Supplementary Materials for

Flux, toxicity and protein expression costs shape genetic interaction in a metabolic pathway

Harry Kemble, Catherine Eisenhauer, Alejandro Couce, Audrey Chapron, Mélanie Magnan, Gregory Gautier, Hervé Le Nagard, Philippe Nghe, Olivier Tenaillon.

Correspondence to: olivier.tenaillon@inserm.fr

This PDF file includes:

Materials and Methods
Figs. S1 to S11
Tables S1 to S5
Captions for Data S1 to S2

Other Supplementary Materials for this manuscript include the following:

Data S1 to S2 [Mutant fitness estimates with their 95% bootstrap confidence intervals and the number of barcodes used for their estimation, Parameter estimates for complete phenotype-fitness model]

Materials and Methods

General microbiology and molecular biology

Lysogeny Broth (LB) powder, agar, salts, sugars, growth supplements, antibiotics and inducers were all purchased from Sigma-Aldrich. Bacteria were cultured in LB, unless otherwise stated. Liquid LB was the standard Lennox formulation, except for when blasticidin-S was included, in which case the Luria low-salt formulation (0.5 g/L NaCl) was used. LB-agar always contained the Luria low-salt formulation. M9 base medium consisted of 1X M9 salts supplemented with 1mM $MgSO_4$ and 100 µM $CaCl_2$. Unless otherwise stated, L-arabinose was used at a concentration of 0.03% w/v. Ampicillin (amp) was used at 100 µg/ml, chloramphenicol (cm) at 10 µg/ml, streptomycin (str) at 50 µg/ml, blasticidin-S (bsd) at 100 µg/ml and erythromycin (erm) at 20 µg/ml. Bacterial cultures were grown at 37°C (with shaking at 200 rpm for liquid cultures; Multitron, Infors HT), unless otherwise stated, and culture stocks were stored at -80°C in LB with 40% glycerol. For electroporation, DNA was added to 50 µl homemade electro-competent cells (unless otherwise stated), transferred to a 1mm-gap electroporation cuvette (VWR) and submitted to a pulse of 1,800 V (Electroporator 2510, Eppendorf). Cells were immediately transferred to fresh LB for recovery at 37°C (unless otherwise stated) with shaking for 30-90 minutes, before being plated on the appropriate selective media and left to grow overnight.

All enzymes and molecular biology reagents were purchased from NEB, unless otherwise stated. Primers were purchased from IDT or Eurofins, and designed with the help of Primer3 (*1*). For sensitive applications like barcoding and NGS library preparation, primers were ordered HPLC-purified, otherwise they were ordered desalted. UltraPure agarose was supplied by Invitrogen, and all agarose gels were stained with SYBR Safe (Thermo Scientific) and visualised with a GelDoc XR+ imager (Bio-Rad). The GeneRuler 1kb Plus ladder (Thermo Scientific) was used for DNA fragment size estimation.

All plasmids used in this study, excluding the mutant library, are detailed in Table S1. DNA fragments used in cloning are detailed in Table S2. Primers, excluding those used for promoter mutagenesis, are provided in Table S3. All strains are detailed in Table S4. Primers used in promoter mutagenesis are provided in Table S5.

Plasmid construction

Our library creation strategy depended on two plasmids, pKH1511c and pKH1511d, which were created in this study. pKH1511c serves as the library "backbone", carrying all the necessary elements of the final plasmid library except for the $P_{LtetO-1}$ and $P_{LlacO-1}$ promoters destined to drive *araA* and *araB* expression, respectively: *p15A* origin-of-replication, *lacI-tetR* repressor cassette (for inducibility of $P_{LtetO-1}$ and $P_{LlacO-1}$), *araA* and *araB*. *araA* and *araB* ORFs (with their upstream ribosome binding site-containing regions) are divergently oriented (with each followed by an artificial transcriptional terminator), and are separated by 2 restriction sites to allow easy insertion of divergently oriented $P_{LtetO-1}$ and $P_{LlacO-1}$ promoters. pKH1511d serves as a template for amplification of a *bsd* blasticidin S-resistance cassette with primers containing the $P_{LtetO-1}$ and $P_{LlacO-1}$ variant sequences, allowing their eventual insertion into pKH1511c (Figure S1). pKH1511d replication is pir-dependent, abolishing the occurrence of false-positive colonies caused by PCR template carryover during library cloning. Plasmids, DNA fragments, PCR primers and bacterial strains used in the construction of these two plasmids are detailed in Tables S1-4, respectively, and the detailed cloning methods follow.

The DNA fragments used to construct pKH1503a, pKH1511c and pKH1511d come from either PCR amplification or from direct restriction digestion of purified plasmid DNA, and were joined by either standard restriction-ligation or by Gibson Assembly (*2*) (in which case, overlaps of ~40 nucleotides were used). PCR amplifications were all performed with Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific) in its High-Fidelity buffer, following the manufacturer's recommendations. Restriction enzymes were used according to the manufacturer's instructions. When found necessary to reduce the occurrence of false-positive colonies, DNA was treated with calf intestinal alkaline phosphatase (to reduce vector self-ligation) and/or DpnI (to digest PCR template). After PCR amplification and/or digestion, DNA fragments were either verified by electrophoresis and column-purified (QIAquick PCR Purification Kit, QIAGEN) or, when necessary, gel purified (QIAquick Gel Extraction Kit, Qiagen). Gel-purification was always followed by a 2nd clean-up (QIAquick PCR Purification Kit, QIAGEN) to improve DNA quality for ligation. For gel extractions, agarose gels were stained with SYBR Safe (Thermo Scientific), and DNA was visualised with blue light to avoid UV-induced DNA damage (Blue Transilluminator, Pearl Biotech). A NanoDrop ND-1000 spectrophotometer (Thermo Scientific) was used to determine DNA concentration for all fragments prior to ligation/Gibson Assembly. Standard ligation and Gibson Assembly were performed using T4 ligase and Gibson Assembly Master Mix (NEB), respectively, according to the manufacturer's recommendations (T4 ligase was then inactivated by heating at 65°C for 10 mins). In both cases, DNA was subsequently microdialysed against water for > 30 mins (MF-Millipore, Merck), and 1-5 µl were electroporated into 50 µl electrocompetent cells. DH5α ΔaraBA was used as the cloning strain except when the plasmid was pir-dependent, in which case PIR1 was used. After electroporation, cells were recovered in 1 ml LB for 30-90 mins at 37°C with shaking at 200 rpm, plated on LB-agar in the presence of the antibiotic indicated in Table S1 and incubated overnight at 37°C. Plasmid DNA was purified from several colonies (QIAquick PCR Purification Kit, QIAGEN) and verified by both restriction analysis and Sanger sequencing of the insert region.

Strain engineering and adaptation

The final library host strain, *E. coli* MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo} (Table S4), was originally designed to possess a rewired D-arabinose metabolism (*3–5*), in which *araB* (but not *araA*) participates. D-arabinose was not used in this study, however, so this feature (D-ara$^{+/evo}$ Δ*fucK*) is not relevant here. In addition, *araA* and *araB* ORFs were removed from the chromosome, to allow them to be expressed exclusively from plasmids (the 3rd gene of the *araBAD* operon, *araD*, was kept on the chromosome under the control of its native L-arabinose-responsive promoter, as were the transcriptional regulator gene, *araC*, and the transporter genes, *araE*, *araFGH* and *araJ*; given the all-or-nothing response of the positive feedback loop governing L-arabinose uptake, all these genes are expected to be maximally induced by internal L-arabinose by the time of fitness measurement (*6*)). Further, *lacIZYA* was replaced by a *cat* chloramphenicol-resistance cassette. This allows the use of IPTG to control the artificial promoter, $P_{LlacO-1}$, in the absence of any effects resulting from induction of the native *lac* operon, and the absence of *lacY* also causes this control to be titratable rather than all-or-nothing (*7*). Finally, this strain was transformed with plasmid pKH1503a (which carries an *araBA* cassette under the control of $P_{LlacO-1;}$ Table S1) and briefly adapted to M9 with alternating D-/L-arabinose (see above) in the presence of a low concentration of IPTG. This adaptation step was included to allow fixation of any mutations conferring a very high fitness advantage to our

engineered strain in our approximate experimental conditions, to avoid them interfering with mutant library competition experiments. Detailed strain engineering methods follow.

Details of the final library host strain, and all intermediates used in its creation, are provided in Table S4. Gene knockouts were performed using the method of Datsenko and Wanner (*8*). The relevant strain was made electrocompetent, electroporated with 10 ng plasmid pKD46 DNA, and transformants were selected on LB-agar with 100 µg/ml ampicillin at 30°C. Several colonies were then re-isolated under the same conditions. The cat chloramphenicol-resistance cassette was PCR-amplified from pKD3 (*8*) using primer pairs KO-araBA-fwd/KO-araBA-rev for araBA, KO-lacIZYA-fwd/KO-lacIZYA-rev for lacIZYA and KO-fucK-fwd/KO-fucK-rev for fucK, and a 2:1 mix of GoTaq/Pfu DNA polymerases (Promega). PCR products were verified by 1% agarose gel electrophoresis, column-purified (QIAquick PCR Purification Kit, QIAGEN) and spectrophotometrically quantified (NanoDrop ND-1000). A pre-culture of a single pKD46-transformed colony was grown overnight (LB-amp) at 30°C and then diluted 100x into LB-amp with 0.2% L-arabinose and grown at 30°C to an OD_{600nm} of ~0.7 (BioMate 3S, Thermo Scientific; 3-5 hours). The culture was made electrocompetent, electroporated with ~200 ng of the purified PCR product, and recombinants were selected on LB-agar with 10 µg/ml chloramphenicol at 37°C, for curing of pKD46. Several colonies were then re-isolated under the same conditions, and tested in parallel for pKD46 curing by plating on LB-amp and checking for colonies after an overnight growth at 30°C. Several of the re-isolated colonies were verified by colony-PCR, using 3 primer pairs for each knockout (*8*). The gene-specific primers are verif-araBA-fwd/verif-araBA-rev for araBA, verif-lacIZYA-fwd/verif-lacIZYA-rev for lacIZYA and verif-fucK-fwd/verif-fucK-rev for fucK, and the common cat primers are c1 and c2 from reference (*8*). For each knockout, the 3 primer pairs were: gene-specific fwd/gene-specific rev, gene-specific fwd/c1 and gene-specific rev/c2. GoTaq DNA polymerase (Promega) was used for amplification, following the manufacturer's recommendations, and PCR products were analysed by agarose gel electrophoresis (1.5%). In the case of *araBA* and *fucK*, we wished to retain function of the remaining genes in their respective operons, and so the cat cassette was removed as described in reference (*8*). For this, a pre-culture of a single recombineered colony was grown overnight (LB-cm, 37°C) and then diluted 100x into LB-cm and grown at 37°C to an OD_{600nm} of ~0.7 (BioMate 3S, Thermo Scientific; 2-4 hours). The culture was made electrocompetent, electroporated with 10 ng plasmid pCP20 DNA, and transformants were selected on LB-agar with 100 µg/ml ampicillin at 30°C. Several colonies were then re-isolated under the same conditions, and then again in the absence of ampicillin at 42°C, to cure pCP20 (*8*). Finally, several colonies were streaked in parallel on LB (37°C, purification), LB-cm (37°C, verify *cat* loss) and LB-amp (30°C, verify pCP20 loss). The loss of the *cat* cassette through FRT recombination was verified molecularly for several clones by colony-PCR, using the same primer pairs and conditions described above for *cat* insertion verification. The PCR products resulting from amplification with the gene-specific primer pairs were also Sanger-sequenced (GATC; using the amplification primers) as a final verification.

Adaptations were performed as described in Table S4. For the initial adaptation step, pre-cultures were grown overnight in LB, washed twice in an equal volume of M9, and 1 ml washed cells were diluted in 100 ml of the appropriate adaptation media. Once growth became apparent, cultures were serially transferred in a volume of 20 ml, being left to grow for ~24 hours between each transfer, at which point they were diluted ~100x into fresh media. After adaptation, colonies were isolated on agar plates containing the same media used for adaptation.

