nature ecology & evolution

ARTICLES

https://doi.org/10.1038/s41559-020-1132-7

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

I. Tomanek[1,4], R. Grah[1,4], M. Lagator[1,2], A. M. C. Andersson[1], J. P. Bollback[3], G. Tkačik[1] and C. C. Guet[1]✉

Organisms cope with change by taking advantage of transcriptional regulators. However, when faced with rare environments, the evolution of transcriptional regulators and their promoters may be too slow. Here, we investigate whether the intrinsic instability of gene duplication and amplification provides a generic alternative to canonical gene regulation. Using real-time monitoring of gene-copy-number mutations in *Escherichia coli*, we show that gene duplications and amplifications enable adaptation to fluctuating environments by rapidly generating copy-number and, therefore, expression-level polymorphisms. This amplification-mediated gene expression tuning (AMGET) occurs on timescales that are similar to canonical gene regulation and can respond to rapid environmental changes. Mathematical modelling shows that amplifications also tune gene expression in stochastic environments in which 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 the expression of any gene, without leaving any genomic signature.

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—including transcription factors and promoter sequences in the DNA—for information processing to occur. However, owing to low single-base-pair mutation rates, complex promoters cannot easily evolve on ecological timescales[4,5].

Gene-copy-number mutations might provide a fundamentally different adaptation strategy that 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, at any given time, a typical bacterial population will contain a large fraction of cells with a duplication somewhere on the chromosome[9,10]. Owing to the long stretches of homology, duplications are highly unstable—at rates (k_{rec}) of between 10^{-3} and 10^{-1} per cell per generation[7,8], *recA*-dependent unequal crossover of the repeated sequence leads to either the 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 substantially increase organismal fitness by increasing gene copy numbers. Owing to their high rates of formation, amplifications provide fast adaptation and facilitate the evolution of functional innovation[14]. By contrast, their high rate of loss makes amplifications transient and difficult to study[14]. Until recently, the impact of this high loss rate on the distribution of copy numbers and associated expression levels in the population, a phenomenon causing antibiotic heteroresistance, has not been appreciated[11,15]. Furthermore, amplifications have been studied only under constant selection for increased expression[16,17], but natural environments are rarely ever constant. Although a large body of research suggests that phenotypic heterogeneity serves as an adaptation to fluctuating environments[18,19], it is not known how the genetic heterogeneity that results from copy-number mutations impacts survival in fluctuating environments.

Here we investigate whether the intrinsic genetic instability of gene amplifications enables 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 that selects for the regulation of a model gene. In this fluctuating environment, we tracked, in real time, copy-number mutations in populations as well as single cells of *E. coli*. Using this system, we tested the ability of GDA to effectively tune levels of gene expression on ecological timescales when environmental perturbations occur at rates that are far too high for transcriptional gene regulation to emerge de novo.

Results

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 environment but detrimental in another environment. As the model gene, we used the dual-selection marker *galK*, which encodes galactokinase. Expression of *galK* is necessary for growth on galactose, but is deleterious in the presence of its chemical analogue 2-deoxy-D-galactose (DOG)[20]. Using *galK* with an arabinose-inducible promoter, we mapped the relationship between the expression level of *galK* and growth in (1) galactose, which selects for high *galK* expression levels and which we refer to as the high-expression environment; and in (2) DOG, which selects for low *galK* expression and which we refer to as the low-expression environment (Fig. 1b). To establish a strong selective trade-off between high and low expression, we used 0.1% galactose for the high-expression environment

[1]Institute of Science and Technology Austria, Klosterneuburg, Austria. [2]School of Biological Sciences, Faculty of Biology, Medicine and Health, University of Manchester, Manchester, UK. [3]Institute of Integrative Biology, Functional and Comparative Genomics, University of Liverpool, Liverpool, UK. [4]These authors contributed equally: I. Tomanek, R. Grah. ✉e-mail: calin@ist.ac.at

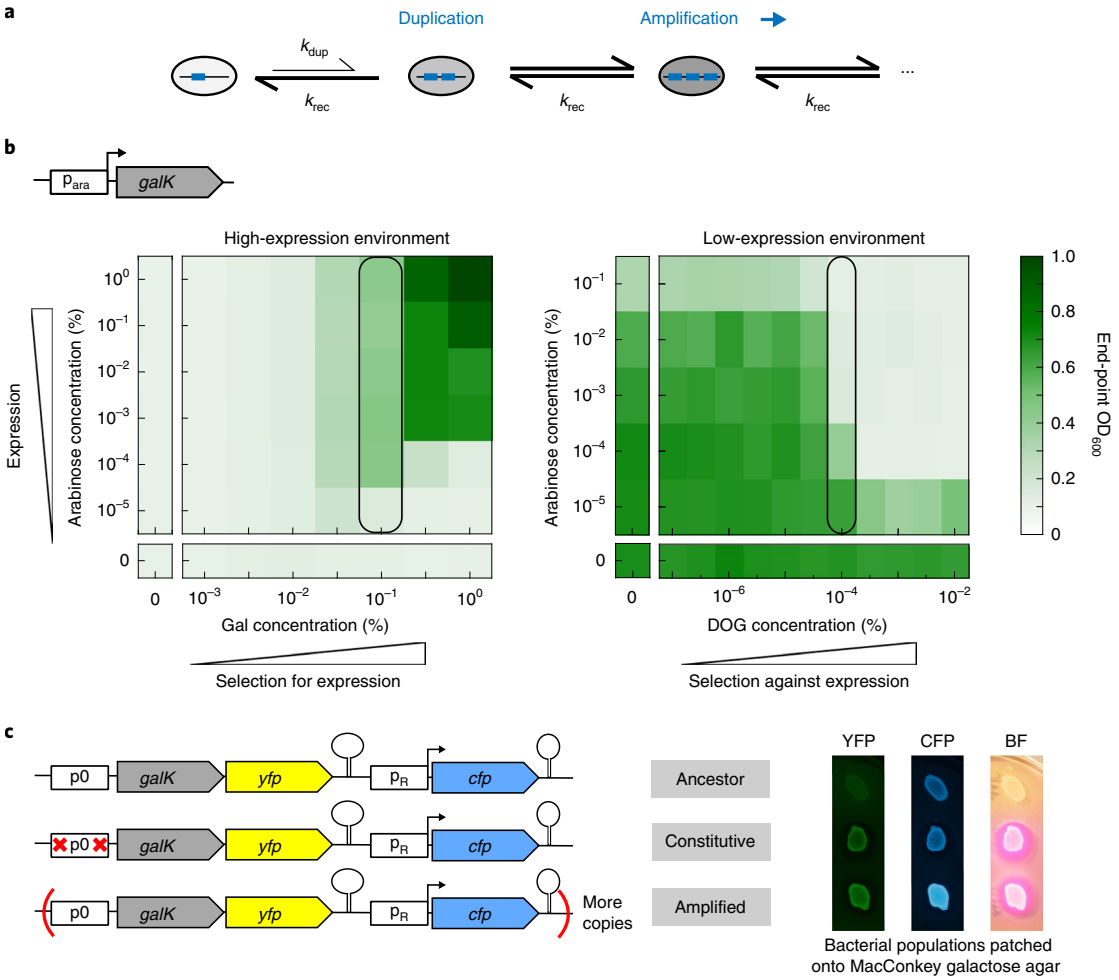

Fig. 1 | An experimental system for monitoring gene copy number under fluctuating selection in real time. a, Schematic of GDA. Genomic loci duplicate at rate $k_{dup} = 10^{-6}$–10^{-2} per cell per generation. The two gene copies oriented in tandem provide long stretches of identical sequence enabling 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. The grey shading of cells shows the amount of gene product generated—increases in copy number result in increased gene expression. **b**, Schematic of the chromosomal cassette used. Expression of the selection marker *galK* is driven by an arabinose-inducible promoter (p_{ara}). Bottom, growth (as measured by end-point measurements of optical density at 600 nm (OD_{600})) in a two-dimensional gradient of arabinose with galactose (high-expression environment) or DOG (low-expression environment). The black boxes indicate concentrations of 0.1% galactose and 0.0001% DOG, which result in a strong selective trade-off between high and low expression and were used for further experiments. **c**, Schematic of the *galK* reporter cassette (p_0 indicates the random sequence/non-promoter, p_R indicates the strong constitutive lambda promoter and the hairpins indicate terminator sequences downstream of *yfp* and *cfp*) 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 compared with the ancestral strain (YFP) and the ability to grow on galactose (bright field (BF) image, white versus pink colonies). The amplified strain shows increased CFP fluorescence (CFP) compared with the ancestral and the constitutive strain, indicating an increase in gene copy number.

and 0.0001% DOG for the low-expression environment in all of the experiments.

We then constructed a reporter-gene cassette on the basis of a previously described construct[21] to monitor expression and copy-number changes of *galK* (Fig. 1c). In this construct, *galK* is not expressed from a promoter but harbours p_0, which is a randomized 188 bp nucleotide sequence that matches the average GC content of *E. coli*[21]. This enabled us to select for increased expression of *galK*. The reporter cassette harbours two fluorophores that enabled us to distinguish between the two principal mechanisms of increasing *galK* expression in evolving populations—promoter mutations and copy-number mutations (Fig. 1c). The promoterless *galK* gene was transcriptionally fused to a yellow fluorescence protein (*yfp*) gene, which reports on *galK* expression. Directly downstream, but separated by a strong terminator sequence, an independently transcribed cyan fluorescence protein (*cfp*) gene provides an estimate of the copy number of the whole cassette (Supplementary Fig. 1a). We inserted this cassette into the bacterial chromosome, close to the origin of replication (*oriC*)—a location with an intermediate tendency for GDA[21]. However, our results hold for a second locus, which is flanked by two identical insertion sequence elements and has a much higher tendency for GDA[21] (Supplementary Fig. 4).

