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 under trimethoprim exposure 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., relative copy number, 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 can be higher by several orders of magnitude 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 stable 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]. Further, 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 NADPH-dependent reduction 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], are all 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 distant 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 *folA*. Several point mutations that confer 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 300 ng/mL of trimethoprim (~0.3 x MIC) for 25 generations. We then isolated and sequenced the genome of a dominant trimethoprim-resistant clone from each population (designated as WTMPR1-5) (Figure 1 A). These sequencing results have been reported earlier [46]. Briefly, all isolates harboured loss-of-function mutations in the *mgrB* gene (Figure 1 B) that resulted in PhoQP-dependent overexpression of *folA* [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 *folA* (Figure 1 B, C). 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 (Figure 1 A). Wild type and *lon*-deficient *E. coli* have similar MIC for trimethoprim, though the *lon*-knockout has a higher IC50 [49]. Among trimethoprim-resistant isolates derived from *E. coli Δlon*, designated as LTMPR1-5, three harboured mutations at the *mgrB* locus (LTMPR1-3, Figure 1 A, B, Table 1). LTMPR4 and LTMPR5 harboured mutations in *pitA*, coding for a metal-phosphate symporter, that have been reported earlier in trimethoprim adapted *E. coli* [51]. LTMPR5 had a deletion encompassing *pitA*, indicating that loss of PitA was beneficial in trimethoprim. A knockout of *pitA* indeed showed greater colony formation on trimethoprim-supplemented media demonstrating that mutations in this gene were adaptive (Figure S1), though the underlying molecular mechanisms are unknown at present. Interestingly, none of the LTMPR isolates 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 B, C, Table 1).

Table 1. Complete list of mutations identified in isolates LTMPR1-5.

Gene/Locus	LTMPR1	LTMPR2	LTMPR3	LTMPR4	LTMPR5	Description
mgrB	IS5 insertion at -46	IS2 insertion at +82	IS5 insertion at -36	-	-	Negative feedback regulator of PhoQ sensor kinase
ybcK	-	IS2 insertion at +118	-	-	-	Prophage recombinase
pitA	-	-	-	$Ser_{425}Pro$	Δ 18 genes [yhiN, *pitA*… *yhiS]*. IS5-mediated	Metal/Phosphate symporter
insH21	Δ *insH21*	Δ *insH21*	Δ *insH21*	Δ *insH21*	Δ *insH21*	IS5 transposase
GDA (encompassing *folA*)	19,796 to 258,462 bp (~2.3x)	15,445 to 609,177 bp (~1.6x)	15,445 to 609,177 bp (~1.9x)	15,445 to 609,177 bp (~1.7x)	-	

Duplicated genomic regions in WTMPR4 and LTMPR1-4 were flanked by IS-elements (Figure 1 C). Seeking to explain the higher frequency of GDA mutations in the trimethoprim-adapted *lon*-knockout, we asked whether IS-transposases were sensitive to proteolysis by Lon. Indeed, heterologously expressed InsB and InsL transposases had greater stability in the *lon*-knockout (Figure S1), suggesting that their accumulation may contribute to higher GDA frequency in *E. coli* Δ*lon*. It would follow that a higher proportion of GDA mutants would be isolated from *E. coli Δlon* during adaptation in other environments as well. In agreement, GDA mutations were detected after evolution of *E. coli Δlon* in 2 of the 5 additional growth conditions tested (Figure 1 D, S2). In all cases duplications were flanked by IS elements (Figure S2). Further, duplicated genomic regions were similar among isolates derived from

the same growth condition, but different across environments (Figure 1 B, S2). On the other hand, wild type *E. coli* did not show GDA mutations in any of the other tested conditions (Figure 1 D).

The duplicated region in trimethoprim-resistant isolates included genes previously associated with resistance to trimethoprim, namely, *folA*, and *acrA* and *B*, which code for components of the AcrAB-TolC efflux pump (Figure 1 C). In one of the isolates, i.e. LTMPR1, the amplified stretch encompassed *folA* but not *acrAB* without affecting trimethoprim IC50 values (Figure 1 B, C). We therefore reasoned that duplication of *folA* was the main effector of adaptation to trimethoprim. Immunoblotting showed that DHFR protein was indeed overproduced in isolates LTMPR1 and WTMPR4 over and above the increase mediated by inactivation of *mgrB* alone (Figure 2 A).

In addition to *folA*, ~173 and ~531 other genes were also duplicated in LTMPR1 and WTMPR4, respectively. To understand the global transcriptional burden imposed by this amplification, we analysed the gene expression profiles of WTMPR4 and LTMPR1 using RNA-sequencing (Supplementary File 1). Since both isolates had loss-of-function mutations in *mgrB*, which cause large changes in global gene expression [51], we also analysed the transcriptomes of *E. coli ΔmgrB* and *E. coli ΔlonΔmgrB* strains for comparison (Supplementary File 1). Transcript level of *folA* was higher in LTMPR1 (2.3-fold) and WTMPR4 (5.7-fold) over respective controls (Supplementary File 1). Other duplicated genes were mildly overexpressed by ~1.7-fold on average in LTMPR1 (over *E. coli ΔlonΔmgrB*) and ~1.3-fold in WTMPR4 (over *E. coli ΔmgrB*) (Figure 2 B, Supplementary File 1). However, there was large variation in expression level among duplicated genes (Figure 2 B), indicating that increase in gene dosage did not necessarily result in enhanced expression levels. Instead, global gene expression profiles of WTMPR4 and LTMPR1 were correlated with *E. coli ΔmgrB* and *E. coli ΔlonΔmgrB* respectively (Figure 2 C, D) and most of the top 50 up/down-regulated genes in WTMPR4 and LTMPR1 could be attributed to loss of *mgrB* alone (Figure S3). In line with this result, strains with and without genomic amplifications had comparable fitness costs (~20 %) in drug-free media, similar in magnitude to the cost of *mgrB* inactivation (Figure 2 E). Further, the fitness cost was alleviated in growth media supplemented with high Mg^{2+}, i.e. by inhibiting the PhoQP pathway, confirming the source of this cost to be loss of *mgrB* rather than the GDA (Figure 2 F). Thus, we concluded that large

GDA mutations encompassing *folA* led to DHFR overexpression and represented an alternate mechanism for trimethoprim-resistance in *E. coli* with undetectable additional cost.

Maintenance of *folA* duplication is linked to antibiotic-generated demand for DHFR expression

We next investigated the evolutionary fate of duplicate *folA* gene copies by evolving the LTMPR1 isolate, which harboured the shortest duplication, in the absence of antibiotic. We also evolved this isolate at sustained drug pressure, i.e. trimethoprim at the same concentration at which it had originally emerged (300 ng/mL). Using three replicate populations in each condition, we asked how DHFR expression and trimethoprim IC50 values changed over ~252 generations of evolution and used genome sequencing to identify the underlying mutational basis. As controls, we performed similar evolution experiments with WTMPR4 (which harboured a GDA and a functional Lon protease) and WTMPR5 (which harboured a *folA* point mutation and a functional Lon protease) (Figure 3 A).

Evolution in trimethoprim-free media

In drug-free conditions, trimethoprim IC50 of LTMPR1 rapidly declined and a concomitant loss of colony formation on trimethoprim-supplemented plates was observed (Figure 3 B, C). Evolved WTMPR4, but not WTMPR5, showed similar loss of colony formation on trimethoprim indicating that resistance mediated by GDA mutations was less stable than point mutations regardless of the presence of Lon (Figure 3 C). For LTMPR1, these trends were explained by a decline in DHFR expression levels over evolution and reversion to a single *folA* copy (Figure 3 D, E, S4). Mutation in *mgrB* was, however, consistently maintained till the end of the experiment (Supplementary File 2). In agreement with genomics data, trimethoprim IC50s of evolving populations stabilized at a value similar to *E. coli Δlon ΔmgrB* (Figure 3 B). Interestingly, a second GDA mutation ("GDA-2", ~2.2 Mb to 3.3 Mb) emerged in these populations (Figure 3 E, S4). The GDA-2 mutation was apparently lost and regained multiple times during the experiment (Figure 3 E, S4), suggesting a dynamic maintenance in the population rather than fixation. Further, the region of the genome amplified in GDA-2 was similar to what we had observed when *E. coli Δlon* had evolved in

drug-free media (Figure S2, S4), indicating that GDA-2 was likely a fitness enhancing mutation unrelated to trimethoprim adaptation.

To further investigate the emergence of GDA-2 during evolution of LTMPR1 in drug-free media we sequenced 6 randomly picked colonies from each of the lineages at the final time point (Figure 3 F). All the sequenced isolates had a single copy of *folA* and a majority of isolates showed the presence of GDA-2 (Figure 3 F, Supplementary File 2). Interestingly, some of the isolates that did not harbour GDA-2 had evolved mutations in *rpoS* (Figure 3 F, Supplementary File 2). Loss-of-function mutations in RpoS are fixed at late time points during adaptation to trimethoprim and enhance the fitness of *E. coli* without conferring resistance [46, 51]. We also sequenced a few randomly picked isolates from WTMPR4 and WTMPR5 evolved lines (2 from each of the 3 replicate lineages in each case) to check for the presence of GDA-2 and *rpoS* mutations. As expected, clones derived from WTMPR4 had reverted to a single copy of *folA*, while WTMPR5 retained its ancestrally inherited *folA* mutation. None of them had evolved the GDA-2 mutation, while a few of the isolates from WTMPR4 and WTMPR5 lines had acquired *rpoS* mutations (Figure 3 G, H).

Consistent with earlier studies, and the role of mutations in *rpoS* in general fitness enhancement in *E. coli*, we found that an LTMPR1-derived isolate with mutant *rpoS* recovered fitness in drug-free media (Figure 4 A). Interestingly, an isolate with the GDA-2 mutation had also recovered fitness (Figure 4 A). Since GDA-2 occurred only in a *lon*-deficient background, and GDA-2 and *rpoS* mutations were mutually exclusive, we wondered whether the fitness enhancing effects of GDA-2 were also mediated by reduced RpoS activity. Lon is known to reduce the levels of RpoS during the exponential growth phase of *E. coli* [52]. Therefore, we reasoned that at least some targets of RpoS would be overexpressed in a *lon*-knockout. To test this prediction, we analysed the transcriptome of *E. coli Δlon* and specifically looked at the expression levels of 168 known targets of RpoS (Supplementary File 3). Of these, 41 genes were overexpressed in the *lon*-knockout by at least 1.5-fold over wild type, including genes induced in the stationary phase and by various stresses such as *poxB*, *aidB* and *cbpA* (Figure 4 B, C). Levels of *rpoS* mRNA were however unaffected in this strain. Next, we compared the expression levels of these 41 RpoS-targets in LTMPR1 and *E. coli Δlon ΔmgrB* with *E. coli Δlon* as the reference and found that most had similar expression levels in these three strains (Figure 4 C). Indeed a few of these genes showed

higher expression levels in LTMPR1 and *E. coli Δlon ΔmgrB*, which was not unexpected since loss of *mgrB* is known to stabilize RpoS [53]. We checked their expression in isolates derived from LTMPR1 which harboured either GDA-2 or *rpoS* mutations. In both isolates RpoS target genes were down-regulated indicating a re-setting of RpoS activity (Figure 4 C). Further, in the isolate harbouring GDA-2, *rpoS* mRNA levels were themselves lowered to 0.49x of *E. coli Δlon* (Supplementary File 3). Finally, deletion of *rpoS* from *E. coli Δlon* enhanced the fitness of the strain (Figure 4 D). Thus, evolution of LTMPR1 in antibiotic-free media was characterized by rapid reversal of GDA-mediated resistance and acquisition of mutations that enhanced fitness by reducing the levels of RpoS-transcribed genes (Figure 4 E).

Evolution at constant trimethoprim pressure

In the presence of trimethoprim (300 ng/mL), antibiotic resistance level and DHFR expression remained high over the course of evolution of LTMPR1 (Figure 5 A, B, C, D). WTMPR4 and WTMPR5 too remained resistant, and in the case of WTMPR4 an increased colony forming efficiency was observed at a late time point (Figure 5 C). These data suggested that under drug pressure high *folA* copy number was likely to be maintained. Indeed, population sequencing of LTMPR1 lineages showed that the ancestral GDA mutation was maintained until 196 generations of evolution in all three replicates (Figure 5 E, S5). By 252 generations, however, a majority of bacteria in Lines A and C had reverted to a single *folA* copy (Figure 5 E, S5) and had acquired a point mutation either in the *folA* gene (Line C) or its promoter (Lines A) (Figure 5 E, Supplementary File 4). Line B, on the other hand, retained high *folA* copy number, but with a reduction in the length of the genomic duplication (Figure 5 D, S5). Line B had also acquired a mutation in the *folA* promoter, however its frequency in the population was low (Figure 5 D, Supplementary File 4).

In agreement with population sequencing, none of the resistant isolates derived from LTMPR1 Line A at the final time point showed the ancestral GDA. Instead, they all harboured a $C_{-35}T$ mutation in the *folA* promoter that is known to up-regulate DHFR expression [51] (Figure 5 F, Supplementary File 4). Among Line B isolates, many had retained the ancestral duplication, without acquiring any additional mutations in *folA*. One of the isolates from this Line that had reverted to a single *folA* copy had gained the $A_{-64}T$ *folA*

promoter mutation (Figure 5 F, Supplementary File 4). In Line C too we found that isolates that had lost the ancestral GDA had evolved a point mutation in the *folA* promoter (+TG_{-30}) or a Trp_{30}Arg mutation in DHFR (Figure 5 F, Supplementary File 4). We found similar results in sequenced isolates from WTMPR4 lineages, i.e. loss of GDA and gain of a point mutation either in *folA* or its promoter (Figure 5 G). WTMPR5-derived isolates did not accumulate additional mutations in *folA* but did acquire mutations in *rpoS* (Figure 5 H). Interestingly, two isolates from LTMPR1 Lines B and C had evolved new GDA mutations encompassing *folA*. In one of the Line B isolates (Isolate 2) the ancestral GDA was replaced by 5x amplification of a shorter genomic region (~12 kb) encompassing *folA* (Figure S6). Both ends of this newly amplified region were flanked by IS1 elements (Figure S6). On the other hand, one of the Line C isolates (Isolate 5) had an expanded duplication beyond what was originally found in LTMPR1 (Figure S6). Also, both copies of *folA* in this isolate coded for DHFR-Trp_{30}Arg (Figure 5 E, S6), suggesting that the original GDA was lost and an extended GDA mutation was gained after fixation of the mutant *folA* allele.

Based on these results we concluded that even under drug pressure, GDA mutations were reversed in most bacteria. Notably, reversion to a single *folA* copy was slower than in drug-free media as it first required the fixation of point mutations at the *folA* locus. However, in a smaller sub-population, more than one *folA* gene copy was maintained. This maintenance was dynamic and likely involved repeated loss and gain of GDA mutations.

To further investigate the replacement of GDA by point mutants, we checked to see if point mutants could invade a population harbouring GDA mutations under drug pressure. We allowed LTMPR1-derived isolates with point mutations in *folA* (marked with Δ*lacZ::Cat*) to compete with a GDA-harbouring isolate from the same time point in evolution and asked how rapidly the former could invade a population of the latter. We performed these competitions with a contemporary GDA-harbouring isolate rather than the ancestor to nullify the effects of other media adaptations such as *rpoS* mutations that may have been acquired over evolution. Since for this assay we used point mutant strains genetically marked with Δ*lacZ::Cat* we verified the neutrality of the Δ*lacZ::Cat* marker by checking if the marked isolate could invade its unmarked parent. In the presence of trimethoprim (300 ng/mL) point mutants were enriched over the GDA mutant in co-culture by about 100-fold (Figure 6 A) recreating the results of our evolution experiment. A possible explanation for this finding may

be higher resistance level conferred by *folA* point mutations than GDA. Indeed, some isolates with *folA* point mutations had higher IC50 values compared to a contemporary GDA-harbouring isolate (Figure 6 B). However, regardless of IC50 values, all isolates also had similar doubling times in the presence of 300 ng/mL trimethoprim (Figure 6 B). Thus, though there were no large observable differences in growth rate between isolates, point mutants could indeed invade and outcompete GDA mutations, reflecting either subtle differences in fitness between them or their relative stabilities in culture.