To cure the plasmid from MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo}, a pre-culture was grown overnight in LB-cm, and dilutions were plated on LB-cm with 2% ribitol and 200 µM IPTG. IPTG induces *araBA* from the plasmid, and AraB converts ribitol to the toxic compound ribitol phosphate (*9*), rendering plasmid-harbouring cells unable to grow. Several colonies were tested and confirmed for plasmid loss by streaking on LB-str and by colony-PCR (primers oKH150401c/oKH150202d, GoTaq (Promega)), with comparison to control colonies grown in the absence of ribitol. The final plasmid-less host strain was also tested once more for its marker-less *ΔaraBA* and *ΔfucK* deletions using colony-PCR (primer pairs verif-araBA-fwd/verif-araBA-rev and verif-fucK-fwd/verif-fucK-rev, as above).

Library creation strategy

On the evolutionary scale, direct changes in the total cellular activity of a particular enzyme can occur through either regulatory mutations, which alter the concentration of active enzyme, or structural mutations, which can effect both active enzyme concentration and kinetic parameters. A common target of regulatory mutations is the promoter (*10*), which controls a protein's expression level by determining transcription rate, and we decided to focus on promoter mutations in this study. We first placed *araA* and *araB* under the control of the well-known artificial, chemically-inducible promoters, $P_{LtetO-1}$ and $P_{LlacO-1,}$ developed by Lutz and Bujard (*11*). They are each regulated by a single transcription factor (*tetR* repressor for $P_{LtetO-1}$ and *lacI* repressor for $P_{LlacO-1}$), and can be specifically induced to different levels by addition of a small, non-metabolisable compound (aTc for $P_{LtetO-1}$ and IPTG for $P_{LlacO-1}$). We focussed mutagenesis on the RNA polymerase-binding sites (-35 and -10 hexamers) of the two promoters, as these sites are known to be the most significant determinants of expression level in the core promoter (*12*, *13*). Conveniently, these sites are identical between $P_{LtetO-1}$ and $P_{LlacO-1}$, coming from phage lambda P_L in both cases (*11*). For each promoter, we constructed all possible single-bp substitutions over this 12bp region (36 mutants for each promoter), along with the wildtype sequence. All 37 sequence variants of the two promoters were combined together, resulting in a plasmid library containing: all 1,296 double-promoter mutants, all 36 single-promoter mutants for each promoter (one promoter is mutated, the other is wildtype) and the full wildtype (both promoters are wildtype). The majority of mutations in the RNA polymerase-binding sites are expected to have little or no effect on repressor binding, and their relative effect on expression should be similar across different inducer concentrations (*14*, *15*). However, one of the -10 bases on $P_{LtetO-1}$ overlaps with a *tet* operator, and three of the -35 bases on $P_{LlacO-1}$ are expected to overlap with a *lac* operator (*11*) (Fig. S1), meaning that the effect on expression of mutations at these positions could depend strongly on inducer concentration (*16*).

The overall structure of the plasmid on which the library is based is shown in Fig. S1. *araA* and *araB* are divergently expressed from $P_{LtetO-1}$ and $P_{LlacO-1}$ promoters, respectively. These two promoters are separated from each other by a short *bsd* blasticidin S resistance cassette (*17*), in order to reduce any physical interactions between them. The presence of a resistance cassette between the promoters also considerably increased cloning efficiency, as explained below, and *bsd* in particular was chosen for its small size (396 bp ORF), making it possible to sequence both promoters on a single amplicon using paired-end Illumina technology (Fig. S1). The promoters' repressors, *tetR* and *lacI*, were included on the plasmid.

Plasmid molecules were also intergenically tagged with unique DNA barcodes, similarly to reference (*18*) (Fig. S1). These were used to help overcome the problem of PCR and sequencing errors and to increase the precision of mutant fitness estimates by providing many

independent frequency trajectories for each mutant (Figs. S2-4). The barcodes thus also allowed us to account for anomalous lineages containing off-target mutations (present in the initial library) and *de novo* mutations (arising during competition assays). They consist of 20 random nucleotides, split into 4 blocks of 5 (*19*) to avoid the creation of restriction sites used in a later sequencing step: $N_5ATN_5ATN_5ATN_5$. Barcodes were inserted downstream of the *lacI-tetR* cassette, far from the $P_{LtetO-1}$ and $P_{LlacO-1}$ promoters, to avoid any effects on *araA* and *araB* expression, and so are expected to be effectively neutral for fitness (Fig. S1). Care was taken throughout to avoid loss of library complexity (Fig. S2), and quality controls were employed at each step of library construction.

The pooled plasmid library was constructed using standard restriction-ligation cloning (Fig. S1). Due to their short length, promoter sequences could be introduced facing outwards on the 5' ends of PCR primers that were used to amplify a *bsd* (*17*) blasticidin S-resistance cassette from plasmid pKH1511d ($P_{LtetO-1}$ on forward primers and $P_{LlacO-1}$ on reverse primers). This was done using primer pools with randomised nucleotides at each of the 12 target positions for each promoter. The primers also contained restriction sites on their 5' extremities, allowing the resulting amplicon pool to be ligated into the library backbone, pKH11511c, in the desired orientation. The resulting plasmid library was transformed into DH5α Δ*araBA* and colonies were selected on blasticidin S. This strategy ensured that the occurrence of false-positive colonies from undigested or self-ligated vector was negligible, as a functional *ori* could only come from pKH11511c (the pKH1511d *ori* is *pir*-dependent), while *bsd* was only present in pKH1511d. Due to the use of fully-randomised nucleotides at each target position and the combinatorial way in which variants of the two promoters were cloned together, the expected genotype frequencies in this initial library are: 1/16 for WT, 1/192 for each of the 72 single-promoter mutants and 1/2304 for each of the 1,296 double-promoter mutants. With this in mind, an estimated 40,000 colonies were harvested in this step to avoid loss of library complexity. Barcodes were added in a 2nd round of restriction-ligation cloning, introduced *via* a randomised PCR primer. The primer, containing fully-randomised nucleotides at 20 positions, was used to amplify the *bla* β-lactamase gene from plasmid pKD3 (*8*), and the resulting amplicon pool was swapped with the *aadA1* streptomycin/spectinomycin resistance gene in the plasmid library backbone. The primer contains restriction sites on its 5' extremity, one of which is used for this ligation, and another of which allows the barcodes to be moved closer to the mutated promoter region in a later step (see *Barcode-promoter association*). The barcoded plasmid library was again transformed into DH5α Δ*araBA* and colonies were this time selected on ampicillin. False-positive colonies were avoided for the same reason as above, as pKD3 also has a pir-dependent *ori*. An estimated 100,000 colonies were harvested during this step, with the vast majority expected to contain a unique barcode. Expected barcode richness was thus: 6,250 for WT, 521 for each single-promoter mutant and 43 for each double-promoter mutant. In a final step, the engineered host strain, MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo}, was transformed with this barcoded plasmid library, and an estimated 600,000 colonies were harvested after selection on ampicillin. Detailed library construction methods follow.

Library creation methods

To create the initial library, two promoter-containing primer sets, oPtetLib-fwd and oPlacLib-rev, were each pooled in equimolar quantity (Table S5). These two primer pools were then used together at a concentration of 0.5 µM each pool to PCR-amplify *bsd* from plasmid pKH1511d, using Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific) in its

High-Fidelity buffer, following the manufacturer's recommendations. Cycling conditions were: 98°C for 30 secs, followed by 35 cycles of 98°C for 10 secs, 60°C for 30 secs and 72°C for 15 secs, with a final extension step of 72°C for 2 mins. PCR product quality was checked by agarose gel electrophoresis, after which the product was column-purified (QIAquick PCR Purification Kit, QIAGEN) and quantified with a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). The purified product and plasmid pKH1511c were then both digested for 90 mins with XhoI and SacI-HF restriction enzymes (NEB CutSmart buffer), and digested DNA was again column-purified (QIAquick PCR Purification Kit, QIAGEN) and quantified with a NanoDrop ND-1000 spectrophotometer. 70ng of the pKH1511c vector fragment was ligated in a 1:3 molar ratio with the *bsd/*promoter-containing insert in a total volume of 20 µl. The ligation was carried out at 16°C overnight using T4 DNA ligase (NEB T4 DNA ligase reaction buffer), which was then deactivated by heating at 65°C for 10 mins. The ligate was microdialysed against water for 30 mins (MF-Millipore, Merck), after which several transformations were performed as follows: 3 µl were electroporated into 50 µl electrocompetent DH5α Δ*araBA* cells; cells were recovered in 500 µl low-salt (Miller) LB for 1 hour at 37°C with shaking at 200 rpm, plated on LB-agar with 100 µg/ml blasticidin-S and incubated overnight at 37°C. Colony-PCR and Sanger sequencing (GATC) of the mutated promoter region was performed on 4 of the resulting colonies as a preliminary test of library quality, and all 4 clones had a unique promoter genotype with a single base substitution in the target region of either one or both promoters, as expected. An estimated 40,000 colonies were scraped off the agar into LB-glycerol (40%), and plasmid DNA was purified from a sample of this cell suspension (QIAprep Spin Miniprep Kit, Qiagen) after thorough mixing.

To barcode the plasmid library, primers oBarcodeBla-fwd and oBarcodeBla-rev (Table S3) were used at a concentration 0.5 µM each to PCR-amplify *bla* from plasmid pKD3 (*8*), using Phusion Hot Start II High-Fidelity DNA Polymerase (Thermo Scientific) in its High-Fidelity buffer, following the manufacturer's recommendations. Cycling conditions were: 98°C for 30 secs, followed by 30 cycles of 98°C for 10 secs, 60°C for 30 secs and 72°C for 25 secs, with a final extension step of 72°C for 3 mins. PCR product quality was checked by agarose gel electrophoresis, after which the product was column-purified (QIAquick PCR Purification Kit, QIAGEN) and quantified with a NanoDrop ND-1000 spectrophotometer (Thermo Scientific). The purified product was then digested for 1 hour with SpeI-HF restriction enzyme (NEB CutSmart buffer), while the purified plasmid library obtained above was digested for 1 hour with BstZ17I and SpeI-HF restriction enzymes (NEB CutSmart buffer). Digested DNA was again column-purified (QIAquick PCR Purification Kit, QIAGEN) and quantified with a NanoDrop ND-1000 spectrophotometer. 60 ng of the digested library was ligated in a 1:4 molar ratio with the *bla*/barcode-containing insert in a total volume of 20 µl. The ligation was carried out at 16°C overnight using T4 DNA ligase (NEB T4 DNA ligase reaction buffer), which was then deactivated by heating at 65°C for 10 mins. The ligate was microdialysed against water for 30 mins (MF-Millipore, Merck), after which several transformations were performed as follows: 1 µl was electroporated into 15µl commercially-prepared ElectroMAX DH5α-E electrocompetent cells (Invitrogen); cells were recovered in 500 µl LB for 30 mins (to minimise cell replication) at 37°C with shaking at 200rpm, plated on LB-agar with 100 µg/ml ampicillin and incubated overnight at 37°C. The use of commercially prepared electrocompetent cells was necessary due to reduced cloning efficiency at this step, possibly due to the ligation reaction involving blunt ends. Plasmid DNA was purified from 3 colonies (QIAquick PCR Purification Kit, QIAGEN) for Sanger sequencing (GATC) of the mutated promoter and barcode regions as a preliminary

test of barcoding efficiency. All 3 colonies were found to possess a unique promoter genotype, as before, along with a unique, correctly-inserted barcode. An estimated 100,000 colonies were scraped off the agar into LB-glycerol (40%), and plasmid DNA was purified from a sample of this cell suspension (QIAprep Spin Miniprep Kit, Qiagen) after thorough mixing.