The ancestral strain carrying the promoterless *galK* construct does not visibly grow in the high-expression environment. After one week of cultivation at 37 °C, mutants with increased *galK* expression appeared (Supplementary Fig. 1b). We randomly selected one evolved clone with increased CFP fluorescence (the amplified strain) and analysed it in detail (see Methods) to confirm its amplification. This amplified strain was then used for further experiments in alternating environments (Fig. 2a–c).

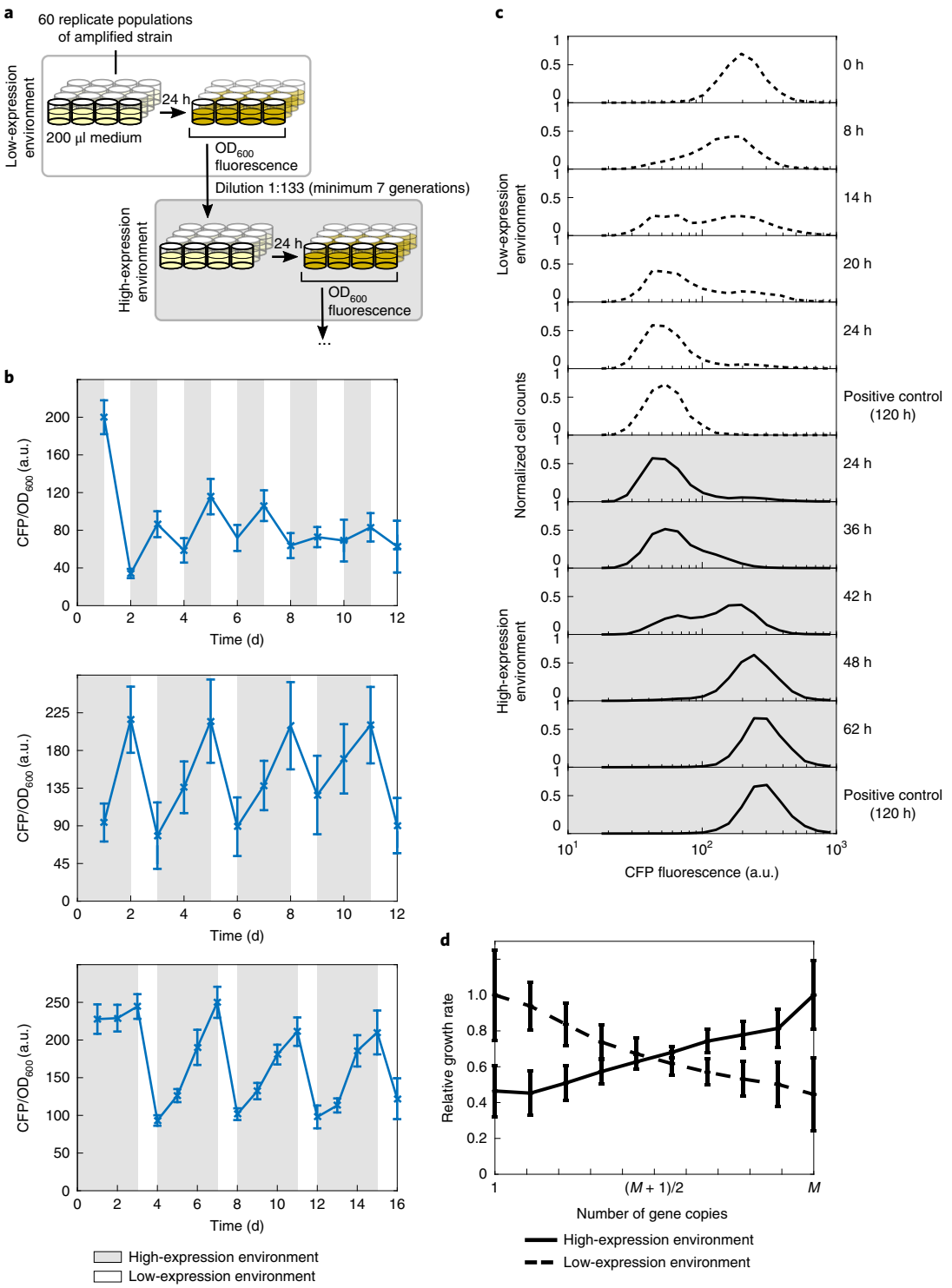

Fig. 2 | AMGET occurs in fluctuating environments. a, Experimental design of alternating selection in 96-well plate batch cultures; a daily dilution of 1:133 was performed. A minimal duration of 24 h per environmental condition (grey shading indicates high-expression environment; no shading indicates low-expression environment) enabled us to measure OD_{600} and fluorescence in populations that reached stationary phase after dividing at least seven times after their last dilution. **b**, Alternating selection of 1 d–1 d (top), 2 d–1 d (middle) and 3 d–1 d (bottom) in high–low-expression environments, respectively. Normalized CFP fluorescence as a proxy for gene copy number of 60, 48 and 60 populations of the amplified strain. Data are mean values and error bars represent s.d. over all of the populations. **c**, Flow cytometry histograms (one out of six replicates from two independent experiments; see **d** for an overview of the full dataset) after the adaptation of an amplified bacterial population to low- and high-expression environments. The positive controls represent populations that were grown in the respective environment for 5 d. **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 change in CFP fluorescence of the population over time (see Methods). *M* denotes the maximum copy number, which we estimate is approximately 10 (bulk measurements of *M* are provided in Supplementary Figs. 1a and 2a, and single-cell-based measurements are provided in Supplementary Fig. 5b). Note that the results do not depend on the precise value of *M*. Data are mean values and error bars represent the s.d. of six replicates from two independent experiments.

For all three alternating regimes, which change on a daily timescale, the mean CFP levels of 60 replicate populations of the amplified strain tracked the environments for the full duration of the experiments. The adaptive change in *galK* copy number (Fig. 2b) occurred within the imposed ecological timescale, rapidly enough to maintain population growth given the daily dilution bottleneck under all three alternating selection regimes (Supplementary Fig. 3a). We confirmed the observed changes in copy number using whole-genome sequencing (Supplementary Fig 2b). To understand these population-level observations, we monitored changes in the expression of *galK* and *cfp* at the single-cell level for two consecutive environmental switches (Fig. 2c). Expression of *galK-yfp* (Supplementary Fig. 3b) was tightly correlated with the observed changes in gene copy number (Supplementary Fig. 3c), indicating that gene expression was effectively tuned by GDA. We refer to this phenomenon as 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 due to 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 Δ*recA* derivative of the amplified strain failed to show a decrease in CFP fluorescence (and, therefore, copy number) in response to increasing concentrations of DOG (Supplementary Fig. 3d). Similarly, Δ*recA* populations were unable to track fluctuating environments, in contrast to their *recA*-wild-type counterparts (Supplementary Fig. 3e).

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 microfluidics device[22,23] and monitored their CFP levels. Mutations in copy number were clearly visible as changes in CFP fluorescence of the mother cell. In approximately 35% of cases, these changes were accompanied by a reciprocal fold change in fluorescence in the daughter cell (Fig. 3b, Supplementary Table 1), as expected due to unequal crossover[24].

To quantify the combined rate of events of gain and loss in copy number by homologous recombination, we analysed the fluorescence time trace of 1,089 mother cells. We found that 55% of traces exhibited constant levels of CFP fluorescence (Fig. 3c, top), indicating stable inheritance of copy number. In about 7% of traces, the constant level of CFP was interrupted by a sudden decrease or increase (Fig. 3c, middle). The corresponding fold changes in fluorescence levels were consistent with gains or losses of entire copies of *cfp*. We estimated that the lower bound for the average number of copy-number mutations k_{rec} was 2.7×10^{-3} per cell per generation, by automatically selecting only clear stepwise transitions in fluorescence, which are indicative of single copy-number-mutation events (see Methods; Supplementary Fig. 5, Supplementary Table 1). Interestingly, 34% of all traces (Supplementary Fig. 5c) exhibit more complex behaviours (Fig. 3c, bottom) and cannot be explained in terms of single-step transitions.

Complex traces are expected to contain more than one duplication or deletion event, even under the expectation that copy-number variations are independent events (Supplementary Fig. 5d). Furthermore, it is conceivable that copy-number mutations are not independent, that is, an increased probability exists for a second mutation after the first increase in copy number occurred. However, we cannot exclude the possibility that most of the complex traces are due to expression noise of one or both fluorophores, especially because CFP expression noise increases with copy number. Moreover, microfluidics experiments showed transient growth defects that were visible as filamentation (Supplementary Table 1). Given that the amplification includes the origin of replication (*oriC*), complex traces might result in part from replication issues. Transiently stalled replication forks could result in overproduction of CFP relative to mCherry, which is located at the phage attachment site *attP21*, almost opposite on the *E. coli* chromosome. Thus, the use of only single clear stepwise 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 that underlies AMGET is selection acting on copy-number polymorphism, we investigated whether it differs from selection acting on single-nucleotide polymorphisms (SNPs). To achieve this, we artificially created a polymorphic population comprising 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; however, because it was not due to amplification, variation was not replenished at high rates. Although the co-culture population tracked short-term environmental fluctuations in a similar manner 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). By contrast, the co-culture population progressively lost the ability to respond to sudden environmental change (Fig. 4d). Although standing variation in the co-culture provided some ability for a population to adapt in the short term, it is replenished at only the rate of point mutations. Thus, this variation—as well as the ability to adapt—is depleted by prolonged selection as the genotype with higher fitness goes to fixation in the population.

