Gene amplification as a form of population-level gene expression regulation

Tomanek I[1]†, Grah R[1]†, Lagator M[1,2], Andersson AMC[1], Bollback JP[3], Tkačik G[1], Guet CC[1]*

[1]Institute of Science and Technology Austria, 3400 Klosterneuburg, Austria.

[2]School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester M13 9PL, United Kingdom.

[3] Institute of Integrative Biology, Functional and Comparative Genomics, University of Liverpool, Liverpool L69 72B, United Kingdom.

*Correspondence to: calin@ist.ac.at.

†Authors contributed equally.

Abstract

Organisms cope with change by employing transcriptional regulators. However, when faced with rare environments, the evolution of transcriptional regulators and their promoters may be too slow. We ask whether the intrinsic instability of gene duplication and amplification provides a generic alternative to canonical gene regulation. By real-time monitoring of gene copy number mutations in *E. coli*, we show that gene duplications and amplifications enable adaptation to fluctuating environments by rapidly generating copy number, and hence expression level, polymorphism. This 'amplification-mediated gene expression tuning' occurs on timescales similar to canonical gene regulation and can deal with rapid environmental changes. Mathematical modeling shows that amplifications also tune gene expression in stochastic environments where transcription factor-based schemes are hard to evolve or maintain. The fleeting nature of gene amplifications gives rise to a generic population-level mechanism that relies on genetic heterogeneity to rapidly tune expression of any gene, without leaving any genomic signature.

Main

Natural environments change periodically or stochastically with frequent or very rare fluctuations and life crucially depends on the ability to respond to such changes. Gene regulatory networks have evolved into an elaborate mechanism for such adjustments as populations were repeatedly required to cope with specific environmental changes[1–3]. Gene regulation requires many dedicated components – transcription factors and promoter sequences on the DNA – for information processing to occur. However, due to low single base-pair mutation rates, complex promoters cannot easily evolve on ecological time scales[4,5].

Gene copy number mutations might provide a fundamentally different adaptation strategy, which neither depends on existing regulation nor requires regulation to evolve. Gene

duplications arise by homologous or illegitimate recombination between sister-chromosomes. Depending on the genomic locus, duplication rates (k_{dup}) can vary between 10^{-6} and 10^{-2} per cell per generation in bacteria[6–9]. This means that a typical bacterial population will contain at any given time a large fraction of cells with a duplication somewhere on the chromosome[9,10]. Due to the long stretches of homology, duplications are highly unstable: at rates (k_{rec}) between 10^{-3} and 10^{-1} per cell per generation[7,8] *recA*-dependent unequal crossover of the repeated sequence leads to deletion of the second copy – restoring the ancestral state – or to further amplification (Fig. 1a). If a gene is under selection for increased expression[11–13], the process of gene duplication and amplification (GDA) can dramatically increase organismal fitness by increasing gene copy numbers. Due to their high rates of formation, amplifications provide fast adaptation and facilitate the evolution of functional innovation[14]. In contrast, their high rate of loss makes amplifications transient and difficult to study[14]. Surprisingly, until recently it has not been appreciated how this high loss rate impacts the distribution of copy numbers and associated expression levels in the population, a phenomenon causing antibiotic heteroresistance[11,15]. Moreover, amplifications have been studied only under constant selection for increased expression[16,17], while natural environments are rarely ever constant. While a large body of work suggests that phenotypic heterogeneity serves as an adaptation to fluctuating environments[18,19], it is not known how the genetic heterogeneity resulting from copy number mutations impacts survival in fluctuating environments.

Here, we ask whether the intrinsic genetic instability of gene amplifications allows bacterial populations to tune gene expression in the absence of evolved regulatory systems. To test this idea experimentally we devised a system of fluctuating environmental selection, which selects for the regulation of a model gene. In this fluctuating environment, we track, in real

time, copy number mutations in populations as well as single cells of *Escherichia coli*. Using this system, we test the ability of GDA to effectively tune gene expression levels on ecological timescales, when environmental perturbations occur at rates far too fast for transcriptional gene regulation to emerge *de novo*.

Results

Amplification-mediated gene expression tuning (AMGET) occurs in fluctuating environments

To test whether GDA can act as a form of gene regulation at the population level, we experimentally introduced environmental fluctuations, such that a given level of expression of a model gene is advantageous in one, but detrimental in another environment. As the model gene, we used the dual selection marker *galK*, encoding galactokinase. Expression of *galK* is necessary for growth on galactose, but deleterious in the presence of its chemical analogue, 2-deoxy-galactose (DOG)[20]. Using *galK* with an arabinose-inducible promoter, we mapped the relationship between *galK* expression level and growth in (i) galactose, which selects for high *galK* expression levels and which we refer to as the 'high expression environment'; and in (ii) DOG, which selects for low *galK* expression and which we refer to as the 'low expression environment' (Fig. 1b). In order to establish a strong selective tradeoff between high and low expression, we used 0.1 % galactose for the high expression environment and 0.0001% DOG for the low expression environment in all experiments.

We then constructed a reporter gene cassette to monitor expression and copy number changes of *galK* (Fig. 1c) based on a previously described construct [21]. In this construct, *galK*

86 is not expressed from a promoter but harbors p_0, a randomized 188 bp nucleotide sequence
87 matching the average GC content of *E. coli* instead[21]. This allowed for the selection of
88 increased expression of *galK*. The reporter cassette harbors two fluorophores that allowed
89 us to distinguish the two principal ways of increasing *galK* expression in evolving
90 populations: promoter mutations and copy number mutations (Fig 1c). The promoterless
91 *galK* gene is transcriptionally fused to a yellow fluorescence protein (*yfp*) gene, which
92 reports on *galK* expression. Directly downstream, but separated by a strong terminator
93 sequence, an independently transcribed cyan fluorescence protein (*cfp*) gene provides an
94 estimate of the copy number of the whole cassette (Fig. S1a). We inserted this cassette into
95 the bacterial chromosome, close to the origin of replication (*oriC*) – a location with an
96 intermediate tendency for GDA[21]. However, our results also hold for a second locus, which is
97 flanked by two identical insertion sequence (IS) elements and has a much higher tendency
98 for GDA[21] (Fig. S4).

99 The ancestral strain carrying the promoterless *galK* construct does not visibly grow in the
100 high expression environment. After one week of cultivation at 37°C, mutants with increased
101 *galK* expression appeared (Fig. S1b). We randomly selected one evolved clone with
102 increased CFP fluorescence ('the amplified strain') and analyzed it in detail (see methods) to
103 confirm its amplification. This amplified strain was then used for further experiments in
104 alternating environments (Fig. 2a-c).

105 In all three alternating regimes, which change on a daily timescale, mean CFP levels of 60
106 replicate populations of the amplified strain tracked the environments for the full duration
107 of the experiments. The adaptive change in *galK* copy number (Fig. 2b) occurred within the
108 imposed ecological timescale, rapidly enough to maintain population growth given the daily

dilution bottleneck under all three alternating selection regimes (Fig. S3a). We confirmed the observed changes in copy number using whole genome sequencing (Fig S2b). To understand these population-level observations, we monitored changes in expression of *galK* and *cfp* at the single cell level for two consecutive environmental switches (Fig. 2c). Expression of *galK-yfp* (Fig. S3b) was tightly correlated with the observed changes in gene copy number (Fig. S3c), indicating that gene expression was effectively tuned by GDA. We refer to this phenomenon as amplification-mediated gene expression tuning (AMGET).

AMGET depends on selection acting on a gene copy number polymorphism

The rapid population dynamics observed during environmental switches (Fig. 2c) might simply be explained by selection acting on gene copy numbers with different fitness (Fig. 2d; Supplementary Note). We therefore hypothesized that AMGET occurs because of the intrinsic genetic instability of gene amplifications, which continuously and rapidly generate copy number polymorphisms that selection can act on. Re-streaking a single bacterial colony of the amplified strain resulted in colonies with different CFP levels, sometimes with sectors of different CFP expression levels within individual colonies (Fig. 3a), demonstrating the intrinsic genetic instability of the amplification. Importantly, this genetic instability is dependent on homologous recombination, as a *ΔrecA* derivative of the amplified strain failed to show a decrease in CFP fluorescence (and thus copy number) in response to increasing concentrations of DOG (Fig. S3d). Similarly, *ΔrecA* populations were not able to track fluctuating environments as their *recA wild-type* counterparts did (Fig. S3e).

To determine the rate at which copy number polymorphisms are generated in an amplified population, we followed individual bacteria over ~40 generations in a mother-machine

132 microfluidic device[22,23] and monitored their CFP levels. Mutations in copy number were
133 clearly visible as changes in CFP fluorescence of the mother cell. In approximately 35% of
134 cases, these changes were accompanied by a reciprocal fold-change of fluorescence in the
135 daughter cell (Fig. 3b, Table S1) as expected from unequal crossover[24].

136 In order to quantify the combined rate of copy number gain and loss events by homologous
137 recombination, we analyzed the fluorescence time trace of 1089 mother cells. 55% of traces
138 exhibit constant levels of CFP fluorescence (Fig. 3c – panel 1) indicating stable inheritance of
139 copy number. In about 7% of traces, the constant level of CFP is interrupted by a sudden
140 decrease or increase (Fig. 3c – panel 2-3). The corresponding fold-changes of fluorescence
141 are consistent with gains or losses of entire copies of *cfp*. We estimated the lower bound for
142 the average number of copy number mutations, k_{rec}, to be 2.7×10^{-3} per cell per generation,
143 by automatically selecting only clear step-wise transitions in fluorescence, which are
144 indicative of single copy-number mutation events (Methods, Fig. S5, Table S1). Interestingly,
145 34% of all traces (Fig. S5c) exhibit more complex behaviors (Fig. 3c – panel 4) and cannot be
146 explained in terms of single step transitions.