To move the barcoded plasmid library into the final host strain, while avoiding the creation of transformants harbouring multiple unique plasmids (*20*), several transformations were performed as follows, with plasmid concentration kept fairly low: 5 ng of the purified barcoded plasmid library obtained above were electroporated into 50 µl electrocompetent MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo} cells; cells were recovered in 500 µl LB for 30 mins at 37°C with shaking at 200rpm, plated on LB-agar with 100 µg/ml ampicillin and incubated overnight at 37°C. An estimated 600,000 colonies were scraped off the agar into LB-glycerol (40%), and this cell suspension was aliquoted and stored at -80°C after thorough mixing.

Barcode-promoter association

To reveal the $P_{LtetO-1}$ and $P_{LlacO-1}$ promoter sequences linked to each barcode sequence, barcodes were first brought closer to the promoters by excision of the intervening region from the plasmid followed by re-circularisation (*18*). PCR-amplification was then used to add the technical sequences necessary for paired-end Illumina MiSeq sequencing of barcode-promoter amplicons (Fig. S1B).

To first move barcodes closer to the promoter region, the purified barcoded plasmid library was digested for 90 mins with XhoI, SalI-HF and SphI restriction enzymes (NEB CutSmart buffer). The largest fragment (~5.5 kb), which contains the mutated promoters and the barcode, was gel-purified (QIAquick Gel Extraction Kit, Qiagen) using a 1% agarose gel and quantified with a NanoDrop ND-1000 spectrophotometer before being self-ligated. XhoI and SalI are isocaudamers, so they create complementary cohesive ends, but the sequence resulting from ligation between these ends is no longer recognised by either enzyme (SphI cuts within the region being discarded, and was simply included to ease gel extraction of the desired fragment). Because of this, they can be included in the reaction mix during self-ligation of the purified fragment to help reduce intermolecular ligation (undesired intermolecular ligation events which recreate XhoI and SalI sites can be reversed, releasing the original monomers and so increasing the efficiency of the desired intramolecular ligation reaction (*18*, *21*)). Due to the inclusion of these restriction enzymes, the self-ligation reaction was carried out in a restriction enzyme buffer, with ATP added for ligase activity. Additionally, the concentration of DNA and ligase was substantially reduced compared to standard ligation reactions to further reduce the occurrence of intermolecular ligation. The self-ligation reaction mix thus consisted of: 1X NEB restriction buffer 2 supplemented with 100 µg/ml BSA and 1 mM ribo-ATP (NEB), 30 ng DNA, 1 U each of XhoI and SalI-HF and 800 U of T4 DNA ligase, in a total volume of 200 µl. Inspired by the strategy of reference (*21*), the reaction was cycled 50 times between 37°C (restriction enzyme and ligase activity optimum) for 5 mins and 16°C (promote annealing of DNA termini) for 15 mins. A final 37°C incubation was carried out for 15 mins to promote digestion of any remaining XhoI and SalI sites, followed by one of 65°C for 20 minutes to inactivate all enzymes. The ligate was concentrated to ~20 µl using a SpeedVac concentrator (Savant DNA 120, Thermo Scientific) and then microdialysed against water for 90 mins (MF-Millipore, Merck). As a preliminary test of the success of this ligation step, a portion of the ligate was used in a transformation to allow isolation and sequencing of several re-circularised plasmids: 2 µl were

electroporated into 50 µl electrocompetent DH5α Δ*araBA* cells; cells were recovered in 500 µl LB for 30 mins at 37°C with shaking at 200 rpm, plated on LB-agar with 100 µg/ml ampicillin and incubated overnight at 37°C; plasmid DNA was purified from 6 colonies (QIAquick PCR Purification Kit, QIAGEN) for Sanger sequencing (GATC) of the ligated region containing the mutated promoters and barcode. All 6 clones were found to possess the expected linking sequence between promoters and barcode, and all plasmids were inferred to be monomeric due to the high Phred scores of the chromatograms (suggesting the presence of a single unique barcode on each re-circularised plasmid).

With the re-circularised DNA placing barcodes in proximity to their respective mutated promoters, this region was then PCR-amplified in a 40 µl reaction using 25 ng of the ligated DNA as template and 0.6 µM each of primers oLinkBarcode-fwd and oLinkBarcode-rev (Table S3). These primers contain adaptors for a 2nd PCR at their 5' extremities, followed by fully randomised hexamers added to increase amplicon diversity to facilitate MiSeq flow-cell clustering. KAPA HiFi HotStart ReadyMixPCR Kit (Kapa Biosystems) was used for amplification, under the following cycling conditions (cycle number was kept low to reduce PCR errors and artefacts): 95°C for 3 mins, followed by 15 cycles of 98°C for 20 secs, 60°C for 30 secs and 68°C for 30 secs, with a final extension step of 68°C for 2 mins. The amplicon (~0.9 kb) was gel-purified (QIAquick Gel Extraction Kit, Qiagen) using a 1.5% agarose gel and quantified fluorometrically (dsDNA HS Assay Kit with a QuBit 2.0, Thermo Scientific). A 2nd 40 µl PCR was then performed using 5 ng of this amplicon as template and 0.6 µM each of a P5 and P7 Nextera Index Kit primer (Illumina) to add Illumina adaptors and multiplexing indexes. KAPA HiFi HotStart ReadyMixPCR Kit (Kapa Biosystems) was again used for amplification, under the following cycling conditions (cycle number was again kept low): 95°C for 30 secs, followed by 12 cycles of 95°C for 10 secs, 55°C for 30 secs and 68°C for 30 secs, with a final extension step of 68°C for 5 mins. The amplicon library (~1 kb) was gel-purified (QIAquick Gel Extraction Kit, Qiagen) using a 1.5% agarose gel and a 20,000X dilution was quantified by qPCR using KAPA Library Quantification Kit for Illumina (Kapa Biosystems) on a LightCycler 480 (Roche), following the manufacturer's recommendations.

The resulting amplicon library is composed of DNA fragments of the structure: P5 - i5 - N_6 PCR tag - $P_{LtetO-1}$ (rev) - *bsd* (rev) - $P_{LlacO-1}$ - N_{20} plasmid barcode - N_6 PCR tag - i7 - P7, which are ~1 kb in size (close to the size-limit for reliable MiSeq sequencing). 300nt paired-end MiSeq sequencing allowed us to sequence the entire $P_{LtetO-1}$ promoter on Read 1 and the plasmid barcode and entire $P_{LlacO-1}$ promoter on Read 2 (note that Reads 1 and 2 do not overlap). For this, a 600-cycle MiSeq Reagent Kit v3 (Illumina) was used, and DNA was loaded at a concentration of 12pM, with a 20% PhiX DNA spike-in (PhiX Control v3, Illumina). Preliminary quality filtering and demultiplexing by the standard MiSeq software package (Illumina) resulted in an output of > 22M read pairs, giving an expected coverage of > 220X for each plasmid barcode.

MiSeq reads were processed using the Mothur (*22*) (version 1.37.6) software package *via* the following steps: reads were quality-filtered by size (>199 bases), number of uncalled bases (<3 Ns) and length of the longest homopolymer stretch, another indicator of overall read quality (<9 bases). Entire $P_{LtetO-1}$ sequences were extracted from Read 1, and barcode sequences and entire $P_{LlacO-1}$ from Read 2, by Needleman alignment to reference sequences (default alignment parameters). Reads for which either the $P_{LtetO-1}$, $P_{LlacO-1}$ or barcode region contained insertions or did not generate a full alignment with the reference were discarded. The Mothur Precluster algorithm was then used to cluster barcode sequences differing by a Hamming distance of 1, with the aim of correcting for PCR and sequencing errors (the potential barcode diversity is so high (>

$1x10^{12}$) that the presence of immediately neighbouring sequences is very likely due to these errors (Fig. S2C)). The algorithm uses sequence abundance to decide the "true" (majority) sequence for each cluster, and to decide where a sequence clusters if it has >1 immediate neighbour. After de-gapping and re-grouping barcode sequences to account for any alignment ambiguities resulting from small deletions, barcode clusters were used to build a dictionary assigning each "true" barcode sequence to a $P_{LtetO-1}$ and $P_{LlacO-1}$ sequence. Due to a high rate of PCR-derived recombination (*23*) being observed (caused by the extensive homology between all fragments, and resulting in some molecules displaying incorrect barcode-promoter associations), a haplotype-based strategy was used for this step rather than one in which each nucleotide is considered independently as in reference (*18*). This is because the small number of mutations expected to be present in each mutant (0-2) means that, at any particular position, the majority of molecules will possess the WT base. If the consensus $P_{LtetO-1}$ and $P_{LlacO-1}$ sequences attached to a particular barcode are computed by considering each nucleotide independently, a high recombination rate can thus result in mutant bases being assigned as the WT base. The haplotype-based strategy, executed in Python (v3.5), consists of the following steps: for each barcode cluster (consisting of reads whose barcode sequences are identical to or the immediate neighbour of the inferred "true" barcode sequence), the associated complete $P_{LtetO-1}$-$P_{LlacO-1}$ concatenate sequences were grouped; the number of occurrences of each of these 108-nt $P_{LtetO-1}$-$P_{LlacO-1}$ sequences was tabulated; if the cluster contained more than 2 read pairs in total, the most abundant concatenate $P_{LtetO-1}$-$P_{LlacO-1}$ sequence is $\geq$ 5x more abundant than the second-most abundant one, and the most abundant concatenate $P_{LtetO-1}$-$P_{LlacO-1}$ sequence contains no Ns (uncalled bases), then this $P_{LtetO-1}$-$P_{LlacO-1}$ sequence is assigned to the "true" barcode sequence for that cluster (else the cluster is discarded). This stringent requirement is aimed at reducing barcode-promoter mis-assignments caused by PCR and sequencing errors, PCR-derived recombination or intermolecular ligation during the first step of barcode-promoter association, as well as to avoid any barcodes that may be linked to multiple promoter genotypes. Only barcodes associated to promoter genotypes for which the entire promoter regions contain no unexpected mutations were considered for further analysis.

Mutant library competition assays

The final mutant library (host strain transformed with barcoded plasmid library) was competed over ~30 mean generations (~3 days) in the presence of L-arabinose and different concentrations of the inducers, aTc and IPTG. Cell density was kept low during competition (OD_{600} < 0.2) by serial transfer into fresh medium, in order to maintain the culture in exponential phase and to avoid large changes in medium composition. Large volumes of media (100 ml) were used to avoid severe population bottlenecks during serial transfer (> 1 x 10^8 cells each transfer). Plasmid DNA was purified from the culture at several time-points for HiSeq sequencing of plasmid barcodes. Plasmid barcode abundance serves as a proxy for the abundance of cells carrying that particular barcode. The change in frequency over time of a barcode thus provides an estimate of competitive fitness for the lineage carrying that barcode (*24*). Since we know the $P_{LtetO-1}$-$P_{LlacO-1}$ sequence associated to each barcode (see *Barcode-promoter association*), this in turn provides us with a distribution of fitness estimates for every mutant.

The base competition medium consisted of M9 + 0.1% casamino acids (for basal growth) + 0.03% L-arabinose, with 100 µg/ml ampicillin to select against plasmid loss. A preliminary competition experiment was performed under inducer concentrations of 20 ng/ml aTc and 30 µM IPTG, expected to endow the wildtype with near-maximal fitness (although this was found to be

inaccurate). A second round of competition experiments was carried out at a later date and was comprised of three different inducer concentration combinations. One duplicated those of the initial experiment to check reproducibility (Figs. S3-4), and the other two were: 5 ng/ml aTc and no IPTG, and 200 ng/ml aTc and no IPTG. No IPTG was chosen to reduce *araB* expression as much as possible, as the preliminary experiments suggested that the wildtype over-expressed *araB* even in the absence of inducer (*25*), due to promoter leakiness. The range of aTc was chosen to explore the full range of achievable *araA* expression.

