-free media
784 calculated using a competitive growth assay using *E. coli ΔlonΔlacZ* as the reference strain.
785 Neutrality of the *ΔlacZ* genetic marker was verified by competition between *E. coli Δlon* and
786 *E. coli ΔlonΔlacZ*. Mean ± S.D. from three independent measurements is plotted. No change
787 in relative fitness compared to the reference (w = 1) is shown as a dotted line for reference. **E.**
788 Model for the effects of Lon deficiency on RpoS levels and bacterial fitness. The roles of
789 mutations within RpoS or the GDA-2 mutations in compensating for the defects of RpoS
790 overproduction are shown.

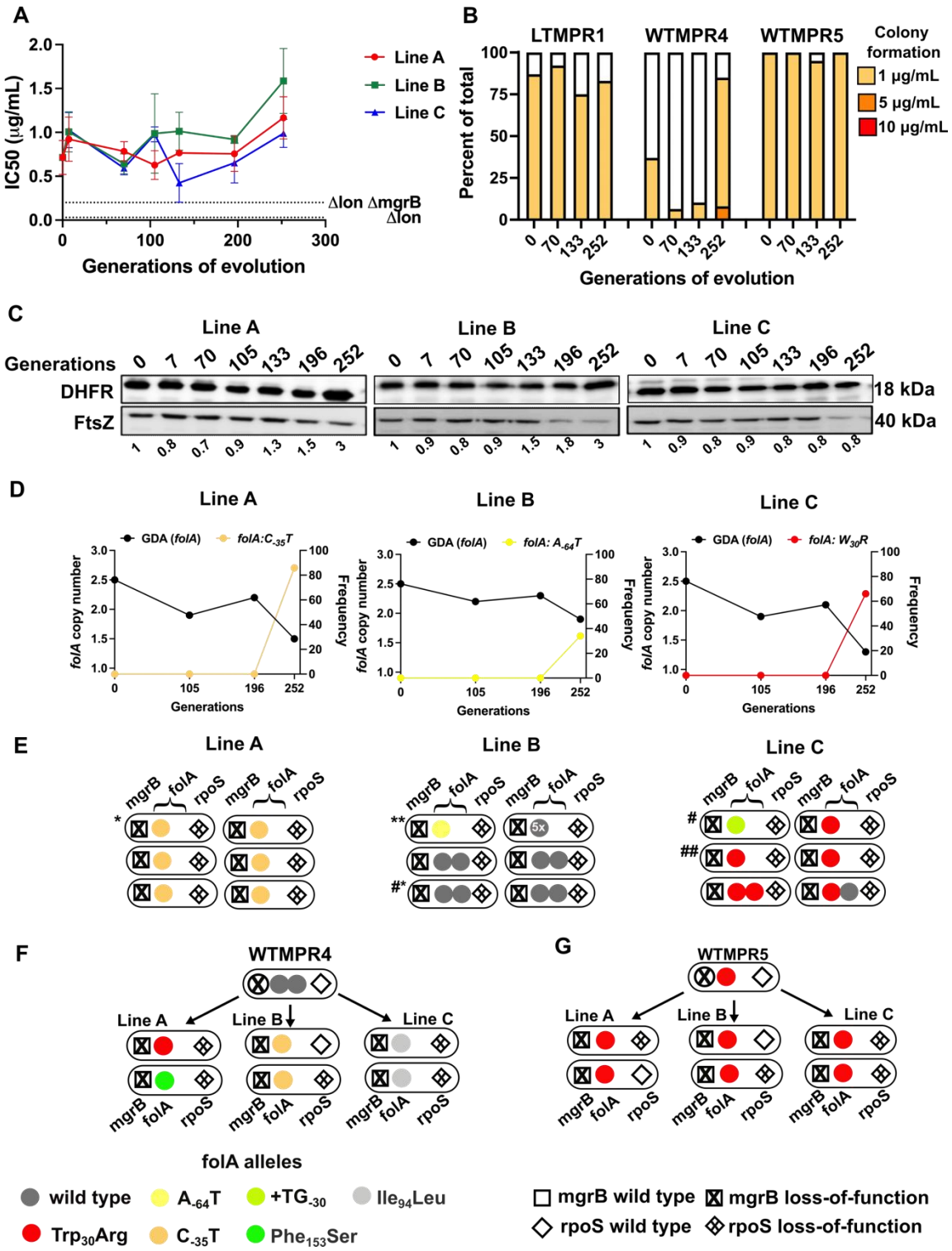


Figure 5. Evolutionary fate of *folA*-encompassing GDA in LTMPR1 at sustained trimethoprim pressure. A. Trimethoprim IC50 values of 3 evolving lineages starting from LTMPR1 (A, B, C) over 252 generations of evolution at 300 ng/mL trimethoprim. Mean ± S.D. from three measurements is plotted at each time point. IC50 values of *E. coli Δlon* and *ΔlonΔmgrB* are shown as dotted lines for reference. **B.** Colony formation of LTMPR1, WTMPR4 and WTMPR5 evolved in 300 ng/mL trimethoprim at indicated time points.

Fraction of the population capable of forming colonies at 1, 5 and 10 μg/mL trimethoprim was calculated across 3 replicate lines. Mean value from the three lines is plotted. The results of similar experiments performed on ancestors (0 generations) are also provided for reference. **C.** DHFR expression in LTMPR1 lineages A, B and C evolved in 300 ng/mL trimethoprim, measured by immunoblotting using anti-DHFR polyclonal antibody. FtsZ was used as a loading control. Quantitation was performed by calculating band intensities using image analysis. DHFR expression at each time point was normalized to the ancestor (i.e. 0 generations, set to 1). Mean of three independent measurements is shown below each lane. **D.** Copy number of *folA* (GDA(*folA*)) at different time points of evolution of LTMPR1 in trimethoprim is plotted on the left y-axis. Copy number was determined by dividing number of reads from an Illumina sequencing experiment corresponding to *folA* by the average number of reads mapping to the rest of the genome. Frequency of various *folA* alleles in the evolving populations at each of the time points is also plotted on the right y-axis. **E.** Point mutations in *folA*, *rpoS* and *mgrB* in 6 randomly picked trimethoprim-resistant colonies from each of the LTMPR1 lines at 252 generations of evolution are shown diagrammatically. Genotypes that were carried forward for further analysis are marked (* or #). **F, G.** Point mutations in *folA*, *mgrB* and *rpoS* in isolates derived from 252 generations of evolution of WTMPR4 (F) and WTMPR5 (G) in 300 ng/mL of trimethoprim. From each of the evolving lineages 2 random isolates were picked for genome sequencing. Various point mutants at the 3 gene loci are represented by appropriate symbols as shown.