AMGET is a general and robust mechanism. Our experimental results qualitatively showed that both gene-copy-number polymorphism as well as selection acting on gene copy number are necessary for AMGET to occur. Using population-genetics theory, we developed a generic mathematical model to quantitatively predict the experimentally observed population dynamics (Fig. 2b). The model describes how gene copy number changes over time in a population that is under selection. Each copy number is treated as a distinct state, and these states differ with respect to growth rates in each of the two environments. Duplication and amplification events are the only source of transition between states. Importantly, all of the model parameters (the strength of selection and the rate at which the copy-number polymorphism is introduced; Fig.1a) were obtained from independent measurements (Supplementary Table 2). Thus, without specifically fitting any parameters, the generic model fully captured the experimentally observed dynamics of AMGET (Fig. 5a, Supplementary Fig. 6a). The good fit between model and experimental data meant that we were able to use the model to expand the understanding of the basic conditions under which AMGET can act as an efficient de facto mechanism of population-level gene regulation.

Qualitatively, the model revealed that, for a population to respond at all to environmental change, two conditions must be met: (1) constant introduction of gene-copy-number variation (that is, non-zero duplication/recombination rate) and (2) selection acting on gene copy number. If either of these are not present, the

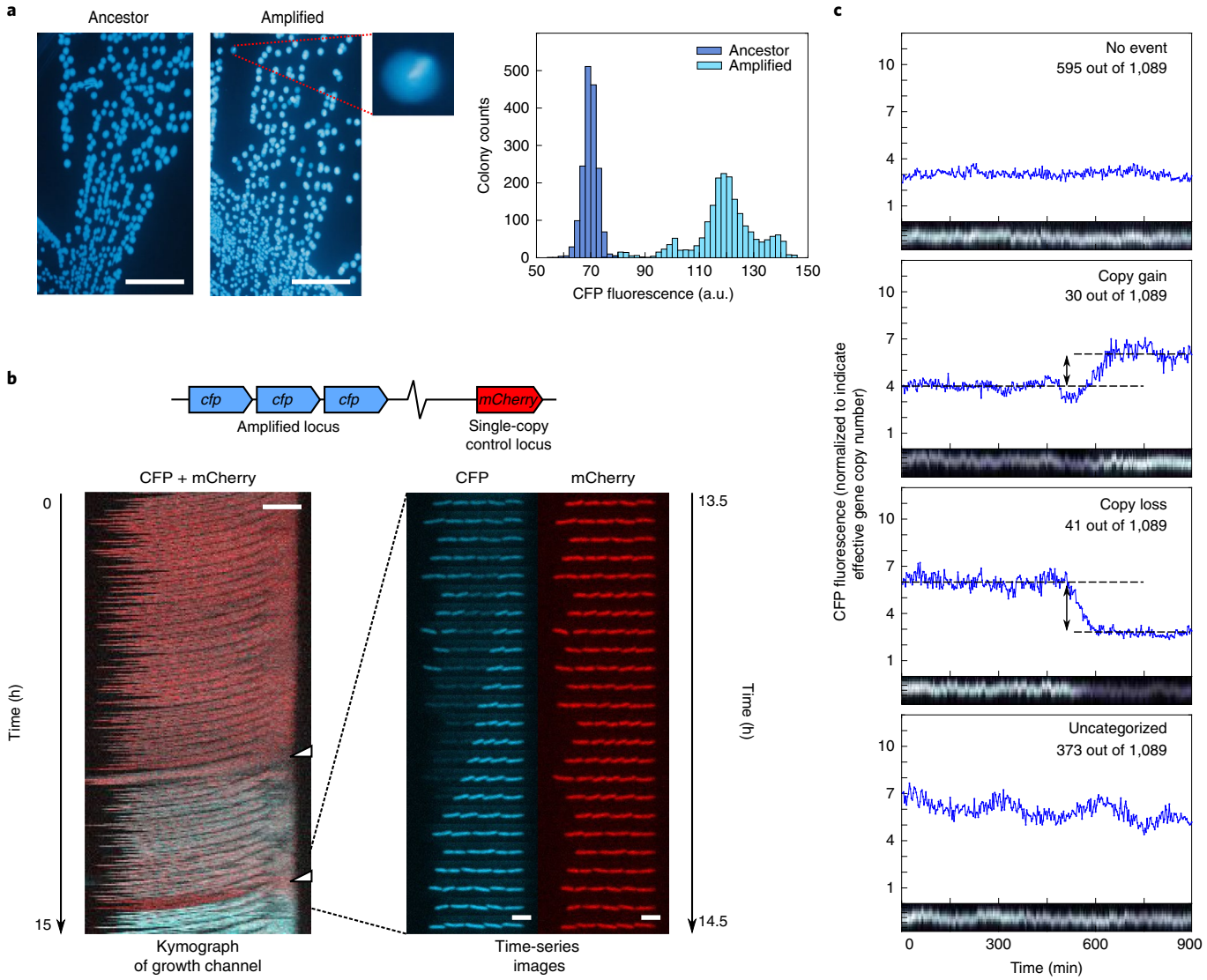

Fig. 3 | High-frequency deletion and duplication events in the amplified locus create gene-copy-number polymorphisms in populations. a, Re-streaks of a single bacterial colony onto non-selective agar. Colonies of the ancestral strain bearing a single copy of *cfp* (left) and the amplified strain (middle) display sectors of different CFP fluorescence (inset, eight-fold magnification). Scale bars, 10 mm. Right, histogram of single-colony mean CFP intensities obtained by resuspending and diluting five ancestral and amplified colonies. **b**, The amplified strain carrying a single copy of *mCherry* in a control locus (top) was grown in a microfluidics device to enable 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 the 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 remains constant during the recombination event. Scale bars, 5 μm. **c**, Examples of single-cell time traces (kymographs and CFP fluorescence sampled from the mother cell) for four representative behaviours: constant expression, stepwise increase and decrease in expression, and complex changes in expression. Frequencies of each behaviour across 1,089 channels from three independent experiments are shown.

population is unable to maintain any long-term response to environmental change.

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 found a sharp switch-like transition from no response to full response for environments that switch typically on timescale of a day or longer (Fig. 5b). In stochastically fluctuating environments, the transition was more gradual (Fig. 5c), albeit no less effective. Furthermore, AMGET maintained its efficiency to tune gene expression in bacterial populations over order-of-magnitude variations in the rates of duplication and recombination, as well as for any fitness cost of expression (Supplementary Fig. 7).