In detail, a sample of the frozen library cell stock was thawed and diluted in 200 ml of M9 + 0.5% casamino acids (with 100 µg/ml ampicillin), in a 500 ml flask, for a final blank-subtracted OD_{600} of 0.12 (200 µl read by Varioskan microplate reader, Thermo Scientific). This common starting-culture was recovered for ~3.5 hours at 37°C with shaking at 200 rpm, reaching an OD_{600} of 0.3, before being washed with 200 ml of M9 + 0.1% casamino acids. Washed cell pellets (each coming from 50 ml of the original culture) were resuspended directly in 100 ml of the different competition media, for an effective 2X dilution of the original culture (OD_{600} of ~0.15; flasks of competition media were always pre-warmed at 37°C to keep temperature constant and detect any contamination, with aTc, IPTG and ampicillin being added at the time of transfer to avoid degradation). These cultures were then acclimatised to their respective competition media for ~2.25 hours (37°C, 200 rpm), reaching an OD_{600} of 0.23-0.28, to allow time for stable induction by aTc, IPTG and L-arabinose. These acclimatised cultures were taken as t_0, and so plasmid DNA was purified from a 50 ml sample of each culture (QIAprep Spin Miniprep Kit, Qiagen) and quantified fluorometrically (dsDNA HS Assay Kit with a QuBit 2.0, Thermo Scientific) for eventual HiSeq sequencing of plasmid barcodes (the rest remaining after this and transfer was pelleted, resuspended in LB-40% glycerol and stored at -80°C as an archive). 3.2 ml of each culture was transferred to 100 ml fresh competition media (~32X dilution) and left to grow (37°C, 200 rpm) to an OD_{600} of ~0.12 (3-4 mean generations). DNA was purified from a 50 ml sample of each culture (t_1), as before, and 3.2 ml of each culture was again transferred to 100 ml of fresh competition media and left to grow to an OD_{600} of ~0.12 (~5 mean generations). This procedure was repeated until t_6 (or t_8 in an initial experiment), for a total of ~29 mean generations of competition (or ~39), over which time the impact of *de novo* mutation appears low (Fig. S3). The precise number of mean generations between each sampling was calculated from OD_{600} values and used for estimating fitness.

Barcode-sequencing of competed mutant library

To track plasmid barcode frequencies throughout the competition experiments, barcodes were PCR-amplified from plasmid DNA in 2 steps, as for *Barcode-promoter association*, to add technical sequences necessary for 100nt overlapping paired-end Illumina HiSeq sequencing. This was performed for time-points t_0, t_1, t_2, t_4, t_6 and t_8 (approximately 0, 4, 9, 19, 29 and 39 mean generations) for the preliminary experiment, and t_1, t_2, t_4 and t_6 for the later experiments. These time-points were chosen with the aim of obtaining precise fitness estimates for both large-effect and small-effect mutations[49].

In detail, at each selected time-point, 20 ng of purified plasmid DNA was PCR-amplified in a 40 µl reaction using 0.6 µM each of primers oBarcodeSeq-fwd and oBarcodeSeq-rev (Table S3). These primers contain adaptors for a 2nd PCR at their 5' extremities, followed by fully randomised hexamers to increase amplicon diversity, as in *Barcode-promoter Association.* In this case, the randomized hexamers were also used to detect PCR duplicates arising from the 2nd PCR[45]. KAPA HiFi HotStart ReadyMixPCR Kit (Kapa Biosystems) was used for amplification,

under the following cycling conditions (cycle number was kept low to reduce PCR errors and artefacts): 95°C for 3 mins, followed by 12 cycles of 98°C for 20 secs, 60°C for 30 secs and 68°C for 30 secs, with a final extension step of 68°C for 2 mins. Amplicons (~200 bp) were gel-purified (QIAquick Gel Extraction Kit, Qiagen) using a 2% agarose gel and quantified fluorometrically (dsDNA HS Assay Kit with a QuBit 2.0, Thermo Scientific). A 2nd 40 µl PCR was then performed using 5-8 ng of each amplicon as template and 0.6 µM each of a P5 and P7 Nextera Index Kit primer (Illumina) to add Illumina adaptors and multiplexing indexes. KAPA HiFi HotStart ReadyMixPCR Kit (Kapa Biosystems) was again used for amplification, under the following cycling conditions: 95°C for 3 mins, followed by 13 cycles of 98°C for 20 secs, 55°C for 30 secs and 68°C for 30 secs, with a final extension step of 68°C for 5 mins. These ~300 bp amplicons, of the structure, P5 - i5 - N_6 PCR tag - N_{20} plasmid barcode - N_6 PCR tag - i7 - P7, were gel-purified (QIAquick Gel Extraction Kit, Qiagen) using a 2% agarose gel and sent to IntegraGen (Evry, France) for qPCR-based quantification, equimolar pooling and 100nt paired-end HiSeq-4000 sequencing (Illumina). Preliminary quality filtering and demultiplexing (Integragen, Evry, France) resulted in ~18 M read pairs per time-point per competition experiment, giving, for each point, an expected barcode coverage of ~200X and an expected mutant coverage of >14,000X.

HiSeq sequencing reads were processed using the Mothur (*22*) (version 1.37.6) software package by the following steps: Forward and reverse reads were joined into contigs using Mothur's make.contigs command with the default parameters. Contigs were then quality-filtered by size (<131bp, as longer contigs imply forward and reverse reads could not be properly overlapped), number of uncalled bases (no Ns) and length of longest homopolymer stretch, an indicator of overall read quality (<9 bases). To remove the majority of PCR duplicates arising from the 2nd PCR (made possible by randomised hexamers introduced on each side of the barcode during the 1st PCR (*19*)), if a particular full contig was present more than once, only one copy was kept. Barcode sequences were then extracted after aligning contigs to the reference sequence (Needleman global alignment). Reads containing insertions or not generating a full alignment with the reference were discarded. Next, the Mothur precluster algorithm was used to cluster barcode sequences differing by a Hamming distance of 1, with the aim of correcting for PCR and sequencing errors, as described in *Barcode-promoter association.* After de-gapping and re-clustering barcode sequences to account for any alignment ambiguities resulting from small deletions, the number of occurrences of each "true" barcode was tabulated across all time-points for each competition experiment. Finally, a custom R (v.3.4.3) script was used to merge these barcode counts tables with the barcode-promoter mutant dictionary generated in *Barcode-promoter association*.

Estimation of competitive fitness and epistasis

We found that competitive fitness was not constant over the course of competition, with, for example, a possible period of physiological adaptation between t_0 and t_2 for certain inducer environments (Fig. S3). By t_6, a substantial number of lower-fitness mutants begin to escape detection completely, and so to avoid any bias in fitness estimates we consider only the frequency changes between t_2 and t_4 (two time-points). We begin by removing outlier barcodes associated to the wildtype genotype, to avoid any systematic biases coming from inaccurate wildtype estimates. This was done by computing the log ratio of t_4 to t_2 counts for all wildtype barcodes and removing those giving values > 1.5x the inter-quartile range above (below) the

upper (lower) quartile. We also removed all barcodes giving < 8 reads at t_2 from our dataset. For every remaining mutant barcode, *i*, we then estimate its log fitness relative to the wildtype as:

$$F_i^{rel} = \frac{\ln\left(\frac{f_i^{t_4}}{\sum f_{wt}^{t_4}}\right) - \ln\left(\frac{f_i^{t_2}}{\sum f_{wt}^{t_2}}\right)}{t_4 - t_2},$$

where f_i is the frequency of a mutant barcode, $\sum f_{wt}$ is the total frequency of all wildtype barcodes and $t_4 - t_2$ is the number of mean generations between the two time-points considered (~9). We now estimate log relative fitness of a mutant *g*, F_g^{rel}, as the median of that of its associated barcodes, $F_{g_i}^{rel}$. We use the median barcode fitness as a fitness estimate for each promoter genotype as a convenient way to filter out the many sources of error in a competition experiment (especially undetected mutations introduced during library construction, *de novo* mutations arising during competition and barcode-promoter misassignments due to PCR and sequencing errors) (Fig. S3A). The number of eligible barcodes for each promoter genotype ranges from a few to thousands (exact numbers are provided in Data S1). Some barcodes disappear from our sequencing sample by t_4, and so are given an F_i^{rel} of *–Inf*. In Env_1 and Env_3, for a very few genotypes this is the case for the majority of their barcodes, and we identify these mutants as being less fit than the wildtype but cannot estimate total/marginal fitness effects or epistasis for them.

To estimate the precision of mutant fitness estimates, we used standard bootstrapping of the eligible mutant and wildtype barcodes (n=1,000), each time computing the mutants' fitness, F_g^{rel}, as the median fitness of their associated barcodes, $F_{g_i}^{rel}$. The same 1,000 sets of randomly sampled wildtype barcodes were used as the references for all mutants. The bootstrap distributions were then used to determine significance (empirical 95% confidence) for non-neutrality of total (F_g^{rel}) and marginal (F_g^{marg}) fitness effects, non-zero epistasis, simple sign epistasis and reciprocal sign epistasis, pairing bootstrap F_g^{rel} estimates by sampled wildtype barcode set when necessary.

The marginal fitness change induced by adding mutation *A* to the genetic background *B* is defined as $F_{A|B}^{marg} = F_{AB}^{rel} - F_B^{rel}$, and epistasis between mutations *A* and *B* is defined as $\varepsilon_{AB} = F_{AB}^{rel} - (F_A^{rel} + F_B^{rel})$ (*26*).

Phenotype-fitness model

We consider a linear metabolic reaction path,

$$S \underset{A}{\rightarrow} X \underset{B}{\rightarrow} Y \underset{C}{\rightarrow} Z \rightarrow \cdots$$

where S is the substrate (L-arabinose) concentration. As shown in references (*27*, *28*), for S and Z fixed, the steady-state flux for non-saturated enzymes and the intermediate concentration are respectively given by:

$$\varphi = \frac{1}{1/A + 1/B + \eta} \quad (1)$$

$$Y = D - \frac{\varphi}{1/A + 1/B} \quad (2)$$

where A and B are proportional to the maximum reaction rates provided by each enzyme, η is the inverse of the maximal flux, φ_{max}, as imposed by the fixed pathway steps, and D is a certain function of S and equilibrium constants (see reference (*28*) for detailed expressions). We note that the flux, φ, is an increasing function of A and B and saturates at φ_{max} for very efficient enzymes A and B, or very high concentrations of them. However, at high fluxes, the hypothesis of unsaturated downstream enzymes breaks down, and a reaction step becomes limiting, such that the concentrations of metabolic intermediates may build up to toxic levels.

To account for such saturation, we extend the model above by considering the full Reversible Michealis-Menten (RMM) form for the step C instead of its first order approximation (similar reasoning applies for longer paths). At steady-state, all reaction speeds must be equal, giving for the third step:

$$\varphi = \frac{\alpha Y - \beta Z}{1 + \gamma Y + \delta Z} \quad (3)$$

where $\alpha, \beta, \gamma, \delta$ are the RMM parameters for C. Equivalently, expressing Y as a function of φ:

$$Y = \frac{\beta Z + (1 + \delta Z)\varphi}{\alpha - \gamma\varphi} \quad (4).$$

We could eliminate Y by combining (2) and (4), Z being fixed, and obtain an exact expression for φ. Note that expression (1) is recovered for $\delta = \gamma = 0$, as this corresponds to the unsaturated case. In the general case, φ would still be an increasing function of A and B and saturate at a certain value, but its expression becomes more complicated.

Instead of using the full expression of φ, we report here an approximation with less parameters, which consistently recovers the monotonicity with A and B, and the limit regimes for unsaturated and saturated downstream steps. For this, we simply keep expression (1) for the flux, and set its saturation by the saturation of the reaction catalysed by C, as obtained in the limit of very high Y in (4):

$$\varphi_{max} = 1/\eta = \alpha/\gamma \, .$$

With this, expression (4) becomes:

$$Y = \frac{P + Q\varphi}{\varphi_{max} - \varphi} \quad (5)$$

where P and Q are functions of the fixed downstream enzyme properties and concentrations. We note in particular that Y diverges when the flux becomes maximal, meaning that the downstream reaction is saturated, leading to an accumulation of Y.