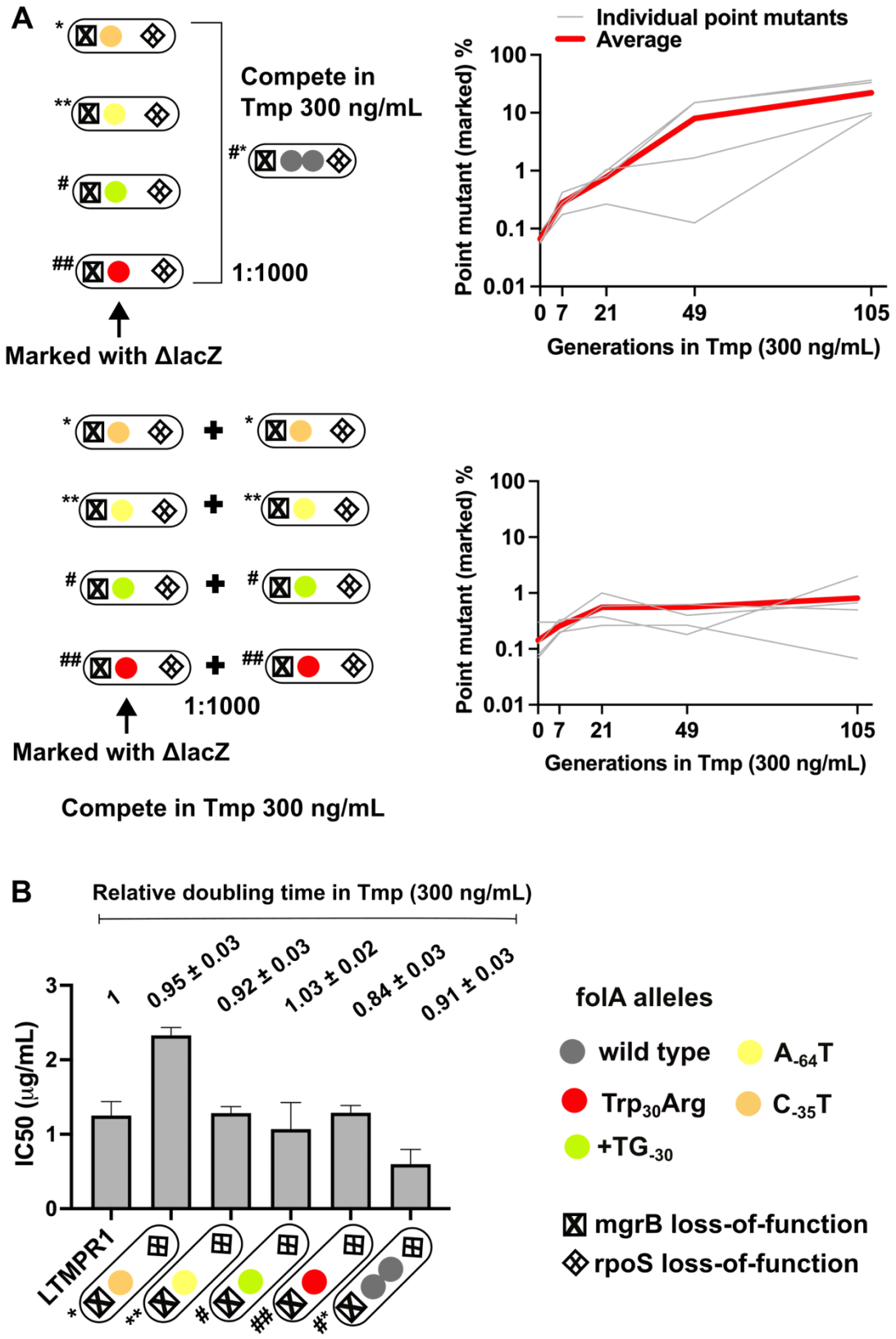


Figure 6. Invasion of GDA-dominant populations by point mutations under trimethoprim pressure. A. Upper panel. Competition between isolates with point mutations in *folA* and an isolate that harboured a *folA* duplication derived from the same time point of LTMPR1 evolution (upper panel). The genotypes of the isolates used are shown

824 diagrammatically (see also Figure 5, Supplementary File 4). The initial mixing ratio was
825 1000:1 in favour of the GDA mutant. Point mutants were marked genetically with
826 *ΔlacZ::Cat*. Neutrality of the *ΔlacZ::Cat* marker was verified by competing marked and
827 unmarked point mutants (lower panel). Percentage of the population constituted by the
828 marked point mutants over ~105 generations of serial transfer in media supplemented with
829 trimethoprim (300 ng/mL) is plotted. Traces for individual mutants are shown in grey.
830 Average of the 4 point mutants is shown in red. **B.** Trimethoprim IC50 values of the LTMPR1
831 ancestor or evolved isolates from 252 generations in trimethoprim with indicated genotypes
832 (see also Figure 5, Supplementary File 4) are plotted as bars. Mean ± S.D. from three
833 independent measurements are plotted. Doubling times relative to LTMPR1 (set to 1) in the
834 presence of 300 ng/mL of trimethoprim are provided above the graph as mean ± S.D. from
835 three independent measurements.

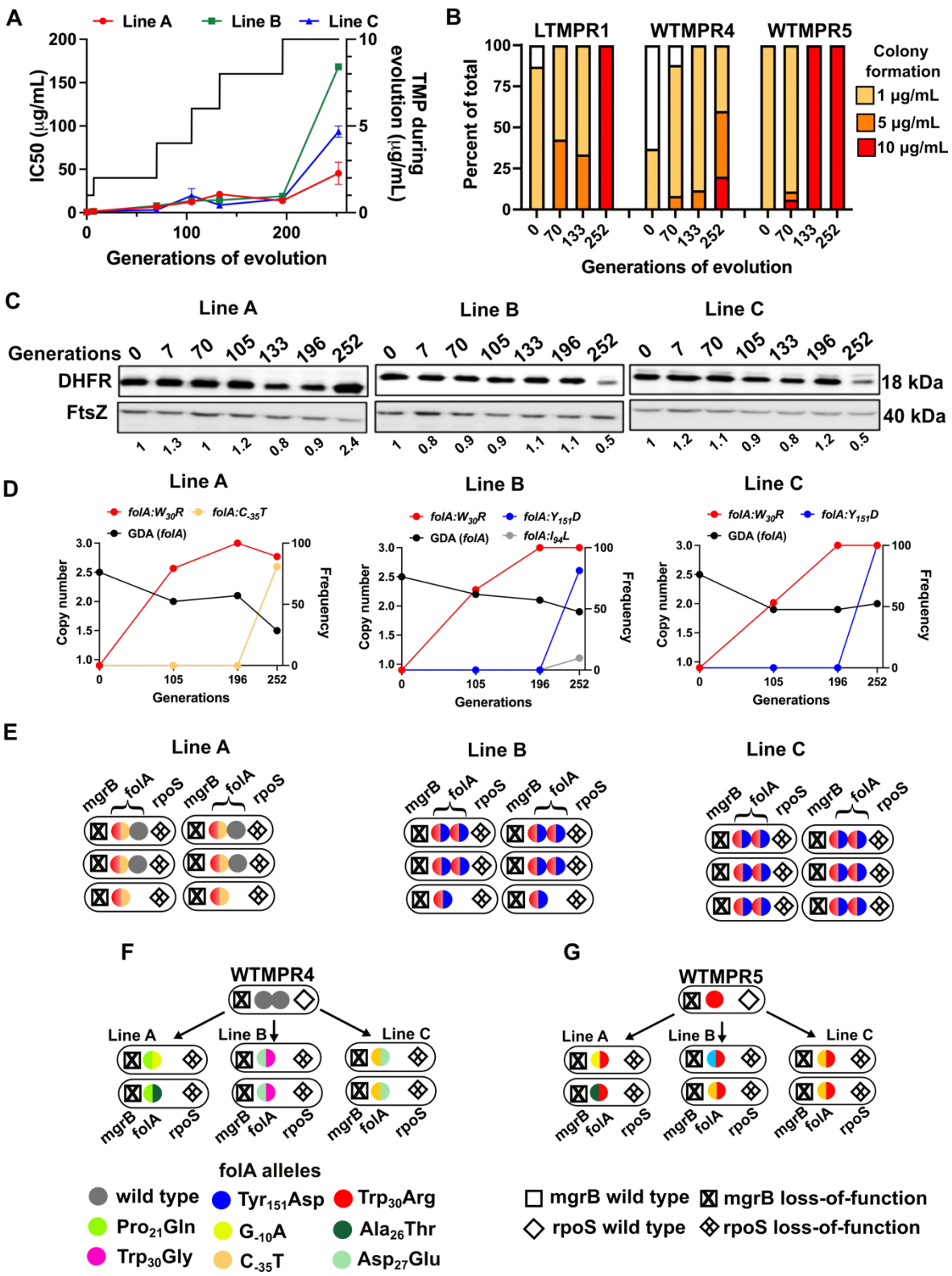


Figure 7. Evolutionary fate of *folA*-encompassing GDA in LTMPR1 at increasing trimethoprim pressure. A. Trimethoprim IC50 values of 3 evolving lineages starting from LTMPR1 (A, B, C) over 252 generations of evolution are plotted on the left Y-axis. Mean ± S.D. from three measurements is plotted at each time point. Trimethoprim concentrations used during evolution are plotted in the right Y-axis. **B.** Colony formation of LTMPR1,

842 WTMPR4 and WTMPR5 evolved in increasing trimethoprim at indicated time points.
843 Fraction of the population capable of forming colonies at 1, 5 and 10 μg/mL trimethoprim
844 was calculated across 3 replicate lines. Mean value from the three lines is plotted. The results
845 of similar experiments performed on ancestors (0 generations) are also provided for
846 reference. **C.** DHFR expression in LTMPR1 lineages A, B and C evolved in increasing
847 trimethoprim, measured by immunoblotting using anti-DHFR polyclonal antibody. FtsZ was
848 used as a loading control. Quantitation was performed by calculating band intensities using
849 image analysis. DHFR expression at each time point was normalized to the ancestor (i.e. 0
850 generations, set to 1). Mean of three independent measurements is shown below each lane. **D.**
851 Copy number of *folA* (GDA(*folA*)) at different time points of LTMPR1 evolution in
852 increasing trimethoprim is plotted on the left y-axis. Copy number was determined by
853 dividing number of reads from an Illumina sequencing experiment corresponding to *folA* by
854 the average number of reads mapping to the rest of the genome. Frequency of various *folA*
855 alleles in the evolving LTMPR1 populations at each of the time points is shown on the right
856 y-axis. **E.** Point mutations in *folA*, *rpoS* and *mgrB* in 6 randomly-picked trimethoprim-
857 resistant colonies from each of the LTMPR1 lines at 252 generations of evolution are shown
858 schematically. **F, G.** Point mutations in *folA*, *mgrB* and *rpoS* in isolates derived from 252
859 generations of evolution of WTMPR4 (F) and WTMPR5 (G) in increasing trimethoprim.
860 From each of the evolving lineages 2 random isolates were picked for genome sequencing.
861 Various point mutants at the 3 gene loci are represented by appropriate symbols as shown.
862