147 Complex traces are expected to contain more than one duplication or deletion event even
148 under the expectation that copy number variations are independent events (Fig. S5d). In
149 addition, it is conceivable that copy number mutations are not independent, i.e., an
150 increased probability exists for a second mutation after the first copy number increase
151 occurred. However, we cannot exclude the possibility that most of the complex traces are
152 due to expression noise of one or both fluorophores, especially since CFP expression noise
153 increases with copy number. Moreover, microfluidics experiments showed transient growth
154 defects visible as filamentation (Table S1). Given that the amplification includes the origin of

replication (*oriC*), complex traces might in part result from replication issues. Transiently stalled replication forks could result in an overproduction of CFP relative to mCherry, which is located at phage attachment site *attP21*, almost opposite on the *E.coli* chromosome. Thus using only single clear step-wise transitions provides a very conservative lower bound for the rate of copy number mutations.

AMGET requires continual generation of gene copy number polymorphisms

Because the mechanism behind AMGET is selection acting on copy number polymorphism, we asked whether it differs from selection acting on single nucleotide polymorphisms (SNPs). To do so, we artificially created a polymorphic population comprised of an equal ratio of two strains — the ancestral strain with no detectable *galK-yfp* expression and a strain with two SNPs in p_0 (Fig. 1c) resulting in constitutive expression of *galK* (Fig. 4a). Importantly, this 'co-culture' contained standing variation in *galK* expression, but because it is not due to amplification, variation is not replenished at high rates. While the 'co-culture' population tracked short-term environmental fluctuations in a manner similar to the amplified population (Fig. 4b), the long-term dynamics of the two populations were crucially different. Despite being grown from a single cell, the amplified population was able to respond to environmental change rapidly after being maintained in a constant high expression environment for increasingly longer periods (Fig. 4c). The 'co-culture' population, in stark contrast, progressively lost the ability to respond to sudden environmental change (Fig. 4d). While standing variation in the 'co-culture' provided some ability for a population to adapt in the short run, it is only replenished at the rate of point mutations. Hence, this

177 variation – as well as the ability to adapt - is depleted by prolonged selection as the
178 genotype with higher fitness goes to fixation in the population.

179

180 **AMGET is a general and robust mechanism**

181 The experimental results have qualitatively shown that both, gene copy number
182 polymorphism and selection acting on it, are necessary for AMGET to occur. Using
183 population genetics theory, we developed a generic mathematical model to quantitatively
184 predict the observed experimentally observed population dynamics (Fig. 2b). The model
185 describes how gene copy number changes over time in a population under selection. Each
186 copy number is treated as a distinct state, and these states differ with respect to growth
187 rates in each of the two environments. Duplication and amplification events are the only
188 source of transition between states. Importantly, all model parameters (the strength of
189 selection and the rate at which the copy-number polymorphism is introduced as shown in
190 Fig.1a) are obtained from independent measurements (Table S2). Thus, without specifically
191 fitting any parameters, the generic model fully captured the experimentally observed
192 dynamics of AMGET (Fig. 5a, Fig. S6a). The good fit between model and experimental data
193 meant that we could use the model to expand the understanding of the basic conditions
194 under which AMGET can act as an efficient *de facto* mechanism of population-level gene
195 regulation.

196 Qualitatively, the model revealed that for a population to respond to environmental change
197 at all, two conditions must be met: (i) constant introduction of gene copy number variation
198 (i.e. non-zero duplication/recombination rate), and (ii) selection acting on it. If either of

these are not present, the population is not able to maintain any long-term response to environmental change.

In order to more quantitatively examine the environmental conditions under which a population can respond to environmental change through AMGET, we defined the response *R* as the maximum fold change in gene expression before and after an environmental change.

We used the model to expand the range of environmental durations beyond those tested in experiment. In periodic environments, we find a sharp, switch-like transition from no response to full response for environments that switch typically on a day or longer timescale (Fig. 5b). In stochastically fluctuating environments, the transition is more gradual (Fig. 5c), yet no less effective. Furthermore, AMGET maintains its efficiency to tune gene expression in bacterial populations over order-of-magnitude variations in the duplication and recombination rates, as well as for any fitness cost of expression (Fig. S7).

AMGET tunes gene expression levels when transcription factor-based schemes are hard to evolve or maintain

Canonical gene regulation is unlikely to evolve or be maintained when a population is exposed to an almost constant environment that is sporadically interrupted by a rare environmental perturbation[3]. We tested if AMGET might provide a generic mechanism of regulating expression under such conditions, by asking how long a population that is fully adapted to one environment needs for responding to a step-like environmental change (Fig. 5b top and side part of heat map; Fig. S6b). Our model results showed very rapid responses to step-like environmental changes on the order of one to six days, for all biologically

relevant parameter values of amplification and duplication rates, as well as fitness cost of expression (Fig. 5d; Fig. S6c-e). AMGET is also a viable mechanism for practically any population size, especially for typical bacterial ones, although its efficiency drops for small populations (Fig. S6f). Therefore, AMGET efficiently tunes gene expression levels across a wide range of environments where transcription factor-mediated regulation would take prohibitively long to evolve[4,5].

Discussion

Biology often relies on messy solutions, be it due to physical limitations or because evolution proceeds by opportunistic tinkering[25,26]. For organisms living in constantly fluctuating environments even the crudest form of gene regulation[27] or gene expression heterogeneity[28] increases fitness compared to not having any regulation at all. Here, we showed that the intrinsic instability of gene amplifications, rapidly tunes gene expression levels when gene regulation is required but no other molecular regulatory mechanism is in place.

Despite resembling canonical gene regulation when observing populations as a whole (Fig. 2b), AMGET does not allow all single cells to change their gene expression concurrently. Instead, only a fraction of the population grows after the environment changes (Table 1). Thus, AMGET may effectively work by allowing bacterial populations to 'hedge their bets' for expression levels that could be required in a future environment. Unlike traditional descriptions of bet-hedging, where genetically identical individuals show variability in their phenotypic states[19], AMGET populations differ in their genotype due to the intrinsic instability of gene amplifications, thus passing on the adaptive state with high probability.

245 Moreover, bet-hedging is typically characterized by switching between a small number of
246 alternative phenotypic states[19], while in an amplified locus, expression can adopt a graded
247 response due to a wide range of copy numbers.

248 Because AMGET enables rapid dynamics and at the same time graded responses, it can be
249 thought of as a form of primitive gene expression regulation at the population level[29].
250 Mechanistically, AMGET bears no resemblance to canonical gene regulation, which employs
251 sensory machinery to alter gene expression in the course of just a single generation. Yet,
252 despite the mechanistic difference, AMGET operates on the time scales of days and thus
253 closer to those of canonical gene regulation, compared to the process of transcriptional
254 rewiring by point mutations, which occur several orders of magnitude less frequently (Table
255 1).

256 AMGET may be one of several ways by which populations can make use of variation in
257 expression levels to rapidly adapt to environmental change. While point mutations occur at
258 lower rates, regulatory rewiring can be surprisingly fast[30], especially when there is pre-
259 existing variation in the precise architecture of regulatory networks. Moreover, noise
260 propagation within gene regulatory networks can create an abundance of different
261 expression levels, which are – in principle – tunable by selection[28]. However, as the results
262 of our co-culture experiment (Fig. 4) show, pre-existing variation can be easily depleted
263 from a population if under strong selection. While it was previously shown that variation
264 can be maintained in the form of multiple plasmid copies[31], our results highlight that
265 multiple copies of a genomic region actively regenerate heterogeneity due to the high
266 recombination rate. Due to this property, AMGET provides a means of tuning expression to

267 rare environmental fluctuations, where canonical gene regulation cannot evolve or be
268 maintained[3].

269 AMGET is fast in bacteria because their generation times are short and their population
270 sizes are usually large. However, our model results show that AMGET is in principle
271 applicable to any other organism, but would take much longer time in relatively small
272 populations (Fig. S6f). A compelling example for the "up-regulation" of a gene on relatively
273 short evolutionary time-scales is that of the salivary amylase in humans, where variation in
274 AMY1 copy number correlates with dietary starch content of human populations[32].

275 Because any genomic region can be potentially amplified, AMGET can act on essentially any
276 bacterial gene, providing regulation when the promoter is lacking altogether or when the
277 existing promoter is not adequately regulated[33,34]. For instance, horizontally transferred
278 genes tend to be poorly regulated, as their integration into endogenous gene regulatory
279 networks can take millions of years[35,36]. At the same time, they are enriched in mobile
280 genetic elements[37,38], providing repetitive sequences for duplication by homologous
281 recombination[14,39]. Indeed, genes with a recent history of horizontal transfer are often
282 amplified[40–42].

283 Similarly, gene amplifications can confer resistance to antibiotics and pesticides, but they
284 are often accompanied by a fitness cost in the absence of the compound[43]. In fact,
285 heteroresistance caused by copy number polymorphisms is much more prevalent than
286 previously thought and can lead to antibiotic treatment failure[11]. Repeated use of
287 antibiotics or pesticides can therefore create alternating selection regimes[44], where AMGET
288 might play an important, yet previously overlooked, role in bacterial adaptation.

289 In spite of their ubiquity, GDA has been underappreciated [14,45]. In principle, fixed
290 amplifications can easily be detected in next generation sequence data by an increase in
291 coverage and mismatches corresponding to the duplication junctions (Fig. S2, Methods).
292 However, duplications revert to the single copy state at high rate without leaving any traces
293 in the genome (Fig. S2a). This implies that populations have to be kept under selection prior
294 to sequencing, a condition that may not typically be met, especially not for environmental
295 isolates[46]. However, despite this challenge, there are many reports of cases where amplified
296 genes have been detected in the sequences of environmental strains and were found
297 associated with adaptation to environmental conditions [33,40,47].