Maintenance of high *folA* gene copy number under drug pressure is driven by unstable mutants of DHFR

Though sustained drug pressure did extend the lifetime of *folA* gene duplication, acquisition of trimethoprim-resistant point mutations facilitated reversion to a single gene copy. We therefore asked whether *folA* duplication could be maintained under conditions of increasing antibiotic pressure, i.e., trimethoprim concentrations starting at 1 μg/mL and ending at 10 μg/mL over 252 generations of evolution (Figure 7 A). Under these conditions, trimethoprim IC50 values increased sharply over the course of evolution eventually reaching ~40 to ~160-fold higher than ancestral LTMPR1 (Figure 7 B). Colony formation at high trimethoprim concentrations also increased at later generations of evolution (Figure 7 C). Similarly evolved WTMPR4 and WTMPR5 also showed enhanced colony formation at high trimethoprim concentrations (Figure 7 C).

Curiously, we observed that the 3 replicate lineages of LTMPR1 showed different trends in DHFR expression. Line A maintained a high level of DHFR protein for the duration of the experiment. On the other hand, Lines B and C showed high DHFR expression until 196 generations, after which it sharply declined (Figure 7 D). Genome sequencing revealed that all 3 lines convergently evolved the $Trp_{30}Arg$ mutation in DHFR early during evolution and had fixed this mutation by 196 generations. Until this time point, all 3 lines retained *folA* duplication (Figure 7 E, S7). Subsequently, Line A evolved an additional $C_{-35}T$ *folA* promoter mutation and had partially lost the ancestral GDA (Figure 7 E, S7, Supplementary File 5). In Lines B and C, a second missense mutation in DHFR, i.e. $Tyr_{151}Asp$, appeared between 196 and 252 generations (Figure 7 D, Supplementary File 5). Line B also showed an $Ile_{94}Leu$ mutation in DHFR at low frequency at generation 252 (Figure 7 D, Supplementary File 5).

Unexpectedly, despite acquiring additional mutations and lower DHFR expression level, LTMPR1 Lines B and C retained two copies of *folA* in contrast to Line A (Figure 7 E, S7, Supplementary File 5).

Individual resistant clones from each of the lineages provided greater insight into the genetic trajectories of evolution. In LTMPR1 Line A, 2 of the 6 sequenced isolates had lost gene duplication and acquired the $Trp_{30}Arg$ coding and $C_{-35}T$ promoter mutations in *folA* (Figure 7 F, Supplementary File 5). Interestingly, the other 4 sequenced isolates from Line A retained the ancestral GDA and harboured the same combination of *folA* point mutations, but only in 1 of the 2 gene copies (Figure 7 F, Supplementary File 5). Ten of the twelve isolates from LTMPR1 Lines B and C, on the other hand, retained *folA* duplication. In all ten isolates, both copies of *folA* had acquired $Trp_{30}Arg$ and $Tyr_{151}Asp$ mutations (Figure 7 F, Supplementary File 5). Unlike LTMPR1, all sequenced isolates of WTMPR4 at 252 generations of evolution had lost their GDA mutation and gained 2 *folA* mutations (Figure 7 G). WTMPR5 too had gained an additional *folA* mutation (Figure 7 H). Notably, none of the isolates from WTMPR4 and WTMPR5 lineages had the $Trp_{30}Arg$ and $Tyr_{151}Asp$ combination of mutations that occurred repeatedly in LTMPR1 lineages.

We next sought to understand why some LTMPR1 bacteria had retained high *folA* copy number despite acquiring multiple point mutations in the drug target. To address this question, we first asked what the advantage of *folA* duplication was in the two different mutational trajectories. Isolates from Line A had similar IC50 values for trimethoprim regardless of whether they had one or two copies of *folA* (Figure 8 A). On the contrary, isolates from Line B that harboured only one copy of mutant *folA* (coding for DHFR-$Trp_{30}Arg$-$Tyr_{151}Asp$) displayed ~3.5-fold lower IC50 than those that harboured two copies (Figure 8 A). Immunoblotting showed that isolates from Line A had similar expression level of DHFR (Figure 8 B). However, DHFR protein level was significantly compromised by the $Trp_{30}Arg$-$Tyr_{151}Asp$ mutation combination and double the number of *folA* gene copies restored this defect (Figure 8 B). Thus, higher gene dosage of DHFR was dispensable once a promoter up-regulatory mutation was fixed. On the other hand, bacteria with alleles of DHFR harbouring multiple missense mutations continued to benefit from overexpression due to *folA* duplication.

Mutations at position Trp_{30} in DHFR render the protein resistant to inhibition by trimethoprim but are accompanied by lower *in vivo* proteolytic stability due to perturbation of the hydrophobic core of the protein [54]. This affects the expression level of DHFR, resulting in lower cellular abundance that limits the resistance level conferred by mutations at Trp_{30} [49, 54]. Interestingly, residue Tyr_{151} lies in the same hydrophobic pocket as Trp_{30} (Figure 8 C). To test the effect of combination of mutations at these sites on trimethoprim resistance and DHFR expression, we first heterologously expressed wild type DHFR, Trp_{30}Arg or Tyr_{151}Asp single mutants and the Trp_{30}Arg-Tyr_{151}Asp double mutant in wild type *E. coli* (Figure 8 D). As expected, DHFR Trp_{30}Arg expression increased the IC50 of trimethoprim. Surprisingly, neither DHFR Tyr_{151}Asp, nor the Trp_{30}Arg-Tyr_{151}Asp double mutant conferred resistance to the antibiotic (Figure 8 E). Indeed, both mutants conferred lower IC50 values than wild type DHFR. This effect was due to loss of hydrophobicity at the Tyr_{151} position, as DHFR-Tyr_{151}Leu and DHFR-Tyr_{151}Phe conferred similar IC50 values as wild type DHFR (Figure 8 E). Since the Trp_{30}Arg-Tyr_{151}Asp mutation combination had evolved in a *lon*-deficient background, we next expressed these mutants in *E. coli ⊿lon* (Figure 8 D). Here too, DHFR Trp_{30}Arg conferred trimethoprim resistance, while the Tyr_{151}Asp mutant did not (Figure 8 E). Strikingly, the combination of Trp_{30}Arg and Tyr_{151}Asp conferred very high-level trimethoprim resistance (Figure 8 E) demonstrating synergistic epistasis that was contingent on the absence of Lon protease. The levels of expression of DHFR mutants in wild type and *lon*-deficient bacteria mirrored trimethoprim resistance levels indicating that clearance by Lon altered their genotype-phenotype relationship (Figure 8 F). Based on these data, we concluded that the action of proteostatic machinery generated additional demand for expression by clearing mutant-DHFR proteins. This additional demand drove the maintenance of *folA* duplications in bacteria evolving at very high drug pressures.

Discussion

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

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

earlier work by Nicoloff and co-workers who found that large IS-element flanked genomic duplications in tetracycline-resistant *E. coli* were commonly preceded by spontaneous mutations that inactivated *lon* [62, 63]. Further, the role of Lon protease in regulating IS-element mobility by a similar mechanism has been previously demonstrated for IS1 and IS903 [64, 65]. We note however that other mechanisms by which Lon may modulate the frequency of IS-mediated GDAs in *E. coli* cannot be ruled out. For instance, Lon protease homologs from several organisms directly bind DNA [66-70], though the role of this binding in DNA transactions is unclear. Similarly, Lon is a regulator of the recombination protein, RecA, by its action on LexA [71] and in *Salmonella* RecA stimulated the rate of gene duplications mediated by IS3 [23]. The second impact of protein quality control on GDA was the action of proteases on trimethoprim-resistant mutants of DHFR, which promoted gene duplication by generating additional demand for expression (Figure 9). Importantly, which of these two effects dominated was determined by a combination of environment, i.e. drug concentrations, and other mutations acquired at the *folA* locus during evolution. In particular, mutations in the *folA* promoter could alleviate expression demand and hence reduce selection for high gene copy number.

The interplay between GDA and point mutations during evolution that we report here agrees in many respects with a recent study by Tomanek and Guet [61]. They use an elegantly designed artificial system [33, 61] in which expression of the *galK* gene in *E. coli* was driven by a very low activity promoter. This strain was then allowed to evolve in different concentrations of galactose. Like the present study, they found that the activity of IS-elements biased evolutionary trajectories in favor of duplications. They also showed that co-occurrence of duplications and point mutations was selected only at high expression demand, in agreement with our findings. Finally, they argued that gene duplication events may constrain evolution by point mutations. Our study corroborates these results at the short term in that the *lon*-knockout which showed a higher propensity for GDA mutations did not evolve point mutations in *folA* at early time points of evolution. However, given the relative low stability of large GDAs, these were inevitably lost over longer duration of evolution and replaced with point mutations. Thus, it is likely that the different mutational profiles of GDA-biased populations may be relevant only at short durations of evolution. It is important to note that though we have focused on the effects of mutations at the *folA* locus, which are directly under 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 up the possibility of alternative explanations for how 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. Moreover, our findings relating proteostasis in general and Lon in particular, to adaptive mutations add to a growing body of literature that define a causal link between protein quality control and mutational paths of evolution. For DHFR itself, decreasing the stringency of protein quality control changes the fitness landscape of single mutations and their combinations, altering the mutational repertoires accessible during adaptive evolution [49, 50, 72, 73]. Similar effects of protein homeostatic machinery on mutant fitness are reported from a diverse range of proteins that influence phenomena like viral evolution [74] and the fate of horizontally acquired genes [75-77]. Indeed, a central role for protein quality control machinery in molecular and organismal evolution is reflected in conserved signatures of the proteostatic network across the different kingdoms of life [78]. These links may be important to consider for a new wave of potential antibacterials that target proteostasis [79-81]. It appears critical to ensure that the impact of changes in evolvability are considered when designing such strategies for therapeutic intervention.

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 2.

Table 2. 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]
7.	*E. coli ΔpitA*	Isogenic knockout of *pitA* gene (*ΔpitA::Kan*)	This study.
Strains/Isolates derived from short-term adaptive laboratory evolution			

1.	WTMPR1, WTMPR2, WTMPR3, WTMPR4 and WTMPR5	Trimethoprim-resistant derivatives of *E. coli* wild-type 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	[46] Note: These 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 $Tyr_{151}Asp$ mutation in frame-with an N-terminal hexa-histidine tag.	This study
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.

Long-term evolution at different trimethoprim pressures

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

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

Measuring trimethoprim resistance

Trimethoprim IC50

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

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

where,

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

Values of IC50 were estimated from the above curve fit and averaged across 3 independent measurements for every strain.

Colony forming efficiency

To calculate the colony forming efficiency of evolving populations at different trimethoprim concentrations, frozen glycerol stocks were first revived in 1 mL LB for 14-15 hours. Saturated cultures were serially diluted (10-fold). Diluted cultures were then spotted (10 mL) on LB agar plates containing 0, 1, 5 and 10 μg/mL of trimethoprim. Spots were allowed to dry completely and the plates were incubated at 37 °C for ~18 hours. Colonies were counted from the least dense spot and multiplied by the appropriate dilution factor to estimate colony forming units per mL (CFUs/mL) of culture.

Growth curves

Appropriate frozen stocks were revived overnight in drug-free LB (1-3 mL) and 1μl of the saturated culture was added to 200 μL of media supplemented with trimethoprim (300 ng/ml) in the wells of a sterile uncoated 96-well plate. Optical density at 600 nm was measured during growth at 37 °C at 15-minute intervals over a 24-hour period on a Perkin-Elmer Ensight Multimode Plate Reader. Doubling time (DT) (in mins) was estimated from log phase, using the following formula:

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

Where OD1 and OD2 represent optical densities at the onset and end of log phase respectively, and T1 and T2 represent time elapsed (in mins), corresponding to OD1 and OD2.

Genome sequencing, identification of mutations and detection of GDA mutations

Genomic DNA was extracted from late log-phase cultures of appropriate strain/resistant isolates/evolving populations using phenol:chloroform:isoamyl alcohol (25:24:1) extraction as previously described [82]. Extracted genomic DNA was then cleaned-up using HiGenoMB

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. All sequenced genomes had an average coverage depth of at least 100x. Variant calling and prediction of new junctions in the genome to identify large structural mutations was done using Breseq [83] 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 [83]. 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:

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

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.

Estimation of DHFR expression by immunoblotting

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

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

Gene expression analysis using RNA-sequencing

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

DGE analyses were $\log_2$(fold change) values for each gene and P-values for statistical significance.

Cloning and site-directed mutagenesis

Cloning of InsB and InsL transposases

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

Site-directed mutagenesis for DHFR

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

folA_$Tyr_{151}Asp$_fwd_*Bgl*II:

5'-CTCTCACAGCGATTGCTTTGAGATCTTGGAGCGGCGGGGCC-3'

folA_$Tyr_{151}Asp$_rev_*Bgl*II:

5'-CGCCGCTCCAAGATCTCAAAGCAATCGCTGTGAGAGTTCTGCG-3'

folA_$Tyr_{151}Leu$_fwd_*Bgl*II:

5'-TCTCACAGCTTATGCTTTGAGATCTTGGAGCGGCGGGGCC-3'

folA_$Tyr_{151}Leu$_rev_*Bgl*II:

5'- CGCCGCTCCAAGATCTCAAAGCATAAGCTGTGAGAGTTCTGCG-3'

folA_$Tyr_{151}Phe$_fwd_*Bgl*II:

5'- CTCACAGCTTTTGCTTTGAGATCTTGGAGCGGCGGGGCC-3'

folA_$Tyr_{151}Phe$_rev_*Bgl*II:

5’- GCCGCTCCAAGATCTCAAAGCAAAAGCTGTGAGAGTTCTGCG-3’

PCR products were digested with *Dpn*I overnight, transformed into chemically competent *E. coli* DH10B and selected on LB agar plates containing ampicillin. Plasmids habouring the desired mutations were screened by restriction digestion and confirmed by sequencing. Sequenced clones were transformed into *E. coli* wild type or *Δlon* for further analyses.

In-cell stability of InsB and InsL

A chloramphenicol-chase assay was used to assess the in-cell stability of heterologously expressed InsB and InsL transposases as described in Matange (2020) [49] with a few modifications. Overnight cultures of *E. coli* wild type or *Δlon* harbouring plasmids pPRO-insB or pPRO-insL were diluted 1% in 5 mL LB supplemented with ampicillin and IPTG (0.1mM). After 4 hours of growth at 37 °C, 1 mL of culture was pelleted. The cell pellet was lysed in 100 μL of 1x Laemmli sample loading buffer by heating at 95°C for 10-20 mins. Chloramphenicol was added to the remaining culture at a final concentration of 25 μg/mL to stop protein synthesis and the culture was placed back in the incubator. After 15, 45, and 90 minutes 1 mL aliquots of the culture were pelleted and lysed. Lysates were subjected to SDS-PAGE followed by immunoblotting to detect the levels of InsB and InsL. Anti-Hexahis monoclonal antibody (Invitrogen) was used as the primary antibody for InsL. Anti-Hexahis polyclonal antibody (Invitrogen) was used as the primary antibody for InsB. Band intensities were calculated in ImageJ and normalised to control (set to 1) to calculate fraction of protein remaining after chloramphenicol treatment.

Acknowledgments

We acknowledge Dr. Manjula Reddy for providing us with the anti-FtsZ polyclonal antibody. We acknowledge Bruhad Dave for his valuable inputs during discussions at early stages of this project and Prof. Satyajit Rath, Prof. Sandhya Visweswariah and Dr. Girish Deshpande for their comments on the manuscript. Funding for this work was provided by DBT/Wellcome Trust India Alliance. NM is a recipient of the India Alliance Intermediate Fellowship. CJ is a recipient of a Senior Research Fellowship from the Council of Scientific and Industrial Research (CSIR), Govt. of India.

Supplementary Files.