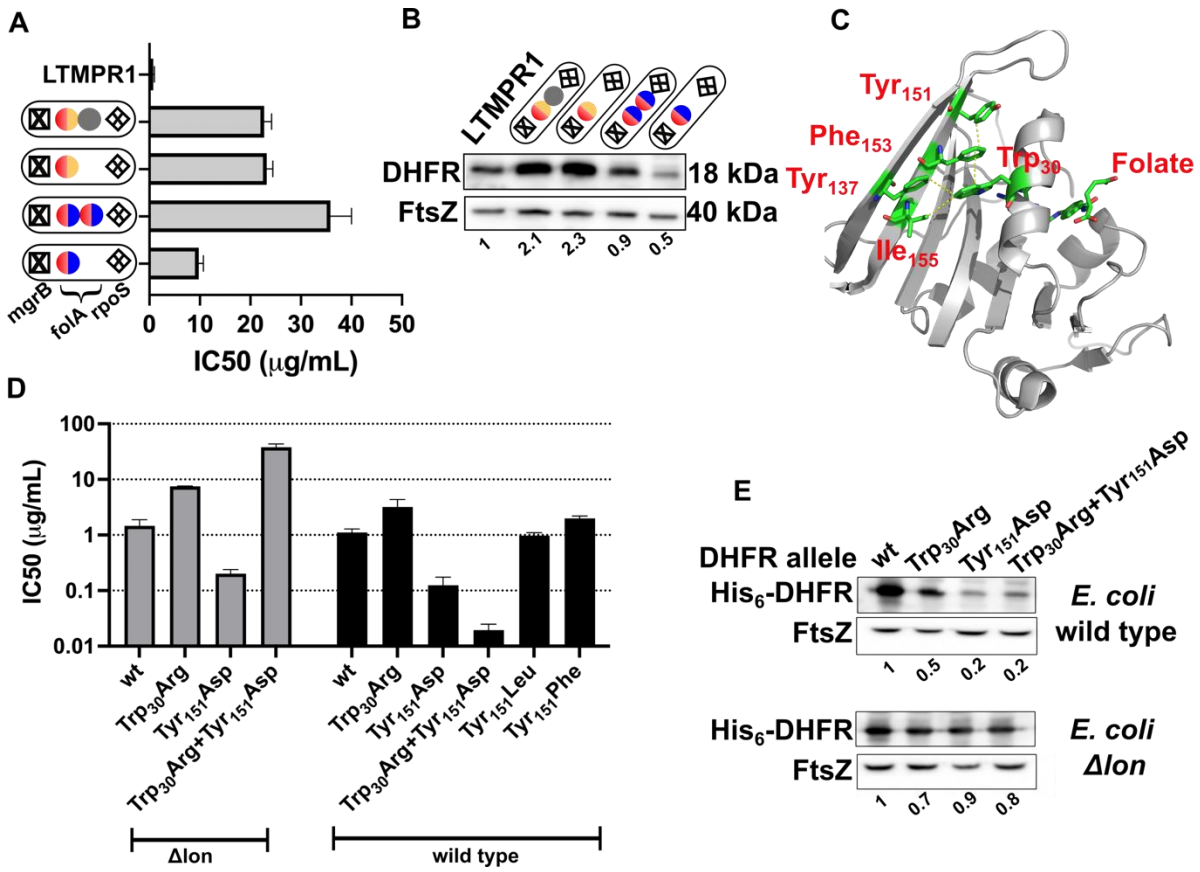
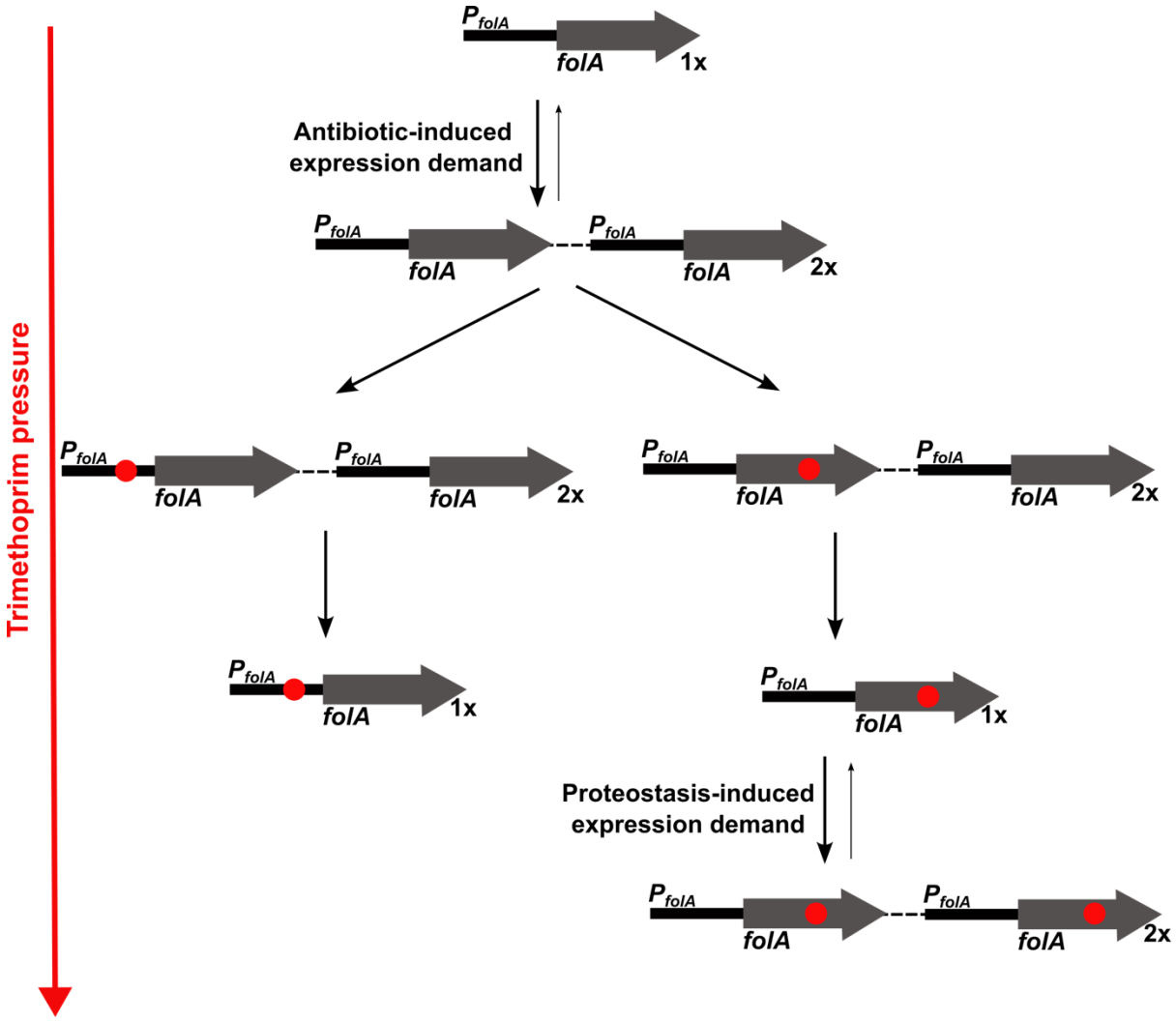


Figure 8. Proteostatic pressure facilitates maintenance of *folA* duplication A. Trimethoprim-IC50 values of resistant isolates with the indicated genotypes (see also Figure 7) derived from LTMPR1 evolution in increasing antibiotic pressure. Mean ± S.D. from three independent measurements is plotted. **B.** DHFR expression level in LTMPR1-derived trimethoprim resistant isolates with indicated genotypes assessed by immunoblotting using anti-DHFR polyclonal antibody. FtsZ was used as a loading control. Quantitation was performed by calculating band intensities using image analysis. DHFR expression was normalized to the ancestor (LTMPR1). Mean of three independent measurements is provided. **C.** Cartoon representation of the structure of *E. coli* DHFR (PDB: 7DFR) bound to folate. Residues Trp_{30}, Tyr_{151}, Phe_{153} and Ile_{155} which form hydrophobic interactions and are required for proteolytic stability are shown as sticks and coloured by element (C: green, O:red, N:blue). Distances of less than or equal to 4 Å, which indicates possible interactions, are shown as dotted yellow traces. Folate bound in the active site of DHFR is shown as sticks. **D.** Trimethoprim IC50 values of *E. coli* wild type or *E. coli Δlon* heterologously expressing DHFR (wt) or its mutants. Mean ± S.D. from three independent measurements is plotted. **E.** Expression level of plasmid-borne DHFR (wt) or its mutants in wild type or *Δlon*