We now assume fitness to be a function of flux and the toxic intermediate (L-ribulose-5-phosphate) concentration, Y, and that there exist constants e and f such that, from (1) and (5):

$$F = e\varphi - fY = e\varphi - f\frac{P + Q\varphi}{\varphi_{max} - \varphi} \quad (6).$$

This expression can be further simplified by considering the low and high flux regimes:
For $\varphi << \varphi_{max}$, (6) behaves as $F = -fP/\varphi_{max} + u\varphi$, with $u = e - f(Q + P/\varphi_{max})/\varphi_{max}$, the offset $-fP/\varphi_{max}$ being determined solely by properties of the fixed downstream enzyme, C. Thus, any fitness change due to mutations in A and B is of the form $u\varphi$.

For $\varphi \sim \varphi_{max}$, the first term of (6) remains finite while the second with numerator $v = f(P + Q\varphi_{max})$ diverges. Thus, replacing e by u as defined in the regime $\varphi << \varphi_{max}$ has a negligible contribution.

Introducing a basal growth rate, ω, supplied by alternative nutrients in the medium (casamino acids), fitness is then well approximated by:

$$F = \omega + u\varphi - \frac{v}{\varphi_{max} - \varphi} \quad (7).$$

In addition to flux and toxic metabolite concentration, gene expression burden can also contribute to fitness changes (*25*, *29–31*). Following the observation that protein expression burden depends on metabolic state (*32*, *33*), we include an expression cost factor in which θ_A and θ_B describe the cost of increasing cellular enzyme activity, including potential contributions from both the amount of expression and the specific enzyme activity constants:

$$F = \left(\omega + u\varphi - \frac{v}{1/\eta - \varphi}\right)(1 - \theta_A A - \theta_B B) \quad (8).$$

This expression is considered valid only when both factors are positive. Expressions (1) and (8) together define a fitness surface in the two-dimensional space of AraA and AraB activities, described by the 6 independent parameters, ω, u, v, θ_A, θ_B and η.
The entire model consists of 83 parameters: the 6 detailed immediately above; 5 defining the "wildtype" activity levels (AraA and AraB activities for the 3 inducer environments, with Env_2 and Env_3 having the same wildtype AraB activity, as both contained the same IPTG concentration); and 72 defining the relative impact of the single mutations (36 for each gene) on enzyme expression/activity. For a given parameter set, the fitness, F^{rel}, of the 72 single mutants and 1,296 double mutants was computed in each of the 3 environments, relative to the respective "wildtype" fitness. The 83-parameter model was fitted on 4,079 data points, corresponding to the computable set of relative fitnesses (Fig. 2A) of the 1,368 mutants measured in 3 different environments.

The model was fit using multiple Monte Carlo Markov Chains (MCMC) (*34*). Parameters were generated randomly from uniform distributions, both initially and at each step of the chain for a randomly chosen parameter (bounds are provided in Fig. S8A and Data S2; bounds for expression effects of inducer concentrations and a few mutations were guided by experimental expression measurements (data not shown)). 800 chains, each of 300,000 steps, were simulated, and for each chain the parameter set giving the best fit with measured fitness values was stored (residuals were weighted to give equal consideration overall to single and double mutants, and were also normalised to the mean fitness effect in the environment from which they came). The distribution of goodness-of-fit values from the 800 chains was multi-modal (*ie*. convergence was

not guaranteed), with ~5-10% of the chains achieving a best fit residing in the lowest peak. We take the best of all these parameter sets as the most likely fit, but the distributions of parameter values from the best 2.5% of chains are also provided in Fig. S8A and Data S2.

Several fitness function variations containing less parameters than the one presented in the main text were fit in the same way, and we conclude that flux, toxicity and gene expression burden must all be accounted for to explain the observed fitness and epistasis values (Fig. S9).

Statistical analyses

All statistical analyses were performed in R (v.3.4.3) and figures were made using the R packages ggplot2 and rgl (for the 3D plot). Lower and upper hinges of box plots correspond to the first and third quartiles. Centre line is the median. Upper and lower whiskers extend from the hinges to the largest and lowest value no further than 1.5× the inter-quartile range away, respectively. Points outside this range are plotted individually.

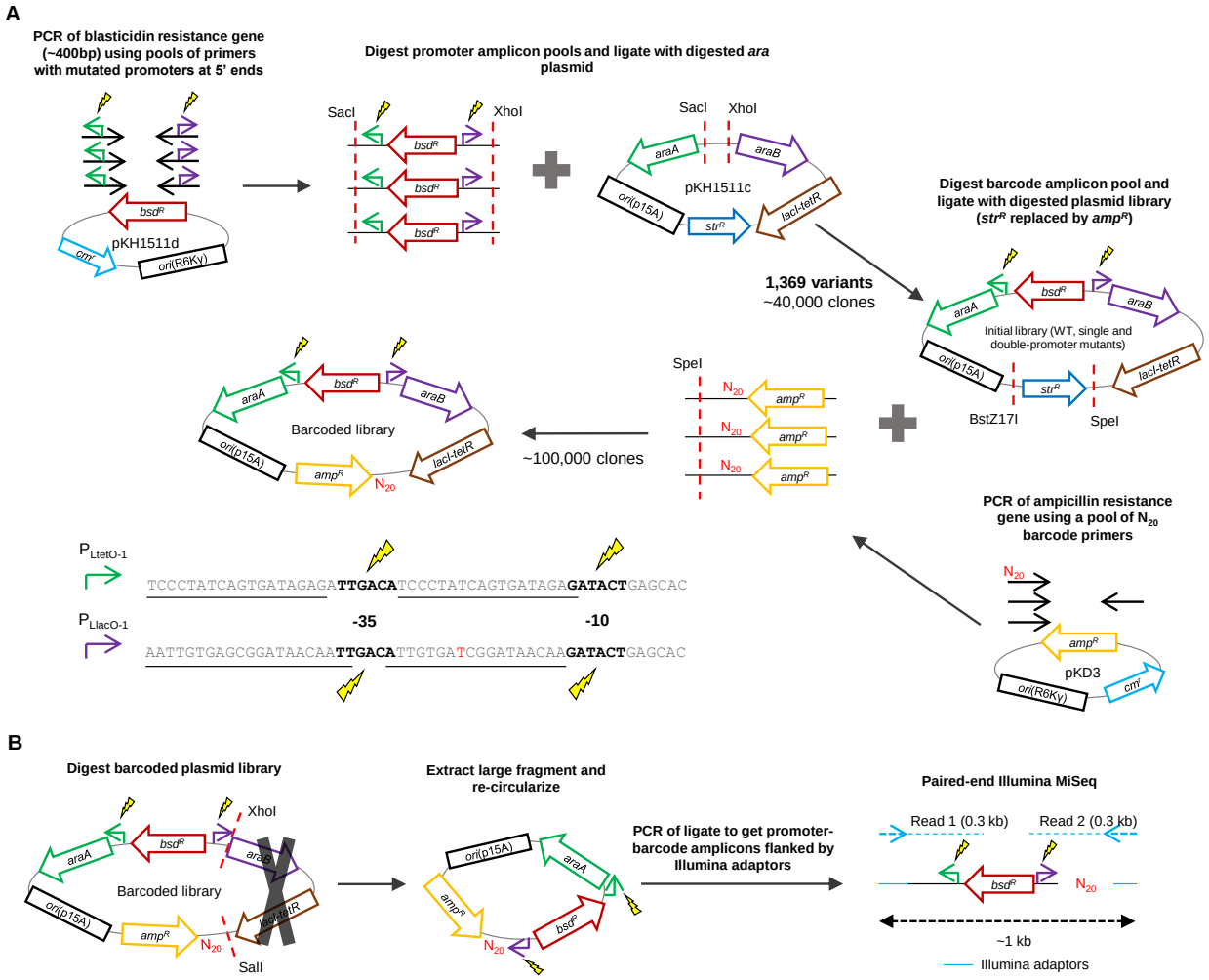

Fig. S1.

Construction and characterisation of barcoded promoter-mutant plasmid library. (**A**) A blasticidin-resistance cassette (*bsd*^{R}) was amplified from pkH1511d using pools of primers carrying variants of the entire $P_{LtetO-1}$ (green arrow) and $P_{LlacO-1}$ (purple arrow) promoters at their 5' ends, flanked by *SacI* and *XhoI* restriction sites. The resulting amplicon pool (containing an expected 1,369 promoter variant combinations – see below) was digested with SacI and XhoI and ligated with a SacI-XhoI digest of plasmid pKH1511c. ~40,000 colonies were harvested after transformation with this ligate, from which plasmid DNA was then purified, giving an initial plasmid library. An ampicillin-resistance cassette (*amp*^{R}) was amplified from pKD3 using for forward priming a pool of primers containing a region of 20 fully randomised nucleotides (the barcode, N_{20}) at their 5' end, flanked by *a* SpeI restriction site. The resulting amplicon pool was digested with SpeI and ligated with a BstZ17I-SpeI digest of the initial plasmid library (BstZ17I creates blunt ends). ~100,000 colonies were harvested after transformation with this ligate, each expected to harbour a plasmid with a unique barcode. Underlined regions of the

$P_{LtetO-1}$ and $P_{LlacO-1}$ sequences are the repressor binding sites reported in reference (*11*). The repressor of $P_{LtetO-1}$ is TetR, and the repressor of $P_{LlacO-1}$ is LacI, both encoded on the constant region of the library plasmid (*lacI-tetR*). The red T in $P_{LlacO-1}$ differs from the original sequence reported in reference (*11*), and was used due to its appearance during an initial adaptation step (this modified sequence still allows titratable control of expression from $P_{LlacO-1}$ using IPTG, as verified by growth and expression measurements – see Table S1). Black letters denote the -35 and -10 RNA-polymerase binding hexamers (note that 1 of the -10 nucleotides in $P_{LtetO-1}$, and 3 of the -35 nucleotides in $P_{LlacO-1}$, overlap with repressor binding sites). These hexamers were targeted for mutation: over these 12 sites, for each promoter, all 36 possible single-nucleotide substitutions were made, along with the wildtype, and the two sets of promoter variants were comprehensively combined. (**B**) To uncover which barcodes were linked to which promoter genotypes, the barcoded plasmid library was first digested with XhoI and SalI to remove the region between the $P_{LtetO-1}$ and $P_{LlacO-1}$ promoters and the barcode. The remaining section of the plasmids was re-circularised by ligation under conditions promoting intramolecular ligation. This ligate was used as template for PCR to amplify the newly created promoter-barcode region while adding Illumina adaptors to the amplicon termini. Finally, non-overlapping paired-end Illumina MiSeq sequencing was used to associate barcode sequences with promoter genotypes.

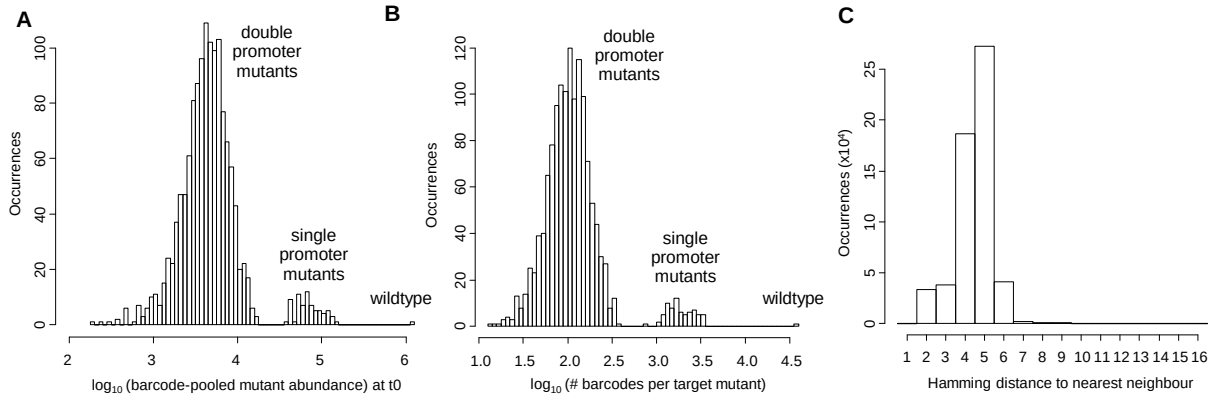

Fig. S2.