Supplementary File 1. Gene expression changes in LTMPR1 and WTMPR4 isolates

Supplementary File 2. Genomic changes associated with evolution of LTMPR1 in drug-free media

Supplementary File 3. Gene expression changes in isolates derived from LTMPR1 evolution in drug-free media containing mutation in *rpoS* or GDA-2

Supplementary File 4. Genomic changes associated with evolution of LTMPR1 in trimethoprim (300 ng/mL)

Supplementary File 5. Genomic changes associated with evolution of LTMPR1 in increasing trimethoprim pressure

Supplementary File 6. Genome sequencing of isolates derived from WTMPR4 evolution at different trimethoprim pressures

Supplementary File 7. Genome sequencing of isolates derived from WTMPR5 evolution at different trimethoprim pressures

Figures

Figure 1. Adaptation to antibiotics by large genomic duplications is favoured in *lon*-deficient *E. coli*. A. Schematic of the experimental pipeline used to compare the mutation repertoire of adaptation to trimethoprim in wild type or *lon*-deficient *E. coli*. **B.** Coverage depth plots and mutations at the *folA* and *mgrB* loci of 5 independently evolved trimethoprim-resistant isolates derived from wild type (WTMPR1-5) or *Δlon E. coli* (LTMPR1-5). Coverage depth plots show the number of reads from Illumina short-read 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. Mutations in *mgrB* and *folA,* and trimethoprim IC50 values (mean ± S.D. from 3 independent measurements) are provided for each isolate below the appropriate coverage depth plot. **C.** Cartoon (not to scale) showing the duplicated genomic stretch and flanking IS-elements in trimethoprim-resistant isolates. Blue arrows represent IS1, while red arrows represent IS186. The positions of *folA* and *acrAB* genes, implicated in trimethoprim 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. **D.** 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. Empty circles represent replicates in which no resistance was detected (dotted perimeter) or where the whole population or resistant isolates were sequenced but no GDA mutation was detected (solid perimeter). Red filled 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 was detected by immunoblotting using an anti-DHFR polyclonal antibody. FtsZ was used as loading control. A representative immunoblot from 3 biological replicates is shown. Quantitation was performed by calculating band intensities using image analysis. Fold change values of DHFR levels over controls (wild type or *E. coli Δlon*, set to 1) is shown as mean ± S.D. from triplicate measurements. Alleles of *mgrB* and *folA* in WTMPR4 and LTMPR1 are

shown diagrammatically. **B.** Expression level of duplicated genes in LTMPR1 and WTMPR4, determined by RNA-sequencing, expressed as fold over *E. coli ΔlonΔmgrB* and *E. coli ΔmgrB* respectively. Each point represents fold change value for a single gene. Mean of the scatter is shown as a black line, and its value is provided along with standard deviation above the plot. **C, D.** Correlation between expression levels of genes in LTMPR1 (C) and WTMPR4 (D) with *E. coli ΔlonΔmgrB* and *E. coli ΔmgrB* respectively. Gene expression levels were estimated using RNA-sequencing and fold changes were calculated using wild type *E. coli* as reference for WTMPR4 and *E. coli ΔmgrB*, and *E. coli Δlon* as reference for LTMPR1 and *E. coli ΔlonΔmgrB*. Each point represents $\log_2$(fold change) value for a single gene. The resulting scatter was fit to a simple linear regression and the obtained R^2 values are provided. **E, F.** Relative fitness (w) of indicated *E. coli* strains calculated using competition experiments with *E. coli ΔlacZ* as reference in the absence (E) or presence (F) of 10 mM $MgSO_4$. Presence of high concentration of Mg^{2+} alleviates the fitness cost of *mgrB* mutations by inhibiting the PhoQP system [51]. For strains derived from *E. coli Δlon, E. coli ΔlonΔlacZ* was used as reference. Neutrality of *ΔlacZ* was established by competition with the unmarked ancestral strains, i.e. wild type (WT) and *E. coli Δlon*. No change in relative fitness compared to the reference strain (w = 1) is indicated by a dotted line. Mean ± S.D. from three independent measurements is plotted. Statistical significance was tested using an unpaired t-test and a p-value of < 0.5 was considered statistically significantly different (*). A p-value ≥ 0.5 was considered not significantly different (ns).

Figure 3. Evolutionary fate of *folA*-duplication in the absence of drug pressure. A. Schematic of the experimental pipeline used to investigate the impact of evolution in drug-free media on resistance level, *folA* copy number and expression level of different trimethoprim resistant *E. coli* populations. **B.** 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. **C.** Colony formation of LTMPR1, WTMPR4 and WTMPR5 evolved in antibiotic-free medium at different generations. 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. **D.** DHFR expression during evolution in the absence of trimethoprim

in LTMPR1 lineages A, B and C, 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. **E.** Copy number of the ancestral GDA encompassing *folA* (GDA(*folA*)) and GDA-2 at different time points of evolution in the absence of trimethoprim are plotted. For GDA(*folA*), 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. **F.** Point mutations in *folA*, *rpoS* and *mgrB*, and the "GDA-2" genomic duplication in 6 randomly picked colonies from each of the LTMPR1 lines at 252 generations of evolution are shown schematically using the appropriate symbols. Asterix (*/**) marks the genotypes that were carried forward for further analyses. **G, H.** Point mutations in *folA*, *mgrB* and *rpoS* in isolates derived from 252 generations of evolution of WTMPR4 (**G**) and WTMPR5 (**H**) in no antibiotic. From each of the evolving lines 2 random isolates were picked for genome sequencing. Various point mutants at the 3 gene loci of interest are represented by appropriate symbols as shown in the legend..

Figure 4. Lower RpoS activity leads to fitness enhancement of LTMPR1 in drug-free media. A. Relative fitness (w) of isolates with the indicated genotypes derived from LTMPR1 evolution in antibiotic-free media (see also Figure 3) calculated using a competitive growth assay using *E. coli ΔlonΔlacZ* as the reference strain. Mean ± S.D. from three independent measurements is plotted. No change in relative fitness compared to the reference (w = 1) is shown as a dotted line for reference. Statistical significance was tested using an unpaired t-test. A p-value of < 0.05 was considered as a statistically significant difference (*). **B.** Expression level of 168 known targets of RpoS in *E. coli Δlon* compared to wild type expressed as $\log_2$(fold change) values determined by RNA sequencing. Dotted lines at 1.5-fold higher and lower than wild type represent cut-offs used to identify overexpressed and under-expressed genes. The pie chart shows the fraction of RpoS targets that were overexpressed in *E. coli Δlon* by at least 1.5-fold. **C.** Expression levels of 41 de-regulated RpoS targets in *E. coli Δlon* compared to wild type are shown as a heat map in the first vertical. The expression levels of these genes in LTMPR1, *E. coli ΔlonΔmgrB* or LTMPR1-derived isolates compared to *E. coli Δlon* are provided in the subsequent verticals. **D.** Relative fitness of *E. coli ΔlonΔrpoS* in antibiotic-free media calculated using a competitive

growth assay using *E. coli ΔlonΔlacZ* as the reference strain. Neutrality of the *ΔlacZ* genetic marker was verified by competition between *E. coli Δlon* and *E. coli ΔlonΔlacZ*. Mean ± S.D. from three independent measurements is plotted. No change in relative fitness compared to the reference (w = 1) is shown as a dotted line for reference. Statistical significance was tested using an unpaired t-test. A p-value of < 0.5 was considered as a statistically significant difference (*) **E.** Model for the effects of Lon deficiency on RpoS levels and bacterial fitness. The roles of mutations within RpoS or the GDA-2 mutation in compensating for the defects of RpoS overproduction are shown.

Figure 5. Evolutionary fate of *folA*-encompassing GDA in LTMPR1 at sustained trimethoprim pressure. A. Schematic of the experimental pipeline used to investigate the impact of evolution in constant drug-pressure on resistance level, *folA* copy number, point mutations in *folA* and expression level of different trimethoprim resistant *E. coli* populations. **B.** 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. **C.** 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. **D.** 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. **E.** 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 plotted on the right y-axis. **F.** 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 #). **G, H.** Point mutations in *folA*, *mgrB* and *rpoS* in isolates derived from 252 generations of evolution of WTMPR4 (**G**) and WTMPR5 (**H**) 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 in the legend.

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 (i.e. alleles at *mgrB*, *folA* and *rpoS* loci) used are shown diagrammatically (see also Figure 5, Supplementary File 4). The initial mixing ratio was 1000:1 in favour of the GDA mutant. Point mutants were marked genetically with *ΔlacZ::Cat*. Neutrality of the *ΔlacZ::Cat* marker was verified by competing marked and unmarked point mutants (lower panel). Percentage of the population constituted by the marked point mutants over ~105 generations of serial transfer in media supplemented with trimethoprim (300 ng/mL) is plotted. Mean ± S.D. from 3 replicates are plotted and traces for individual mutants are appropriately colored. **B.** Trimethoprim IC50 values of the LTMPR1 ancestor or evolved isolates from 252 generations in trimethoprim with indicated genotypes (see also Figure 5, Supplementary File 4) are plotted as bars. Mean ± S.D. from three independent measurements are plotted. Statistical significance was tested using an unpaired t-test and LTMPR1 was used as the reference. A p-value < 0.05 was considered significantly different (*) while ≥ 0.5 was considered not statistically significant (ns). Doubling times relative to LTMPR1 (set to 1) in the presence of 300 ng/mL of trimethoprim are provided above the graph as mean ± S.D. from three independent measurements.

Figure 7. Evolutionary fate of *folA*-encompassing GDA in LTMPR1 at increasing trimethoprim pressure. A. Schematic of the experimental pipeline used to investigate the impact of evolution in increasing drug-pressure on resistance level, *folA* copy number, point mutations in *folA* and expression level of different trimethoprim resistant *E. coli* populations. **B.** 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. **C.** Colony formation of LTMPR1, WTMPR4 and WTMPR5 evolved in increasing 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. **D.** DHFR expression in LTMPR1 lineages A, B and C evolved in increasing 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. **E.** Copy number of *folA* (GDA(*folA*)) at different time points of LTMPR1 evolution in increasing 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 LTMPR1 populations at each of the time points is shown on the right y-axis. **F.** 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 schematically. **G, H.** Point mutations in *folA*, *mgrB* and *rpoS* in isolates derived from 252 generations of evolution of WTMPR4 (G) and WTMPR5 (H) in increasing 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 in the legend.

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. Statistical significance was tested using an unpaired t-test. A p-value of <0.05 was considered statistically significant (*). A p-value > 0.05 was considered not statistically significant (ns). **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.** Schematic of the experiment used to test trimethoprim resistance and expression level of various DHFR mutants in *E. coli* wild type (WT) or *Δlon*. **E.** 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. Statistical significance was tested using an unpaired t-test. Comparisons were between mutant and wild type DHFR, unless otherwise indicated. A p-value of <0.05 was considered statistically significant (*). A p-value > 0.05 was considered not statistically significant (ns). **F.** Expression level of plasmid-borne DHFR (wt) or its mutants in wild type or *Δlon E. coli* 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. Expression of mutants was normalized to wild type DHFR (set to 1). Mean of three independent measurements is shown.

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

Supplementary Figures

Figure S1. A. Loss of *mgrB* or *pitA* genes is beneficial in trimethoprim. Colony formation of *E. coli* wild type, *ΔmgrB* and *ΔpitA* on media supplemented with trimethoprim at the indicated concentrations. Mean ± S.D. from three independent replicates are plotted.

Statistical significance was tested using a t-test and p-value < 0.05 is indicated with an asterisk (*). **B.** Stability of heterologously expressed InsB and InsL transposases in wild type or *Δlon E. coli* measured using a chloramphenicol (CMP) chase assay. Levels of the plasmid-expressed transposase were measured using immunoblotting with an anti-His monoclonal/polyclonal antibody in lysates prepared from cells at indicated time points following treatment with 50 μg/mL of CMP. A representative image from three independent experiments is shown. Band intensities were quantitated by image analyses and normalized to 0' (set to 1). The fraction of protein remaining is plotted in blue for wild type and red for *E. coli Δlon*. Values from 3 independent experiments are plotted as mean ± S.D. Statistical significance was tested using an unpaired t-test. A p-value of <0.05 was considered statistically significant (*). A p-value > 0.05 was considered not statistically significant (ns).

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

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

Figure S4. Coverage depth plots for population sequencing at 105, 196 and 252 generations of the 3 lineages of LTMPR1 (A, B, C) evolving in trimethoprim-free media. 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.

Figure S5. Coverage depth plots for population sequencing at 105, 196 and 252 generations of the 3 lineages of LTMPR1 (A, B, C) evolving in trimethoprim-supplemented (300 ng/mL) media. 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.

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

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.

References

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

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

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

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

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

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

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

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

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

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

21. von Rozycki T, Nies DH. *Cupriavidus metallidurans*: evolution of a metal-resistant bacterium. Antonie Van Leeuwenhoek. 2009;96(2):115-39. Epub 20081001. doi: 10.1007/s10482-008-9284-5. PubMed PMID: 18830684.

22. Bi X, Liu LF. RecA-independent and recA-dependent intramolecular plasmid recombination: Differential homology requirement and distance effect. J Mol Biol. 1994;235(2):414-23. doi: 10.1006/jmbi.1994.1002. PubMed PMID: 8289271.

23. Reams AB, Kofoid E, Kugelberg E, Roth JR. Multiple Pathways of Duplication Formation with and Without Recombination (RecA) in *Salmonella enterica*. Genetics. 2012;192(2):397-415. doi: 10.1534/genetics.112.142570. PubMed PMID: WOS:000309547400006.

24. Reams AB, Kofoid E, Savageau M, Roth JR. Duplication frequency in a population of *Salmonella enterica* rapidly approaches steady state with or without recombination. Genetics. 2010;184(4):1077-94. Epub 20100118. doi: 10.1534/genetics.109.111963. PubMed PMID: 20083614; PubMed Central PMCID: PMCPMC2865909.

25. Reams AB, Neidle EL. Genome plasticity in *Acinetobacter*: new degradative capabilities acquired by the spontaneous amplification of large chromosomal segments. Molecular Microbiology. 2003;47(5):1291-304. doi: DOI 10.1046/j.1365-2958.2003.03342.x. PubMed PMID: WOS:000181056400009.

26. Min YH, Kwon AR, Yoon JM, Yoon EJ, Shim MJ, Choi EC. Molecular analysis of constitutive mutations in ermB and ermA selected in vitro from inducibly MLSB-resistant enterococci. Arch Pharm Res. 2008;31(3):377-80. Epub 20080413. doi: 10.1007/s12272-001-1167-8. PubMed PMID: 18409053.

27. Sun S, Berg OG, Roth JR, Andersson DI. Contribution of gene amplification to evolution of increased antibiotic resistance in *Salmonella typhimurium*. Genetics. 2009;182(4):1183-95. Epub 20090527. doi: 10.1534/genetics.109.103028. PubMed PMID: 19474201; PubMed Central PMCID: PMCPMC2728858.

28. Vinella D, Cashel M, D'Ari R. Selected amplification of the cell division genes ftsQ-ftsA-ftsZ in *Escherichia coli*. Genetics. 2000;156(4):1483-92. doi: 10.1093/genetics/156.4.1483. PubMed PMID: 11102351; PubMed Central PMCID: PMCPMC1461353.

29. Nilsson AI, Zorzet A, Kanth A, Dahlstrom S, Berg OG, Andersson DI. Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. Proc Natl Acad Sci U S A. 2006;103(18):6976-81. Epub 20060424. doi: 10.1073/pnas.0602171103. PubMed PMID: 16636273; PubMed Central PMCID: PMCPMC1459004.

30. Andersson DI, Nicoloff H, Hjort K. Mechanisms and clinical relevance of bacterial heteroresistance. Nature Reviews Microbiology. 2019;17(8):479-96. doi: 10.1038/s41579-019-0218-1. PubMed PMID: WOS:000475479100007.

31. Nicoloff H, Hjort K, Levin BR, Andersson DI. The high prevalence of antibiotic heteroresistance in pathogenic bacteria is mainly caused by gene amplification. Nat Microbiol. 2019;4(3):504-14. doi: 10.1038/s41564-018-0342-0. PubMed PMID: WOS:000459201400019.

32. Sun S, Ke R, Hughes D, Nilsson M, Andersson DI. Genome-wide detection of spontaneous chromosomal rearrangements in bacteria. Plos One. 2012;7(8):e42639. Epub 20120803. doi: 10.1371/journal.pone.0042639. PubMed PMID: 22880062; PubMed Central PMCID: PMCPMC3411829.

33. Tomanek I, Grah R, Lagator M, Andersson AMC, Bollback JP, Tkacik G, et al. Gene amplification as a form of population-level gene expression regulation. Nat Ecol Evol. 2020;4(4):612-25. doi: 10.1038/s41559-020-1132-7. PubMed PMID: WOS:000519008300005.

34. Ahmad SI, Kirk SH, Eisenstark A. Thymine metabolism and thymineless death in prokaryotes and eukaryotes. Annu Rev Microbiol. 1998;52:591-625. doi: 10.1146/annurev.micro.52.1.591. PubMed PMID: 9891809.

35. Gangjee A, Jain HD, Kurup S. Recent advances in classical and non-classical Antifolates as antitumor and antiopportunistic infection agents: part I. Anti-Cancer Agent Me. 2007;7(5):524-42. doi: Doi 10.2174/187152007781668724. PubMed PMID: WOS:000249469700005.

36. Gangjee A, Jain HD, Kurup S. Recent advances in classical and non-classical antifolates as antitumor and antiopportunistic infection agents: Part II. Anti-Cancer Agent Me. 2008;8(2):205-31. doi: Doi 10.2174/187152008783497064. PubMed PMID: WOS:000255810200007.