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

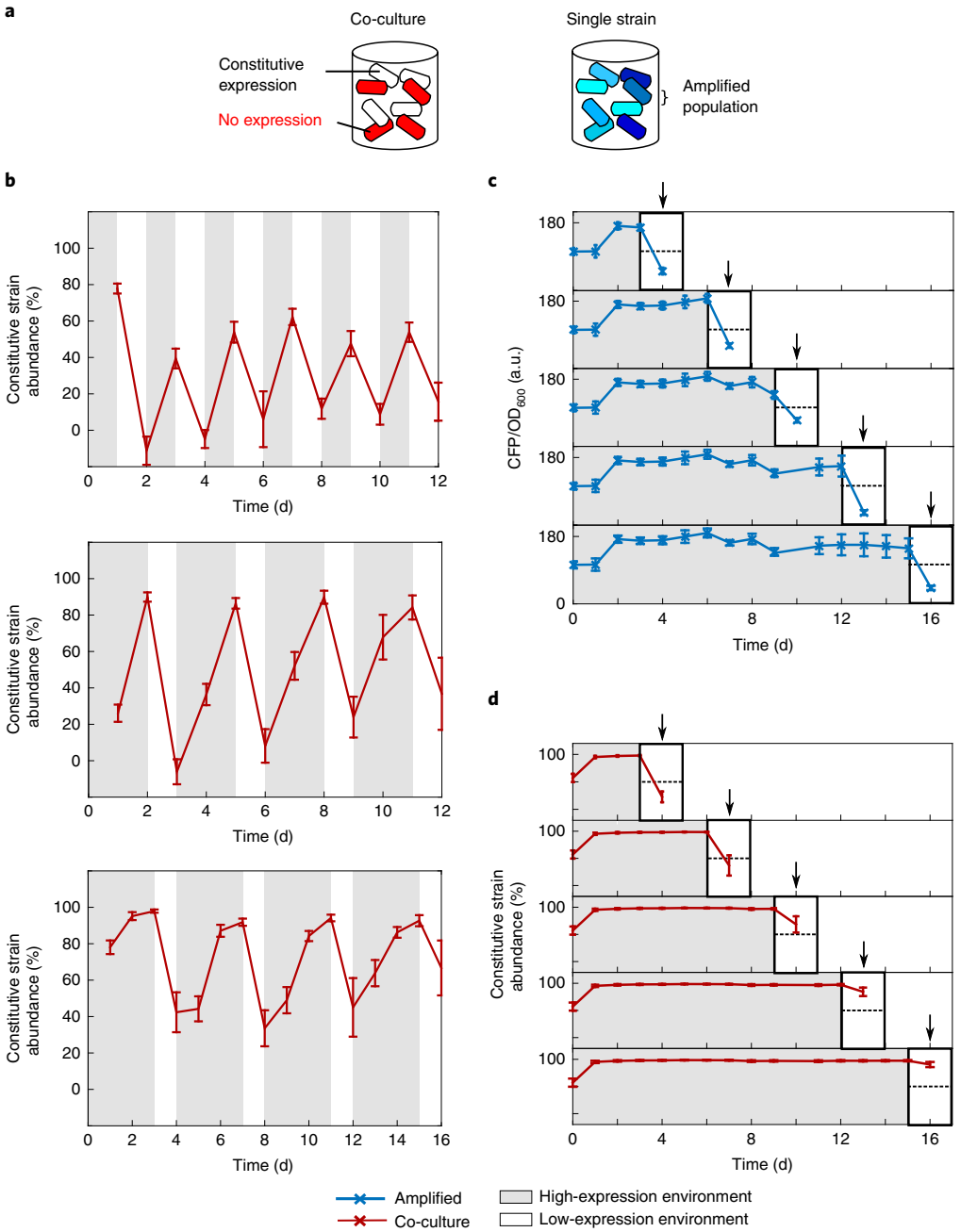

Fig. 4 | AMGET requires continual generation of gene-copy-number polymorphisms. a, Schematic of a co-culture composed of the ancestral strain without *galK* expression and a strain with two SNPs in p_0 (Supplementary Fig. 1c) resulting in high *galK* expression (left). Fluorescently labelling the ancestor enabled us to monitor relative strain abundance (see Methods). Right, a population consisting of a single amplified strain contains cells with different copy numbers of *galK* and, accordingly, different expression levels. **b**, Alternating selection following the schemes 1 d–1 d, 2 d–1 d and 3 d–1 d in high–low-expression environments, respectively. The abundance of the constitutive strain in 18 co-culture populations tracks environments; the non-expressing strain is abundant in the low-expression environment and the constitutive strain is abundant in the high-expression environment. Data are mean values and error bars represent the s.d. of 18 replicates. **c**,**d**, To estimate the ability of a population to respond to a change in the environment, periods of increasing length spent in the high-expression environment were followed by 1 d in the low-expression environment. **c**, The copy number of amplified populations, measured according to CFP fluorescence, adjusted to the low-expression environment (black arrows) even after prolonged growth in the high-expression environment. **d**, By contrast, the response of the co-culture to the low-expression environment after prolonged growth in the high-expression environment decreased with time spent in the high-expression environment. The mean response on day 16 (1.11 for co-culture; 4 for amplified) differs significantly (two-sided *t*-test, $P < 10^{-3}$) between populations of co-culture (**d**) and amplified (**c**) strains (see Methods). The error bars represent the s.d. of 36 replicates.

sporadically interrupted by a rare environmental perturbation[3]. We tested whether AMGET provides a generic mechanism of regulating expression under such conditions by investigating the time that a population that is fully adapted to one environment needs to respond to a step-like environmental change (Fig. 5b, heat map; Supplementary Fig. 6b). Our model results showed very rapid responses to step-like environmental changes on the order of 1–6 d for all of the biologically relevant parameter values of amplification

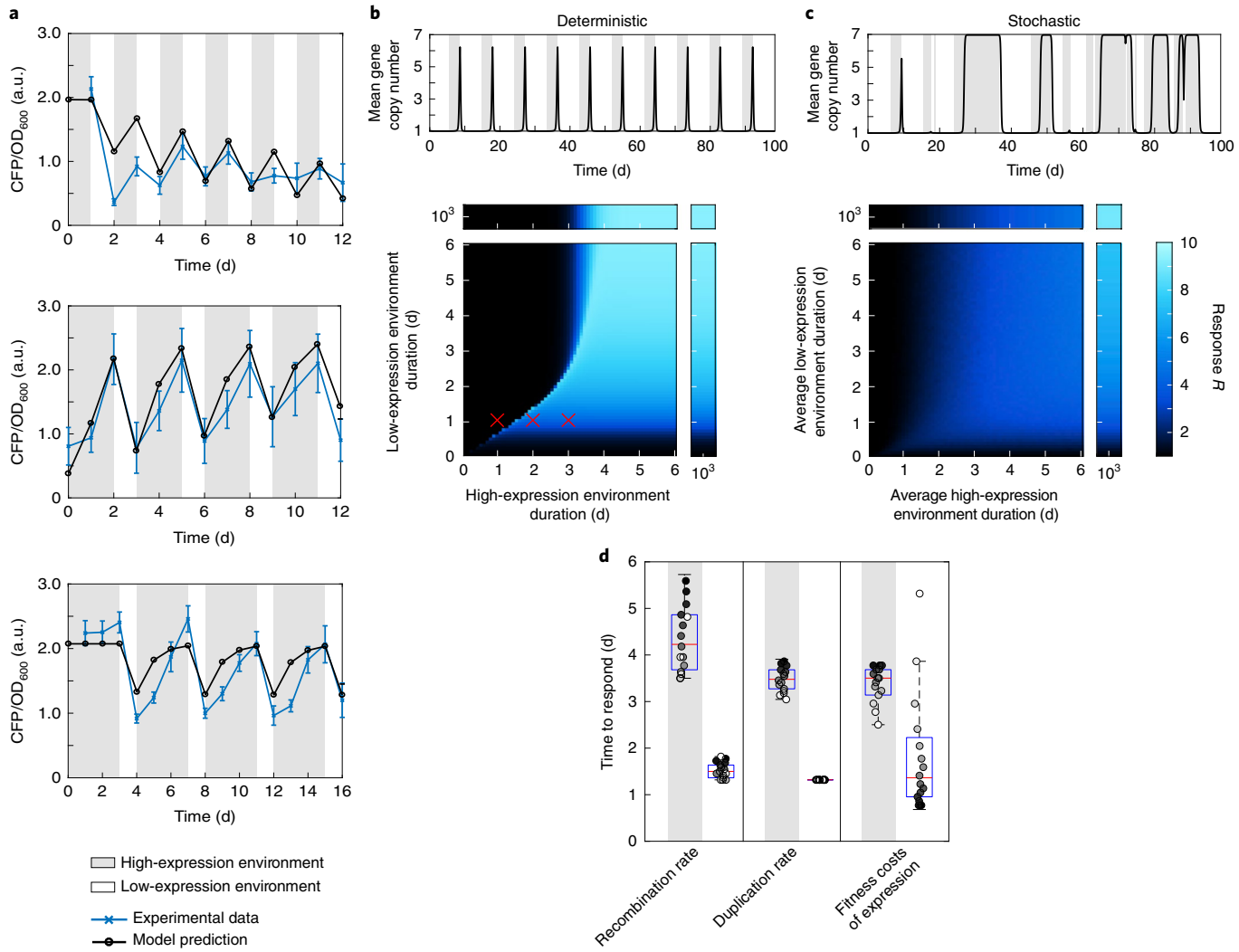

Fig. 5 | AMGET is a robust strategy for tuning population-level gene expression across a range of environments. a, Comparison of model predictions (with all of the parameters derived from independent calibration experiments; see Methods) and experimental data for three different environmental durations. Pearson correlation between data and model: 0.72 (top), 0.92 (middle) and 0.87 (bottom). Details about parameter sensitivity are provided in Supplementary Fig. 6a. Data are mean values and error bars represent s.d. over 60 (top), 48 (middle) and 60 (bottom) bacterial populations. **b**,**c**, Example of a gene expression time trace for deterministic (**b**) and stochastic (**c**) environment durations (top). Bottom, the response R (maximum expression fold change before and after the environmental change), shown in colour, as a function of the two environment durations. The red crosses (from left to right) in **b** indicate environments shown in **a** (from top to bottom). The gradual increase in response in **c** occurred due to averaging across responses, which are deterministic for each individual environmental transition (**c**, top). **d**, Variation in response time when uniformly sampling sets of parameters (black circles) in the range of 10^{-4}–5×10^{-2}, 10^{-5}–10^{-3} and 0.1–1 for recombination rate, duplication rate and fitness costs of expression, respectively (Supplementary Fig. 6c–e). The median (red line) and the 25th and 75th percentiles (blue box) are indicated. For all of the plots, when not varied, we use recombination and duplication rates $k_{rec}{}^{0}=1.34\times10^{-2}$ and $k_{dup}=10^{-4}$, respectively. For all of the rates, units are per cell per generation. In our setup, the one-day timescale is equivalent to between 10 and 23 generations (the 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; Fig. 2d, Supplementary Table 2).

and duplication rates, as well as fitness cost of expression (Fig. 5d; Supplementary Fig. 6c–e). AMGET is also a viable mechanism for practically any population size, especially for typical bacterial population sizes, although its efficiency drops for small populations (Supplementary Fig. 6f). Thus, AMGET efficiently tunes gene expression levels across a wide range of environments in which transcription-factor-mediated regulation would take a prohibitively long time 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 that live in constantly fluctuating environments, even the crudest form of gene regulation[27] or gene expression heterogeneity[28] increases fitness compared with not having any regulation at all. Here we showed that the intrinsic instability of gene amplifications rapidly tunes levels of gene expression when gene regulation is required but when no other molecular regulatory mechanism is in place.

Although AMGET resembles canonical gene regulation when populations are observed 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 function by

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

	Regulation	Amplification	Adaptation (rewiring through point mutations)	Bet-hedging strategies
Mechanism	Hard-wired response of individual cells	Mutation	Mutation	Phenotypic differences between genetically identical cells
Rate of switching expression on	1	10^{-6}–10^{-2} per cell per generation[6–9]	10^{-9} bp^{-1} per cell per generation[59,60]	More than 10^{-5} variants per total cells[61]
Rate of switching expression off	1	10^{-3}–10^{-1} per cell per generation[7,8]	10^{-9} bp^{-1} per cell per generation[59,60]	
Active sensing machinery required	Yes	No	No	No
Can substitute for regulation on ecological timescales	–	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

allowing bacterial populations to 'hedge their bets' for expression levels that could be required in a future environment. In contrast to traditional descriptions of bet hedging, in which genetically identical individuals show variability in their phenotypic states[19], AMGET populations differ in their genotype due to the intrinsic instability of gene amplifications and, therefore, pass on the adaptive state with high probability. Moreover, bet hedging is typically characterized by switching between a small number of alternative phenotypic states[19]; by contrast, in an amplified locus, expression can adopt a graded response due to a wide range of copy numbers.