298 The notion that GDA "might be thought of as a rather crude regulatory mechanism"[29] is
299 more than 40 years old. However, so far almost all experimental work has focused on the
300 benefits of amplification in constant, stable environments, thereby selecting for increased
301 expression only[16,48]. Here, we demonstrated how flexible GDA is in rapidly altering gene
302 expression levels of populations in response to a wide range of environmental fluctuations.
303 AMGET is thus a critical, and a critically underappreciated, mechanism of bacterial survival.

304

305 **Methods**

306 Bacterial strain background construction

307 Except when noted otherwise, all changes to the *E.coli* chromosome were introduced by
308 pSIM6-mediated recombineering[49]. All recombinants were selected on either 25µg/ml
309 kanamycin or 10µg/ml chloramphenicol, to ensure single-copy integration. All resistance
310 markers introduced by recombineering were flipped by transforming plasmid pCP20 and
311 streaking transformants on LB at the non-permissive temperature of 37°C[50]. We used strain

MG1655 for all experiments, except for testing galactose and DOG concentrations (Fig.1c). For that purpose, we placed *galK* under control of the *pBAD* promoter and used strain BW27784, which allows relatively linear induction of the pBAD promoter over a 1000 fold range of arabinose concentration[51]. In both strain backgrounds the genes *galK*, *mglBAC* and *galP* were altered in order to allow galactose- and DOG-selection.

Endogenous *galK* was deleted by P1-transduction of *galK*::kan from the Keio-collection[52]. The *mglBAC* operon was deleted to avoid selective import of galactose but not DOG[53]. To express *galP* for DOG to be imported in the absence of galactose, its endogenous promoter was replaced by constitutive promoter J23100[54]. For this, the fragment BBa_K292001 (available at the Registry of Biological Parts, http://parts.igem.org/Part:BBa_K292001) was cloned into pKD13[50] yielding plasmid pMS1 with FRT-kan-FRT upstream of J23100. The cassette FRT-kan-FRT-J23100 was used for recombineering.

Assembly of the chromosomal gene cassettes

The chromosomal reporter gene cassette used for experimental evolution (p_0-RBS-galK-RBS-yfp-p_R-cfp; Fig. 1c) was assembled on plasmid pMS6* using standard cloning techniques. Plasmid pMS6* is based on plasmid pMS7, which contains the 'evo-cassette' (p_0-RBS-tetA-yfp-p_R-cfp)[21]. To obtain pMS6* we replaced the translational fusion of tetA-yfp on pMS7 with *galK* from MG1655 in a transcriptional fusion with *yfp venus*, originally derived from pZA21-yfp[55]. In addition, XmaI and XhoI restriction sites were added directly upstream and downstream of p_0 by two consecutive inverse PCRs.

The chromosomal gene cassette for testing galactose and DOG concentrations (*pBAD-galK*, Fig. 1b) was assembled on plasmid pIT07, which was obtained by cloning *galK-yfp* as well as a chloramphenicol resistance flanked by FRT sites from pMS6* into pBAD24 [56]. Gene

cassettes were integrated into chromosomal loci 1 and 2 (corresponding to locus D and E in Ref. [21]) by recombineering[49] and checked by PCR with flanking primers and sequencing of the full-length construct.

Strain modification for microfluidics

The amplification of locus 1 was moved from the evolved strain IT028-EE1-D8 to the ancestral background (IT028) by P1 transduction to isolate it from the effect of other potential mutations in the evolved background, including a sticky phenotype, which clogged the microfluidic devices. In order to obtain a single copy control locus p_R-mCherry from our lab collection was introduced into the phage 21 attachment site (*attP21*) by P1-transduction[22].

RecA deletion in amplified strain locus 1 (Fig. S3d,e)

RecA was deleted in the amplified strain by replacing it with the kanamycin cassette from pKD13[50]. In order to maintain the amplified state, recombinants were selected on M9 0.1% galactose medium supplemented with 25µg/ml kanamycin and verified by sequencing.

Culture conditions

All experiments were conducted in M9 medium supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and different carbon sources (all Sigma-Aldrich, St. Louis, Missouri). For evolution experiments 0.1% galactose (high expression environment) or 1% glycerol combined with 0.0001% 2-deoxy-d-galactose (DOG) (low expression environment), respectively, were added as carbon sources. For microfluidics experiments M9 medium was supplemented

with 0.2% glucose and 1% casein hydrolysate and 0.01% Tween20 (Sigma-Aldrich, St. Louis, Missouri) was added as surfactant prior to filtering the medium (0.22 µm).

All bacterial cultures were grown at 37°C. Growth and fluorescence measurements in liquid cultures were performed in clear flat-bottom 96-well plates using a Biotek H1 platereader (Biotek, Vinooski, Vermont).

Mapping the relationship between *galK* expression level and growth growth

For the 2D gradients of arabinose and galactose or DOG (Fig. 1b), respectively, an overnight culture of the test-cassette strain was diluted 1:200 into 96-well plates containing 200 µl of M9 supplemented with carbon sources, DOG and the inducer arabinose, as indicated in Fig. 1b. Cultures were grown in the platereader with continuous orbital shaking.

Evolution experiments

For all evolution experiments (1. experimental evolution of the amplified strains in the high expression environment and 2. alternating selection experiments), cultures were grown in 200µl liquid medium in 96-well plates and shaken in a Titramax plateshaker (Heidolph, Schwabach, Germany, 750 rpm). Populations were transferred to fresh plates using a VP407 pinner (V&P SCIENTIFIC, INC., San Diego, California) resulting in a dilution of ~ 1:133.

1. Evolution of the amplified strains in the high expression environment

To obtain the amplified strains of locus 1 and 2, respectively, an overnight culture inoculated from a single colony of the ancestral strain carrying the reporter gene cassette in the respective loci (IT028; Fig. S1b-c) or 2 (IT030; Fig. S4b) was started in LB-medium. Cells were pelleted, washed twice and diluted 1:100 into M9 0.1% galactose (locus 1) or M9 0.1% galactose supplemented with 0.1% casamino acids (locus 2). For locus 1, the timing of each dilution into fresh medium (~1:133) was chosen such as to maximize the number of rescued

populations and to minimize the amount of time spent in stationary phase for grown populations. The transfers happened at days 10, 13, 15, 17, 18 and 19 (Fig. S1c). The first signs of growth were detected in several wells only after approximately one week of cultivation in minimal galactose medium (Fig S1b). The evolving populations were monitored by spotting them onto MacConkey galactose agar in 128 x 86mm omnitray plates prior to transfer. For locus 2, the evolving populations were transferred daily (~1:133, corresponding to seven generations) and spotted on to LB plates supplemented with 0.5% charcoal (Fig. S4b) to improve fluorescence quantification. Colony fluorescence of all experiments was recorded using a custom-made macroscope set-up (https://openwetware.org/wiki/Macroscope)[57]. For the isolation of clones, evolved populations were streaked twice for purification on LB agar and grown in M9 galactose medium prior to freezing. For both locus 1 and 2, respectively, all further experiments were started from the original freezer stock of the amplified strain. This was done for two practical reasons: i) to save the time needed for duplications (and higher order amplifications) to evolve (one week in M9 galactose medium used for locus 1 and one day in M9 medium supplemented with casaminoacids used for locus 2), and more importantly, ii) to allow interpretation and reproducibility of the fluorescence data of the alternating selection experiments. As the reporter gene cassette allows selecting for increased *galK* expression but not for amplification itself, it is necessary to screen mutants with increased *galK* expression for increased CFP fluroescence. During amplification the initial duplication step is rate-limiting and break-points differ between evolving populations. We therefore limited ourselves to two amplified strains (locus 1 and 2), which we analyzed in detail. Amplified populations were thus started from single colonies, which were grown non-selectively on LB (Lennox) agar by streaking the original freezer stock. Due to the high rate

of recombination, any given streak of the original amplified freezer stock contains colonies with a single copy of *galK* (Fig. 3a, right panel). In order to pick only amplified colonies, we examined CFP fluorescence using the macroscope.

We characterized evolved amplified strains by Sanger sequencing of the p_0 region, amplification junctions and the *rho* gene, which was found mutated in a previous study using the same locus [21]. For the strain amplified in locus 1 (IT028-EE1-D8), increased *galK* expression is achieved by increased *galK* copy number as evident from increased CFP fluorescence (Fig. 1c), as well as through a missense mutation in the termination factor *rho* (S265>A), allowing for baseline-expression via transcriptional read-through from the upstream *rsmG* into *galK*[21]. The amplified region spans 16 kb from *atpB* at the left replicore over the origin of replication to *rbsD* into the right replicore.

For the strain amplified in locus 2 (IT030-EE11-D4), *galK* expression comes solely from the increase in copy number (no mutations in p_0 were detected). In this case, inverse PCR and sequencing confirmed that two identical IS elements (*IS1B* and *IS1C*) form the junction of the amplified segment[21]. Whole genome sequencing of both amplified strains confirmed amplification junctions and the *rho* mutation detected with PCR and Sanger sequencing and revealed two additional single nucleotide changes in the amplified strain locus 1 (*coaA*, pos. 4174770, C>T, resulting in R>H; *wcaF*, pos. 2128737, C>A, resulting in G>V).

2. Alternating selection experiments

For the experiments in Fig. 2b, a pre-culture of the amplified strain (IT028-EE1-D8) was grown in M9 0.1% galactose overnight, which was then inoculated 1:200 into the medium as indicated. For the experiment alternating two days in high and one day in low expression

environment (Fig. 2b – middle panel), populations were first subjected to a scheme of daily alternating selection for six days prior to switching to the 2-1 scheme.