37. Huennekens FM. The methotrexate story: a paradigm for development of cancer chemotherapeutic agents. Adv Enzyme Regul. 1994;34:397-419. doi: 10.1016/0065-2571(94)90025-6. PubMed PMID: 7942284.

38. Plowe CV. Malaria chemoprevention and drug resistance: a review of the literature and policy implications. Malar J. 2022;21(1):104. Epub 20220324. doi: 10.1186/s12936-022-04115-8. PubMed PMID: 35331231; PubMed Central PMCID: PMCPMC8943514.

39. Gleckman R, Blagg N, Joubert DW. Trimethoprim: mechanisms of action, antimicrobial activity, bacterial resistance, pharmacokinetics, adverse reactions, and

therapeutic indications. Pharmacotherapy. 1981;1(1):14-20. doi: 10.1002/j.1875-9114.1981.tb03548.x. PubMed PMID: 6985448.

40. Brochet M, Couve E, Zouine M, Poyart C, Glaser P. A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in Streptococcus agalactiae. J Bacteriol. 2008;190(2):672-80. Epub 20071116. doi: 10.1128/JB.01357-07. PubMed PMID: 18024520; PubMed Central PMCID: PMCPMC2223700.

41. Goker E, Waltham M, Kheradpour A, Trippett T, Mazumdar M, Elisseyeff Y, et al. Amplification of the Dihydrofolate-Reductase Gene Is a Mechanism of Acquired-Resistance to Methotrexate in Patients with Acute Lymphoblastic-Leukemia and Is Correlated with P53 Gene-Mutations. Blood. 1995;86(2):677-84. doi: DOI 10.1182/blood.V86.2.677.bloodjournal862677. PubMed PMID: WOS:A1995RH80100030.

42. Inselburg J, Bzik DJ, Horii T. Pyrimethamine resistant Plasmodium falciparum: overproduction of dihydrofolate reductase by a gene duplication. Mol Biochem Parasitol. 1987;26(1-2):121-34. doi: 10.1016/0166-6851(87)90136-8. PubMed PMID: 3323903.

43. Morales C, Garcia MJ, Ribas M, Miro R, Munoz M, Caldas C, et al. Dihydrofolate reductase amplification and sensitization to methotrexate of methotrexate-resistant colon cancer cells. Mol Cancer Ther. 2009;8(2):424-32. Epub 20090203. doi: 10.1158/1535-7163.MCT-08-0759. PubMed PMID: 19190117.

44. Baym M, Lieberman TD, Kelsic ED, Chait R, Gross R, Yelin I, et al. Spatiotemporal microbial evolution on antibiotic landscapes. Science. 2016;353(6304):1147-51. doi: 10.1126/science.aag0822. PubMed PMID: 27609891; PubMed Central PMCID: PMCPMC5534434.

45. Palmer AC, Toprak E, Baym M, Kim S, Veres A, Bershtein S, et al. Delayed commitment to evolutionary fate in antibiotic resistance fitness landscapes. Nat Commun. 2015;6:7385. Epub 20150610. doi: 10.1038/ncomms8385. PubMed PMID: 26060115; PubMed Central PMCID: PMCPMC4548896.

46. Patel V, Matange N. Adaptation and compensation in a bacterial gene regulatory network evolving under antibiotic selection. Elife. 2021;10. Epub 20210930. doi: 10.7554/eLife.70931. PubMed PMID: 34591012; PubMed Central PMCID: PMCPMC8483737.

47. Watson M, Liu JW, Ollis D. Directed evolution of trimethoprim resistance in *Escherichia coli*. Febs J. 2007;274(10):2661-71. Epub 20070419. doi: 10.1111/j.1742-4658.2007.05801.x. PubMed PMID: 17451440.

48. Toprak E, Veres A, Michel JB, Chait R, Hartl DL, Kishony R. Evolutionary paths to antibiotic resistance under dynamically sustained drug selection. Nat Genet. 2011;44(1):101-5. Epub 20111218. doi: 10.1038/ng.1034. PubMed PMID: 22179135; PubMed Central PMCID: PMCPMC3534735.

49. Matange N. Highly Contingent Phenotypes of Lon Protease Deficiency in *Escherichia coli* upon Antibiotic Challenge. J Bacteriol. 2020;202(3). Epub 20200115. doi: 10.1128/JB.00561-19. PubMed PMID: 31740490; PubMed Central PMCID: PMCPMC6964741.

50. Rodrigues JV, Bershtein S, Li A, Lozovsky ER, Hartl DL, Shakhnovich EI. Biophysical principles predict fitness landscapes of drug resistance. Proc Natl Acad Sci U S A. 2016;113(11):E1470-8. Epub 20160229. doi: 10.1073/pnas.1601441113. PubMed PMID: 26929328; PubMed Central PMCID: PMCPMC4801265.

51. Vinchhi R, Yelpure C, Balachandran M, Matange N. Pervasive gene deregulation underlies adaptation and maladaptation in trimethoprim-resistant *E. coli*. mBio. 2023:e0211923. Epub 20231130. doi: 10.1128/mbio.02119-23. PubMed PMID: 38032208.

52. Ranquet C, Gottesman S. Translational regulation of the Escherichia coli stress factor RpoS: a role for SsrA and Lon. J Bacteriol. 2007;189(13):4872-9. Epub 20070420. doi: 10.1128/JB.01838-06. PubMed PMID: 17449615; PubMed Central PMCID: PMCPMC1913435.

53. Xu J, Li T, Gao Y, Deng J, Gu J. MgrB affects the acid stress response of Escherichia coli by modulating the expression of iraM. FEMS Microbiol Lett. 2019;366(11). doi: 10.1093/femsle/fnz123. PubMed PMID: 31158277.

54. Matange N, Bodkhe S, Patel M, Shah P. Trade-offs with stability modulate innate and mutationally acquired drug resistance in bacterial dihydrofolate reductase enzymes. Biochem J. 2018;475(12):2107-25. Epub 20180629. doi: 10.1042/BCJ20180249. PubMed PMID: 29871875.

55. Ohno S. Evolution by gene duplication. London, New York: Allen & Unwin; Springer-Verlag; 1970. xv, 160 p. p.

56. Bergthorsson U, Andersson DI, Roth JR. Ohno's dilemma: Evolution of new genes under continuous selection. P Natl Acad Sci USA. 2007;104(43):17004-9. doi: 10.1073/pnas.0707158104. PubMed PMID: WOS:000250487600051.

57. Long MY, VanKuren NW, Chen SD, Vibranovski MD. New Gene Evolution: Little Did We Know. Annual Review of Genetics, Vol 47. 2013;47:307-33. doi: 10.1146/annurev-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

Microbiol. 2020;22(7):2968-88. doi: 10.1111/1462-2920.15067. PubMed PMID: WOS:000537012700001.

67. Karlowicz A, Wegrzyn K, Gross M, Kaczynska D, Ropelewska M, Siemiatkowska M, et al. Defining the crucial domain and amino acid residues in bacterial Lon protease for DNA binding and processing of DNA-interacting substrates. J Biol Chem. 2017;292(18):7507-18. doi: 10.1074/jbc.M116.766709. PubMed PMID: WOS:000400761300019.

68. Liu T, Lu B, Lee I, Ondrovicová G, Kutejová E, Suzuki CK. DNA and RNA binding by the mitochondrial Lon protease is regulated by nucleotide and protein substrate. J Biol Chem. 2004;279(14):13902-10. doi: 10.1074/jbc.M309642200. PubMed PMID: WOS:000220478500077.

69. Lu B, Yadav S, Shah PG, Liu T, Tian B, Pukszta S, et al. Roles for the human ATP-dependent Lon protease in mitochondrial DNA maintenance. J Biol Chem. 2007;282(24):17363-74. Epub 20070409. doi: 10.1074/jbc.M611540200. PubMed PMID: 17420247.

70. Sonezaki S, Okita K, Oba T, Ishii Y, Kondo A, Kato Y. Protein substrates and heat shock reduce the DNA binding ability of *Escherichia coli* Lon protease. Appl Microbiol Biot. 1995;44(3-4):484-8. PubMed PMID: WOS:A1995TM49800033.

71. Breidenstein EB, Bains M, Hancock RE. Involvement of the lon protease in the SOS response triggered by ciprofloxacin in Pseudomonas aeruginosa PAO1. Antimicrob Agents Chemother. 2012;56(6):2879-87. Epub 20120326. doi: 10.1128/AAC.06014-11. PubMed PMID: 22450976; PubMed Central PMCID: PMCPMC3370746.

72. Guerrero RF, Scarpino SV, Rodrigues JV, Hartl DL, Ogbunugafor CB. Proteostasis Environment Shapes Higher-Order Epistasis Operating on Antibiotic Resistance. Genetics. 2019;212(2):565-75. Epub 20190423. doi: 10.1534/genetics.119.302138. PubMed PMID: 31015194; PubMed Central PMCID: PMCPMC6553834.

73. Thompson S, Zhang Y, Ingle C, Reynolds KA, Kortemme T. Altered expression of a quality control protease in E. coli reshapes the in vivo mutational landscape of a model enzyme. Elife. 2020;9. Epub 20200723. doi: 10.7554/eLife.53476. PubMed PMID: 32701056; PubMed Central PMCID: PMCPMC7377907.

74. Yoon J, Patrick JE, Ogbunugafor CB, Shoulders MD. Viral Evolution Shaped by Host Proteostasis Networks. Annu Rev Virol. 2023;10(1):77-98. Epub 20230418. doi: 10.1146/annurev-virology-100220-112120. PubMed PMID: 37071930; PubMed Central PMCID: PMCPMC10543606.

75. Bershtein S, Serohijos AW, Bhattacharyya S, Manhart M, Choi JM, Mu W, et al. Protein Homeostasis Imposes a Barrier on Functional Integration of Horizontally Transferred Genes in Bacteria. Plos Genet. 2015;11(10):e1005612. Epub 20151020. doi: 10.1371/journal.pgen.1005612. PubMed PMID: 26484862; PubMed Central PMCID: PMCPMC4618355.

76. Brown NG, Pennington JM, Huang W, Ayvaz T, Palzkill T. Multiple global suppressors of protein stability defects facilitate the evolution of extended-spectrum TEM beta-lactamases. J Mol Biol. 2010;404(5):832-46. Epub 20101016. doi: 10.1016/j.jmb.2010.10.008. PubMed PMID: 20955714; PubMed Central PMCID: PMCPMC3032993.

77. Standley M, Blay V, Beleva Guthrie V, Kim J, Lyman A, Moya A, et al. Experimental and In Silico Analysis of TEM beta-Lactamase Adaptive Evolution. ACS Infect Dis. 2022;8(12):2451-63. Epub 20221114. doi: 10.1021/acsinfecdis.2c00216. PubMed PMID: 36377311; PubMed Central PMCID: PMCPMC9745794.

78. Koutsandreas T, Felden B, Chevet E, Chatziioannou A. Protein homeostasis imprinting across evolution. NAR Genom Bioinform. 2024;6(1):lqae014. Epub 20240215. doi: 10.1093/nargab/lqae014. PubMed PMID: 38486886; PubMed Central PMCID: PMCPMC10939379.

79. Khodaparast L, Khodaparast L, Wu GQ, Michiels E, Gallardo R, Houben B, et al. Exploiting the aggregation propensity of beta-lactamases to design inhibitors that induce enzyme misfolding. Nature Communications. 2023;14(1). doi: ARTN 5571 10.1038/s41467-023-41191-z. PubMed PMID: WOS:001066679400020.

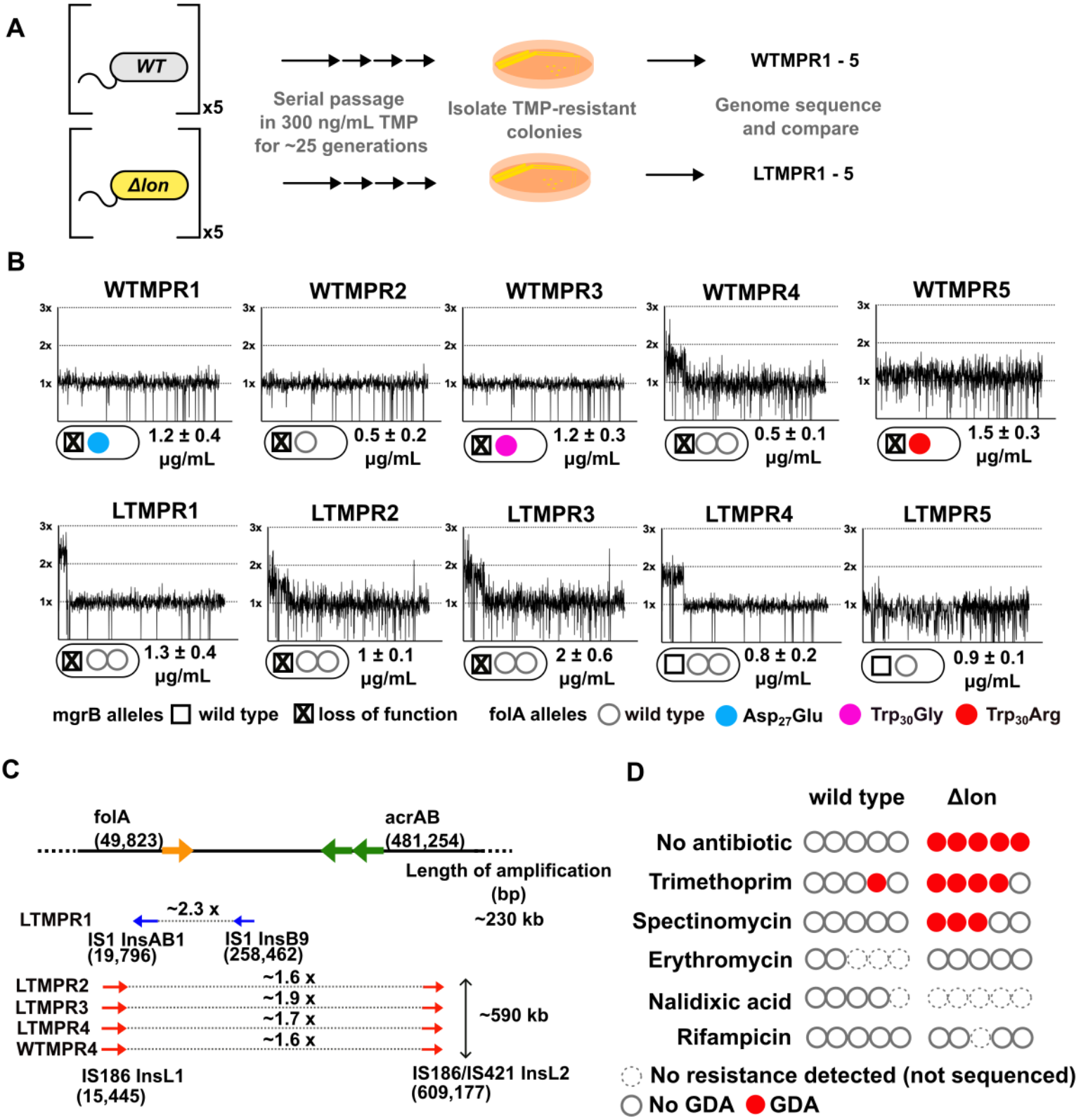
80. Lupoli TJ, Vaubourgeix J, Burns-Huang K, Gold B. Targeting the Proteostasis Network for Mycobacterial Drug Discovery. ACS Infect Dis. 2018;4(4):478-98. Epub 20180302. doi: 10.1021/acsinfecdis.7b00231. PubMed PMID: 29465983; PubMed Central PMCID: PMCPMC5902792.