Because AMGET enables rapid dynamics and, at the same time, graded responses, it can be thought of as a form of primitive gene expression regulation at the population level[29]. Mechanistically, AMGET bears no resemblance to canonical gene regulation, in which sensory machinery alters gene expression in the course of a single generation. However, despite the mechanistic differences, AMGET operates on a timescale of days that is therefore more similar to the timescale of canonical gene regulation compared with the process of transcriptional rewiring by point mutations, which occur several orders of magnitude less frequently (Table 1).

AMGET may be one of several mechanisms by which populations can make use of variation in expression levels to rapidly adapt to environmental changes. Although point mutations occur at lower rates, regulatory rewiring can be surprisingly fast[30], especially when there is pre-existing variation in the precise architecture of regulatory networks. Moreover, noise propagation within gene regulatory networks can create an abundance of different expression levels that are—in principle—tunable by selection[28]. However, as the results of our co-culture experiment show (Fig. 4), pre-existing variation can be easily depleted from a population if under strong selection. Although it was previously shown that variation can be maintained in the form of multiple plasmid copies[31], our results highlight that multiple copies of a genomic region actively regenerate heterogeneity due to the high recombination rate. Owing to this property, AMGET provides a means of tuning expression to rare environmental fluctuations, in which canonical gene regulation cannot evolve or be maintained[3].

AMGET is fast in bacteria because their generation times are short and their population sizes are usually large. However, our model results show that AMGET is applicable, in principle, to any other organism, but would take much longer time in relatively small populations (Supplementary Fig. 6f). A compelling example for the upregulation of a gene on relatively short evolutionary timescales is that of the salivary amylase in humans—the variation in the copy numbers of *AMY1* is correlated with the dietary starch content of human populations[32].

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

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

Despite their ubiquity, GDA has been underappreciated[14,45]. In principle, fixed amplifications can be easily detected in next-generation sequencing data on the basis of an increase in coverage and mismatches that correspond to the duplication junctions (Supplementary Fig. 2; see Methods). However, duplications revert to the single copy state at high rate without leaving any traces in the genome (Supplementary Fig. 2a). This implies that populations have to be kept under selection before sequencing, a condition that may not typically be met, especially not for environmental isolates[46]. However, despite this challenge, there are many reports of cases in which amplified genes have been detected in the sequences of environmental strains and were found to be associated with adaptation to environmental conditions[33,40,47].

The notion that GDA "might be thought of as a rather crude regulatory mechanism"[29] is more than 40 years old. However, to date, almost all experimental work has focused on the benefits of amplification in constant, stable environments and, therefore, selected for increased expression only[16,48]. Here we demonstrated the flexibility of GDA in rapidly altering gene expression levels of populations in response to a wide range of environmental fluctuations. Thus, AMGET is an essential, and a critically underappreciated, mechanism of bacterial survival.

Methods

Construction of the bacterial strain background. Except where noted otherwise, all changes to the *E. coli* chromosome were introduced using pSIM6-mediated recombineering[49]. All of the recombinants were selected on either 25 μg ml^{-1} kanamycin or 10 μg ml^{-1} chloramphenicol to ensure single-copy integration. All of the resistance markers introduced by recombineering were flipped by transforming plasmid pCP20 and streaking transformants onto LB medium at the non-permissive temperature of 37 °C (ref. [50]). We used the strain MG1655 for all of the experiments, except for testing concentrations of galactose and DOG (Fig.1c). For that purpose, we placed *galK* under control of the p*BAD* promoter and used strain BW27784, which enables relatively linear induction of the p*BAD* promoter over a 1,000-fold range of arabinose concentration[51]. In both strain backgrounds, the genes *galK*, *mglBAC* and *galP* were altered to enable 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 (ref. [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 (ref. [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 with plasmid pMS6* using standard cloning techniques. The 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*, which was originally derived from pZA21-yfp[55]. Moreover, XmaI and XhoI restriction sites were added directly upstream and downstream of p_0 using two consecutive inverse PCRs.

The chromosomal gene cassette for testing concentrations of galactose and DOG (p*BAD-galK*; Fig. 1b) was assembled in the plasmid pIT07, which was obtained by cloning *galK-yfp* as well as a chloramphenicol resistance sequence flanked by FRT sites from pMS6* into pBAD24 (ref. [56]). Gene cassettes were integrated into chromosomal loci 1 and 2 (which correspond to loci D and E in ref. [21]) by recombineering[49] and were checked by PCR-amplifying using flanking primers and then sequencing the full-length construct.

Strain modification for microfluidics experiments. 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. To obtain a single-copy control locus, p_R-mCherry from our laboratory collection was introduced into the phage 21 attachment site (*attP21*) by P1-transduction[22].

***RecA* deletion in amplified strain locus 1.** *RecA* was deleted in the amplified strain by replacing it with the kanamycin cassette from pKD13 (ref. [50]). To maintain the amplified state, recombinants were selected on M9 0.1% galactose medium supplemented with 25 μg ml^{-1} kanamycin and verified by sequencing (Supplementary Fig. 3d,e).

Culture conditions. All of the experiments were conducted in M9 medium supplemented with 2 mM $MgSO_4$, 0.1 mM $CaCl_2$ and different carbon sources (Sigma-Aldrich). For evolution experiments, 0.1% galactose (high-expression environment) or 1% glycerol combined with 0.0001% DOG (low-expression environment) was added as a carbon source. For microfluidics experiments, M9 medium was supplemented with 0.2% glucose and 1% casein hydrolysate and 0.01% Tween 20 (Sigma-Aldrich) was added as surfactant before filtering the medium (0.22 μm).

All of the 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 plate reader (BioTek).

Mapping the relationship between *galK* expression level and growth. For the two-dimensional gradients of arabinose and galactose or DOG (Fig. 1b), 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. The cultures were grown on the plate reader with continuous orbital shaking.

Evolution experiments. For all of the evolution experiments (experimental evolution of the amplified strains in the high-expression environment and alternating selection experiments), cultures were grown in 200 μl liquid medium in 96-well plates and shaken in a Titramax plate shaker (Heidolph; 750 r.p.m.). Populations were transferred to fresh plates using a VP407 pinner (V&P Scientific) resulting in a dilution of ~1:133.

Evolution of the amplified strains in the high-expression environment. To obtain the amplified strains of locus 1 and 2, an overnight culture inoculated from a single colony of the ancestral strain carrying the reporter-gene cassette in locus 1 (IT028; Supplementary Figs. 1b–c) or 2 (IT030; Supplementary Fig. 4b) 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 to maximize the number of rescued populations and to minimize the amount of time that grown populations spent in stationary phase. The transfers were performed on days 10, 13, 15, 17, 18 and 19 (Supplementary Fig. 1c). The first signs of growth were detected in several wells after only approximately one week of cultivation in minimal galactose medium (Supplementary Fig. 1b). The evolving populations were monitored by spotting them onto MacConkey galactose agar in 128×86 mm omnitray plates before transfer. For locus 2, the evolving populations were transferred daily (~1:133, corresponding to seven generations) and spotted onto LB plates supplemented with 0.5% charcoal (Supplementary Fig. 4b) to improve fluorescence quantification. Colony fluorescence for all of the experiments was recorded using a custom-made macroscope set up (https://openwetware.org/wiki/Macroscope)[57]. To isolate clones, evolved populations were streaked twice for purification onto LB agar and grown in M9 galactose medium before freezing. For both locus 1 and 2, all further experiments were started from the original freezer stock of the amplified strain. This was done for two practical reasons: (1) to save the time needed for duplications (and higher order amplifications) to evolve (1 week in M9 galactose medium was used for locus 1 and 1 d in M9 medium supplemented with casamino acids was used for locus 2) and, importantly, (2) to enable interpretation and reproducibility of the fluorescence data of the alternating selection experiments. As the reporter-gene cassette enables selecting for increased *galK* expression but not for amplification itself, it is necessary to screen mutants with increased *galK* expression for increased CFP fluorescence. 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 analysed in detail. Amplified populations were therefore started from single colonies, which were grown non-selectively on LB (Lennox) agar by streaking the original freezer stock. Owing to the high rate of recombination, any given streak of the original amplified freezer stock contained colonies with a single copy of *galK* (Fig. 3a, right). 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 to be mutated in a previous study using the same locus[21]. For the strain amplified in locus 1 (IT028-EE1-D8), increased *galK* expression was 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* (S265A), enabling baseline expression via transcriptional read-through from the upstream *rsmG* into *galK* (ref. [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 insertion sequence 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*, position 4174770, C>T, resulting in R>H; *wcaF*, position 2128737, C>A, resulting in G>V).

Alternating selection experiments. For the experiments in Fig. 2b, a pre-culture of the amplified strain (IT028-EE1-D8) was grown overnight in M9 0.1% galactose, and was then inoculated 1:200 into the medium as indicated. For the experiment alternating between 2 d in the high-expression environment and 1 d in the low-expression environment (Fig. 2b, middle), populations were first treated with a scheme of daily alternating selection for 6 d before switching to the 2 d–1 d scheme.

For the co-culture experiments (Fig. 4), a preculture of the amplified strain (IT028-EE1-D8) was grown overnight in M9 0.1% galactose. 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 at a ratio of 1:1. We labelled 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) initially evolved in parallel to the amplified strain and resulted 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 (1) the ancestral strains (2) the amplified strains after initial selection in the high-expression environment (galactose) as well as (3) the amplified strains

after overnight selection in the low-expression environment (DOG), for locus 1 and locus 2. 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, whereas M9 0.1% galactose was used for the galactose-selected clones. A whole-genome library was prepared and sequenced by Microsynth AG using an Illumina NextSeq system (with a mean read length of 75 bp). FASTQ files were assembled to the MG1655 genome (GenBank, U00096.3) using the Geneious alignment algorithm with the default options in Geneious Prime v.2019.2.1. SNPs were analysed using the variant finding tool of Geneious.