For the co-culture experiments (Fig. 4), a pre-culture of the amplified strain (IT028-EE1-D8) was grown in M9 0.1% galactose overnight. In parallel, the ancestral strain carrying a single silent copy of *galK* in locus 1 (IT028) and a strain constitutively expressing *galK* in locus 1 (IT028-H5r), were grown overnight in M9 1% glycerol and mixed in a 1:1 ratio. We labeled the ancestral strain by transduction of *attP21*::p_R-*mCherry* (IT034). The constitutive strain was obtained by oligo-recombineering two point mutations into p_0 of the ancestral strain and selecting recombinants on M9 0.1% galactose agar. These two point mutations (-29 A>T and -37 G>T) have initially evolved in parallel to the amplified strain and result in a similar level of *galK* expression (Fig. 1c).
To quantify the relative abundance of the two strains in the co-culture, we calculated the expression ratio of the two strains, using an exchange rate between CFP and mCherry units from the ancestral strain expressing both fluorophores (IT034).

Whole genome sequencing

We isolated gDNA from overnight cultures of single clones of i) the ancestral strains ii) the amplified strains after initial selection in the high expression environment (galactose) as well as iii) the amplified strains after overnight selection in the low expression environment (DOG), for Locus 1 and Locus 2, respectively. In all cases overnight cultures were inoculated from colonies grown non-selectively on LB agar. For the overnight culture M9 1% glycerol was used for the ancestral and DOG-selected clones, while M9 0.1% galactose was used for the galactose-selected clones. A whole genome library was prepared and sequenced by

Microsynth AG (Balgach, Switzerland) on an Illumina Next.Seq (with a mean read length of 75 bp). Fastq files were assembled to the MG1655 genome (Genbank accession number U00096.3) using the Geneious alignment algorithm with default options of the software Geneious Prime version 2019.2.1. SNPs were analyzed using the variant finding tool of Genious.

Flow Cytometry

Three colonies of the amplified strain and the constitutive control strain, respectively, were inoculated into culture tubes with 2ml M9 0.1% galactose (high expression environment) and grown for three days with transfers every 24h. This population was inoculated into M9 + 1% glycerol + 0.0001% DOG (low expression environment). OD_{600} was monitored to assure continuous exponential growth by regular dilutions. Samples for flow cytometry were frozen at the indicated time points (Fig. 2c). After 24h in the low expression environment, the populations were transferred back to the high expression environment with dilution and sampling occurring in the same manner. In parallel, the positive controls were grown for five days in both selection environments, respectively, with transfers occurring every 24h. Fluorescence was measured using a BD FACSCanto™ II system (BD Biosciences, San Jose, CA) equipped with FACSDiva software. Fluorescence from the Pacific Blue channel (CFP) was collected through a 450/50nm band-pass filter using a 405nm laser. Fluorescence of the FITC channel (YFP) was collected through a 510/50 band-pass filter using a 488nm laser. The bacterial population was gated on the FSC and SSC signal resulting in approximately 6000 events analyzed per sample, out of 10,000 recorded events.

Microfluidics experiments

For the microfluidics experiments, a single colony of the amplified strain was picked and grown overnight in nonselective LB (Lennox) medium.

Microfluidics devices were prepared as described previously[22]. Briefly, devices had dimensions 23 µm × 1.3 µm × 1.3 µm (*l*, *w*, *h*) for the growth channels with 5 µm spacing along a trench for growth medium. Devices were fabricated by curing degassed polydimethylsiloxane (Sylgard 184, 1:10 catalyst:resin) inside epoxy replicate master molds produced from primary wafer-molded devices. Microscopy was performed on an inverted Nikon Ti-Eclipse microscope and with a previously described set-up[22]. Per experiment, multiple positions of a single mother machine were imaged using a 60x 1.4 NA oil immersion objective lens. To image constitutive mCherry, the green LED (549+/-15nm) was used at a light intensity of 670µW and an exposure time of 170-200ms. To image CFP, the cyan LED (475+/-28nm) at a light intensity of 270µW and an exposure time of 90-100ms was used.

Analysis of microfluidics data

The mother machine allowed tracing of mother cells for ~ 38 divisions, thereby following the fate of arising copy number mutations in the absence of selection. In three experiments, we analyzed 336, 369 and 384 mother cell lineages, respectively, equaling a total of approximately 40.000 cell divisions (with a division time of 23.6 (+- 1.5) min as determined by counting septation lines in growth channel kymographs).

Microfluidics data analysis was based on mother cell time traces (Fig. 4c). To this end, we used Fiji/ImageJ to create kymographs, by laying a line through the middle of mother cells perpendicular to the growth channel using the built-in Multi-Kymograph tool with a pixel width of 9. Kymographs of CFP and mCherry were then analyzed using MATLAB.

Determining what data to include

To minimize the influence of three unknown factors (maturation rate and bleaching of the two fluorophores, and the degree of bleedthrough between channels on the microfluidic chip), we were restrictive with the colonies we included.

1. We excluded all fluorescence changes that occured when the cells were dying. Only colonies (mother cell lineages) that continuously grew until the end of the experiment were included. Specifically, the last 10 frames of mean mCherry fluorescence of mother cells needed to exceed the background threshold (68%, 76%, 82% of total colonies included, respectively, for the three experiments).

2. Some colonies exhibited a large variation in growth rate, due to temporary slowdown and/or filamentation. In the kymographs this was seen as a large variance in the constitutive mCherry channel. We excluded colonies with a variance > 1.5 times the mCherry experiment-wide variance (thus including 96%, 96%, 96% of total colonies included for the three experiments, respectively).

3. In some cases there was significant bleedthrough between adjacent colonies. To avoid double counting transitions, the colony that was less bright was removed from the data set if two adjacent colonies had a correlation of 0.6 or higher (99%, 98%, 98% of total colonies included, respectively, for the three experiments).

For the identified colonies the maximum fluorescence value per time point was extracted for both, mCherry and CFP channels. These were plotted against each other and a rectangular area, bounded by a manually selected max and min for each channel was

chosen such as to include all but extreme outliers (Fig. S5a). Accordingly, 99% of data points were included in all three experiments.

Normalization

To correct for slow temporal drift in the signal of CFP and mCherry, a time average over all colonies was taken and a 7th degree polynomial fitted. All time points were divided by the corresponding polynomial estimates.

Furthermore, mCherry fluorescence was flat-field corrected based on the expectation that mCherry is roughly constant across all colonies. To do so, a line was fitted to the coordinate to get an estimate of the backround of each location. The data was divided by the corresponding estimated value.

Probability density function

For the probability density function (PDF) in Fig. S5b we normalized for differential growth rate by dividing the CFP fluorescence by the constitutively expressed mCherry fluorescence. To reduce noise, a median filter (MATLAB medfilt1) was applied to the ratio of CFP and mCherry over 20 data points.

To get an estimate of the PDF of the CFP/mCherry single cell fluorescence, we used a kernel density estimation (KDE) (MATLAB function ksdensity). To estimate a proxy for copy numbers, we found points where the first and second derivative of the PDF is zero. These points were set as initial conditions for a pairwise fitting of peak mean and variance. All but the first and the last peak had two estimates for mean and variance. For the mean, the average of the two was taken and for the variance the smaller one was chosen. To assign boundaries for states, the estimated variance was halved. For plotting, the height of each

peak was set to match the peak height. No weight was fitted. The mean inter-peak distance for each PDF was used as a proxy of copy numbers for plotting in Fig. 4c.

Estimation of nS2R2 for classification of single cell traces

We have classified the single cell traces using a normalized R squared, the proportion of variance explained, which we call nS2R2. In this adjustment, each element in both the residual and the total sum of squares is normalized by the predicted value: nS2R2 = $1-S^{norm}_{res}/S^{norm}_{total}$, where $S^{norm}_{res} = \sum_i (y_i-f_i)^2/f_i^2$, $S^{norm}_{res} = \sum_i (y_i-y_0)^2/f_i^2$, where y_i, f_i, and y_0 represent measurements, fitted/predicted values, and mean of the measurements, respectively. This normalization takes into account that the intrinsic noise increases with expression and thus penalizes it less. Next, the algorithm fits one constant to the start and one constant to the end value of the CFP/mCherry trace, and reports this estimation parameter (nS2R2) based on which it classifies traces as shown in the pie charts of Fig. S5c. Clear transitions exhibit an nS2R2 score of >0.5 and were verified by eye analyzing microfluidics movies in detail (Table S1). The algorithm classifies no-events ("flat lines") if the nS2R2 score lies between 0 and 0.5. Traces, which cannot be classified unambiguously neither as clear transition nor as a clear no-event, i.e. with nS2R2 below 0, are classified as "complex traces". This occurs if the start and end of CFP/mCherry trace values are similar but vary significantly in between.

Quantitative PCR

For qPCR, DNA was isolated using Wizard Genomic DNA purification kit (Promega, Madison, Wisconsin) from 50 ul of frozen samples from different time points (1,4,9,10,11, gal 10, single copy control, DOG 8, DOG 10) of one flow cytometry experiment grown for 4-5

574 generations in LB. To quantify fluorescence, the same cultures were patched onto LB agar
575 supplemented with 0.5% charcoal and imaged using the macroscope.
576 We performed qPCR using Promega qPCR 2x Mastermix (Promega, Madison, Wisconsin) and
577 a C1000 instrument (Bio-Rad, Hercules, California). To quantify the copy number of samples
578 of an evolving population, we designed one primer within *cfp* (target) and used one primer
579 within *rbsB* as a close reference, which lies outside the amplified region. We compared the
580 ratios of the target and the reference loci to the ratio of the same two loci in the single copy
581 control. Using dilution series of one of the gDNA extracts as template, we calculated the
582 efficiency of primer pairs to be 89.01% and 92.57%, for *cfp* and *rbsB*, respectively. We
583 quantified the copy number of *cfp* in each sample employing the Pfaffl method, which takes
584 amplification efficiency into account[58]. qPCR was done in three technical replicates.
585
586 Measurement of colony fluorescence (Fig. S1c, Fig. S4b, Fig. 3a)
587 Colonies were grown without selection and imaged using the macroscope set up.
588 To obtain mean colony CFP fluorescence intensity, a region of interest was determined
589 using the ImageJ plugin 'Analyze Particles' (settings: 200px-infinity, 0.5-1.0 roundness) to
590 identify colonies on 16-bit images with threshold adjusted according to the default value.
591 The region of interest including all colonies was then used to measure intensity.
592
593 Mathematical model
594 A simple mathematical model recapitulates the change in *galK* copy number of the
595 amplified population (Fig. 5a). Importantly, the parameters for the model were estimated
596 purely from calibration measurements (growth rates, fitness in the two environments with
597 respect to copy number (flow cytometry experiments), number of generations spent in each

environment, and recombination rate, k_{rec}) and the literature (k_{dup}, [14]). Their values are listed in Table S2. No parameter was fit to reproduce the measurements in Fig. 5a.