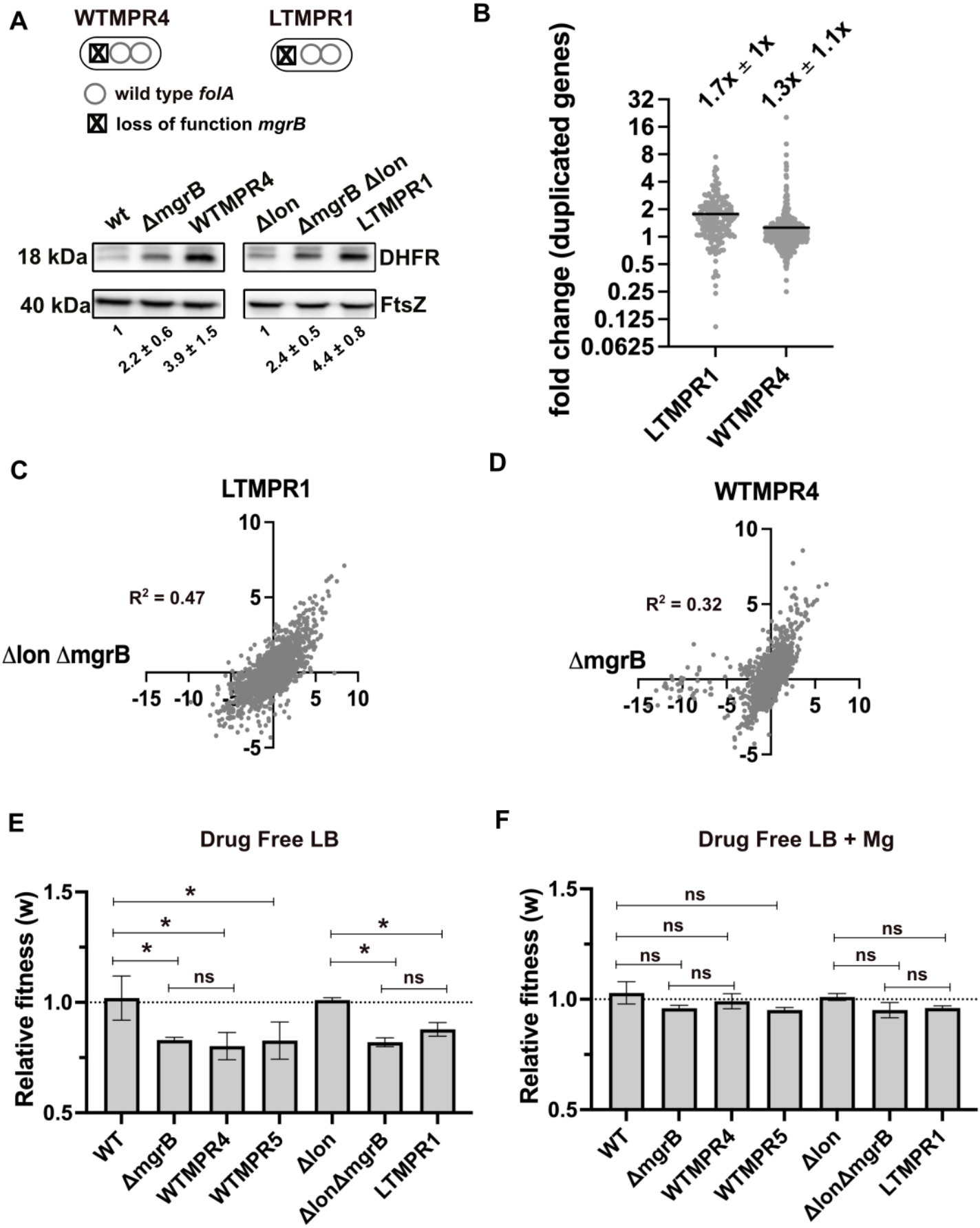
81. Won HI, Zinga S, Kandror O, Akopian T, Wolf ID, Schweber JTP, et al. Targeted protein degradation in mycobacteria uncovers antibacterial effects and potentiates antibiotic efficacy. Nat Commun. 2024;15(1):4065. Epub 20240514. doi: 10.1038/s41467-024-48506-8. PubMed PMID: 38744895; PubMed Central PMCID: PMCPMC11094019.

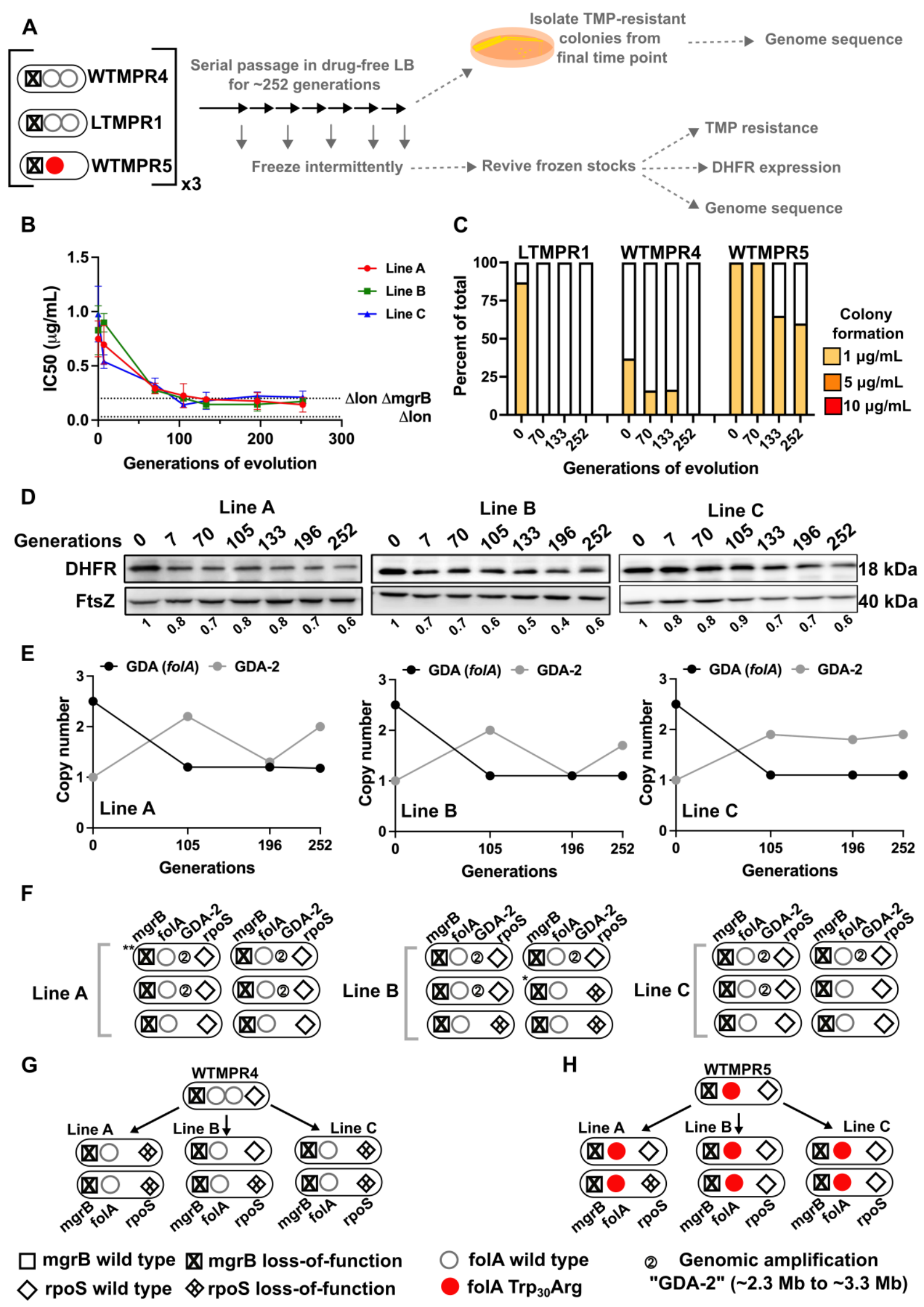
82. Matange N, Hegde S, Bodkhe S. Adaptation Through Lifestyle Switching Sculpts the Fitness Landscape of Evolving Populations: Implications for the Selection of Drug-Resistant Bacteria at Low Drug Pressures. Genetics. 2019;211(3):1029-44. Epub 20190122. doi:

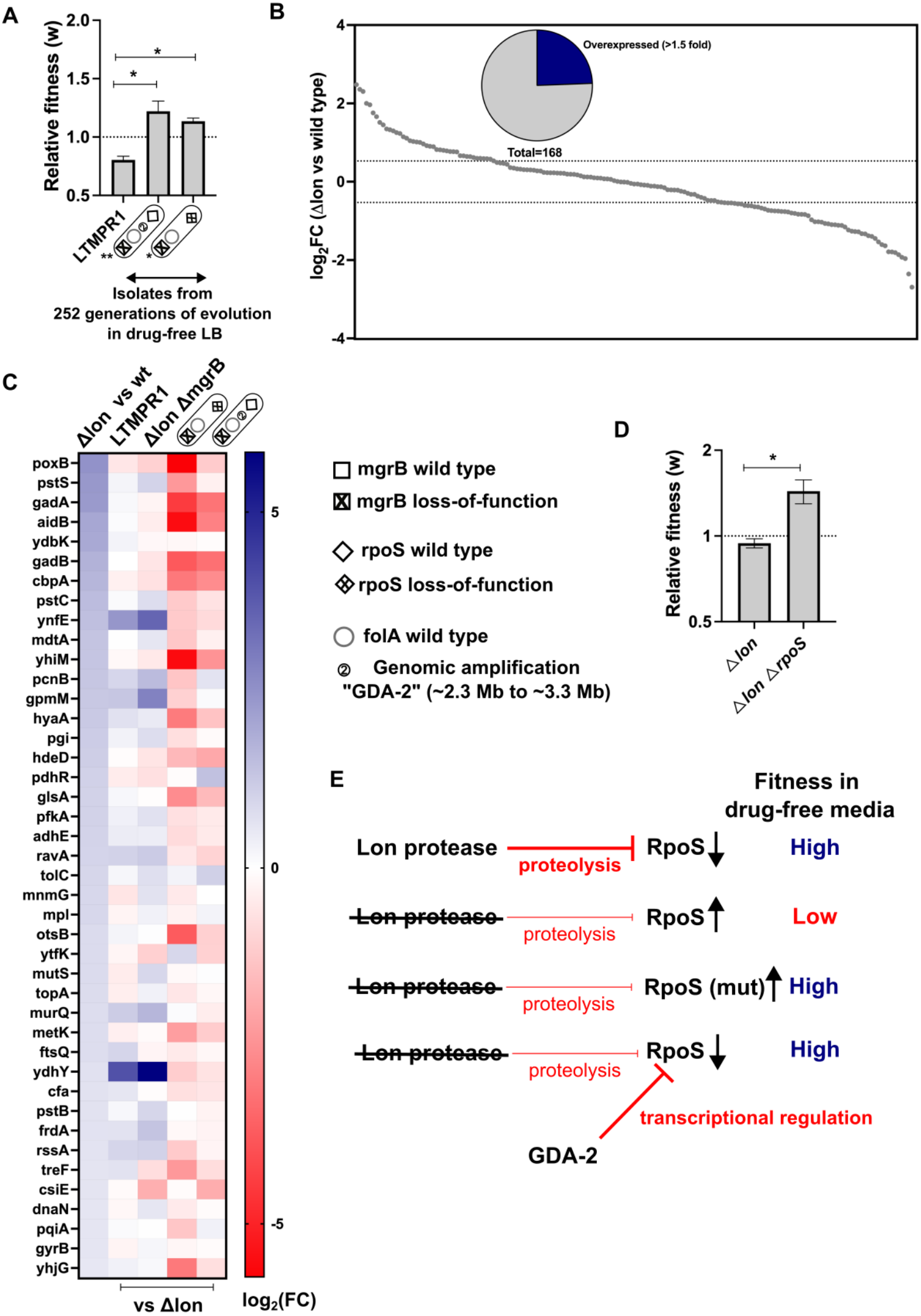
10.1534/genetics.119.301834. PubMed PMID: 30670539; PubMed Central PMCID: PMCPMC6404241.

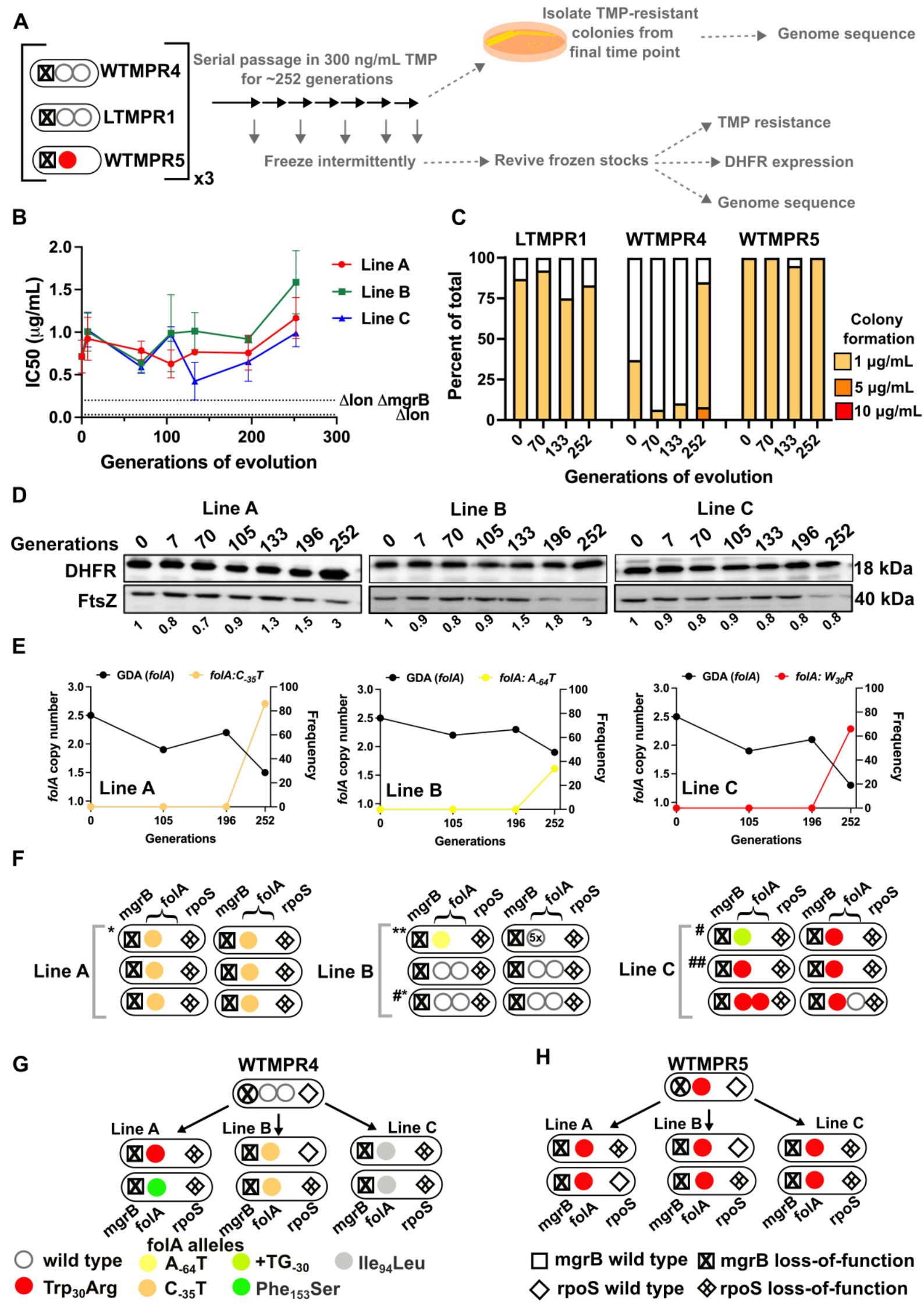
83. Deatherage DE, Barrick JE. Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol Biol. 2014;1151:165-88. doi: 10.1007/978-1-4939-0554-6_12. PubMed PMID: 24838886; PubMed Central PMCID: PMCPMC4239701.