Sequencing coverage and quality of barcoded mutant library. Data from t_0 of the preliminary competition experiment. (**A**) The total coverage (after pooling barcode counts) of each genotype is on the order of 10^3 for double mutants, 10^5 for single mutants and 10^6 for the "wildtype". These different ranges result directly from the library creation strategy. (**B**) The number of unique barcodes associated to each genotype is on the order of 10^2 for double mutants, 10^3 for single mutants and 10^4 for the wildtype. These different ranges also result directly from the library creation strategy. (**C**) Over all barcode sequences observed, the mean Hamming distance to a barcode's nearest neighbor is 4.5. The complete absence of immediately neighbouring sequences is due to the preclustering analysis, in which immediately neighbouring sequences were assumed to be the result of PCR and sequencing errors.

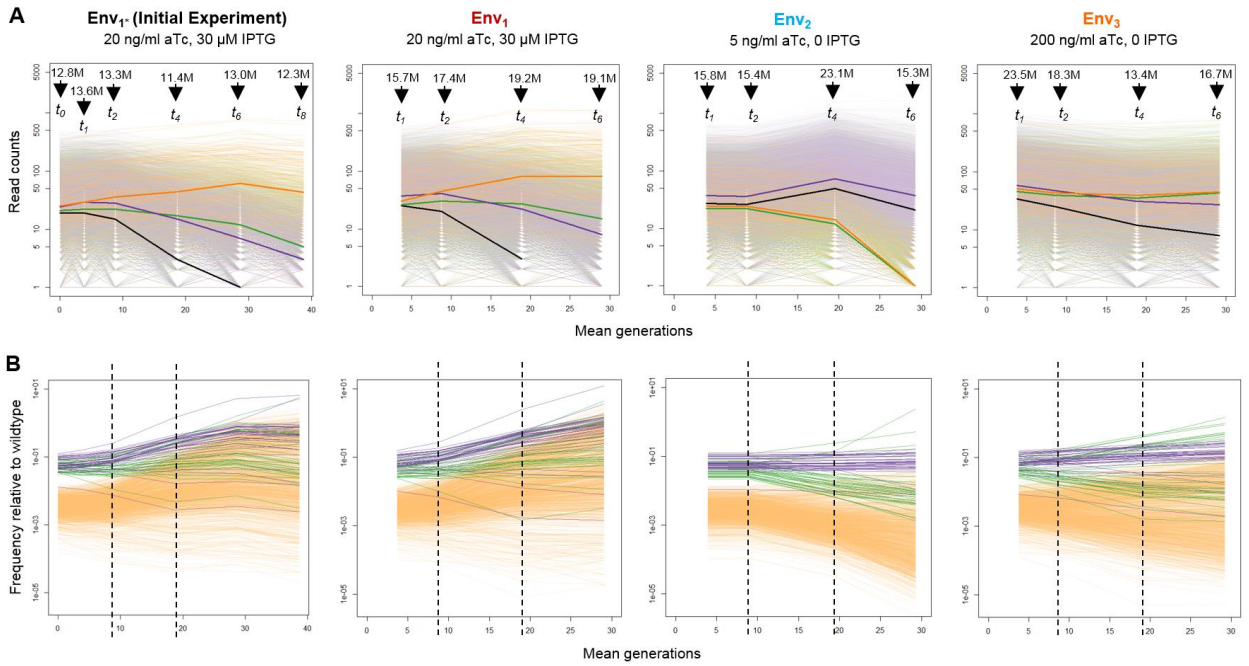

Fig. S3.

Mutant dynamics during pooled competition assays under different inducer concentrations. (**A**) Example trajectories are shown for all barcodes associated to the wildtype (black), a single $P_{LtetO-1}$-*araA* mutant (green), a single $P_{LlacO-1}$-*araB* mutant (purple) and the resulting double mutant (orange). Thick lines show median read counts. Numbers are the total number of HiSeq reads obtained at each sampled time-point. (**B**) Barcode-grouped trajectories are shown for all 1,368 mutants relative to the wildtype. Colours as in **A.** At every time-point, read counts for all barcodes belonging to a particular mutant have been summed and normalized to WT read counts. Dashed lines indicate time-window chosen for fitness estimation.

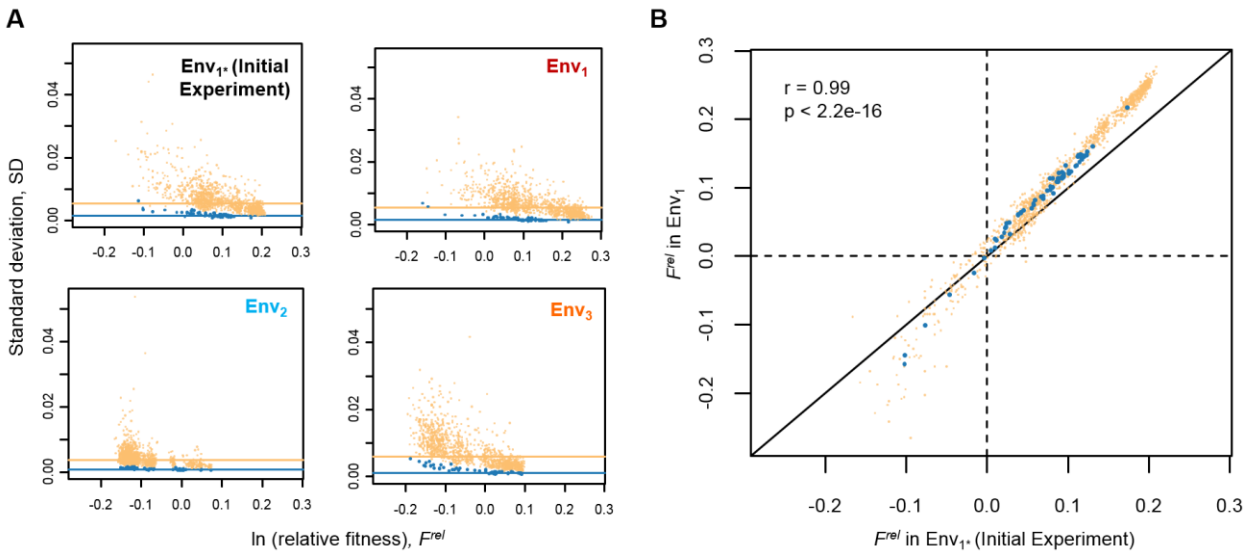

Fig. S4.

Measurement precision and reproducibility. (**A**) Fitness estimates are plotted against their corresponding bootstrap standard deviations (SD) for the different competition assays. Single mutants (blue) yield more precise estimates as they are associated to more barcodes than double mutants (orange). Precision is lower for less-fit genotypes due to their more rapidly decreasing abundances and so higher counting noise. Lines show median SDs. (**B**) F^{rel} estimates are compared between two replicate experiments (Env_1 conditions; same mutant library stock). Colors as in **A**. Reproducibility is high (Pearson's r = 0.99, n = 1,344 mutants), but systematic differences are apparent, likely due to small differences in media composition.

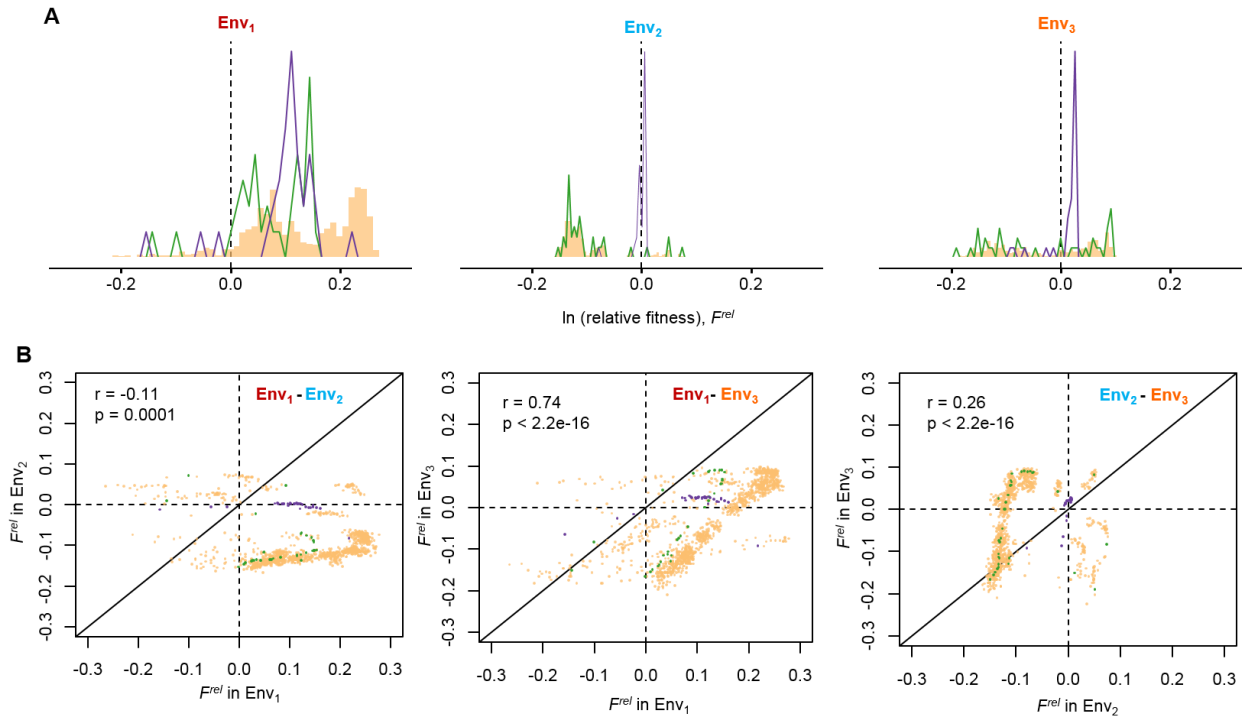

Fig. S5.

Fitness effects of single and double mutations across environments. (**A**) Density distributions of fitness effects (F^{rel}) of single $P_{LtetO-1}$-*araA* mutants (green), single $P_{LlacO-1}$-*araB* mutants (purple) and double mutants (orange). (**B**) Correlations between mutant F^{rel} in different environments range from strongly positive to weakly positive and weakly negative, and can show strong signs of non-monotonicity. Pearson's r is shown, with n = 1,345, 1,345 and 1,366 mutants, left-right. Colours as in **A**.

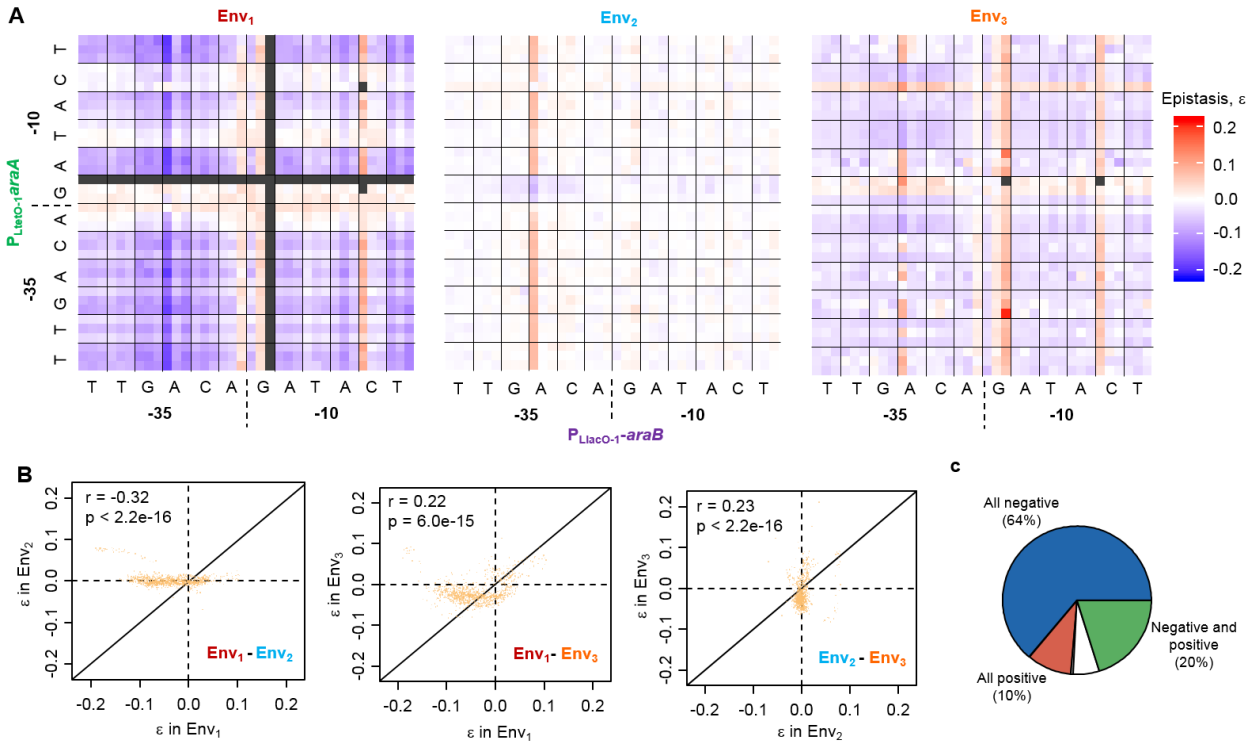

Fig. S6.