The model describes the time evolution of a population where cells with different gene copy numbers are represented by distinct states. The duplication and amplification events are the only source of transition between states. The time evolution proceeds iteratively, with discrete times representing synchronous cell divisions in the population.

The size of subpopulation N_j of cells with gene copy number j at time $t+1$ equals:

$$N_j(t+1) = \overbrace{\underbrace{(1-k_{rec}s_j)N_j(t)}_{\text{no duplication of amplification event}}}^{\text{daughter 1}} + \overbrace{\underbrace{(1-k_{rec}-k_{dup}\delta_{j,1})s_jN_j(t)}_{\text{no duplication of amplification event}}}^{\text{daughter 2}} + \underbrace{\sum_{k=2}^{M} k_{rec}P_{kj}s_kN_k(t)}_{\text{amplification event}} + \underbrace{k_{dup}s_1N_1(t)\delta_{j,2}}_{\text{duplication event}} \qquad (1)$$

where s_j is the relative growth rate of the subpopulation with j gene copies in the given environment (taken from Fig. 2d), δ_{jk} a Kronecker delta which equals 1 if $j=k$ and 0 otherwise. The equation for single and double gene copy numbers (j=1 or j=2, respectively) has an additional term to reflect duplication events. As we assume that the rate of recombination per copy is constant, the overall recombination is proportional to the number of gene copies k; $k_{rec}=k\ k_{rec}^0$ (ref [8]). P_{kj} represents the transition probabilities given an amplification event and is computed in the following way: assuming a homologous recombination between sister chromosomes occurs somewhere in the gene, we computed all possible combinations of how genes can be recombined to form different number of gene copies between the two daughter cells. P_{kj} then represents the probability that, given a recombination event, a daughter cell obtains j gene copies with its mother having k of them before the event. For example, starting with three gene copies, there is 22% probability to obtain four gene copies, or 22% probability to have one copy in the daughter (Fig. S6h). We have observed in microfluidics experiments that most (65%) copy number changes happen

only in the mother cell while the daughter cell remains unchanged. Therefore, we do not model recombination as a reciprocal event.

Based on platereader bulk experiments, observations indicated an upper limit for the copy number a cell can have. Thus, in our model, a cell can have up to *M* gene copies; if that number is exceeded, the cell stops dividing. This upper limit for gene copy number was confirmed in microfluidics and qPCR experiments, indicating to be between 6 and 12. Our single cell analysis showed that *M*=10 is a good estimate (Fig. S5b, according to number of states in the probability density function, see *Analysis of the microfluidics data*). However, the results of the mathematical model do not depend on the precise value within the measured range, as all results remain qualitatively the same for any value in the range of 6 and 12. Fig. S6g shows that relative growth rates, obtained from flow cytometry experiments, are independent of *M*.

Measurements of model parameters (Table S2)

T1 & T2, generations per day in 96 well plates

In order to model the alternating selection experiment (Fig. 5a), we needed to know the maximal growth rate of the amplified strain (IT028-EE1-D8) in the high and low expression environments, respectively. Because the exact details of cultivation (such as culture volume, shaking speed and temperature fluctuations) strongly affected growth rate, we were unable to measure growth curves while keeping cultures under the conditions of the original experiment. Hence, we estimated growth rate indirectly without perturbing the experiment, by determining the maximal number of generations possible in 24h (number of generations = 24[h] * growth rate[1/h]/log(2)) from a dilution series experiment. Populations pre-adapted to the respective environment were grown to carrying capacity of the respective

medium and diluted by a factor of approximately 2^n (with n ranging between 7 and 28). We sought the maximal dilution that could still be compensated by growth (by requiring after 24h of growth the OD_{600} to reach the OD_{600} of the stationary phase). All dilutions of equal to or less than $1:2^{22}$ and $1:2^{23}$ were able to reach stationary phase in the high and low expression environment, respectively, yielding model parameters T1=22 and T2=23 for the maximal possible number of generations.

T10 & T20, generations per day in culture tubes

Parameters T10 and T20 were necessary for obtaining the fitness landscape in Fig. 2d (and the resulting relative growth rates s_j). T10 and T20 generations per day, measured under the exact conditions of the flow cytometry experiment (Fig. 2c), namely exponential growth in culture tubes with 2ml volume of M9 0.1% galactose or M9 1% glycerol + 0.0001% DOG, respectively. We measured OD_{600} with a WPA Biowave spectrophotometer (Biochrom, UK).

Determining fitness landscape and relative growth rates s_j

The relative growth rates for each genotype (copy number state) in the high and low expression environments, respectively, were computed from flow cytometry time series experiments assuming exponential growth with no duplication/amplification event (k_{dup}=0, k_{rec}=0). This is a valid approximation as long as the two rates are small enough, such that the population structure consists of all copy number types, i.e., that each subpopulation is much larger than the additional cells created by a single amplification or duplication event.

The flow cytometry measurements of the distribution of CFP expression at different times were split in *M* equal-width bins. The lowest and highest bins were set according to the equilibrium fluorescence distribution in DOG and galactose, respectively. For the lowest bin,

669 we took the values of fluorescence <85, while for the high bin we took the mode
670 fluorescence values of the measured distributions, corresponding to >160 for the first, and
671 >245 for the second set of flow cytometry experiments. Each bin represents a given gene
672 copy number. The distributions between different times were then compared using iterative
673 exponential growth model:
674
675 $$N_j(t_2) = (1+s_j)^{(t_2-t_1)/t_{1/2}} N_j(t_1) \qquad (2)$$
676
677 where N_j is the population size with *j* gene copy number, $t_{1/2}$ is the doubling time, t_1 and t_2
678 are two measurement times, and s_j represents the relative growth of cells with *j* gene
679 copies. The population distributions for all time points were obtained from the flow
680 cytometry data given the binning described above. Using this model, we obtained growth
681 rates s_j for each pair of consecutive distributions at times t_i and t_{i+1} in the following way:
682 given population distribution at time *i*, we predicted the new distribution given Eq. (2). We
683 found such s_j values that minimize the Euclidian difference between the predicted and
684 observed population distribution at time *i+1*. We repeated this for all pairs of consecutive
685 distributions (at times t_i and t_{i+1}) and different replicates to obtain a set of solutions for s_j.
686 Using this approach, we acquired only relative growth rates, which still allowed constants to
687 be added to the growth rates. To tackle this, we added such constants to each growth rates
688 in order to i) minimize the χ^2 of the differences between each growth rate solution and the
689 mean of all solutions, which optimally removes the replicate-to-replicate variability (error
690 bars in Fig 2D) on the inferred relative growth rates but does not affect their mean value;
691 and ii) force the average growth rate of the adapted state to be 1 (i.e., for *j*=1 in low
692 expression environment and *j*=M is high expression environment, s_j=1) by adding a term to

693 the χ^2 error function of the form (adapted state expression - 1)^2. Fixing *s* to be 1 in a
694 reference environment is a convention that mathematically will not affect any subsequent
695 results.

696 The absolute maximal growth rates in the two environments were measured in
697 populations grown in high and low expression environments for 120h, respectively. Thus,
698 they represent the growth rates of populations with the highest and lowest possible copy
699 number (Fig. 2c, positive controls). The estimated fitness values for both high expression
700 environment (s_j^{HEE}) and low expression environment (s_j^{LEE}) can be found in Table S2.

701

702 **Estimation of recombination rate k_{rec} from microfluidics data**

703 We obtained a conservative estimate for the lower bound for the average number of copy
704 number mutations from single step transitions in the pie charts (Fig. S5c). Out of 72 mother
705 cell time traces classified as clear transition events, we verified 67 by detailed analysis of
706 microscopy images (Table S1). We accordingly calculated the lower bound for the mutation
707 rate as 67 events/1089 lineages/22.7 generations yielding $k_{rec} = 2.7*10^{-3}$ (+-$7.4*10^{-4}$) per cell
708 per generation.

709 To estimate the mean recombination rate to be used in the model, two corrections have to
710 be made: i) because our model assumes that the recombination rate is proportional to the
711 number of gene copies [8], we had to take into account that cells with higher initial gene copy
712 number are more likely to undergo a recombination event; and ii) as our experimental setup
713 only allowed us to see if there has been a change in gene copy numbers or not, we had to
714 take into account that there are some recombination events that do not change the gene
715 copy number.

To account for i), we first computed the probability distribution that a given number of independent recombination events occur (Fig. S5d): given the assumed independence of recombination events, the probability of observing a certain number of recombination events for a given cellular trace is approximately Poisson distributed, with the parameter being the expected number of events per microfluidic experiment duration (i.e., the effective recombination rate times the number of generations). The total number of observed generations was: 37.7, 36.3, and 41.3 for the three microfluidics experiments, respectively. Our approach is an approximation, namely it assumes a constant effective recombination rate for each trace throughout the experiment, which can be violated if more than one recombination event occurs. For example, the first recombination event can change the gene copy number, which in turn changes the probability of subsequent recombination events happening. While it is in principle possible to take this into account, it substantially complicates the inference of the recombination rate from data and makes it strongly model dependent.