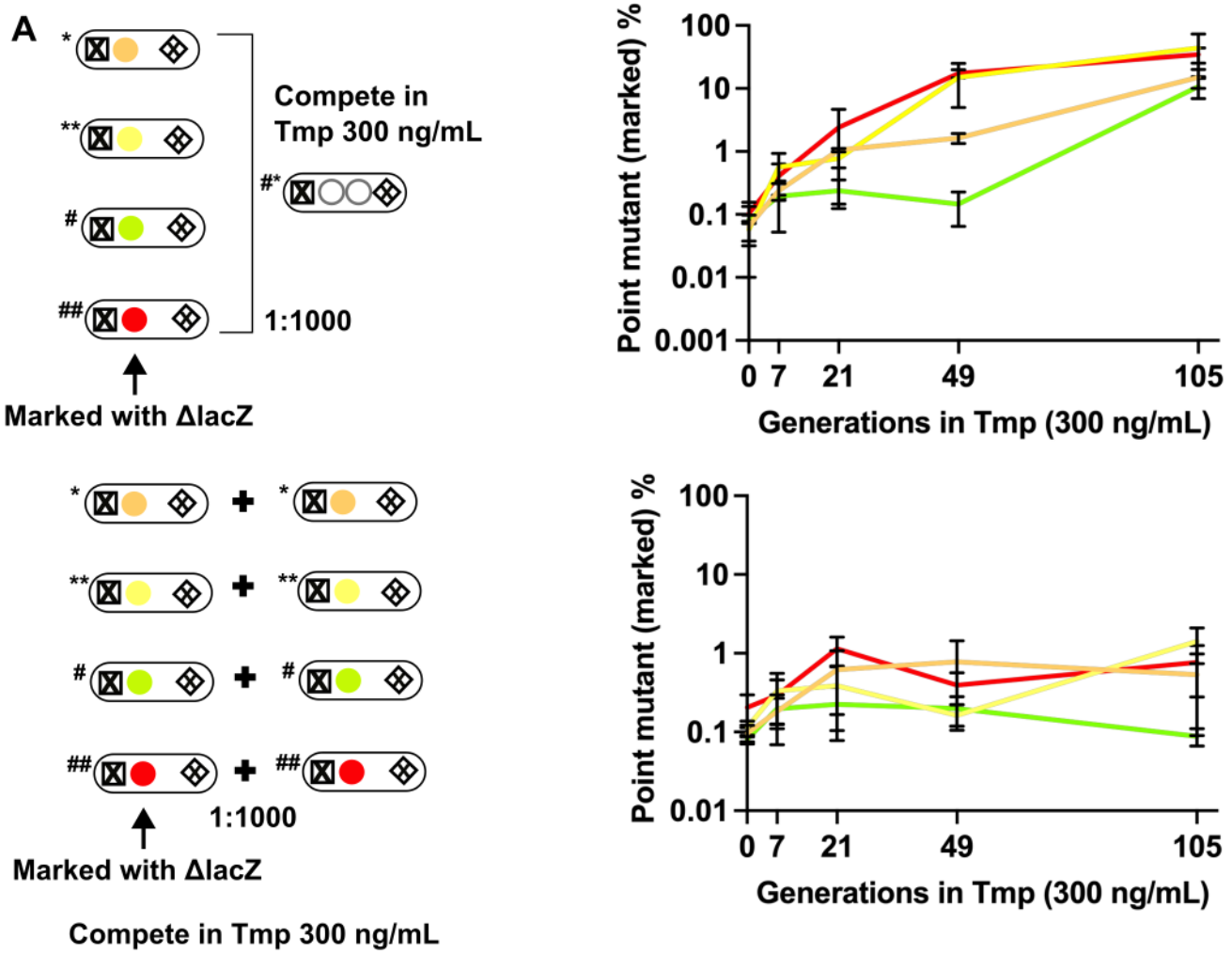

A
WT
x5
Δlon
x5
Serial passage
in 300 ng/mL TMP
for ~25 generations
Isolate TMP-resistant
colonies
WTMPR1 - 5
Genome sequence
and compare
LTMPR1 - 5
B
WTMPR1
WTMPR2
WTMPR3
WTMPR4
WTMPR5
3x
2x
1x
1.2 ± 0.4
μg/mL
0.5 ± 0.2
μg/mL
1.2 ± 0.3
μg/mL
0.5 ± 0.1
μg/mL
1.5 ± 0.3
μg/mL
LTMPR1
LTMPR2
LTMPR3
LTMPR4
LTMPR5
1.3 ± 0.4
μg/mL
1 ± 0.1
μg/mL
2 ± 0.6
μg/mL
0.8 ± 0.2
μg/mL
0.9 ± 0.1
μg/mL
mgrB alleles wild type loss of function
folA alleles wild type Asp27Glu Trp30Gly Trp30Arg
C
folA
(49,823)
acrAB
(481,254)
Length of amplification
(bp)
~230 kb
LTMPR1
~2.3 x
IS1 InsAB1
(19,796)
IS1 InsB9
(258,462)
LTMPR2
LTMPR3
LTMPR4
WTMPR4
~1.6 x
~1.9 x
~1.7 x
~1.6 x
~590 kb
IS186 InsL1
(15,445)
IS186/IS421 InsL2
(609,177)
D
wild type
Δlon
No antibiotic
Trimethoprim
Spectinomycin
Erythromycin
Nalidixic acid
Rifampicin
No resistance detected (not sequenced)
No GDA
GDA

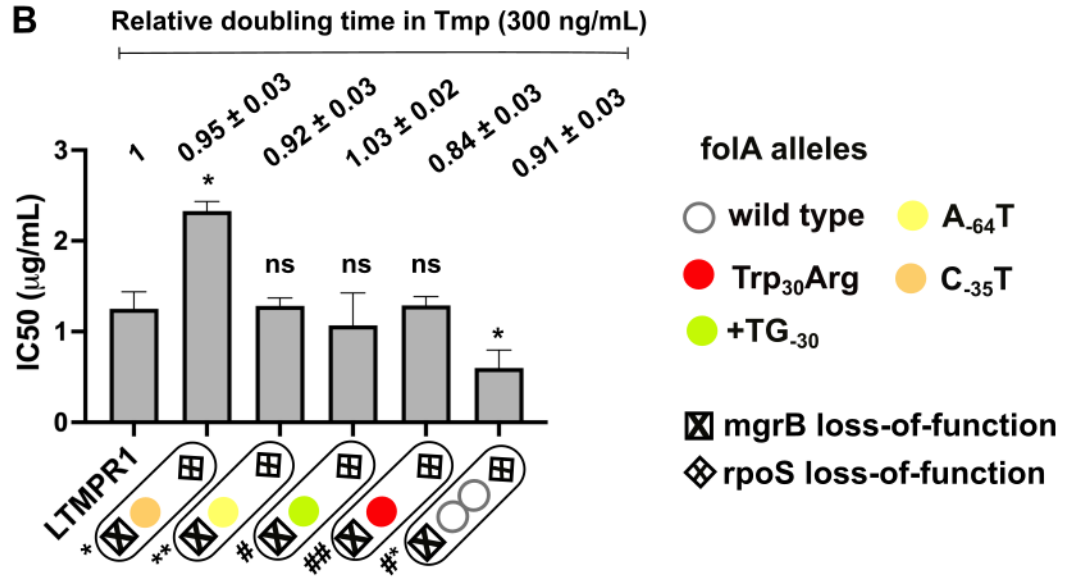

A
WTMPR4
LTMPR1
wild type folA
loss of function mgrB
wt
ΔmgrB
WTMPR4
Δlon
ΔmgrB Δlon
LTMPR1
18 kDa
40 kDa
DHFR
FtsZ
1
2.2 ± 0.6
3.9 ± 1.5
1
2.4 ± 0.5
4.4 ± 0.8
B
fold change (duplicated genes)
1.7x ± 1x
1.3x ± 1.1x
32
16
8
4
2
1
0.5
0.25
0.125
0.0625
LTMPR1
WTMPR4
C
LTMPR1
R² = 0.47
Δlon ΔmgrB
D
WTMPR4
R² = 0.32
ΔmgrB
E
Drug Free LB
Relative fitness (w)
1.5
1.0
0.5
*
ns
WT
ΔmgrB
WTMPR4
WTMPR5
Δlon
ΔlonΔmgrB
LTMPR1
F
Drug Free LB + Mg
Relative fitness (w)
ns

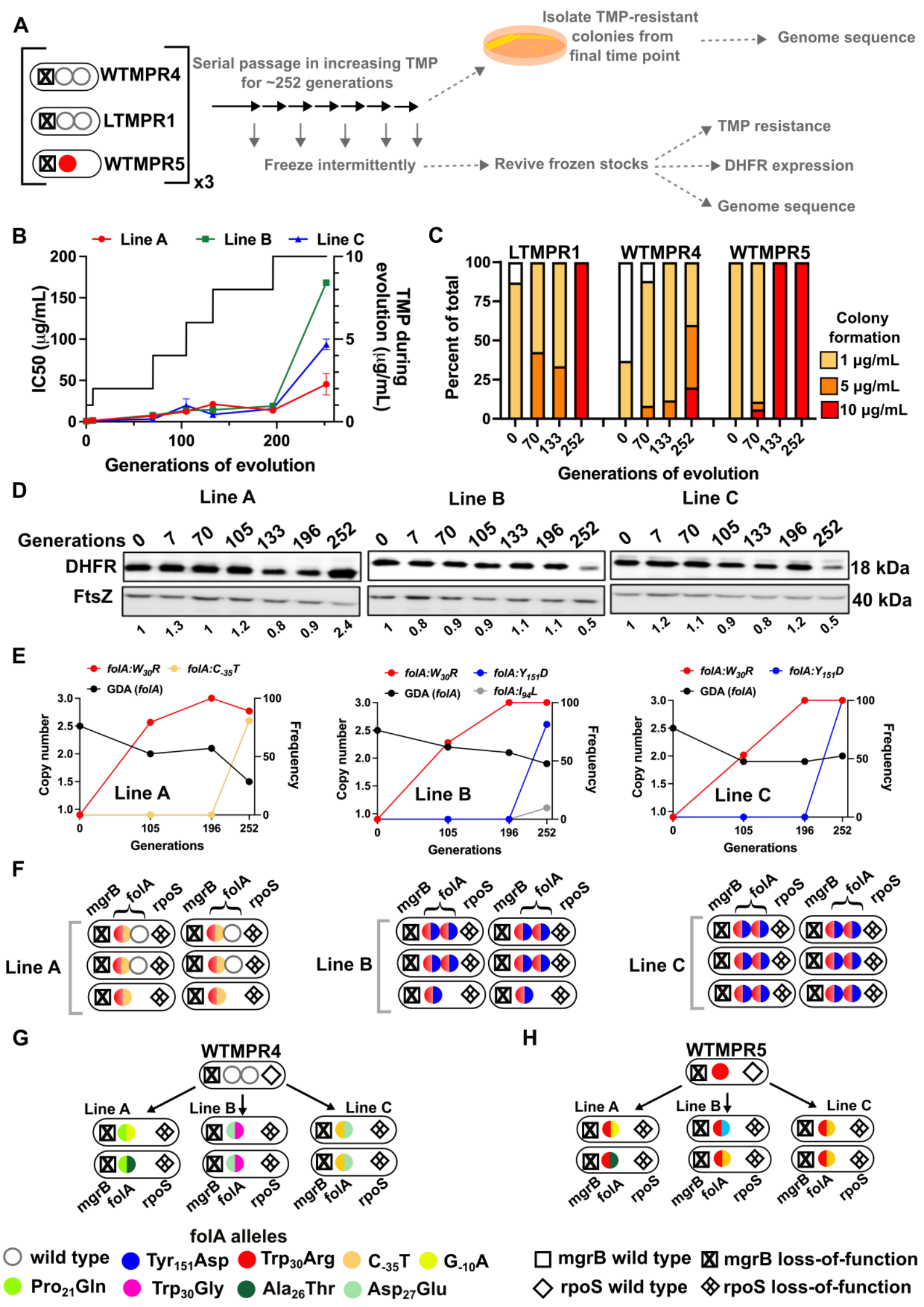

A
WTMPR4
LTMPR1
WTMPR5
x3
Serial passage in drug-free LB for ~252 generations
Isolate TMP-resistant colonies from final time point
Genome sequence
Freeze intermittently
Revive frozen stocks
TMP resistance
DHFR expression
Genome sequence
B
IC50 (μg/mL)
Line A
Line B
Line C
ΔIon ΔmgrB
ΔIon
Generations of evolution
C
LTMPR1
WTMPR4
WTMPR5
Percent of total
Colony formation
1 μg/mL
5 μg/mL
10 μg/mL
Generations of evolution
D
Line A
Line B
Line C
Generations
DHFR
FtsZ
18 kDa
40 kDa
E
GDA (folA)
GDA-2
Copy number
Line A
Line B
Line C
Generations
F
mgrB
folA
GDA-2
rpoS
Line A
Line B
Line C
G
WTMPR4
Line A
Line B
Line C
H
WTMPR5
Line A
Line B
Line C
mgrB wild type
mgrB loss-of-function
folA wild type
Genomic amplification "GDA-2" (~2.3 Mb to ~3.3 Mb)
rpoS wild type
rpoS loss-of-function
folA Trp30Arg

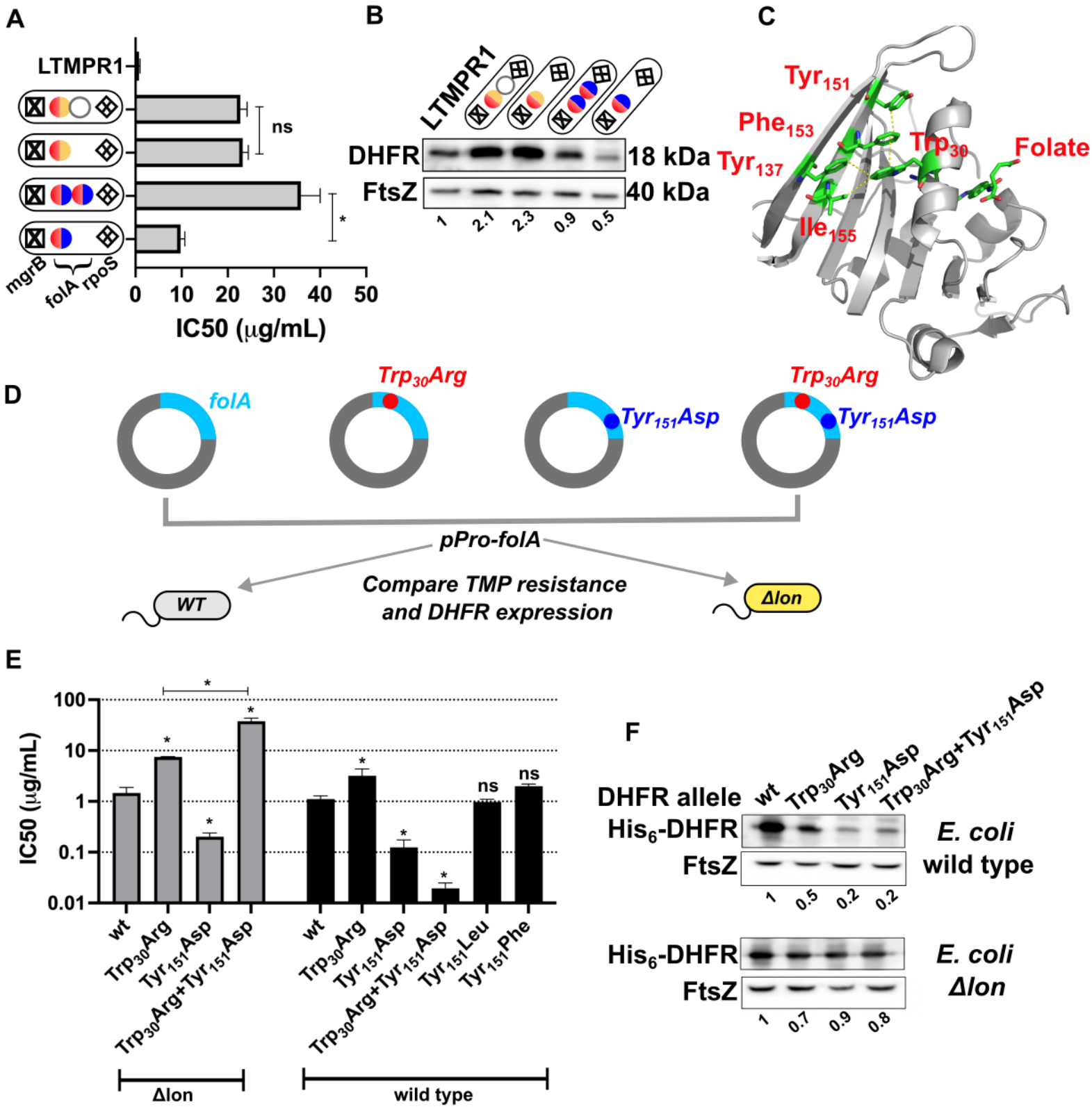

A
Relative fitness (w)
2.0
1.5
1.0
0.5
*
*
LTMPR1
**
*
Isolates from
252 generations of evolution
in drug-free LB
B
log2FC (Δlon vs wild type)
4
2
0
-2
-4
Overexpressed (>1.5 fold)
Total=168
C
Δlon vs wt
LTMPR1
Δlon ΔmgrB
poxB
pstS
gadA
aidB
ydbK
gadB
cbpA
pstC
ynfE
mdtA
yhiM
pcnB
gpmM
hyaA
pgi
hdeD
pdhR
glsA
pfkA
adhE
ravA
tolC
mnmG
mpl
otsB
ytfK
mutS
topA
murQ
metK
ftsQ
ydhY
cfa
pstB
frdA
rssA
treF
csiE
dnaN
pqiA
gyrB
yhjG
5
0
-5
vs Δlon
log2(FC)
mgrB wild type
mgrB loss-of-function
rpoS wild type
rpoS loss-of-function
folA wild type
Genomic amplification
"GDA-2" (~2.3 Mb to ~3.3 Mb)
D
Relative fitness (w)
2
1
0.5
*
Δlon
Δlon ΔrpoS
E
Fitness in
drug-free media
Lon protease
proteolysis
RpoS
High
Lon protease
proteolysis
RpoS
Low
Lon protease
proteolysis
RpoS (mut)
High
Lon protease
proteolysis
RpoS
High
transcriptional regulation
GDA-2