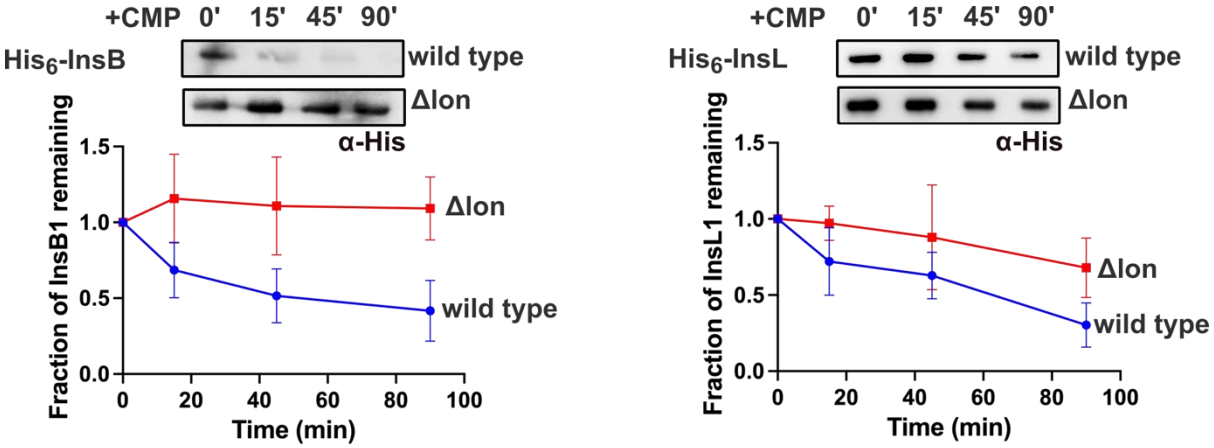
881 *E. coli* assessed by immunoblotting using anti-DHFR polyclonal antibody. FtsZ was used as a
882 loading control. Quantitation was performed by calculating band intensities using image
883 analysis. Expression of mutants was normalized to wild type DHFR (set to 1). Mean of three
884 independent measurements is shown.
885



886
887 **Figure 9. Model of expression demand induced selection of *folA* gene duplication,**
888 **followed by replacement by phenotypically equivalent point mutants.** Expression demand
889 for DHFR is generated by trimethoprim pressure, which results in the selection of *folA* gene
890 duplications. Gene duplication is unstable and reverses to a single copy of wild type *folA*
891 when drug-pressure is withdrawn. Under drug pressure, *folA* gene duplication is maintained
892 until a point mutation that confers resistance arises in one of the copies of *folA* (promoter or
893 coding region). Upon acquisition of a point mutation, reversal of gene duplication ensues
894 unless additional expression demand is generated by the action of proteostatic machinery on
895 unstable drug-resistant DHFR mutants.

896 **Supplementary Figures**

897

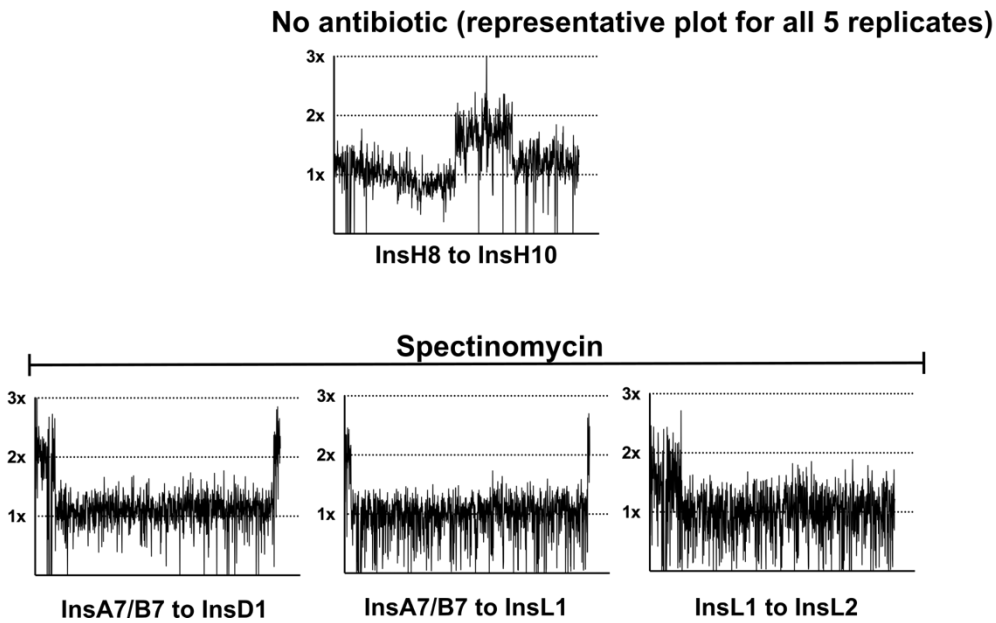


898

899 **Figure S1.** Stability of heterologously expressed InsB and InsL transposases in wild type or
900 *Δlon E. coli* measured using a chloramphenicol (CMP) chase assay. Levels of the plasmid-
901 expressed transposase were measured using immunoblotting with an anti-His
902 monoclonal/polyclonal antibody in lysates prepared from cells at indicated time points
903 following treatment with 50 μg/mL of CMP. A representative image from three independent
904 experiments is shown. Band intensities were quantitated by image analyses and normalized to
905 0' (set to 1). The fraction of protein remaining is plotted in blue for wild type and red for *E.*
906 *coli Δlon*. Values from 3 independent experiments are plotted as mean ± S.D.

907

Representative coverage plots of *E. coli Δlon* evolved in no antibiotic or spectinomycin



908

909

910 **Figure S2.** Coverage plots and flanking IS elements for *E. coli Δlon* populations evolving in
911 no antibiotic or spectinomycin that showed GDA mutations. For no-antibiotic, only one of the
912 5 replicates is shown as similar coverage plots were obtained for all 5 replicates.
913