As per our model assumption, the effective recombination rate is equal to the initial number of gene copies times the basal recombination rate. Therefore, we used all single cell traces to estimate a starting gene copy distribution. To do this, we averaged the normalized fluorescence (as a proxy for the starting effective gene copy number, see Fig. 3c) over the time points 20 through 50. Next, we computed a Poisson probability distribution of obtaining k events (k=0,1,…) in the time of the experiment for each individual trace, with the basal recombination rate multiplied with the starting gene copy number (Fig. S5d). For example, if a single cell trace started with 4 gene copies, the expected number of events per experiment would be 4 times the basal recombination rate times the number of generations. Next, we averaged over all computed Poisson probability distributions,

obtained from all single cell traces. This effectively means obtaining a total probability distribution for seeing 0, 1, or more recombination events over all recorded single-cell traces, taking into account point i).

Next, we consider point ii), taking into account the effect of recombination events that do not change the gene copy. We know from the *Pkj* matrix that the probability of keeping the gene copy numbers is the reciprocal of the initial gene copy number. Therefore, we took into account all events that would be seen as zero or single events (but are not) and adjusted the probability distributions. For this, we defined two probability distributions: the distribution of observed events, $p_{observed}$, which we are trying to find; and the distribution of "actual" number of events, p_{actual}, which we computed as described above. For example, in the observed distribution that is compared with experimental data, we classified as single events all double events where one of the recombination events leaves the copy number unchanged, all triple events where two events keep the copy numbers unchanged, etc. Therefore, the probability of observed events also includes the actual probability from states with $k>0$ in which recombination did not change the copy number: $p_{observed}(k=0) = p_{actual}(k=0) + \sum_j p_{actual}(j) / \varepsilon_0^j$, for all $j>0$, with $p(j)$ being the probability of having j recombination events, and ε_0 being the initial gene copy number in the given single cell trace (estimated from experimental single cell traces). The $(1/\varepsilon_0)^j$ represents the probability of having j consecutive recombination events, all of which leave the gene copy number unchanged. Analogously, the observed probability for a single event ($k=1$) to occur is: $p_{observed}(k=1) = p_{actual}(k=1) + \sum_j (j-1)p_{actual}(j) / \varepsilon_0^{j-1}$, for all $j>1$. The prefactor (j-1) comes from the number of different possibilities of having events that keep the gene copy number unchanged. For example, having 3 recombination events, there are 3 different ways of having two events that keep the gene copy number unchanged while one event changes it.

After taking both corrections into account, we obtain a probability distribution of observing *k* recombination events (Fig. S5d). The estimate of the basal recombination rate, k_{rec}^{0}, is based on the proportion of traces classified by our algorithm as no mutation events. We looked for such a recombination rate that best matched the number of no-events in the probability distribution (Fig. S5c-d). We obtained k_{rec}^{0} as 0.01434 per cell per generation, which is approximately 5x larger than the conservative lower bound.

Model comparison with experimental data

For comparison of the model with the experimental data (Fig. 5a), we simulated the full experimental protocol (for parameter values, see Table S2):

1) We exposed a single copy, ancestral population to a week of high expression environment, driving the population structure close to equilibrium. This mimicked the evolution of the amplified strain in the high expression environment such that both experimental and simulated population started with the same degree of copy number polymorphism.

2) The population spent one day in the low environment (for details on procedure in each day, see below).

3) For the experiment shown in Fig. 5a top panel, the population was additionally exposed to three daily oscillations between high and low expression environment.

4) The population was exposed to the environments indicated in Fig. 5a.

5) For every experiment, bacterial culture was diluted by a factor of D=133 every day, thus limiting growth. This growth limitation was enforced by multiplying all growth rates by $g(c) = (1 - \min(c/133,0))^{0.01}$, with *c* being the number of cells, relative to the number of

cells after each dilution. The exponent 0.01 was chosen such that *g(c)* was smooth but nearly a step function.

6) To compare the units of experimental and simulated data, we obtained a common reference point. We took this to be the expression value after one week in the high expression environment, when the population has already equilibrated. We aligned these two points to have the same expression value. This value varies between different experiments.

The simulation of one day consisted of (for parameter values see Table S2):

1) Given the recombination rate and number of states *M*, we computed the transition matrix P_{kj} (see Eq. 1) in the following way: given *k* copy numbers, the probability of going from *k* to *j*<*k* copy numbers equals j/k^2, while probability for *k* to *j*>= equals $(2k-j)/k^2$ [8]. Furthermore, we assumed that no transitions that increase copy numbers beyond *M* are allowed. We implemented this by setting all probabilities that go over *M* gene copies to zero.

2) Next, to update the current population structure following Eq. 1, we used the current population structure, N_j, selection on the states (growth rates) in the given environment, s_j (Fig. 2d), transition matrix, P_{kj} (probability of having j copies given k copies), the duplication and recombination rate (k_{dup} and k_{rec}, respectively), and the dilution factor D. First, we computed the total population growth since the last dilution, i.e., the ratio of population size of current time point and the size after last dilution. Second, we computed *g(c)* (taking into account the saturation of the population) and multiplied it with each of the selection values s_j in Eq. 1. Then, we used these new values to compute N_j at the new time point.

3) We repeated the step 2 for 23 or 22 times for low or high expression environment, respectively. These numbers represent the number of cell divisions per day and were determined experimentally. Steps 2-3 represent time evolution of the population over the period of one day.

4) We diluted the population by a factor of D=133.

5) We repeated the steps 2-4 according to the environment the population is exposed at on the new day (selection different between the two environments). With this step, we simulate different days, diluting after each (step 4).

6) For each time point, we computed expression as the average gene copy number: $E=\sum j w_j$, where w_j is the the proportion of cells with j gene copies and sum goes over all gene copy numbers.

7) At the end, we returned the population distribution and expression at each time point.

For simulation of the stochastic environmental durations, we followed the same procedure as for the deterministic ones, except that the environment durations here were randomly drawn from an exponential distribution.

Finite size population model

To compute the response times for a finite size population (Fig. S6f), we used the Wright-Fisher model where the population size is kept constant. The procedure was:

1) Given all parameters of the system and using the infinite size population model (Eq. 1), we obtained the equilibrium distribution of the population in the starting environment. We computed the equilibrium distribution of copy numbers in the

infinite population size limit by computing the eigenvector corresponding to the largest eigenvalue of the transition matrix (obtained from r.h.s. of Eq. 1), and obtained the starting finite population as a multinomial draw of N individuals from this equilibrium distribution.

2) After the environmental transition, we updated the distribution after each division. The new distribution was computed using the Eq. 1.

3) We computed the new population, as a multinomial draw of N individuals, randomly drawn from the new population distribution.

4) After each division, we computed the expression of the population.

5) We repeated steps 3-5 until response R=M/2 has been reached. The number of generations until this point represents the time to response. We define response as the ratio of mean copy numbers before and after the environmental switch.

Fig. S6f shows the response time as the average over 100 replicate simulations of the algorithm above.

Quantification and Statistical Analysis

Statistical details of individual experiments, including number of replicate experiments, mean values, and standard deviations, are described in the figure legends and indicated in the figures.

For the *t*-test in Fig. 4c-d we computed the response as the fold change between mean expression of days 1-15 in the high expression environment and mean expression in the low expression environment on day 16 for amplified populations (Fig. 4c). For the co-culture populations (Fig. 4d), we analogously computed the response as fold change between mean

859 constitutive strain abundance of days 1-15 in the high expression environment and mean
860 constitutive strain abundance in the low expression environment on day 16.
861 We used a two-sided *t*-test (Matlab function ttest2) to compute the p-value ($2.6*10^{-68}$) for
862 the difference in mean response between amplified (Fig. 4c) and co-culture populations (Fig.
863 4d).
864 For measuring the linear dependence between the experimental data and model prediction
865 in Fig. 5a, we computed the Pearson correlation coefficient using the inbuilt Matlab function
866 corrcoef.

867

868 **Author contributions**

869 CCG, RG, ML, GT, IT conceived study. IT carried out experiments. AMCA, RG, IT analyzed
870 data. RG, GT did the formal analysis. RG, IT wrote original draft and revised with AMCA, JPB,
871 CCG, ML, GT.

872

873 **Data availability**

874 Experimental data that support the findings of this study have been deposited in IST
875 DataRep and are publicly available at https://doi.org/10.15479/AT:ISTA:7016.

876

877 **Code availability**

878 All simulation and analysis scripts are available from the corresponding author upon
879 request.

880

Competing interests

Authors declare no competing interests.

Acknowledgments

We thank Laurence Hurst, Nick Barton, Maros Pleska, Magdalena Steinrück, Bor Kavcic, and Anna Staron for input on the manuscript, and Tobias Bergmiller and Remy Chait for help with microfluidics experiments. IT is a recipient the OMV fellowship. RG is the recipient of a DOC (Doctoral Fellowship Programme of the Austrian Academy of Sciences) Fellowship of the Austrian Academy of Sciences.

References

1. Moxon, E. R., Rainey, P. B., Nowak, M. A. & Lenski, R. E. Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr. Biol.* **4**, 24–33 (1994).

2. Savageau, M. A. Genetic regulatory mechanisms and the ecological niche of Escherichia coli. *Proc. Natl. Acad. Sci. U. S. A.* **71**, 2453–5 (1974).

3. Gerland, U. & Hwa, T. Evolutionary selection between alternative modes of gene regulation. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8841–6 (2009).

4. Tuğrul, M., Paixão, T., Barton, N. H. & Tkačik, G. Dynamics of Transcription Factor Binding Site Evolution. *PLOS Genet.* **11**, e1005639 (2015).

5. Berg, J., Willmann, S. & Lässig, M. Adaptive evolution of transcription factor binding sites. *BMC Evol. Biol.* **4**, 1–12 (2003).