Flow cytometry. Three colonies of each the amplified strain and the constitutive control strain were inoculated into culture tubes with 2 ml M9 0.1% galactose (high-expression environment) and grown for 3 d with transfers every 24 h. This population was inoculated into M9 + 1% glycerol + 0.0001% DOG (low-expression environment). OD_{600} was monitored to ensure continuous exponential growth by regular dilutions. The samples for flow cytometry analysis were frozen at the indicated time points (Fig. 2c). After 24 h in the low-expression environment, the populations were transferred back to the high-expression environment, and dilution and sampling were performed in the same manner. In parallel, the positive controls were grown for 5 d in both selection environments with transfers occurring every 24 h. Fluorescence was measured using a BD FACSCanto II system (BD Biosciences) equipped with FACSDiva software. Fluorescence from the Pacific Blue channel (CFP) was collected through a 450/50 nm band-pass filter using a 405 nm laser. Fluorescence of the FITC channel (YFP) was collected through a 510/50 nm band-pass filter using a 488 nm laser. The bacterial population was gated on the FSC and SSC signal resulting in approximately 6,000 events analysed 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 non-selective LB (Lennox) medium.

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

Analysis of microfluidics data. The mother machine enabled us to trace mother cells for ~38 divisions and, therefore, follow the fate of arising copy-number mutations in the absence of selection. In three experiments, we analysed 336, 369 and 384 mother cell lineages, respectively, equalling 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 were based on mother cell time traces (Fig. 4c). To this end, we used Fiji and 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 analysed using MATLAB.

Determining which 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 that we included.

1. We excluded all of the changes in fluorescence that occurred 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% and 82% of total colonies were included for the three experiments, respectively).
2. Some of the 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 of >1.5× the mCherry experiment-wide variance (therefore including 96%, 96% and 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 dataset if two adjacent colonies had a correlation of 0.6 or higher (99%, 98% and 98% of total colonies included for the three experiments, respectively).

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, which was bounded by a manually selected maximum and minimum for each channel, was chosen to include all except for extreme outliers (Supplementary Fig. 5a). 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 of the colonies was taken and a 7th degree polynomial was fitted. All of the time points were divided by the corresponding polynomial estimates.

Furthermore, mCherry fluorescence was flat-field corrected on the basis of the expectation that mCherry is roughly constant across all of the colonies. To achieve this, a line was fitted to the coordinate to obtain an estimate of the background of each location. The data were divided by the corresponding estimated value.

Probability density function. For the probability density function (PDF) in Supplementary Fig. 5b, we normalized for differential growth rate by dividing the CFP fluorescence by the constitutively expressed mCherry fluorescence. To reduce noise, a median filter (MATLAB function medfilt1) was applied to the ratio of CFP and mCherry over 20 data points.

To obtain an estimate of the PDF of the CFP/mCherry single-cell fluorescence, we used a kernel density estimation (MATLAB function ksdensity). To estimate a proxy for copy numbers, we found points at which the first and second derivative of the PDF was zero. These points were set as initial conditions for a pairwise fitting of peak mean and variance. All except for the first and the last peak had two estimates for mean and variance. For the mean, the average of the two was taken and the smaller value was chosen for the variance. 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 classified the single-cell traces using a normalized R^2, 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:

$$\begin{aligned} \mathrm{nS2R2} &= 1 - S_{\mathrm{res}}^{\mathrm{norm}}/S_{\mathrm{total}}^{\mathrm{norm}} \text{ where } S_{\mathrm{res}}^{\mathrm{norm}} \\ &= \textstyle\sum_i (y_i - f_i)^2/f_i^2,\ S_{\mathrm{res}}^{\mathrm{norm}} \sum_i (y_i - y_0)^2/f_i^2 \end{aligned}$$

where y_i, f_i, and y_0 represent measurements, fitted/predicted values and the mean of the measurements, respectively. This normalization takes into account that the intrinsic noise increases with expression and, therefore, penalizes it less. Next, the algorithm fits one constant to the start value and one constant to the end value of the CFP/mCherry trace and reports this estimation parameter (nS2R2). On the basis of this parameter, the algorithm classifies traces as shown in the pie charts of Supplementary Fig. 5c. Clear transitions exhibit an nS2R2 score of >0.5 and were verified visually, and the microfluidics videos were analysed in detail (Supplementary Table 1). The algorithm classified no-events (flat lines) if the nS2R2 score was between 0 and 0.5. Traces that could not be classified unambiguously as either a clear transition or a clear no-event, that is, with nS2R2 below 0, were classified as complex traces. This occurs if the start and end of CFP/mCherry trace values are similar but vary significantly in between.

qPCR. For quantitative PCR (qPCR), DNA was isolated using the Wizard Genomic DNA purification kit (Promega) from 50 µl of frozen samples that were obtained from different time points (0 h, 8 h, 14 h, 20 h, 24 h, positive control in galactose after 120 h, positive control in DOG after 120 h) of one flow cytometry experiment and a single copy control strain, all grown for 4–5 generations in LB. To quantify fluorescence, the same cultures were patched onto LB agar supplemented with 0.5% charcoal and imaged using the macroscope.

We performed qPCR using the Promega qPCR 2× Mastermix (Promega) and a C1000 instrument (Bio-Rad). To quantify the copy number of samples of an evolving population, we designed one primer within *cfp* (target) and used one primer within *rbsB* as a close reference, which lies outside the amplified region. We compared the ratios of the target and the reference loci to the ratio of the same two loci in the single copy control. Using dilution series of one of the genomic DNA extracts as template, we calculated that the efficiency of primer pairs was 89.01% and 92.57% for *cfp* and *rbsB*, respectively. We quantified the copy number of *cfp* in each sample using the Pfaffl method, which takes amplification efficiency into account[58]. qPCR was performed in three technical replicates.

Measurement of colony fluorescence. Colonies were grown without selection and imaged using the macroscope set-up.

To obtain mean colony CFP fluorescence intensity, a region of interest was determined using the ImageJ plugin 'Analyze Particles' (settings: 200px-infinity, 0.5–1.0 roundness) to identify colonies on 16-bit images with the threshold adjusted according to the default value. The region of interest including all of the colonies was then used to measure intensity (Fig. 3a, Supplementary Figs. 1c and 4b).

Mathematical model. A simple mathematical model recapitulates the change in *galK* copy number of the amplified population (Fig. 5a). Importantly, the parameters for the model were estimated purely from calibration measurements (growth rates, fitness in the two environments with 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 Supplementary Table 2. No parameter was fit to reproduce the measurements in Fig. 5a.

The model describes the time evolution of a population in which 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; discrete times represent synchronous cell divisions in the population. The size of subpopulation N_j of cells with gene copy number j at time $t+1$ equals:

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

where s_j is the relative growth rate of the subpopulation with j gene copies in the given environment (Fig. 2d), δ_{jk} is a Kronecker delta that 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 as follows: assuming a homologous recombination between sister chromosomes occurs somewhere in the gene, we computed all of the possible combinations of how genes can be recombined to form a 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 a 22% probability to obtain four gene copies or 22% probability to have one copy in the daughter (Supplementary Fig. 6h). We have observed in microfluidics experiments that most (65%) copy-number changes happen only in the mother cell while the daughter cell remains unchanged. Thus, we do not model recombination as a reciprocal event.

On the basis of plate reader bulk experiments, observations indicated an upper limit for the copy number that a cell can have. In our model, a cell can therefore 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, which indicated that the limit is between 6 and 12. Our single-cell analysis showed that $M=10$ is a good estimate (according to number of states in the probability density function; see the 'Analysis of the microfluidics data' section; Supplementary Fig. 5b). However, the results of the mathematical model do not depend on the precise value within the measured range, as all of the results remain qualitatively the same for any value in the range of 6 to 12. Supplementary Fig. 6g shows that the relative growth rates, obtained from flow cytometry experiments, are independent of M.

Measurements of model parameters. T1 and T2 generations per day in 96-well plates. 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. We therefore estimated growth rate indirectly without perturbing the experiment by determining the maximal number of generations possible in 24 h (number of generations $=24[h]\times$ growth rate$[1/h]/\log_2$) from a dilution series experiment. Populations that were 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 that the OD_{600} reaches the OD_{600} of the stationary phase after 24 h of growth). 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. All model parameters are listed in Supplementary Table 2.

T10 and T20 generations per day in culture tubes. The parameters T10 and T20 were necessary to obtain the fitness landscape in Fig. 2d (and the resulting relative growth rates s_j). T10 and T20 generations per day were measured under the exact conditions of the flow cytometry experiment (Fig. 2c); the conditions were as follows: exponential growth in culture tubes with 2 ml of M9 with 0.1% galactose or M9 1% glycerol + 0.0001% DOG, respectively. We measured OD_{600} using a WPA Biowave spectrophotometer (Biochrom).