Epistasis across environments. (**A**) Genotype-epistasis maps. "-35" and "-10" denote the RNA polymerase-binding hexamers. Letters show the wildtype base at each position. The three mutants at each position are ordered alphabetically, as in Fig. 2A. Grey denotes incomputable epistasis coefficients. (**B**) Correlations between epistasis coefficients in different environments, with Pearson's r (n = 1,223, 1,223 and 1,294 mutation pairs, left-right). (**C**) The fraction of mutation pairs (n=1,296) for which, across environments, epistasis can be positive but never negative (red), negative but never positive (blue), or both positive and negative (green). Pairs exhibiting no detectable epistasis in any environment are shown in grey, and those for which epistasis could not be computed in all environments are white.

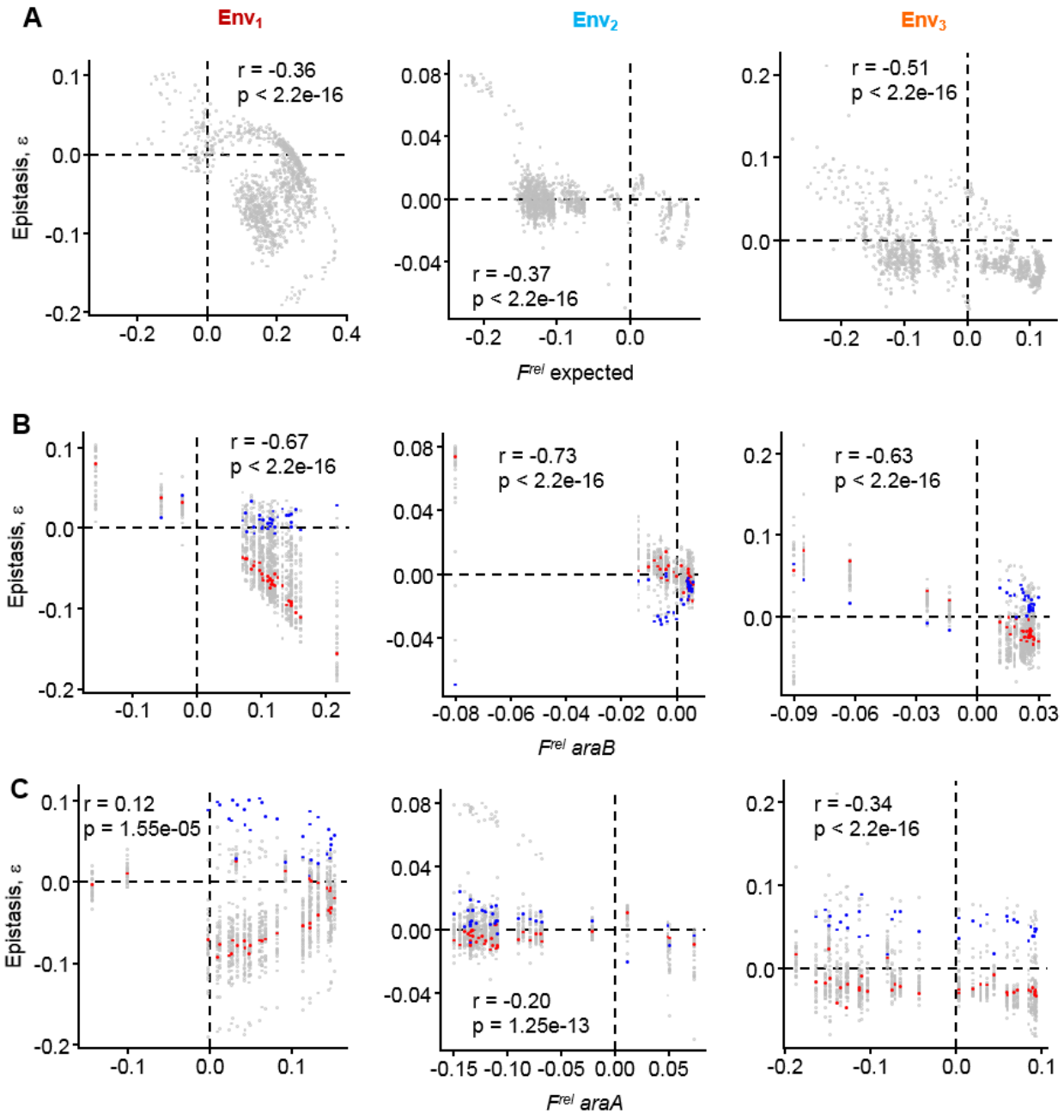

Fig. S7.

Correlations between individual fitness effects and epistasis. (**A**) In all environments, the sum of the fitness effects of two individual mutations (F^{rel} expected) correlates negatively with the epistasis they experience when combined, a trend of diminishing returns and losses (Pearson's r, n = 1,223, 1,296 and 1,294 mutation pairs, $Env_{1\text{-}3}$). The relationship appears complex, however. (**B**) When $P_{LlacO\text{-}1}$-*araB* is considered alone, the negative correlation between fitness effects and epistasis is stronger, but in Env_2 and Env_3 there is evidence of non-monotonicity (Pearson's r, number of mutation pairs as for **A**). Different $P_{LtetO\text{-}1}$-*araA* alleles can cause different trends within an environment, and the same $P_{LtetO\text{-}1}$-*araA* allele can cause different trends across

environments (coloured alleles as for Fig. 3B, top panel). (**C**) When $P_{LtetO-1}$-*araA* is considered alone, the negative correlation between fitness effects and epistasis is weaker, and in Env_1 it even becomes positive, albeit strongly non-monotonous (Pearson's r, number of mutation pairs as for **A**). Different $P_{LlacO-1}$-*araB* alleles can cause different trends within an environment, and the same $P_{LlacO-1}$-*araB* allele can cause different trends across environments (coloured alleles as for Fig. 3B, bottom panel).

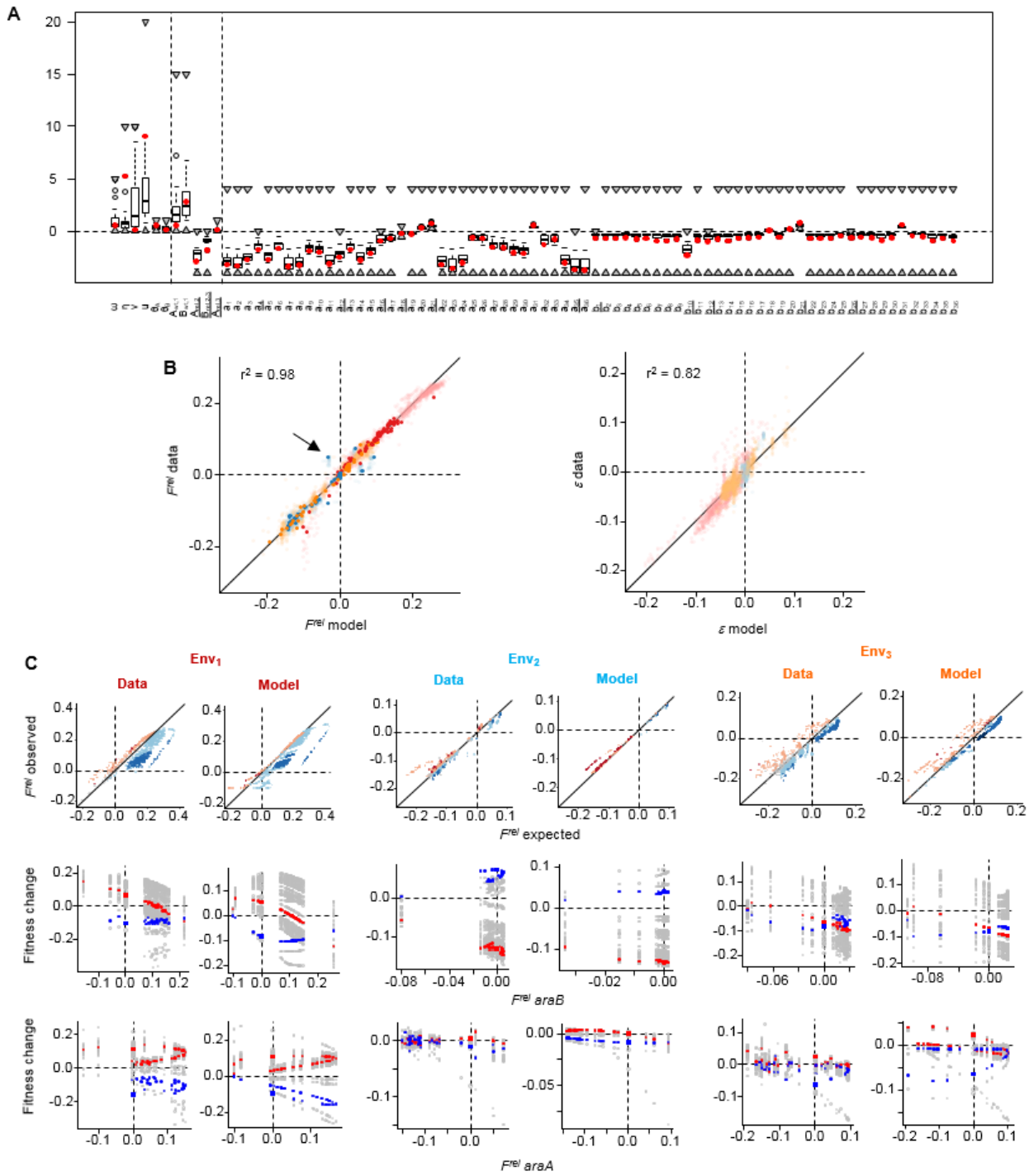

Fig. S8.

Performance of flux-toxicity-expression burden model. (**A**) Parameter estimates. Boxplots show distributions from the best 2.5% of Markov chains (n = 800 chains). Red points show parameter estimates from the best chain. Triangles show bounds of the uniform prior distributions. Parameter descriptions are given in Data S2. Vertical dashed lines separate the fitness function parameters, parameters describing wildtype expression levels across environments, and the expression effect (natural logarithm) of mutations (ordered as in Fig. 2B), from left to right. Prior

bounds of underlined expression effect parameters were guided by expression measurements. The majority of mutations in both promoters are predicted to decrease expression (expression effect < 0), which is not surprising as the (identical) "wildtype" RNA polymerase-binding sequences are a Hamming distance of only 2 away from the bacterial consensus sequence, indicating near-maximal binding strength. (**B**) Correlations between observed values and those predicted by the model. Left – fitness (n = 4,079 mutant measurements); right – epistasis (n = 3,813 mutation pair measurements); $p < 2.2e\text{-}16$ for both. Opaque points are single-mutants. Points are coloured by environment, as in Fig. 4A. Arrow points to genotypes containing a qualitative outlier mutation, $P_{LtetO\text{-}1}$-*araA* G7A, which is also the only mutation to be beneficial in all environments (Fig. 2B), presumably because its effect on *expression* depends on the environment (supported by the fact that it lies in a repressor binding site (Fig. S1A)). (**C**) Comparison of epistatic trends from experimental data and model, across environments. Top row – as for Fig. 3A; lower two rows – as for Fig. 3B (same 4 alleles coloured in all environments). Looping is explained by single-mutants lying on two sides of a phenotypic optimum.