6. Anderson, P. & Roth, J. Spontaneous tandem genetic duplications in Salmonella

typhimurium arise by unequal recombination between rRNA (rrn) cistrons. *Proc. Natl. Acad. Sci.* **78**, 3113–3117 (1981).

7. Reams, A. B., Kofoid, E., Savageau, M. & Roth, J. R. Duplication Frequency in a Population of Salmonella enterica Rapidly Approaches Steady State With or Without Recombination. *Genetics* **184**, 1077–1094 (2010).

8. Pettersson, M. E., Sun, S., Andersson, D. I. & Berg, O. G. Evolution of new gene functions: simulation and analysis of the amplification model. *Genetica* **135**, 309–324 (2009).

9. Sun, S., Ke, R., Hughes, D., Nilsson, M. & Andersson, D. I. Genome-Wide Detection of Spontaneous Chromosomal Rearrangements in Bacteria. *PLoS One* **7**, e42639 (2012).

10. Roth, J. R. *et al.* Rearrangements of the Bacterial Chromosome: Formation and Applications. *Escherichia coli Salmonella Cell. Mol. Biol.* 2256–2276 (1996). doi:10.1126/science.3045970

11. Nicoloff, H., Hjort, K., Levin, B. R. & Andersson, D. I. The high prevalence of antibiotic heteroresistance in pathogenic bacteria is mainly caused by gene amplification. *Nat. Microbiol.* **4**, 504–514 (2019).

12. Bass, C. & Field, L. M. Gene amplification and insecticide resistance. *Pest Manag. Sci.* **67**, 886–90 (2011).

13. Albertson, D. G. Gene amplification in cancer. *Trends Genet.* **22**, 447–455 (2006).

14. Andersson, D. I. & Hughes, D. Gene Amplification and Adaptive Evolution in Bacteria. *Annu. Rev. Genet.* **43**, 167–195 (2009).

15. Hjort, K., Nicoloff, H. & Andersson, D. I. Unstable tandem gene amplification

924 generates heteroresistance (variation in resistance within a population) to colistin in
925 Salmonella enterica. *Mol. Microbiol.* **102**, 274–289 (2016).

926 16. Näsvall, J., Sun, L., Roth, J. R. & Andersson, D. I. Real-time evolution of new genes by
927 innovation, amplification, and divergence. *Science* **338**, 384–7 (2012).

928 17. Elde, N. C. *et al.* Poxviruses deploy genomic accordions to adapt rapidly against host
929 antiviral defenses. *Cell* **150**, 831–41 (2012).

930 18. Kussell, E. & Laibler. Phenotypic Diversity, Population Growth, and Information in
931 Fluctuating Environments. *Science (80-.).* **309**, 2075–2078 (2005).

932 19. Veening, J.-W., Smits, W. K. & Kuipers, O. P. Bistability, Epigenetics, and Bet-Hedging
933 in Bacteria. *Annu. Rev. Microbiol.* **62**, 193–210 (2008).

934 20. Barkan, D., Stallings, C. L. & Glickman, M. S. An improved counterselectable marker
935 system for mycobacterial recombination using galK and 2-deoxy-galactose. *Gene* **470**,
936 31–36 (2011).

937 21. Steinrueck, M. & Guet, C. C. Complex chromosomal neighborhood effects determine
938 the adaptive potential of a gene under selection. *Elife* **6**, 1–26 (2017).

939 22. Bergmiller, T. *et al.* Biased partitioning of the multidrug efflux pump AcrAB-TolC
940 underlies long-lived phenotypic heterogeneity. *Science* **356**, 311–315 (2017).

941 23. Wang, P. *et al.* Robust Growth of Escherichia coli. *Curr. Biol.* **20**, 1099–1103 (2010).

942 24. Reams, A. B. & Roth, J. R. Mechanisms of gene duplication and amplification. *Cold*
943 *Spring Harb. Perspect. Biol.* **7**, a016592 (2015).

944 25. Tawfik, D. S. Messy biology and the origins of evolutionary innovations. *Nat. Chem.*
945 *Biol.* **6**, 692–696 (2010).

946 26. Jacob, F. Evolution and tinkering. *Science (80-.).* **196**, 1161 LP – 1166 (1977).

947 27. Troein, C., Ahrén, D., Krogh, M. & Peterson, C. Is transcriptional regulation of
948 metabolic pathways an optimal strategy for fitness? *PLoS One* **2**, e855 (2007).

949 28. Wolf, L., Silander, O. K. & van Nimwegen, E. Expression noise facilitates the evolution
950 of gene regulation. *Elife* **4**, 1–48 (2015).

951 29. Anderson, R. P. & Roth, J. R. Tandem Genetic Duplications in Phage and Bacteria.
952 *Annu. Rev. Microbiol.* **31**, 473–505 (1977).

953 30. Taylor, T. B. *et al.* Evolutionary resurrection of flagellar motility via rewiring of the
954 nitrogen regulation system. *Science (80-.).* **347**, 1014–1017 (2015).

955 31. Rodriguez-Beltran, J. *et al.* Multicopy plasmids allow bacteria to escape from fitness
956 trade-offs during evolutionary innovation. *Nat. Ecol. Evol.* **2**, 873–881 (2018).

957 32. Perry, G. H. *et al.* Diet and the evolution of human amylase gene copy number
958 variation. *Nat. Genet.* **39**, 1256–1260 (2007).

959 33. Gil, R., Sabater-Muñoz, B., Perez-Brocal, V., Silva, F. J. & Latorre, A. Plasmids in the
960 aphid endosymbiont Buchnera aphidicola with the smallest genomes. A puzzling
961 evolutionary story. *Gene* **370**, 17–25 (2006).

962 34. Latorre, A., Gil, R., Silva, F. J. & Moya, A. Chromosomal stasis versus plasmid plasticity
963 in aphid endosymbiont Buchnera aphidicola. *Heredity (Edinb).* **95**, 339–347 (2005).

964 35. Lercher, M. J. & Pál, C. Integration of horizontally transferred genes into regulatory
965 interaction networks takes many million years. *Mol. Biol. Evol.* **25**, 559–67 (2008).

966 36. Pál, C., Papp, B. & Lercher, M. J. Adaptive evolution of bacterial metabolic networks
967 by horizontal gene transfer. *Nat Genet* **37**, 1372–1375 (2005).

968 37. Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. Genomic islands in pathogenic
969 and environmental microorganisms. *Nat. Rev. Microbiol.* **2**, 414–24 (2004).

970 38. Juhas, M. *et al.* Genomic islands: tools of bacterial horizontal gene transfer and
971 evolution. *FEMS Microbiol. Rev.* **33**, 376–93 (2009).

972 39. Pettersson, M. E., Andersson, D. I., Roth, J. R. & Berg, O. G. The Amplification Model
973 for Adaptive Mutation. *Genetics* **169**, 1105–1115 (2005).

974 40. Gusev, O. *et al.* Comparative genome sequencing reveals genomic signature of
975 extreme desiccation tolerance in the anhydrobiotic midge. *Nat. Commun.* **5**, 4784
976 (2014).

977 41. Hooper, S. D. & Berg, O. G. Duplication is more common among laterally transferred
978 genes than among indigenous genes. *Genome Biol.* **4**, R48 (2003).

979 42. Eme, L., Gentekaki, E., Curtis, B., Archibald, J. M. & Roger, A. J. Lateral Gene Transfer
980 in the Adaptation of the Anaerobic Parasite Blastocystis to the Gut. *Curr. Biol.* **27**,
981 807–820 (2017).

982 43. Nguyen, T. N., Phan, Q. G., Duong, L. P., Bertrand, K. P. & Lenski, R. E. Effects of
983 carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of
984 Escherichia coli K12. *Mol. Biol. Evol.* **6**, 213–25 (1989).

985 44. Gladman, S. L. *et al.* Large tandem chromosome expansions facilitate niche
986 adaptation during persistent infection with drug-resistant Staphylococcus aureus.
987 *Microb. Genomics* **1**, e000026 (2015).

988 45. Elliott, K. T., Cuff, L. E. & Neidle, E. L. Copy number change: evolving views on gene
989 amplification. *Future Microbiol.* **8**, 887–899 (2013).

990 46. Eydallin, G., Ryall, B., Maharjan, R. & Ferenci, T. The nature of laboratory
991 domestication changes in freshly isolated Escherichia coli strains. *Environ. Microbiol.*
992 **16**, 813–828 (2014).

993 47. Greenblum, S., Carr, R. & Borenstein, E. Extensive Strain-Level Copy-Number
994 Variation across Human Gut Microbiome Species. *Cell* **160**, 583–594 (2015).

995 48. Dhar, R., Bergmiller, T. & Wagner, A. INCREASED GENE DOSAGE PLAYS A
996 PREDOMINANT ROLE IN THE INITIAL STAGES OF EVOLUTION OF DUPLICATE TEM-1
997 BETA LACTAMASE GENES. *Evolution (N. Y).* **68**, 1775–1791 (2014).

998 49. Datta, S., Costantino, N. & Court, D. L. A set of recombineering plasmids for gram-
999 negative bacteria. *Gene* **379**, 109–115 (2006).

1000 50. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in
1001 Escherichia coli K-12 using PCR products. *Proc. Natl. Acad. Sci.* **97**, 6640–6645 (2000).

1002 51. Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L. & Keasling, J. D.
1003 Homogeneous expression of the P(BAD) promoter in Escherichia coli by constitutive
1004 expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147**,
1005 3241–7 (2001).

1006 52. Baba, T. *et al.* Construction of Escherichia coli K-12 in-frame, single-gene knockout
1007 mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).

1008 53. Nagelkerke, F. & Postma, P. W. 2-Deoxygalactose, a specific substrate of the
1009 Salmonella typhimurium galactose permease: Its use for the isolation of galP
1010 mutants. *J. Bacteriol.* **133**, 607–613 (1978).