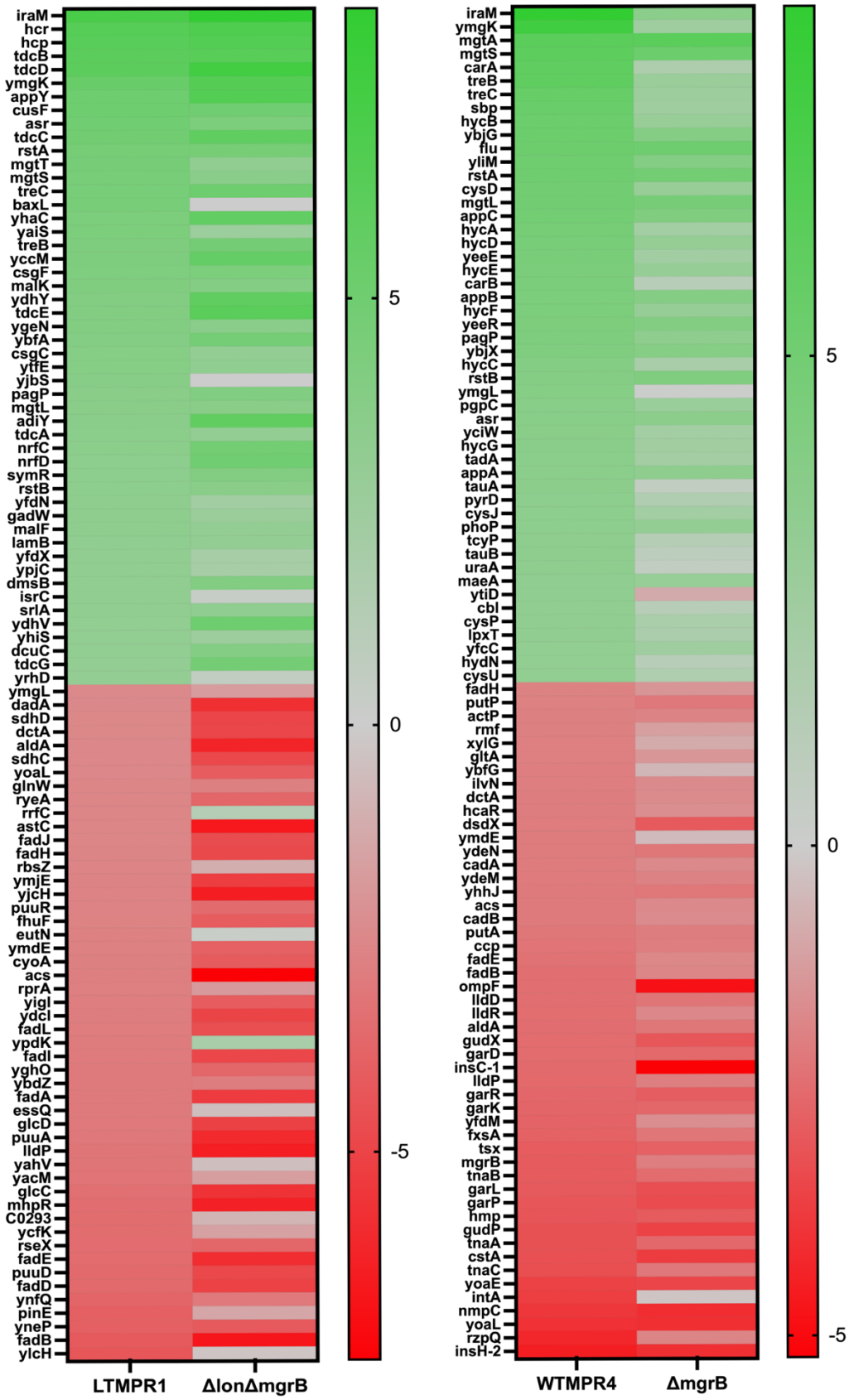
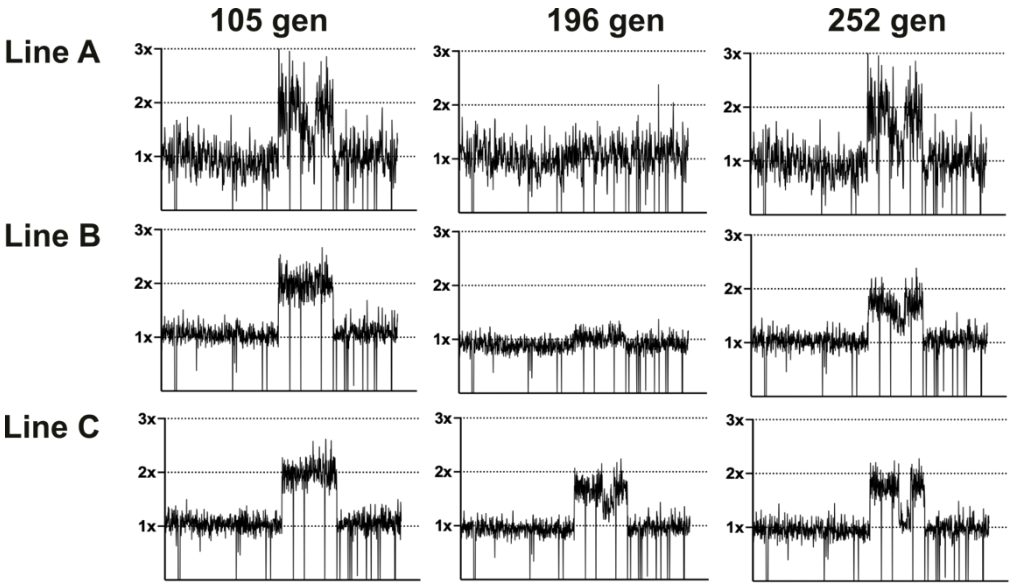
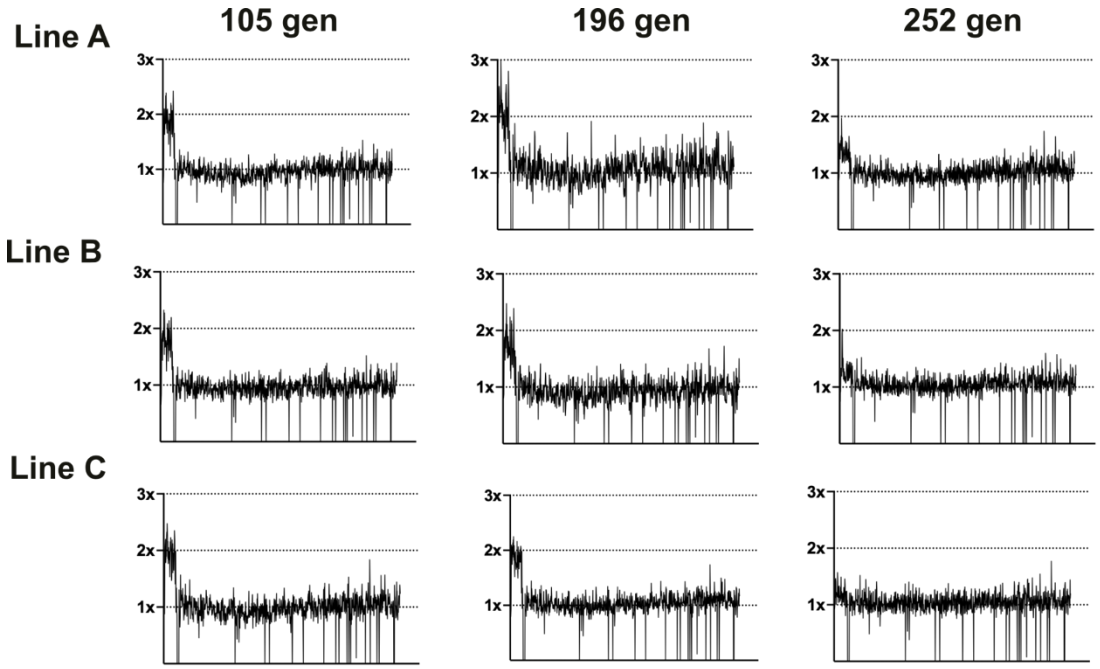


Figure S3. Gene expression changes associated with *folA*-encompassing duplication. Heat map comparing the top 50 up/downregulated genes in LTMPR1 and WTMPR4 with *E.*

917 *coli ΔlonΔmgrB* and *E. coli ΔmgrB* respectively. Log_2(Fold change) values were calculated
918 based on RNA-sequencing with respect to wild type or *Δlon E. coli* and represented on a
919 green/red scale as indicated. Gene names are shown adjacent to the appropriate cell. Genes
920 are ranked according to their fold change in LTMPR1 and WTMPR4.
921



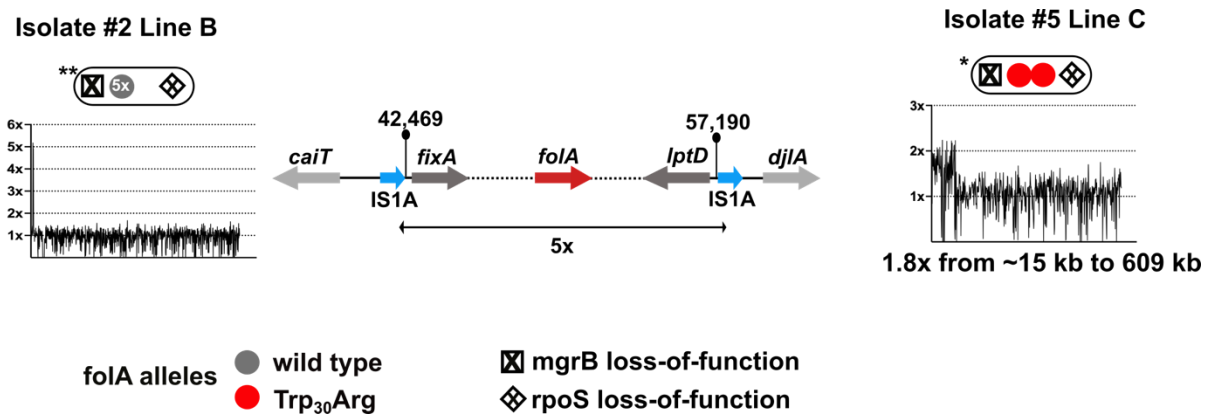
922
923 **Figure S4.** Coverage depth plots for population sequencing at 105, 196 and 252 generations
924 of the 3 lineages of LTMPR1 (A, B, C) evolving in trimethoprim-free media. Coverage depth
925 plots show the number of reads from Illumina sequencing (y-axis) mapped to each genomic
926 coordinate of the reference genome (x-axis). Coverage of 1x, 2x and 3x are marked by dotted
927 lines for reference.
928



929

930 **Figure S5.** Coverage depth plots for population sequencing at 105, 196 and 252 generations
931 of the 3 lineages of LTMPR1 (A, B, C) evolving in trimethoprim-supplemented (300 ng/mL)
932 media. Coverage depth plots show the number of reads from Illumina sequencing (y-axis
933 mapped) to each genomic coordinate of the reference genome (x-axis). Coverage of 1x, 2x
934 and 3x are marked by dotted lines for reference.

935



936

937 **Figure S6.** Coverage plots for Line C Isolate 5 and Line B Isolate 2. Line B Isolate 3 showed
938 5x amplification of a shorter genomic stretch encompassing *folA* which is shown
939 diagrammatically next to the coverage plot. Line C Isolate 5 showed an expanded GDA
940 encompassing *folA* compared to the LTMPR1 ancestor, the coordinates for which are below
941 the coverage plot.