Determining the 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 events ($k_{dup}=0$, $k_{rec}=0$). This is a valid approximation provided that the two rates are small enough, such that the population structure consists of all copy-number types, that is, 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, we took the values of fluorescence of <85, whereas, for the high bin, we took the mode fluorescence values of the measured distributions, corresponding to >160 for the first and >245 for the second set of flow cytometry experiments. Each bin represents a given gene copy number. The distributions between different times were then compared using an iterative exponential growth model:

$$N_j(t_2) = \left(1+s_j\right)^{(t_2-t_1)/t_{1/2}} N_j(t_1) \quad (2)$$

where N_j is the population size with j gene copy number, $t_{1/2}$ is the doubling time, t_1 and t_2 are two measurement times and s_j represents the relative growth of cells with j gene copies. The population distributions for all of the time points were obtained from the flow cytometry data given the binning described above. Using this model, we obtained growth rates s_j for each pair of consecutive distributions at times t_i and t_{i+1} as follows: given population distribution at time i, we predicted the new distribution given equation (2). We found such s_j values that minimize the Euclidian difference between the predicted and observed population distribution at time $i+1$. We repeated this for all of the pairs of consecutive distributions (at times t_i and t_{i+1}) and different replicates to obtain a set of solutions for s_j. Using this approach, we acquired only relative growth rates, which still allowed constants to be added to the growth rates. To address this, we added such constants to each growth rates to (1) minimize the χ^2 of the differences between each growth rate solution and the mean of all of the solutions, which optimally removes the replicate-to-replicate variability (Fig. 2d, error bars) on the inferred relative growth rates but does not affect their mean value; and (2) force the average growth rate of the adapted state to be 1 (that is, for $j=1$ in low-expression environment and $j=M$ is high-expression environment, $s_j=1$) by adding a term to the χ^2 error function of the form (adapted state expression -1)^2. Fixing s to be 1 in a reference environment is a convention that mathematically will not affect any subsequent results.

The absolute maximal growth rates in the two environments were measured in populations grown in high- and low-expression environments for 120 h, respectively. These rates therefore represent the growth rates of populations with the highest and lowest possible copy numbers (Fig. 2c, positive controls). The estimated fitness values for both high-expression environment (s_j^{HEE}) and low-expression environment (s_j^{LEE}) are provided in Supplementary Table 2.

Estimation of recombination rate k_{rec} from microfluidics data. We obtained a conservative estimate for the lower bound of the average number of copy-number mutations from single-step transitions in the pie charts (Supplementary Fig. 5c). Out of 72 mother-cell time traces that were classified as clear transition events, we verified 67 by performing detailed analysis of microscopy images (Supplementary Table 1). Accordingly, we calculated the lower bound for the mutation rate as 67 events/1,089 lineages/22.7 generations yielding $k_{rec}=2.7\times10^{-3}$ ($\pm7.4\times10^{-4}$) per cell per generation.

To estimate the mean recombination rate to be used in the model, two corrections have to be made: (1) as our model assumes that the recombination rate is proportional to the number of gene copies[8], we had to take into account that cells with a higher initial gene copy number are more likely to undergo a recombination event; and (2) as our experimental setup only allowed us to see if there has been a change in gene copy numbers or not, we had to take into account that there are some recombination events that do not change the gene copy number.

To account for (1), we first computed the probability distribution that a given number of independent recombination events occur (Supplementary Fig. 5d): 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 (that is, the effective recombination rate × 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, that is, it assumes a constant effective recombination rate for each trace throughout the experiment that 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 occurring. Although 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 × the basal recombination rate. We therefore used all

of the single-cell traces to estimate a starting gene copy distribution. To achieve this, we averaged the normalized fluorescence (as a proxy for the starting effective gene copy number; Fig. 3c) over the time points 60–150 min. We next computed a Poisson probability distribution of obtaining k events ($k=0,1,\ldots$) in the time of the experiment for each individual trace, with the basal recombination rate multiplied with the starting gene copy number (Supplementary Fig. 5d). For example, if a single-cell trace started with 4 gene copies, the expected number of events per experiment would be 4 × the basal recombination rate × the number of generations. We next averaged over all of the computed Poisson probability distributions that were obtained from all of the single-cell traces. We effectively obtained a total probability distribution for seeing 0, 1 or more recombination events over all of the recorded single-cell traces, taking into account point (1).

Next, we consider point (2), taking into account the effect of recombination events that do not change the gene copy number. We know from the P_{kj} matrix that the probability of keeping the gene copy numbers is the reciprocal of the initial gene copy number. We therefore took into account all of the 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 the 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 of the double events in which one of the recombination events leaves the copy number unchanged, all of the triple events in which two events keep the copy numbers unchanged and so on. The probability of observed events therefore also includes the actual probability from states with $k>0$ in which recombination did not change the copy number: $p_{\text{observed}}(k=0)=p_{\text{actual}}(k=0)+\sum_j p_{\text{actual}}(j)/\varepsilon_0^j$, for all $j>0$, where $p(j)$ is the probability of having j recombination events and ε_0 is 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: $p_{\text{observed}}(k=1)=p_{\text{actual}}(k=1)+\sum_j (j-1)p_{\text{actual}}(j)/\varepsilon_0^{j-1}$, 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, if there are three recombination events, there are three different ways in which there are two events where the gene copy number remains unchanged and one event in which it is changed.

After taking both corrections into account, we obtained a probability distribution of observing k recombination events (Supplementary Fig. 5d). The estimate of the basal recombination rate, $k_{\text{rec}}{}^0$, 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 (Supplementary Fig. 5c–d). We obtained $k_{\text{rec}}{}^0$ as 0.01434 per cell per generation, which is approximately 5× larger than the conservative lower bound.

Model comparison with experimental data. To compare the model with the experimental data (Fig. 5a), we simulated the full experimental protocol (the parameter values are provided in Supplementary Table 2):

(1) We exposed a single-copy ancestral population to a week of the 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 populations started with the same degree of copy-number polymorphism.
(2) The population spent 1 d in the low-expression environment (details on the procedure for each day are provided below).
(3) For the experiment shown in Fig. 5a (top), the population was also exposed to three daily oscillations between the high- and low-expression environments.
(4) The population was exposed to the environments indicated in Fig. 5a.
(5) For every experiment, the bacterial culture was diluted by a factor of $D=133$ every day, therefore limiting growth. This growth limitation was enforced by multiplying all of the growth rates by $g(c)=1-\min(c/133.0))^{0.01}$ where c is 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 1 d consisted of (parameter values are provided in Supplementary Table 2):

(1) Given the recombination rate and number of states M, we computed the transition matrix P_{kj} (equation (1)) as follows: given k copy numbers, the probability of going from k to $j<k$ copy numbers equals j/k^2, while the probability of going from k to $j\geq k$ equals $(2k-j)/k^2$ (ref. [8]). Furthermore, we assumed that no transitions that increase copy numbers beyond M are allowed. We implemented this by setting all of the probabilities that go over M gene copies to zero.
(2) Next, to update the current population structure following equation (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, that is, the ratio of population size of the current time point and the size after the last dilution. Second, we computed $g(c)$ (taking into account the saturation of the population) and multiplied it by each of the selection values s_j in equation (1). We then used these new values to compute N_j at the new time point.
(3) Step 2 was repeated 23 or 22 times for the 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 1 d.
(4) We diluted the population by a factor of $D=133$.
(5) We repeated steps 2–4 according to the environment that the population is exposed to 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 proportion of cells with j gene copies and sum is taken over all different gene copy numbers j.
(7) At the end, we returned the population distribution and expression at each time point.

To simulate 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 population of a finite size (Supplementary Fig. 6f), we used the Wright–Fisher model, where the population size is kept constant. The procedure was:

(1) Given all of the parameters of the system and using the infinite size population model (equation (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 the right-hand side of equation (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 equation (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.

Supplementary Fig. 6f 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 the number of replicate experiments, mean values and s.d., 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 the mean constitutive strain abundance of days 1–15 in the high-expression environment and the mean constitutive strain abundance in the low-expression environment on day 16.

We used a two-sided t-test (Matlab function ttest2) to compute the P value (2.6×10^{-68}) for the difference in mean response between the amplified (Fig. 4c) and co-culture populations (Fig. 4d).

To measure the linear dependence between the experimental data and model prediction in Fig. 5a, we computed the Pearson correlation coefficient using the inbuilt Matlab function corrcoef.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

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

Code availability

The scripts for our mathematical model and for the analysis of microfluidics time traces have been deposited in IST DataRep and are publicly available at https://research-explorer.app.ist.ac.at/record/7383.

Received: 5 August 2019; Accepted: 29 January 2020;
Published online: 09 March 2020