1011 54. Zhou, L. *et al.* Chromosome engineering of Escherichia coli for constitutive production

of salvianic acid A. *Microb. Cell Fact.* **16**, 84 (2017).

55. Lutz, R. & Bujard, H. Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–10 (1997).

56. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose PBAD promoter. *J. Bacteriol.* **177**, 4121–30 (1995).

57. Chait, R., Shrestha, S., Shah, A. K., Michel, J. B. & Kishony, R. A differential drug screen for compounds that select against antibiotic resistance. *PLoS One* **5**, e15179 (2010).

58. Pfaffl, M. W. A new mathematical model for relative quantification in real-time RT-PCR. *Nucleic Acids Res.* **29**, 45e – 45 (2001).

59. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl. Acad. Sci.* **88**, 7160–7164 (1991).

60. Elez, M. *et al.* Seeing Mutations in Living Cells. *Curr. Biol.* **20**, 1432–1437 (2010).

61. Bayliss, C. D. Determinants of phase variation rate and the fitness implications of differing rates for bacterial pathogens and commensals. *FEMS Microbiol. Rev.* **33**, 504–520 (2009).

Figure Legends

Fig. 1. An experimental system for monitoring gene copy number under fluctuating selection in real time.

a, Gene duplication and amplification (GDA). Genomic loci duplicate at rate (k_{dup}) 10^{-6} - 10^{-2} per cell per generation. The two gene copies oriented in tandem provide long stretches of identical sequence allowing for homologous recombination at rate (k_{rec}) 10^{-4} - 10^{-1} per cell per generation with *recA*-dependent unequal crossover leading to further duplication (amplification) or deletion. Grey shading of cells symbolizes the amount of gene product made: increases in copy number result in increased gene expression.

b, Schematic of chromosomal cassette used. Expression of the selection marker, *galK*, is driven by an arabinose-inducible promoter (p_{ara}). Growth (as measured by end point OD_{600}) in a 2D gradient of arabinose with galactose (high expression environment) or DOG (low expression environment), respectively. Boxes mark concentrations of 0.1% galactose and 0.0001% DOG, which result in a strong selective tradeoff between high and low expression and were used for further experiments. **c**, Schematic showing *galK* reporter cassette (p_0 = random sequence/'non-promoter', p_R = strong constitutive lambda promoter, terminator sequences downstream of *yfp* and *cfp*, respectively) and genetic changes of strains evolved in the high expression environment with resulting phenotypes on MacConkey galactose agar. Both evolved strains show increased *galK-yfp* expression over the ancestral strain (YFP) and the ability to grow on galactose (BF = bright field image, white versus pink colonies). The amplified strain shows increased CFP fluorescence (CFP) over the ancestral and the constitutive strain, indicating a gene copy number increase.

Fig. 2. Amplification-mediated gene expression tuning (AMGET) occurs in fluctuating environments. a, Experimental design of alternating selection in 96-well plate batch cultures, with a daily dilution of 1:133. A minimal duration of 24h per environmental condition (no shading = low expression environment, grey shading = high expression environment) allows measuring OD_{600} and fluorescence in populations that have reached stationary phase after dividing at least seven times after their last dilution. **b**, Alternating selection of 1 day - 1 day, 2 days – 1 day and 3 days – 1 day in high and low expression environment, respectively. Normalized CFP fluorescence as proxy for gene copy number of 60, 48 and 60 populations of the amplified strain. Error bars represent standard deviation (SD) over all populations. **c**, Flow cytometry histograms (one of six replicates from two independent experiments; see d. for an overview of the full dataset) following the adaptation of an amplified bacterial population to low and high expression environments. Positive controls represent populations grown in respective environment for 5 days**.** **d**, Fitness as a function of copy number in the two environments. Growth rates relative to those of maximally adapted populations (positive controls in **c**) as a proxy for fitness were calculated from the population´s shift in CFP fluorescence over time (see Methods). *M* denotes the maximum copy number, which we estimate to be approximately 10 (see bulk measurements of *M* in Fig. S1a and Fig. S2a, and single cell-based measurements in Fig. S5b). Note that results do not depend on the precise value of *M*). Error bars represent the standard deviation of six replicates from two independent experiments.

Fig. 3. High-frequency deletion/duplication events in the amplified locus create gene copy number polymorphism in populations. a, Re-streaks of a single bacterial colony on nonselective agar. Ancestral strain bearing a single copy of *cfp* (left), amplified strain (middle) colonies display sectors of different CFP fluorescence (inset). Scale bars, 10 mm. Histogram of single-colony mean CFP intensities obtained by resuspending and diluting five ancestral and amplified colonies, respectively (right). **b**, The amplified strain carrying a single copy of *mCherry* in a control locus (top) was grown in a microfluidics device to allow tracking of cell lineages in the absence of selection. Overlay of kymographs of CFP and mCherry fluorescence for one microfluidics growth channel (left). Two recombination events are visible as pronounced changes in CFP relative to mCherry fluorescence (white arrows). Time series images of CFP and mCherry fluorescence (right) of the same channel during the second amplification event. An increase in CFP fluorescence of the mother cell (rightmost position in the growth channel) occurs concomitantly with reciprocal loss of CFP fluorescence in its first daughter cell. As mother and daughter cell divide again, their altered level of CFP fluorescence is inherited by their respective daughter cells. mCherry fluorescence of the control locus stays constant during the recombination event. Scale bars, 5µm. **c**, Examples of single-cell time traces (kymographs and CFP fluorescence sampled from the mother cell) for four representative behaviors: constant expression, stepwise increase and decrease in expression, and complex expression changes. Frequencies of each behavior across 1089 channels from three independent experiments are shown in figure panels.

Fig. 4. AMGET requires continual generation of gene copy number polymorphisms. **a**, Schematic of a co-culture composed of the ancestral strain without *galK* expression and a strain with two SNPs in p_0 (Fig. S1C) resulting in high *galK* expression (left). Fluorescently labeling the ancestor allows monitoring relative strain abundance (Methods). A population consisting of a single amplified strain (right) contains cells with different *galK* copy numbers and, accordingly, expression levels. **b**, Alternating selection following the scheme 1 day - 1 day, 2 days – 1 day and 3 days – 1 day in high and low expression environment, respectively. Constitutive strain abundance of 18 co-culture populations tracks environments, with the non-expressing strain being abundant in the low expression environment and the constitutive strain being abundant in the high expression environment. Error bars represent the SD of 18 replicates. **c-d**, To estimate a population´s ability to respond to a change in the environment, periods of increasing length spent in the high expression environment are followed by one day in the low expression environment. **c**, Copy number of amplified populations as measured by CFP fluorescence is adjusted to the low expression environment (black arrows) even after prolonged growth in the high expression environment. **d**, In contrast, response of the co-culture to the low expression environment after prolonged growth in the high expression environment decreases with time spent in the high expression environment. The mean response on day 16 (1.11 for co-culture, 4 for amplified) differs significantly ($p<10^{-3}$, two-sided t-test) between populations of co-culture (**d**) and amplified (**c**) (see Methods). Error bars represent the SD of 36 replicates.

Fig. 5. AMGET is a robust strategy for population level gene expression tuning across a range of environments. **a**, Comparison of model predictions (with all parameters derived from independent calibration experiments; see Methods) and experimental data for three different environmental durations. Pearson correlation between data and model: 0.72 (top), 0.92 (middle), 0.87 (bottom). See Fig. S6a for parameter sensitivity. Error bars represent standard deviation (SD) over of 60, 48 and 60 bacterial populations, respectively. **b-c**, Top: example of gene expression time trace for deterministic (**b**) and stochastic (**c**) environment durations. Bottom: response R (maximum expression fold change before and after the environmental change), shown in color, as a function of the two environment durations. Red crosses in **b** mark environments shown in **a**. The gradual increase in response in **c** occurs because of averaging across responses, which are deterministic for each individual environmental transition (**c** top). **d**, Variation of response time when uniformly sampling sets of parameters (black circles) in the range of 10^{-4} - $5x10^{-2}$, 10^{-5} - 10^{-3}, and 0.1 - 1 for recombination rate, duplication rate, and fitness costs of expression, respectively (Fig. S6c-e). The plot shows the median (red line) with the 25^{th} and 75^{th} percentile (blue box). In all plots, when not varied, we use recombination and duplication rates $k_{rec}^{0}=1.34x10^{-2}$ and $k_{dup}=10^{-4}$, respectively. All rates have units of $cell^{-1}$ $generation^{-1}$. In our setup, one-day timescale is equivalent to between 10 and 23 generations (lower and upper bound, respectively; the bounds are estimated from the minimum and maximum growth rate of the least and best adapted copy number types, Table S2, Fig. 2d).

Table 1. Comparison of regulation, amplification, adaptation and bet-hedging strategies.

	regulation	**amplification**	**adaptation (rewiring via point mutations)**	**bet-hedging strategies**
mechanism	hard-wired response of individual cells	mutation	mutation	phenotypic differences between genetically identical cells
rate ON	1	10^{-6} - 10^{-2} $cell^{-1}$ $gen.^{-1}$ [6–9]	10^{-9} bp^{-1} $cell^{-1}$ $gen.^{-1}$ [59,60]	>10^{-5} variants per total cells [61]
rate OFF	1	10^{-3} - 10^{-1} $cell^{-1}$ $gen.^{-1}$ [7,8]	10^{-9} bp^{-1} $cell^{-1}$ $gen.^{-1}$ [59,60]	
active sensing machinery required	yes	no	no	no
can substitute for regulation on ecological time scales	-	yes	no	yes
expression state genetically heritable	no	yes	yes	no
tuning (allows graded expression)	typically not	yes	yes, but very long timescales	typically not
High reversibility (rate OFF > rate ON)	yes	yes	no	yes
suitable for rare stresses	no	yes	probably not, due to slow reversibility	depends on cost and rate