942

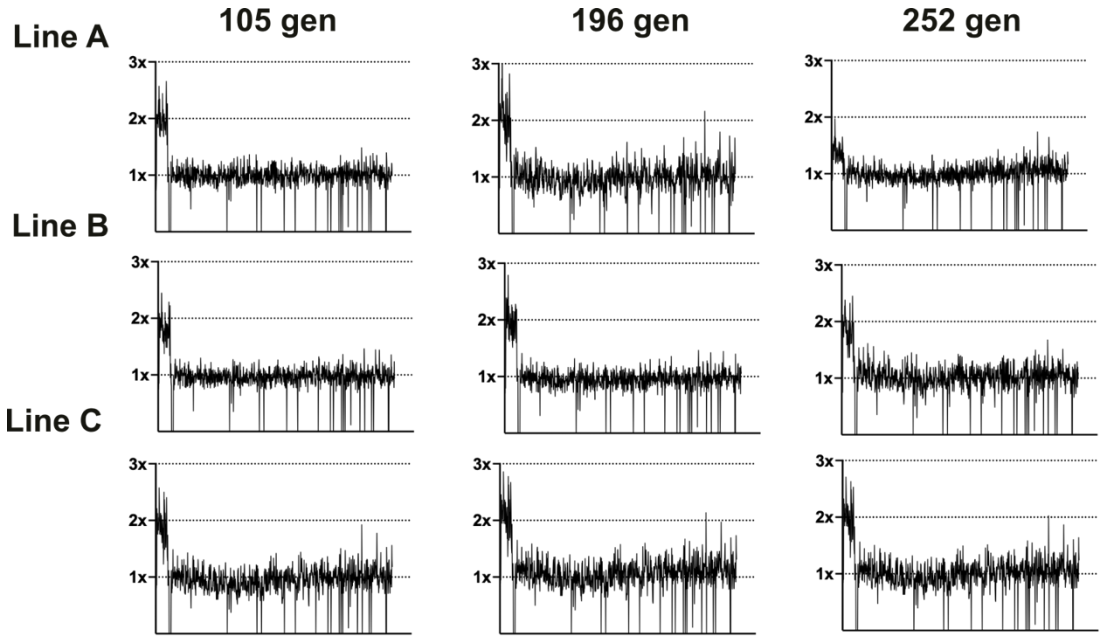


Figure S7. Coverage depth plots for population sequencing at 105, 196 and 252 generations of the 3 lineages of LTMPR1 (A, B, C) evolving in increasing trimethoprim concentrations. Coverage depth plots show the number of reads from Illumina sequencing (y-axis) mapped to each genomic coordinate of the reference genome (x-axis). Coverage of 1x, 2x and 3x are marked by dotted lines for reference.

952 **References**

953 1. Andersson DI, Hughes D. Gene Amplification and Adaptive Evolution in Bacteria.
954 Annu Rev Genet. 2009;43:167-95. doi: 10.1146/annurev-genet-102108-134805. PubMed
955 PMID: WOS:000273580300008.

956 2. Elliott KT, Cuff LE, Neidle EL. Copy number change: evolving views on gene
957 amplification. Future Microbiol. 2013;8(7):887-99. doi: 10.2217/fmb.13.53. PubMed PMID:
958 WOS:000325822800015.

959 3. Kondrashov FA. Gene duplication as a mechanism of genomic adaptation to a
960 changing environment. P Roy Soc B-Biol Sci. 2012;279(1749):5048-57. doi:
961 10.1098/rspb.2012.1108. PubMed PMID: WOS:000310999000025.

962 4. Dürrbaum M, Storchová Z. Effects of aneuploidy on gene expression: implications for
963 cancer. Febs J. 2016;283(5):791-802. doi: 10.1111/febs.13591. PubMed PMID:
964 WOS:000372005600002.

965 5. Jiang Z, Li CX, Li FY, Wang XS. EGFR Gene Copy Number as a Prognostic Marker
966 in Colorectal Cancer Patients Treated with Cetuximab or Panitumumab: A Systematic Review
967 and Meta Analysis. Plos One. 2013;8(2). doi: ARTN e56205
968 10.1371/journal.pone.0056205. PubMed PMID: WOS:000315159200024.

969 6. Spielmann M, Lupiáñez DG, Mundlos S. Structural variation in the 3D genome. Nat
970 Rev Genet. 2018;19(7):453-67. doi: 10.1038/s41576-018-0007-0. PubMed PMID:
971 WOS:000435393700009.

972 7. Pascual-Anaya J, D'Aniello S, Kuratani S, Garcia-Fernàndez J. Evolution of gene
973 clusters in deuterostomes. Bmc Dev Biol. 2013;13. doi: Artn 26
974 10.1186/1471-213x-13-26. PubMed PMID: WOS:000321576400001.

975 8. Taylor JS, Van de Peer Y, Meyer A. Genome duplication, divergent resolution and
976 speciation. Trends Genet. 2001;17(6):299-301. doi: Doi 10.1016/S0168-9525(01)02318-6.
977 PubMed PMID: WOS:000170644200002.

978 9. Ting CT, Tsaur SC, Sun S, Browne WE, Chen YC, Patel NH, et al. Gene duplication
979 and speciation in *Drosophila*: Evidence from the *Odysseus* locus. P Natl Acad Sci USA.
980 2004;101(33):12232-5. doi: 10.1073/pnas.0401975101. PubMed PMID:
981 WOS:000223410100059.

982 10. Jiao WB, Patel V, Klasen J, Liu F, Pecinkova P, Ferrand M, et al. The Evolutionary
983 Dynamics of Genetic Incompatibilities Introduced by Duplicated Genes in *Arabidopsis*
984 *thaliana*. Mol Biol Evol. 2021;38(4):1225-40. doi: 10.1093/molbev/msaa306. PubMed
985 PMID: 33247726; PubMed Central PMCID: PMCPMC8042742.

11. Maclean CJ, Greig D. Reciprocal Gene Loss Following Experimental Whole-Genome Duplication Causes Reproductive Isolation in Yeast. Evolution. 2011;65(4):932-45. doi: 10.1111/j.1558-5646.2010.01171.x. PubMed PMID: WOS:000289158300003.

12. Zuellig MP, Sweigart AL. Gene duplicates cause hybrid lethality between sympatric species of *Mimulus*. Plos Genet. 2018;14(4). doi: ARTN e1007130 10.1371/journal.pgen.1007130. PubMed PMID: WOS:000431115700001.

13. Sandegren L, Andersson DI. Bacterial gene amplification: implications for the evolution of antibiotic resistance. Nat Rev Microbiol. 2009;7(8):578-88. doi: 10.1038/nrmicro2174. PubMed PMID: 19609259.

14. Anderson P, Roth J. Spontaneous tandem genetic duplications in Salmonella typhimurium arise by unequal recombination between rRNA (rrn) cistrons. Proc Natl Acad Sci U S A. 1981;78(5):3113-7. doi: 10.1073/pnas.78.5.3113. PubMed PMID: 6789329; PubMed Central PMCID: PMCPMC319510.

15. Seaton SC, Elliott KT, Cuff LE, Laniohan NS, Patel PR, Neidle EL. Genome-wide selection for increased copy number in *Acinetobacter baylyi* ADP1: locus and context-dependent variation in gene amplification. Mol Microbiol. 2012;83(3):520-35. Epub 20111229. doi: 10.1111/j.1365-2958.2011.07945.x. PubMed PMID: 22211470.

16. Sonti RV, Roth JR. Role of gene duplications in the adaptation of *Salmonella typhimurium* to growth on limiting carbon sources. Genetics. 1989;123(1):19-28. doi: 10.1093/genetics/123.1.19. PubMed PMID: 2680755; PubMed Central PMCID: PMCPMC1203782.

17. Hahn MW. Distinguishing Among Evolutionary Models for the Maintenance of Gene Duplicates. J Hered. 2009;100(5):605-17. doi: 10.1093/jhered/esp047. PubMed PMID: WOS:000269206800012.

18. Innan H, Kondrashov F. The evolution of gene duplications: classifying and distinguishing between models. Nat Rev Genet. 2010;11(2):97-108. doi: 10.1038/nrg2689. PubMed PMID: WOS:000273688200004.

19. Kuzmin E, Taylor JS, Boone C. Retention of duplicated genes in evolution. Trends Genet. 2022;38(1):59-72. doi: 10.1016/j.tig.2021.06.016. PubMed PMID: WOS:000732466400007.

20. Hufnagel DA, Choby JE, Hao S, Johnson AF, Burd EM, Langelier C, et al. Antibiotic-Selected Gene Amplification Heightens Metal Resistance. mBio. 2021;12(1). Epub 20210119. doi: 10.1128/mBio.02994-20. PubMed PMID: 33468696; PubMed Central PMCID: PMCPMC8545094.