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. USA*. **71**, 2453–2455 (1974).
3. Gerland, U. & Hwa, T. Evolutionary selection between alternative modes of gene regulation. *Proc. Natl Acad. Sci. USA* **106**, 8841–8846 (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**, 42 (2004).
6. Anderson, P. & Roth, J. Spontaneous tandem genetic duplications in Salmonella typhimurium arise by unequal recombination between rRNA (rrn) cistrons. *Proc. Natl Acad. Sci. USA* **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. in *Escherichia coli and Salmonella: Cellular and Molecular Biology* (ed. Neidhardt, F. C.) 2256–2276 (American Society for Microbiology, 1996).
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**, 886–890 (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 generates heteroresistance (variation in resistance within a population) to colistin in *Salmonella enterica*. *Mol. Microbiol.* **102**, 274–289 (2016).
16. Näsvall, J., Sun, L., Roth, J. R. & Andersson, D. I. Real-time evolution of new genes by innovation, amplification, and divergence. *Science* **338**, 384–387 (2012).
17. Elde, N. C. et al. Poxviruses deploy genomic accordions to adapt rapidly against host antiviral defenses. *Cell* **150**, 831–841 (2012).
18. Kussell, E. & Laibler. Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309**, 2075–2078 (2005).
19. Veening, J.-W., Smits, W. K. & Kuipers, O. P. Bistability, epigenetics, and bet-hedging in bacteria. *Annu. Rev. Microbiol.* **62**, 193–210 (2008).
20. Barkan, D., Stallings, C. L. & Glickman, M. S. An improved counterselectable marker system for mycobacterial recombination using *galK* and 2-deoxy-galactose. *Gene* **470**, 31–36 (2011).
21. Steinrueck, M. & Guet, C. C. Complex chromosomal neighborhood effects determine the adaptive potential of a gene under selection. *eLife* **6**, e25100 (2017).
22. Bergmiller, T. et al. Biased partitioning of the multidrug efflux pump AcrAB-TolC underlies long-lived phenotypic heterogeneity. *Science* **356**, 311–315 (2017).
23. Wang, P. et al. Robust growth of *Escherichia coli*. *Curr. Biol.* **20**, 1099–1103 (2010).
24. Reams, A. B. & Roth, J. R. Mechanisms of gene duplication and amplification. *Cold Spring Harb. Perspect. Biol.* **7**, a016592 (2015).
25. Tawfik, D. S. Messy biology and the origins of evolutionary innovations. *Nat. Chem. Biol.* **6**, 692–696 (2010).
26. Jacob, F. Evolution and tinkering. *Science* **196**, 4295 (1977).
27. Troein, C., Ahrén, D., Krogh, M. & Peterson, C. Is transcriptional regulation of metabolic pathways an optimal strategy for fitness? *PLoS ONE* **2**, e855 (2007).
28. Wolf, L., Silander, O. K. & van Nimwegen, E. Expression noise facilitates the evolution of gene regulation. *eLife* **4**, e05856 (2015).
29. Anderson, R. P. & Roth, J. R. Tandem genetic duplications in phage and bacteria. *Annu. Rev. Microbiol.* **31**, 473–505 (1977).
30. Taylor, T. B. et al. Evolutionary resurrection of flagellar motility via rewiring of the nitrogen regulation system. *Science* **347**, 1014–1017 (2015).
31. Rodriguez-Beltran, J. et al. Multicopy plasmids allow bacteria to escape from fitness trade-offs during evolutionary innovation. *Nat. Ecol. Evol.* **2**, 873–881 (2018).
32. Perry, G. H. et al. Diet and the evolution of human amylase gene copy number variation. *Nat. Genet.* **39**, 1256–1260 (2007).
33. Gil, R., Sabater-Muñoz, B., Perez-Brocal, V., Silva, F. J. & Latorre, A. Plasmids in the aphid endosymbiont *Buchnera aphidicola* with the smallest genomes. A puzzling evolutionary story. *Gene* **370**, 17–25 (2006).
34. Latorre, A., Gil, R., Silva, F. J. & Moya, A. Chromosomal stasis versus plasmid plasticity in aphid endosymbiont *Buchnera aphidicola*. *Heredity* **95**, 339–347 (2005).
35. Lercher, M. J. & Pál, C. Integration of horizontally transferred genes into regulatory interaction networks takes many million years. *Mol. Biol. Evol.* **25**, 559–567 (2008).
36. Pál, C., Papp, B. & Lercher, M. J. Adaptive evolution of bacterial metabolic networks by horizontal gene transfer. *Nat. Genet.* **37**, 1372–1375 (2005).
37. Dobrindt, U., Hochhut, B., Hentschel, U. & Hacker, J. Genomic islands in pathogenic and environmental microorganisms. *Nat. Rev. Microbiol.* **2**, 414–424 (2004).
38. Juhas, M. et al. Genomic islands: tools of bacterial horizontal gene transfer and evolution. *FEMS Microbiol. Rev.* **33**, 376–393 (2009).
39. Pettersson, M. E., Andersson, D. I., Roth, J. R. & Berg, O. G. The amplification model for adaptive mutation. *Genetics* **169**, 1105–1115 (2005).
40. Gusev, O. et al. Comparative genome sequencing reveals genomic signature of extreme desiccation tolerance in the anhydrobiotic midge. *Nat. Commun.* **5**, 4784 (2014).
41. Hooper, S. D. & Berg, O. G. Duplication is more common among laterally transferred genes than among indigenous genes. *Genome Biol.* **4**, R48 (2003).
42. Eme, L., Gentekaki, E., Curtis, B., Archibald, J. M. & Roger, A. J. Lateral gene transfer in the adaptation of the anaerobic parasite blastocystis to the gut. *Curr. Biol.* **27**, 807–820 (2017).
43. Nguyen, T. N., Phan, Q. G., Duong, L. P., Bertrand, K. P. & Lenski, R. E. Effects of carriage and expression of the Tn*10* tetracycline-resistance operon on the fitness of *Escherichia coli* K12. *Mol. Biol. Evol.* **6**, 213–225 (1989).
44. Gladman, S. L. et al. Large tandem chromosome expansions facilitate niche adaptation during persistent infection with drug-resistant *Staphylococcus aureus*. *Microb. Genomics* **1**, e000026 (2015).
45. Elliott, K. T., Cuff, L. E. & Neidle, E. L. Copy number change: evolving views on gene amplification. *Future Microbiol.* **8**, 887–899 (2013).
46. Eydallin, G., Ryall, B., Maharjan, R. & Ferenci, T. The nature of laboratory domestication changes in freshly isolated *Escherichia coli* strains. *Environ. Microbiol.* **16**, 813–828 (2014).
47. Greenblum, S., Carr, R. & Borenstein, E. Extensive strain-level copy-number variation across human gut microbiome species. *Cell* **160**, 583–594 (2015).
48. Dhar, R., Bergmiller, T. & Wagner, A. Increased gene dosage plays a predominant role in the initial stages of evolution of duplicate TEM-1 beta lactamase genes. *Evolution* **68**, 1775–1791 (2014).
49. Datta, S., Costantino, N. & Court, D. L. A set of recombineering plasmids for gram-negative bacteria. *Gene* **379**, 109–115 (2006).
50. Datsenko, K. A. & Wanner, B. L. One-step inactivation of chromosomal genes in *Escherichia coli* K-12 using PCR products. *Proc. Natl Acad. Sci. USA* **97**, 6640–6645 (2000).
51. Khlebnikov, A., Datsenko, K. A., Skaug, T., Wanner, B. L. & Keasling, J. D. Homogeneous expression of the P_{BAD} promoter in *Escherichia coli* by constitutive expression of the low-affinity high-capacity AraE transporter. *Microbiology* **147**, 3241–3247 (2001).
52. Baba, T. et al. Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection. *Mol. Syst. Biol.* **2**, 2006.0008 (2006).
53. Nagelkerke, F. & Postma, P. W. 2-Deoxygalactose, a specific substrate of the *Salmonella typhimurium* galactose permease: its use for the isolation of galP mutants. *J. Bacteriol.* **133**, 607–613 (1978).
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/I_1-I_2 regulatory elements. *Nucleic Acids Res.* **25**, 1203–1210 (1997).
56. Guzman, L. M., Belin, D., Carson, M. J. & Beckwith, J. Tight regulation, modulation, and high-level expression by vectors containing the arabinose P_{BAD} promoter. *J. Bacteriol.* **177**, 4121–4130 (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 (2001).
59. Drake, J. W. A constant rate of spontaneous mutation in DNA-based microbes. *Proc. Natl Acad. Sci. USA* **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).

Acknowledgements

We thank L. Hurst, N. Barton, M. Pleska, M. Steinrück, B. Kavcic and A. Staron for input on the manuscript, and To. Bergmiller and R. Chait for help with microfluidics experiments. I.T. is a recipient the OMV fellowship. R.G. is a recipient of a DOC (Doctoral Fellowship Programme of the Austrian Academy of Sciences) Fellowship of the Austrian Academy of Sciences.

Author contributions

C.C.G., R.G., M.L., G.T. and I.T. conceived the study. I.T. performed experiments. A.M.C.A., R.G. and I.T. analysed data. R.G. and G.T. performed the formal analysis. R.G. and I.T. wrote the original draft and revised with A.M.C.A., J.P.B., C.C.G., M.L. and G.T.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at https://doi.org/10.1038/s41559-020-1132-7.

Correspondence and requests for materials should be addressed to C.C.G.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

nature research

Corresponding author(s): Calin Guet

Last updated by author(s): 20/01/20

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☐	☒	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☐	☒	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	FACS diva software (BD); NIS (Nikon Eclipse Microscope Software); Gen5 (Biotek Synergy H1 platereader Software)
Data analysis	ImageJ (Fiji) for Image analysis; custom matlab code for i) microfluidics data analysis, ii) used for modeling part; Geneious Prime version 2019.2.1 used to analyze whole genome sequence data

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

All data generated or analysed during this study are either included in this manuscript (and its supplementary information files) or will be made available from the corresponding author on request.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	evolution experiments were carried out at maximum possible replication number (full 96-well plate). Replicate sample sizes were determined based on previous studied in the lab of similar kind
Data exclusions	No data were excluded;
Replication	Evolution experiments were done with 18-60 replicate populations; Evolution experiments were highly reproducible (parallel evolutionary outcome) due to the high rate of copy number mutations involved.
Randomization	Measurements of all replicates were taken at independent time points, and were hence inherently randomized
Blinding	Experiments involved comparisons of only a single group in biological and technical replicates, not requiring blinding by the experimenters

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☒	☐ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☐	☒ Flow cytometry
☒	☐ MRI-based neuroimaging

Flow Cytometry

Plots

Confirm that:

☒ The axis labels state the marker and fluorochrome used (e.g. CD4-FITC).

☒ The axis scales are clearly visible. Include numbers along axes only for bottom left plot of group (a 'group' is an analysis of identical markers).

☒ All plots are contour plots with outliers or pseudocolor plots.

☐ A numerical value for number of cells or percentage (with statistics) is provided.

Methodology

Sample preparation	Exponentially growing E.coli cultures were sampled along indicated timepoints, frozen and measured without any staining.
Instrument	FACS Canto II (BD)
Software	FACS Diva Software (BD)
Cell population abundance	No sorting was done.

Gating strategy | Cells were gated for FSC/SSC against filtered medium.

☐ Tick this box to confirm that a figure exemplifying the gating strategy is provided in the Supplementary Information.