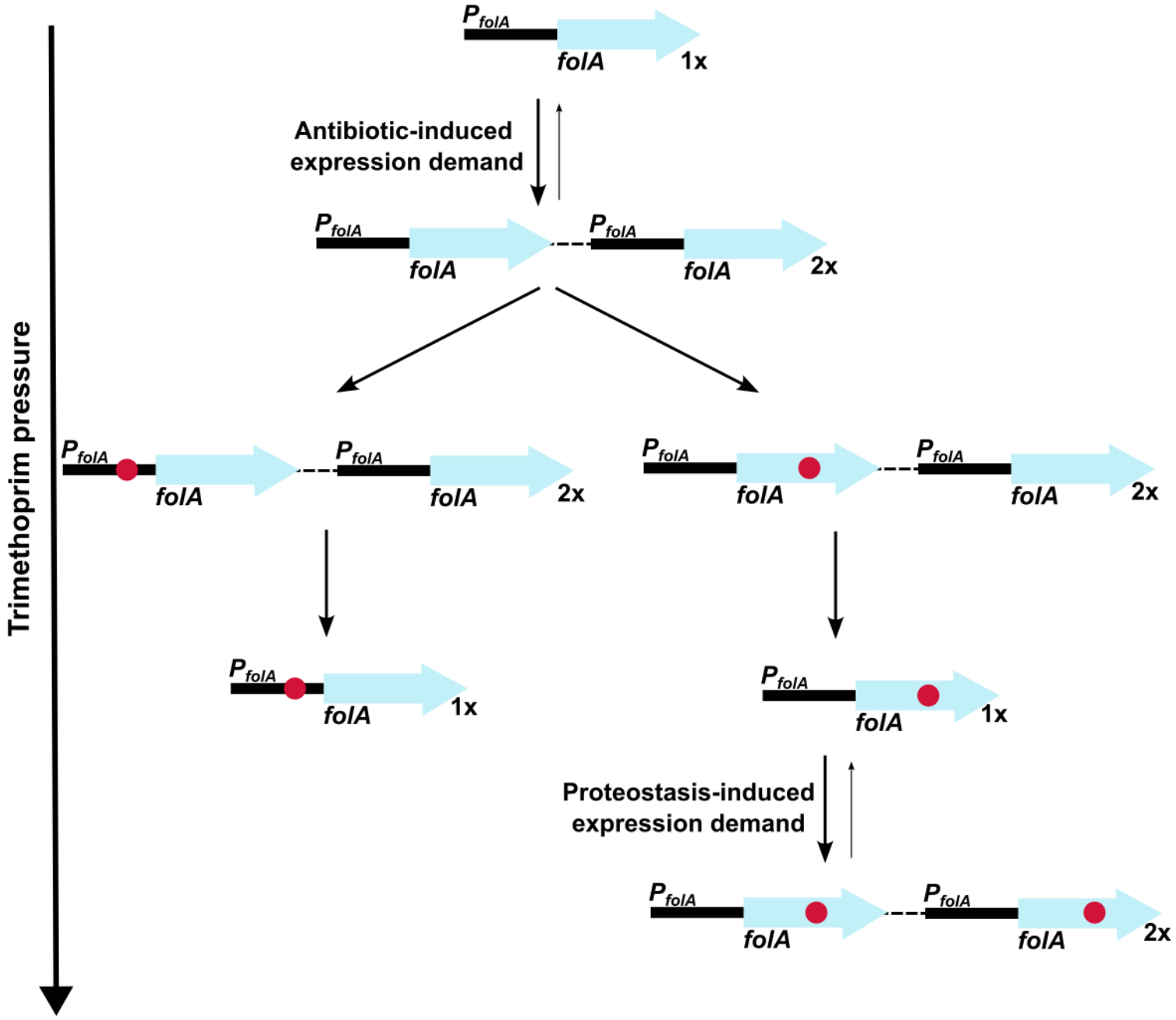

A
WTMPR4
LTMPR1
WTMPR5
x3
Serial passage in 300 ng/mL TMP for ~252 generations
Freeze intermittently
Isolate TMP-resistant colonies from final time point
Genome sequence
Revive frozen stocks
TMP resistance
DHFR expression
Genome sequence
B
IC50 (μg/mL)
Generations of evolution
Line A
Line B
Line C
ΔIon ΔmgrB
ΔIon
C
LTMPR1
WTMPR4
WTMPR5
Percent of total
Colony formation
1 μg/mL
5 μg/mL
10 μg/mL
Generations of evolution
D
Line A
Line B
Line C
Generations
DHFR
FtsZ
18 kDa
40 kDa
E
GDA (folA)
folA:C-35T
folA: A-64T
folA: W30R
folA copy number
Frequency
Generations
Line A
Line B
Line C
F
mgrB
folA
rpoS
Line A
Line B
Line C
5x
G
WTMPR4
Line A
Line B
Line C
folA alleles
wild type
A-64T
+TG-30
Ile94Leu
Trp30Arg
C-35T
Phe153Ser
H
WTMPR5
Line A
Line B
Line C
mgrB wild type
mgrB loss-of-function
rpoS wild type
rpoS loss-of-function


A
Compete in
Tmp 300 ng/mL
1:1000
Marked with ΔlacZ
Point mutant (marked) %
100
10
1
0.1
0.01
0.001
0 7 21 49 105
Generations in Tmp (300 ng/mL)
1:1000
Marked with ΔlacZ
Compete in Tmp 300 ng/mL
Point mutant (marked) %
100
10
1
0.1
0.01
0 7 21 49 105
Generations in Tmp (300 ng/mL)

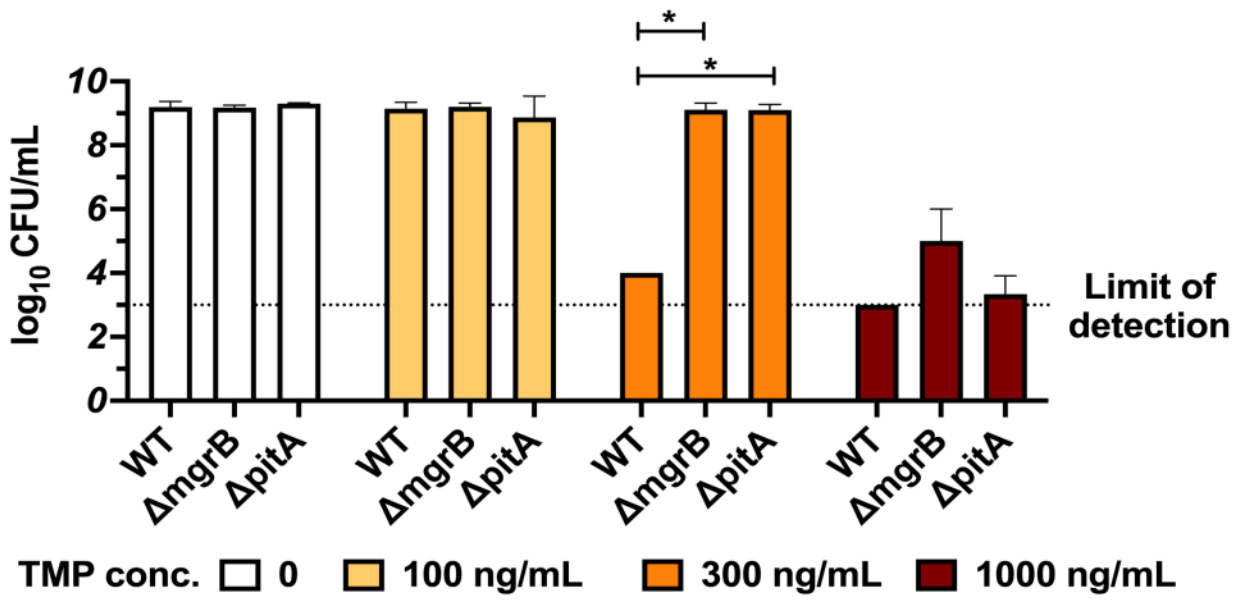

B
Relative doubling time in Tmp (300 ng/mL)
1
0.95 ± 0.03
0.92 ± 0.03
1.03 ± 0.02
0.84 ± 0.03
0.91 ± 0.03
IC50 (μg/mL)
3
2
1
0
*
ns
ns
ns
*
LTMPR1
folA alleles
wild type
A-64T
Trp30Arg
C-35T
+TG-30
mgrB loss-of-function
rpoS loss-of-function


A
WTMPR4
LTMPR1
WTMPR5
x3
Serial passage in increasing TMP for ~252 generations
Freeze intermittently
Isolate TMP-resistant colonies from final time point
Genome sequence
Revive frozen stocks
TMP resistance
DHFR expression
Genome sequence
B
Line A
Line B
Line C
IC50 (μg/mL)
TMP during evolution (μg/mL)
Generations of evolution
C
LTMPR1
WTMPR4
WTMPR5
Percent of total
Colony formation
1 μg/mL
5 μg/mL
10 μg/mL
Generations of evolution
D
Line A
Line B
Line C
Generations
DHFR
FtsZ
18 kDa
40 kDa
E
folA:W30R
folA:C-35T
GDA (folA)
folA:Y151D
folA:I94L
Copy number
Frequency
Generations
Line A
Line B
Line C
F
mgrB
folA
rpoS
Line A
Line B
Line C
G
WTMPR4
Line A
Line B
Line C
H
WTMPR5
folA alleles
wild type
Tyr151Asp
Trp30Arg
C-35T
G-10A
Pro21Gln
Trp30Gly
Ala26Thr
Asp27Glu
mgrB wild type
mgrB loss-of-function
rpoS wild type
rpoS loss-of-function

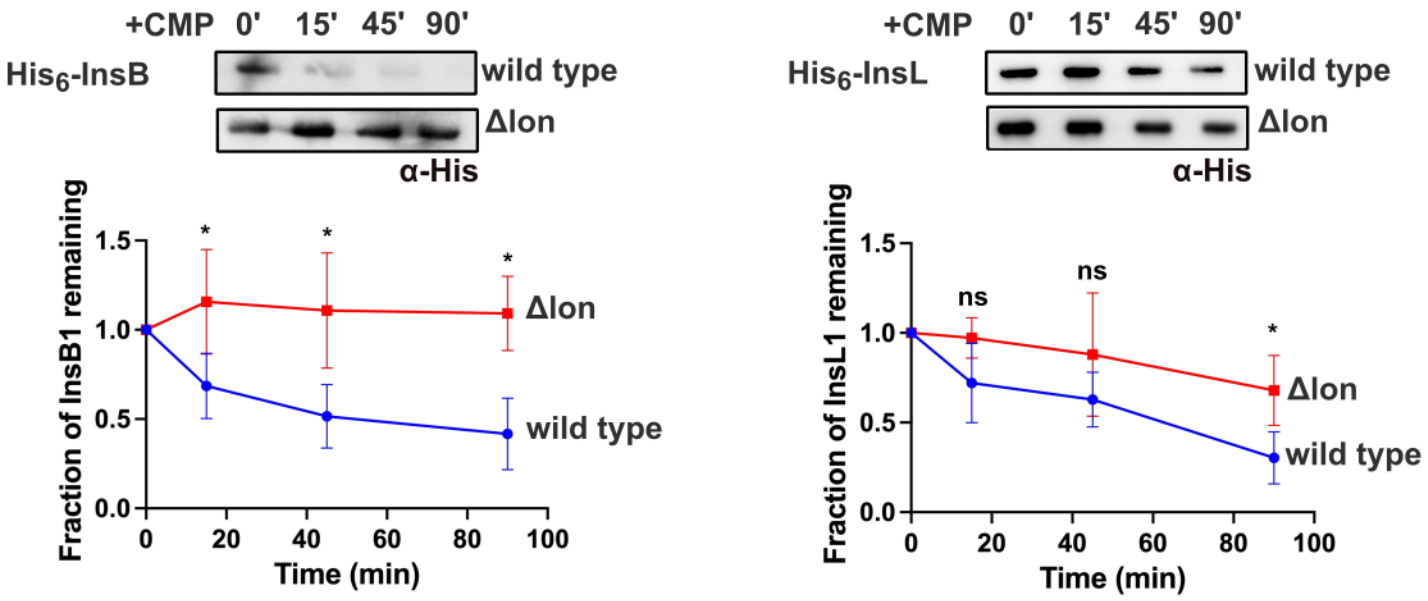

A
LTMPR1
ns
*
mgrB
folA
rpoS
0
10
20
30
40
50
IC50 (μg/mL)
B
LTMPR1
DHFR
18 kDa
FtsZ
40 kDa
1
2.1
2.3
0.9
0.5
C
Tyr151
Phe153
Tyr137
Trp30
Folate
Ile155
D
folA
Trp30Arg
Tyr151Asp
Trp30Arg
Tyr151Asp
pPro-folA
Compare TMP resistance
and DHFR expression
WT
Δlon
E
IC50 (μg/mL)
100
10
1
0.1
0.01
wt
Trp30Arg
Tyr151Asp
Trp30Arg+Tyr151Asp
wt
Trp30Arg
Tyr151Asp
Trp30Arg+Tyr151Asp
Tyr151Leu
Tyr151Phe
ns
ns
Δlon
wild type
F
DHFR allele
wt
Trp30Arg
Tyr151Asp
Trp30Arg+Tyr151Asp
His6-DHFR
FtsZ
E. coli
wild type
1
0.5
0.2
0.2
His6-DHFR
FtsZ
E. coli
Δlon
1
0.7
0.9
0.8


P_{folA}
folA
1x
Antibiotic-induced expression demand
P_{folA}
folA
P_{folA}
folA
2x
P_{folA}
folA
P_{folA}
folA
2x
P_{folA}
folA
P_{folA}
folA
2x
P_{folA}
folA
1x
P_{folA}
folA
1x
Proteostasis-induced expression demand
P_{folA}
folA
P_{folA}
folA
2x
Trimethoprim pressure

A

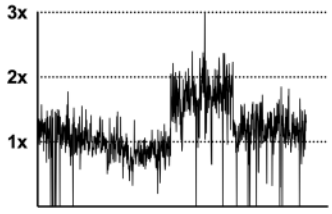

*
*
10
8
6
4
2
0
log10 CFU/mL
Limit of detection
WT
ΔmgrB
ΔpitA
WT
ΔmgrB
ΔpitA
WT
ΔmgrB
ΔpitA
WT
ΔmgrB
ΔpitA
TMP conc.
0
100 ng/mL
300 ng/mL
1000 ng/mL

B


+CMP 0' 15' 45' 90'
His6-InsB
wild type
Δlon
α-His
Fraction of InsB1 remaining
1.5
1.0
0.5
0.0
0
20
40
60
80
100
Time (min)
*
*
*
Δlon
wild type
+CMP 0' 15' 45' 90'
His6-InsL
wild type
Δlon
α-His
Fraction of InsL1 remaining
1.5
1.0
0.5
0.0
0
20
40
60
80
100
Time (min)
ns
ns
*
Δlon
wild type

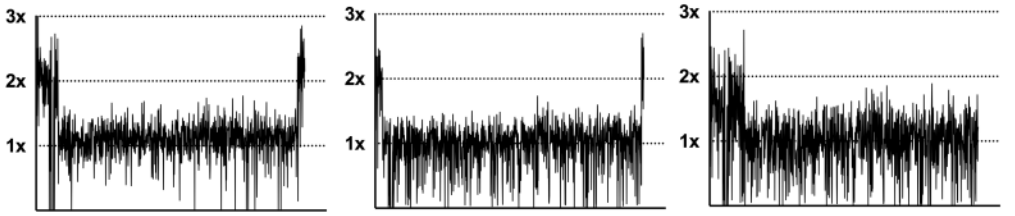
Representative coverage plots of *E. coli ΔIon* evolved in no antibiotic or spectinomycin