1020 21. von Rozycki T, Nies DH. *Cupriavidus metallidurans*: evolution of a metal-resistant
1021 bacterium. Antonie Van Leeuwenhoek. 2009;96(2):115-39. Epub 20081001. doi:
1022 10.1007/s10482-008-9284-5. PubMed PMID: 18830684.
1023 22. Bi X, Liu LF. RecA-independent and recA-dependent intramolecular plasmid
1024 recombination: Differential homology requirement and distance effect. J Mol Biol.
1025 1994;235(2):414-23. doi: 10.1006/jmbi.1994.1002. PubMed PMID: 8289271.
1026 23. Reams AB, Kofoid E, Kugelberg E, Roth JR. Multiple Pathways of Duplication
1027 Formation with and Without Recombination (RecA) in *Salmonella enterica*. Genetics.
1028 2012;192(2):397-415. doi: 10.1534/genetics.112.142570. PubMed PMID:
1029 WOS:000309547400006.
1030 24. Reams AB, Kofoid E, Savageau M, Roth JR. Duplication frequency in a population of
1031 *Salmonella enterica* rapidly approaches steady state with or without recombination. Genetics.
1032 2010;184(4):1077-94. Epub 20100118. doi: 10.1534/genetics.109.111963. PubMed PMID:
1033 20083614; PubMed Central PMCID: PMCPMC2865909.
1034 25. Reams AB, Neidle EL. Genome plasticity in *Acinetobacter*: new degradative
1035 capabilities acquired by the spontaneous amplification of large chromosomal segments.
1036 Molecular Microbiology. 2003;47(5):1291-304. doi: DOI 10.1046/j.1365-2958.2003.03342.x.
1037 PubMed PMID: WOS:000181056400009.
1038 26. Min YH, Kwon AR, Yoon JM, Yoon EJ, Shim MJ, Choi EC. Molecular analysis of
1039 constitutive mutations in ermB and ermA selected in vitro from inducibly MLSB-resistant
1040 enterococci. Arch Pharm Res. 2008;31(3):377-80. Epub 20080413. doi: 10.1007/s12272-001-
1041 1167-8. PubMed PMID: 18409053.
1042 27. Sun S, Berg OG, Roth JR, Andersson DI. Contribution of gene amplification to
1043 evolution of increased antibiotic resistance in *Salmonella typhimurium*. Genetics.
1044 2009;182(4):1183-95. Epub 20090527. doi: 10.1534/genetics.109.103028. PubMed PMID:
1045 19474201; PubMed Central PMCID: PMCPMC2728858.
1046 28. Vinella D, Cashel M, D'Ari R. Selected amplification of the cell division genes ftsQ-
1047 ftsA-ftsZ in *Escherichia coli*. Genetics. 2000;156(4):1483-92. doi:
1048 10.1093/genetics/156.4.1483. PubMed PMID: 11102351; PubMed Central PMCID:
1049 PMCPMC1461353.
1050 29. Nilsson AI, Zorzet A, Kanth A, Dahlstrom S, Berg OG, Andersson DI. Reducing the
1051 fitness cost of antibiotic resistance by amplification of initiator tRNA genes. Proc Natl Acad
1052 Sci U S A. 2006;103(18):6976-81. Epub 20060424. doi: 10.1073/pnas.0602171103. PubMed
1053 PMID: 16636273; PubMed Central PMCID: PMCPMC1459004.

1054 30. Andersson DI, Nicoloff H, Hjort K. Mechanisms and clinical relevance of bacterial
1055 heteroresistance. Nature Reviews Microbiology. 2019;17(8):479-96. doi: 10.1038/s41579-
1056 019-0218-1. PubMed PMID: WOS:000475479100007.
1057 31. Nicoloff H, Hjort K, Levin BR, Andersson DI. The high prevalence of antibiotic
1058 heteroresistance in pathogenic bacteria is mainly caused by gene amplification. Nat
1059 Microbiol. 2019;4(3):504-14. doi: 10.1038/s41564-018-0342-0. PubMed PMID:
1060 WOS:000459201400019.
1061 32. Sun S, Ke R, Hughes D, Nilsson M, Andersson DI. Genome-wide detection of
1062 spontaneous chromosomal rearrangements in bacteria. Plos One. 2012;7(8):e42639. Epub
1063 20120803. doi: 10.1371/journal.pone.0042639. PubMed PMID: 22880062; PubMed Central
1064 PMCID: PMCPMC3411829.
1065 33. Tomanek I, Grah R, Lagator M, Andersson AMC, Bollback JP, Tkacik G, et al. Gene
1066 amplification as a form of population-level gene expression regulation. Nat Ecol Evol.
1067 2020;4(4):612-25. doi: 10.1038/s41559-020-1132-7. PubMed PMID:
1068 WOS:000519008300005.
1069 34. Ahmad SI, Kirk SH, Eisenstark A. Thymine metabolism and thymineless death in
1070 prokaryotes and eukaryotes. Annu Rev Microbiol. 1998;52:591-625. doi:
1071 10.1146/annurev.micro.52.1.591. PubMed PMID: 9891809.
1072 35. Gangjee A, Jain HD, Kurup S. Recent advances in classical and non-classical
1073 Antifolates as antitumor and antiopportunistic infection agents: part I. Anti-Cancer Agent Me.
1074 2007;7(5):524-42. doi: Doi 10.2174/187152007781668724. PubMed PMID:
1075 WOS:000249469700005.
1076 36. Gangjee A, Jain HD, Kurup S. Recent advances in classical and non-classical
1077 antifolates as antitumor and antiopportunistic infection agents: Part II. Anti-Cancer Agent
1078 Me. 2008;8(2):205-31. doi: Doi 10.2174/187152008783497064. PubMed PMID:
1079 WOS:000255810200007.
1080 37. Huennekens FM. The methotrexate story: a paradigm for development of cancer
1081 chemotherapeutic agents. Adv Enzyme Regul. 1994;34:397-419. doi: 10.1016/0065-
1082 2571(94)90025-6. PubMed PMID: 7942284.
1083 38. Plowe CV. Malaria chemoprevention and drug resistance: a review of the literature
1084 and policy implications. Malar J. 2022;21(1):104. Epub 20220324. doi: 10.1186/s12936-022-
1085 04115-8. PubMed PMID: 35331231; PubMed Central PMCID: PMCPMC8943514.
1086 39. Gleckman R, Blagg N, Joubert DW. Trimethoprim: mechanisms of action,
1087 antimicrobial activity, bacterial resistance, pharmacokinetics, adverse reactions, and

1088 therapeutic indications. Pharmacotherapy. 1981;1(1):14-20. doi: 10.1002/j.1875-
1089 9114.1981.tb03548.x. PubMed PMID: 6985448.
1090 40. Brochet M, Couve E, Zouine M, Poyart C, Glaser P. A naturally occurring gene
1091 amplification leading to sulfonamide and trimethoprim resistance in Streptococcus agalactiae.
1092 J Bacteriol. 2008;190(2):672-80. Epub 20071116. doi: 10.1128/JB.01357-07. PubMed PMID:
1093 18024520; PubMed Central PMCID: PMCPMC2223700.
1094 41. Goker E, Waltham M, Kheradpour A, Trippett T, Mazumdar M, Elisseyeff Y, et al.
1095 Amplification of the Dihydrofolate-Reductase Gene Is a Mechanism of Acquired-Resistance
1096 to Methotrexate in Patients with Acute Lymphoblastic-Leukemia and Is Correlated with P53
1097 Gene-Mutations. Blood. 1995;86(2):677-84. doi: DOI
1098 10.1182/blood.V86.2.677.bloodjournal862677. PubMed PMID: WOS:A1995RH80100030.
1099 42. Inselburg J, Bzik DJ, Horii T. Pyrimethamine resistant Plasmodium falciparum:
1100 overproduction of dihydrofolate reductase by a gene duplication. Mol Biochem Parasitol.
1101 1987;26(1-2):121-34. doi: 10.1016/0166-6851(87)90136-8. PubMed PMID: 3323903.
1102 43. Morales C, Garcia MJ, Ribas M, Miro R, Munoz M, Caldas C, et al. Dihydrofolate
1103 reductase amplification and sensitization to methotrexate of methotrexate-resistant colon
1104 cancer cells. Mol Cancer Ther. 2009;8(2):424-32. Epub 20090203. doi: 10.1158/1535-
1105 7163.MCT-08-0759. PubMed PMID: 19190117.
1106 44. Baym M, Lieberman TD, Kelsic ED, Chait R, Gross R, Yelin I, et al. Spatiotemporal
1107 microbial evolution on antibiotic landscapes. Science. 2016;353(6304):1147-51. doi:
1108 10.1126/science.aag0822. PubMed PMID: 27609891; PubMed Central PMCID:
1109 PMCPMC5534434.
1110 45. Palmer AC, Toprak E, Baym M, Kim S, Veres A, Bershtein S, et al. Delayed
1111 commitment to evolutionary fate in antibiotic resistance fitness landscapes. Nat Commun.
1112 2015;6:7385. Epub 20150610. doi: 10.1038/ncomms8385. PubMed PMID: 26060115;
1113 PubMed Central PMCID: PMCPMC4548896.
1114 46. Patel V, Matange N. Adaptation and compensation in a bacterial gene regulatory
1115 network evolving under antibiotic selection. Elife. 2021;10. Epub 20210930. doi:
1116 10.7554/eLife.70931. PubMed PMID: 34591012; PubMed Central PMCID:
1117 PMCPMC8483737.
1118 47. Watson M, Liu JW, Ollis D. Directed evolution of trimethoprim resistance in
1119 *Escherichia coli*. Febs J. 2007;274(10):2661-71. Epub 20070419. doi: 10.1111/j.1742-
1120 4658.2007.05801.x. PubMed PMID: 17451440.

1121 48. Toprak E, Veres A, Michel JB, Chait R, Hartl DL, Kishony R. Evolutionary paths to
1122 antibiotic resistance under dynamically sustained drug selection. Nat Genet. 2011;44(1):101-
1123 5. Epub 20111218. doi: 10.1038/ng.1034. PubMed PMID: 22179135; PubMed Central
1124 PMCID: PMCPMC3534735.
1125 49. Matange N. Highly Contingent Phenotypes of Lon Protease Deficiency in Escherichia
1126 coli upon Antibiotic Challenge. J Bacteriol. 2020;202(3). Epub 20200115. doi:
1127 10.1128/JB.00561-19. PubMed PMID: 31740490; PubMed Central PMCID:
1128 PMCPMC6964741.
1129 50. Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI.
1130 Biophysical principles predict fitness landscapes of drug resistance. Proc Natl Acad Sci U S
1131 A. 2016;113(11):E1470-8. Epub 20160229. doi: 10.1073/pnas.1601441113. PubMed PMID:
1132 26929328; PubMed Central PMCID: PMCPMC4801265.
1133 51. Vinchhi R, Yelpure C, Balachandran M, Matange N. Pervasive gene deregulation
1134 underlies adaptation and maladaptation in trimethoprim-resistant *E. coli*. mBio.
1135 2023:e0211923. Epub 20231130. doi: 10.1128/mbio.02119-23. PubMed PMID: 38032208.
1136 52. Ranquet C, Gottesman S. Translational regulation of the Escherichia coli stress factor
1137 RpoS: a role for SsrA and Lon. J Bacteriol. 2007;189(13):4872-9. Epub 20070420. doi:
1138 10.1128/JB.01838-06. PubMed PMID: 17449615; PubMed Central PMCID:
1139 PMCPMC1913435.
1140 53. Xu J, Li T, Gao Y, Deng J, Gu J. MgrB affects the acid stress response of Escherichia
1141 coli by modulating the expression of iraM. FEMS Microbiol Lett. 2019;366(11). doi:
1142 10.1093/femsle/fnz123. PubMed PMID: 31158277.
1143 54. Matange N, Bodkhe S, Patel M, Shah P. Trade-offs with stability modulate innate and
1144 mutationally acquired drug resistance in bacterial dihydrofolate reductase enzymes. Biochem
1145 J. 2018;475(12):2107-25. Epub 20180629. doi: 10.1042/BCJ20180249. PubMed PMID:
1146 29871875.
1147 55. Ohno S. Evolution by gene duplication. London, New York: Allen & Unwin;
1148 Springer-Verlag; 1970. xv, 160 p. p.
1149 56. Bergthorsson U, Andersson DI, Roth JR. Ohno's dilemma: Evolution of new genes
1150 under continuous selection. P Natl Acad Sci USA. 2007;104(43):17004-9. doi:
1151 10.1073/pnas.0707158104. PubMed PMID: WOS:000250487600051.
1152 57. Long MY, VanKuren NW, Chen SD, Vibranovski MD. New Gene Evolution: Little
1153 Did We Know. Annual Review of Genetics, Vol 47. 2013;47:307-33. doi: 10.1146/annurev-
1154 genet-111212-133301. PubMed PMID: WOS:000329816000014.

58. Holloway AK, Palzkill T, Bull JJ. Experimental evolution of gene duplicates in a bacterial plasmid model. J Mol Evol. 2007;64(2):215-22. Epub 20070108. doi: 10.1007/s00239-006-0087-x. PubMed PMID: 17211548.

59. Kugelberg E, Kofoid E, Reams AB, Andersson DI, Roth JR. Multiple pathways of selected gene amplification during adaptive mutation. P Natl Acad Sci USA. 2006;103(46):17319-24. doi: 10.1073/pnas.0608309103. PubMed PMID: WOS:000242249400047.

60. Naseeb S, Ames RM, Delneri D, Lovell SC. Rapid functional and evolutionary changes follow gene duplication in yeast. Proc Biol Sci. 2017;284(1861). doi: 10.1098/rspb.2017.1393. PubMed PMID: 28835561; PubMed Central PMCID: PMCPMC5577496.

61. Tomanek I, Guet CC. Adaptation dynamics between copy-number and point mutations. Elife. 2022;11. doi: ARTN e82240
10.7554/eLife.82240. PubMed PMID: WOS:000912674700001.

62. Nicoloff H, Perreten V, Levy SB. Increased genome instability in *Escherichia coli* lon mutants: relation to emergence of multiple-antibiotic-resistant (Mar) mutants caused by insertion sequence elements and large tandem genomic amplifications. Antimicrob Agents Chemother. 2007;51(4):1293-303. Epub 20070112. doi: 10.1128/AAC.01128-06. PubMed PMID: 17220404; PubMed Central PMCID: PMCPMC1855481.

63. Nicoloff H, Perreten V, McMurry LM, Levy SB. Role for tandem duplication and lon protease in AcrAB-TolC-dependent multiple antibiotic resistance (Mar) in an *Escherichia coli* mutant without mutations in marRAB or acrRAB. Journal of Bacteriology. 2006;188(12):4413-23. doi: 10.1128/Jb.01502-05. PubMed PMID: WOS:000238173900027.

64. Derbyshire KM, Kramer M, Grindley ND. Role of instability in the cis action of the insertion sequence IS903 transposase. Proc Natl Acad Sci U S A. 1990;87(11):4048-52. doi: 10.1073/pnas.87.11.4048. PubMed PMID: 2161528; PubMed Central PMCID: PMCPMC54044.

65. Rouquette C, Serre MC, Lane D. Protective role for H-NS protein in IS1 transposition. J Bacteriol. 2004;186(7):2091-8. doi: 10.1128/JB.186.7.2091-2098.2004. PubMed PMID: 15028694; PubMed Central PMCID: PMCPMC374409.

66. Hua CF, Wang TT, Shao XL, Xie YP, Huang H, Liu JG, et al. *Pseudomonas psyringae* dual-function protein Lon switches between virulence and metabolism by acting as both DNA-binding transcriptional regulator and protease in different environments. Environ

1188 Microbiol. 2020;22(7):2968-88. doi: 10.1111/1462-2920.15067. PubMed PMID:
1189 WOS:000537012700001.
1190 67. Karlowicz A, Wegrzyn K, Gross M, Kaczynska D, Ropelewska M, Siemiatkowska M,
1191 et al. Defining the crucial domain and amino acid residues in bacterial Lon protease for DNA
1192 binding and processing of DNA-interacting substrates. J Biol Chem. 2017;292(18):7507-18.
1193 doi: 10.1074/jbc.M116.766709. PubMed PMID: WOS:000400761300019.
1194 68. Liu T, Lu B, Lee I, Ondrovicová G, Kutejová E, Suzuki CK. DNA and RNA binding
1195 by the mitochondrial Lon protease is regulated by nucleotide and protein substrate. J Biol
1196 Chem. 2004;279(14):13902-10. doi: 10.1074/jbc.M309642200. PubMed PMID:
1197 WOS:000220478500077.
1198 69. Lu B, Yadav S, Shah PG, Liu T, Tian B, Pukszta S, et al. Roles for the human ATP-
1199 dependent Lon protease in mitochondrial DNA maintenance. J Biol Chem.
1200 2007;282(24):17363-74. Epub 20070409. doi: 10.1074/jbc.M611540200. PubMed PMID:
1201 17420247.
1202 70. Sonezaki S, Okita K, Oba T, Ishii Y, Kondo A, Kato Y. Protein substrates and heat
1203 shock reduce the DNA binding ability of *Escherichia coli* Lon protease. Appl Microbiol Biot.
1204 1995;44(3-4):484-8. PubMed PMID: WOS:A1995TM49800033.
1205 71. Breidenstein EB, Bains M, Hancock RE. Involvement of the lon protease in the SOS
1206 response triggered by ciprofloxacin in Pseudomonas aeruginosa PAO1. Antimicrob Agents
1207 Chemother. 2012;56(6):2879-87. Epub 20120326. doi: 10.1128/AAC.06014-11. PubMed
1208 PMID: 22450976; PubMed Central PMCID: PMCPMC3370746.
1209 72. Matange N, Hegde S, Bodkhe S. Adaptation Through Lifestyle Switching Sculpts the
1210 Fitness Landscape of Evolving Populations: Implications for the Selection of Drug-Resistant
1211 Bacteria at Low Drug Pressures. Genetics. 2019;211(3):1029-44. Epub 20190122. doi:
1212 10.1534/genetics.119.301834. PubMed PMID: 30670539; PubMed Central PMCID:
1213 PMCPMC6404241.
1214 73. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved
1215 microbes from next-generation sequencing data using breseq. Methods Mol Biol.
1216 2014;1151:165-88. doi: 10.1007/978-1-4939-0554-6_12. PubMed PMID: 24838886; PubMed
1217 Central PMCID: PMCPMC4239701.
1218
1219