No antibiotic (representative plot for all 5 replicates)

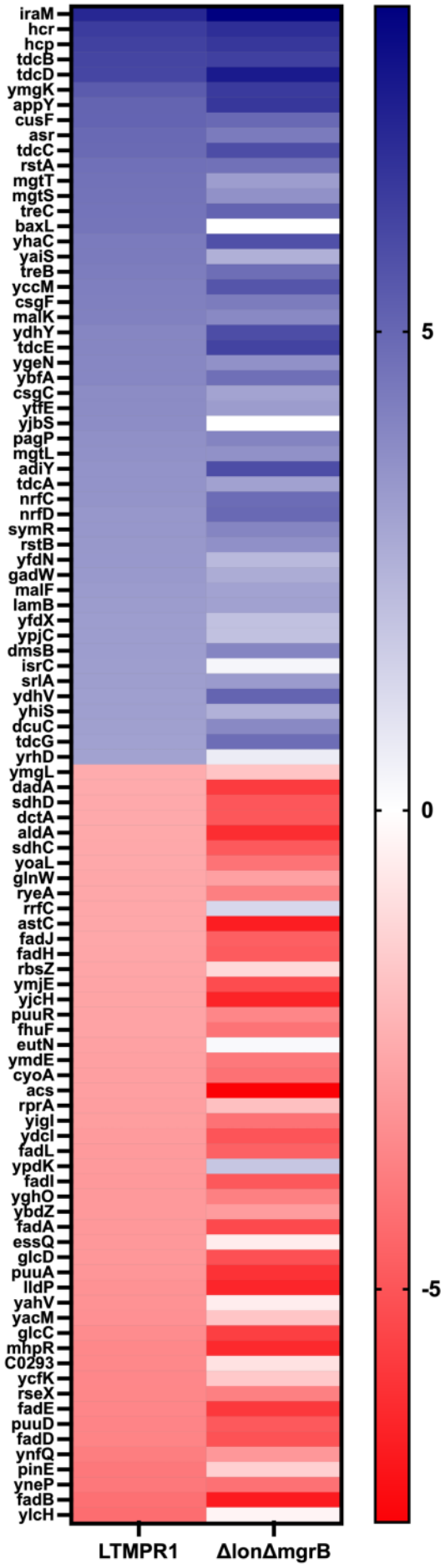


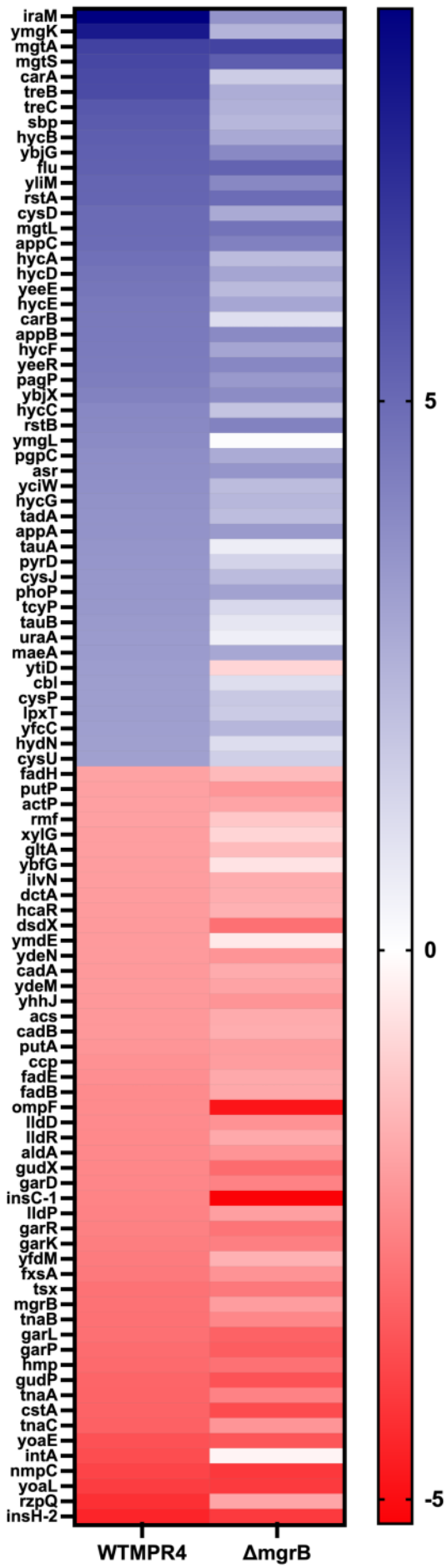
InsH8 to InsH10

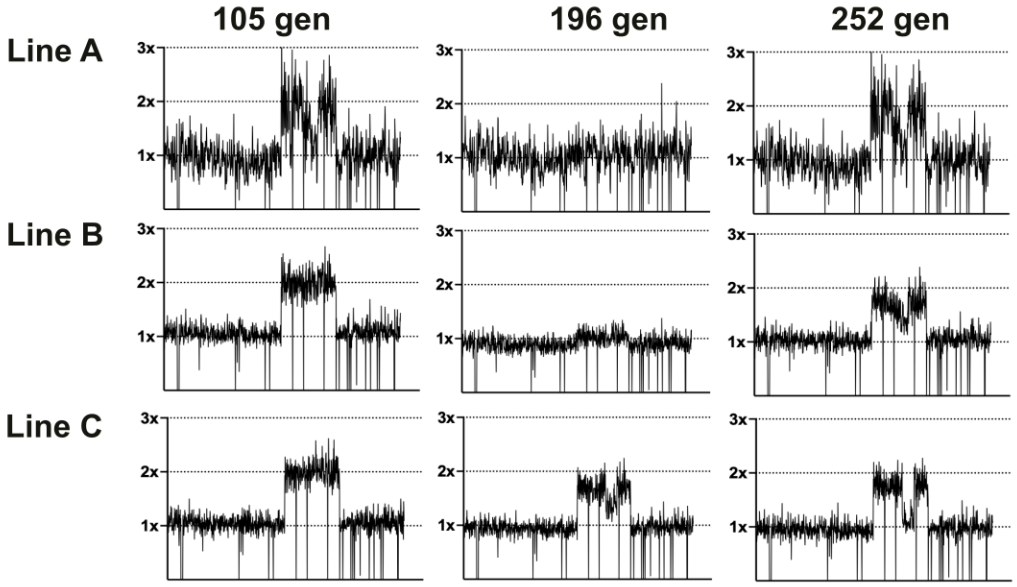
Spectinomycin

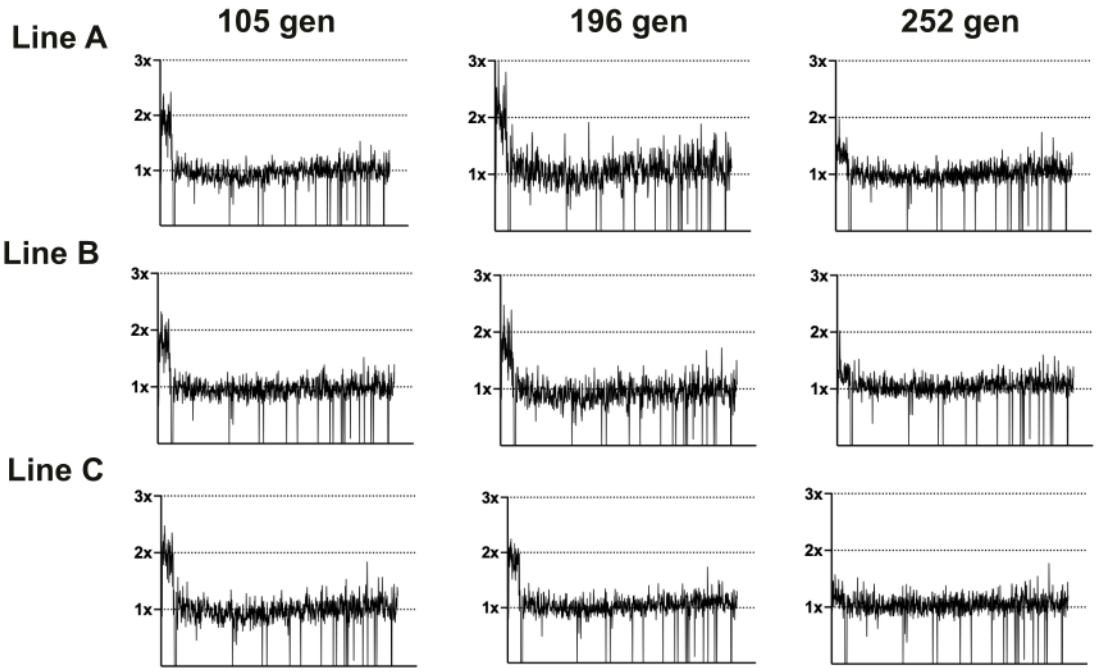


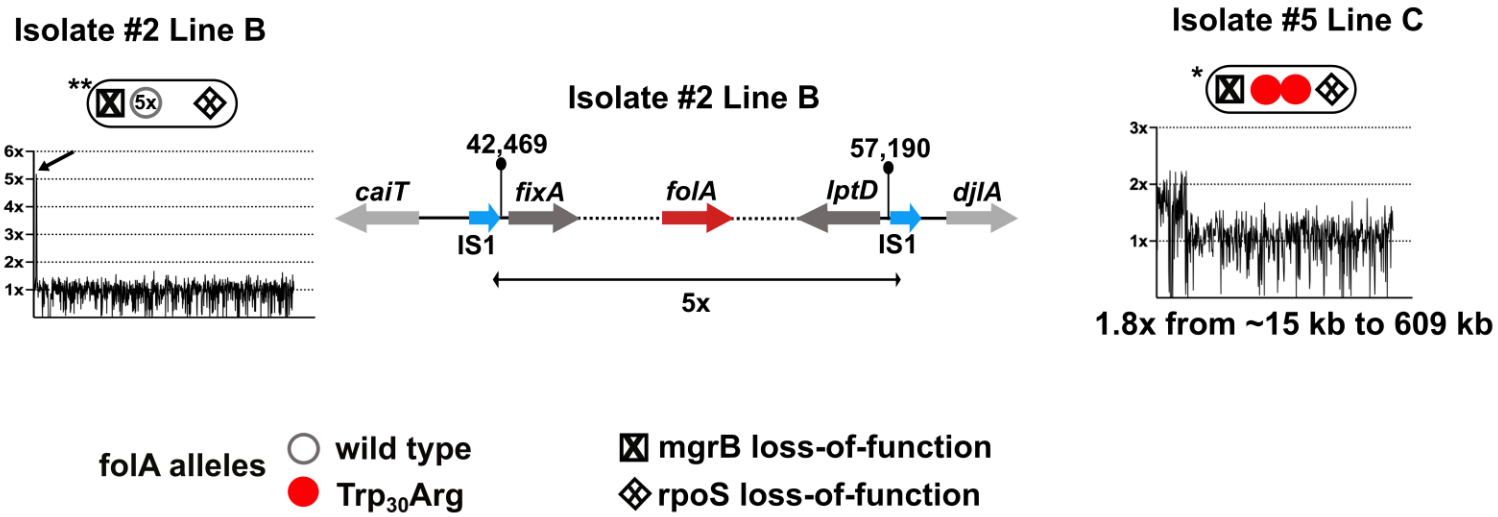
InsA7/B7 to InsD1 **InsA7/B7 to InsL1** **InsL1 to InsL2**

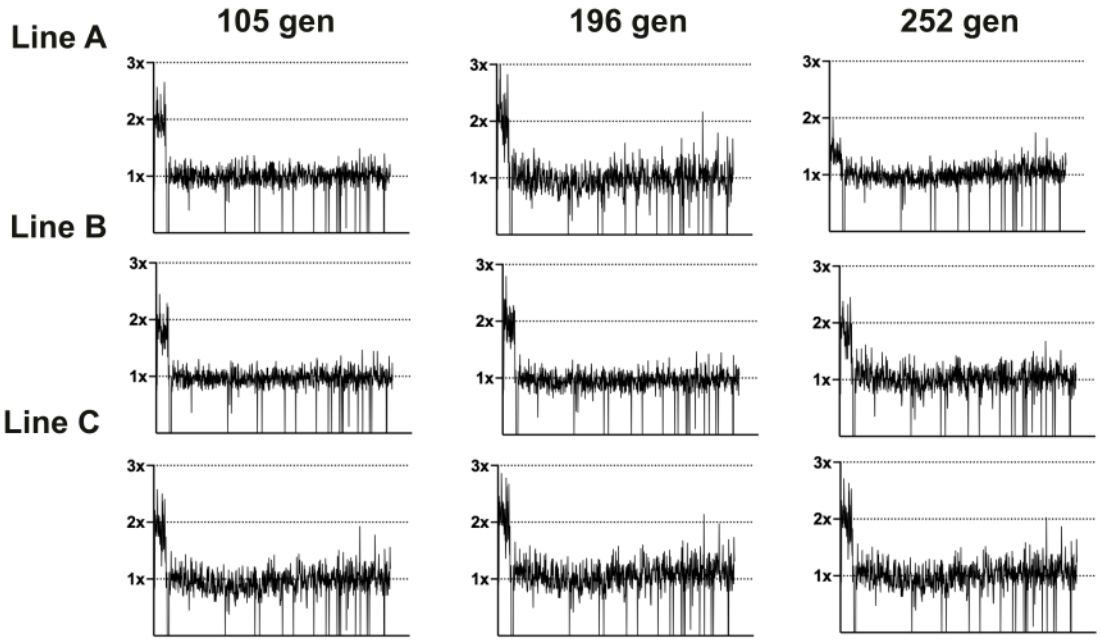

iraM
hcr
hcp
tdcB
tdcD
ymgK
appY
cusF
asr
tdcC
rstA
mgtT
mgtS
treC
baxL
yhaC
yaiS
treB
yccM
csgF
malK
ydhY
tdcE
ygeN
ybfA
csgC
yffE
yjbS
pagP
mgtL
adiY
tdcA
nrfC
nrfD
symR
rstB
yfdN
gadW
malF
lamB
yfdX
ypjC
dmsB
isrC
srlA
ydhV
yhiS
dcuC
tdcG
yrhD
ymgL
dadA
sdhD
dctA
aldA
sdhC
yoaL
glnW
ryeA
rrfC
astC
fadJ
fadH
rbsZ
ymjE
yjcH
puuR
fhuF
eutN
ymdE
cyoA
acs
rprA
yigI
ydcI
fadL
ypdK
fadI
yghO
ybdZ
fadA
essQ
glcD
puuA
lldP
yahV
yacM
glcC
mhpR
C0293
ycfK
rseX
fadE
puuD
fadD
ynfQ
pinE
yneP
fadB
ylcH
5
0
-5
LTMPR1
ΔlonΔmgrB


iraM
ymgK
mgtA
mgtS
carA
treB
treC
sbp
hycB
ybjG
flu
yliM
rstA
cysD
mgtL
appC
hycA
hycD
yeeE
hycE
carB
appB
hycF
yeeR
pagP
ybjX
hycC
rstB
ymgL
pgpC
asr
yciW
hycG
tadA
appA
tauA
pyrD
cysJ
phoP
tcyP
tauB
uraA
maeA
ytiD
cbl
cysP
lpxT
yfcC
hydN
cysU
fadH
putP
actP
rmf
xylG
gltA
ybfG
ilvN
dctA
hcaR
dsdX
ymdE
ydeN
cadA
ydeM
yhhJ
acs
cadB
putA
ccp
fadE
fadB
ompF
lldD
lldR
aldA
gudX
garD
insC-1
lldP
garR
garK
yfdM
fxsA
tsx
mgrB
tnaB
garL
garP
hmp
gudP
tnaA
cstA
tnaC
yoaE
intA
nmpC
yoaL
rzpQ
insH-2
5
0
-5
WTMPR4
ΔmgrB


105 gen
196 gen
252 gen
Line A
Line B
Line C
3x
2x
1x


105 gen
196 gen
252 gen
Line A
Line B
Line C
3x
2x
1x


Isolate #2 Line B
**
5x
6x
5x
4x
3x
2x
1x
Isolate #2 Line B
42,469
57,190
caiT
fixA
folA
lptD
djlA
IS1
IS1
5x
Isolate #5 Line C
*
3x
2x
1x
1.8x from ~15 kb to 609 kb
folA alleles
wild type
Trp30Arg
mgrB loss-of-function
rpoS loss-of-function


105 gen
196 gen
252 gen
Line A
Line B
Line C
3x
2x
1x