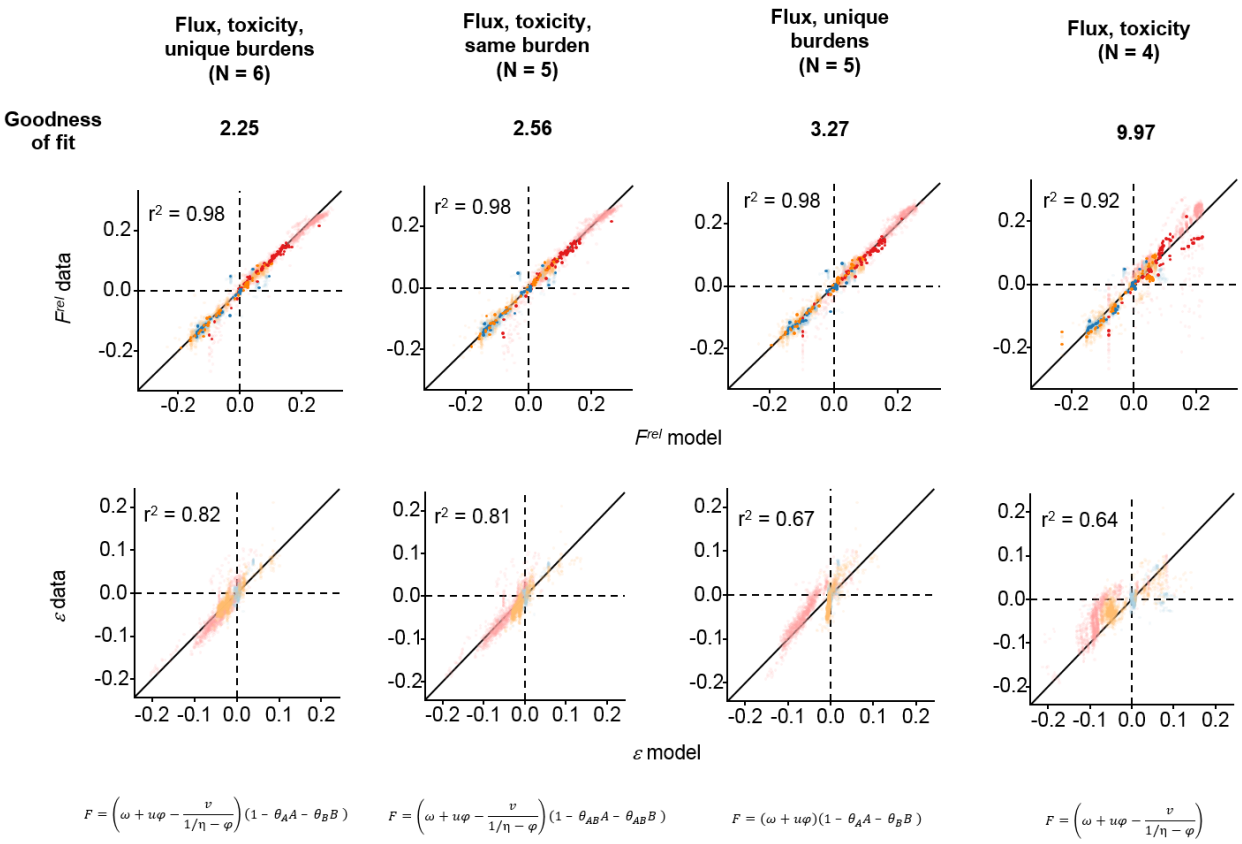

Fig. S9.

Goodness of fit comparison of different phenotype-fitness models. Correlations between observed values and those predicted by different model variations. Top row – fitness (n = 4,079 mutant measurements); bottom row – epistasis (n = 3,813 mutation pair measurements); $p < 2.2e\text{-}16$ for all. Opaque points are single-mutants. Points are coloured by environment, as in Fig. 4A. Goodness of fit is calculated as the sum of the squared differences between all observed fitness effects and epistasis coefficients and those predicted by the models (n = 7,892). N is the number of parameters defining the fitness function for each model. From left to right: complete model used in main text; as complete model, except that expression burden per activity unit is the same for both proteins; as complete model, but no toxicity; as complete model, but no expression burden.

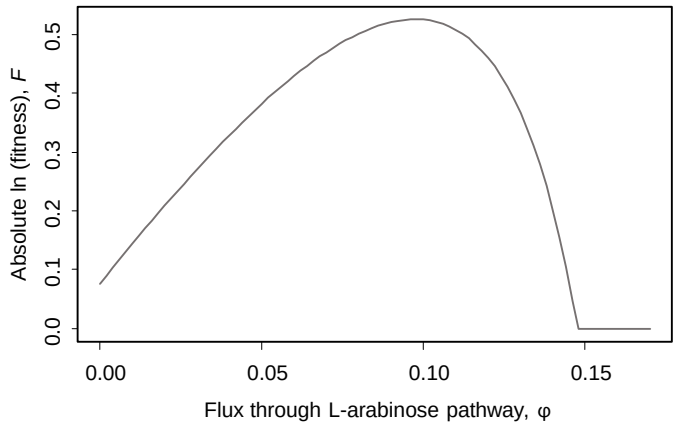

Fig. S10.

Flux-fitness relationship predicted by model. The fitted model results in the existence of a particular flux that is optimal for fitness (*27, 35*). As the flux exceeds this optimum, the rapid accumulation of the toxic intermediate, L-ribulose-5-phosphate, causes a steep fitness decline. The flux-fitness function diverges at very high fluxes (above the predicted range of our dataset), presumably as one or more of the simplifying assumptions underlying the enzyme activity-flux function starts to break down.

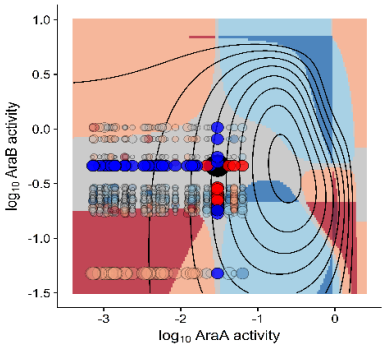

Fig. S11.

Fitness surface coloured by predicted epistasis category in Env_2 **(**as for Fig. 4C). The vast majority of interactions in this environment are predicted, and observed, to be weak (see blue points in Fig. S8C, right panel).

Plasmid name	Description	DNA fragments used for construction (this study)	Construction method / Supplier	Antibiotic used for selection	Accidental mutations / Sequence conflicts
pKD3 (*8*)	PCR template plasmid for Datsenko-Wanner (*8*) gene deletion, containing a *cat* Cm-resistance cassette flanked by *FRT* sites and an R6Kγ *pir*-dependent *ori*. Also used as PCR template for *bla* amplification in library barcoding step	-	Lab stocks	Cm	-
pKD46 (*8*)	Plasmid with L-arabinose-inducible λ Red expression cassette for Datsenko-Wanner (*8*) recombineering; temperature-sensitive *ori* (repA101ts) for easy curing	-	Lab stocks	Amp	-
pCP20 (*8*)	Plasmid with yeast *FLP* recombinase expression cassette for Datsenko-Wanner (*8*) resistance-gene excision; temperature-sensitive *ori* (repA101ts) for easy curing	-	Lab stocks	Amp	-
pSkunk3-BLA (*36*)	Phagemid containing *p15A* and *f1 ori*s, *bla* β-lactamase gene and *aadA1* Str/Sp-resistance gene. Used for backbone (*f1* phage *ori* not exploited in this study)	-	A. Birgy	Str	-
pZS4Int-1 (*11*)	*pSC101 ori*, *lacI* and *tetR* repressor genes under constitutive promoters, *attP* phage λ attachment site and *aadA1* Str/Sp -resistance gene. Used for *lacI* and *tetR*	-	A. Decrulle and I. Matic	Sp	G -> C at +246 of *tetR* ORF, causing Lys82 -> Asn82 (reported in other constructs, including reference (*37*)); 2 small insertions between *tetR* stop codon and its T1 terminator
pKH1503a	pSkunk3-BLA backbone, with *bla* replaced by: *araBA* under $P_{LlacO-1}$ inducible promoter (*11*) and *lacI* and *tetR* repressor genes under constitutive promoters (*11*)	pSkunk-bkb, aKH150312a, aKH150312b	Gibson Assembly	Str	-
pKH1503a^{evo}	Plasmid purified from a single colony (MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo} *[pKH1503a^{evo}]*) isolated after adaptation to alternating D- and L-arabinose. Sanger sequencing of *araBA*, *tetR* and *lacI*, along with their regulatory regions, revealed a single G –> C substitution in the 2nd *lac*O1 operator (-23 from TSS, in notation of reference (*11*)). This was found in 3/3 colonies tested from the evolved population, and was deliberately included in all future $P_{LlacO-1}$ -containing plasmids of this study (it was found through growth and expression measurements to still allow titratable expression control by IPTG)	-	Purified from a single colony isolated after MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat [pKH1503a]* adaptation to alternating D- and L-arabinose	Str	-
pKH1511c	pKH1503a^{evo} backbone (rather than pKH1503a backbone, to exploit any unseen adaptive mutations arising during adaptation), with $P_{LlacO-1}$-*araBA* replaced by *araA* and *araB* in divergent orientation and promoter-less, separated by SacI and XhoI restriction sites to allow easy insertion of divergent promoters	aKH151120a, aKH151120b, aKH151120c	Restriction-ligation	Str	C -> A substitution (synonymous) at +1638 of *araB* ORF
pSW23T::*attP* (*38*)	*ori*$V_{R6Kγ}$ (*pir*-dependent replication), *attP* phage λ attachment site, *cat* Cm-resistance gene. Used for *pir*-dependent backbone to avoid template plasmid carryover during cloning	-	A. Soler and D. Mazel	Cm	-
pBSK-BSD1	pBluescript SK phagemid containing *pUC* and *f1 ori*s, *bsd* Bsd-resistance cassette and *bla* β-lactamase gene. Used for *bsd*	-	A. Couce (gene synthesis by Epoch Life Science, Inc, TX, USA)	Amp	-
pKH1511d	pSW23T::*attP* with *bsd* Bsd-resistance cassette inserted into multiple cloning site. Used to avoid plasmid carryover during future *bsd* cloning	pSW23T::*attP*-bkb, aKH151126a	Gibson Assembly	Cm	-

Table S1.

Plasmids used in this study. Amp: ampicillin (100 µg/ml); Bsd: blasticidin; Cm: chloramphenicol (10 µg/ml); Spec: spectinomycin (50 µg/ml); Str: streptomycin (50 µg/ml).

DNA fragment name	Description/Creation	PCR template or digested plasmid	Primers used for PCR (blank if fragment comes directly from plasmid digestion)	Restriction enzymes used (either post-PCR or directly on plasmid)
pSkunk-bkb	pSkunk3 backbone, containing *ori*s and *aadA1* Str/Sp-resistance gene. Double-digest of pSkunk3-BLA to excise *bla*, followed by gel-extraction of backbone fragment	pSkunk3-BLA (*36*)	-	EcoRV, SpeI
aKH150312a	*lacI-tetR* constitutive expression cassette (*inc*. T1 terminator), with a downstream extension overlapping the SpeI extremity of pSkunk-bkb. PCR-amplification; overlap introduced on reverse primer	pZS4Int-1 (*11*)	oKH150312a, oKH150312b	-
aKH150312b	$P_{LlacO-1}$-*araBA* bicistronic cassette (*inc*. BBa_B1002 artificial terminator (BioBrick Foundation)), with an upstream extension overlapping the EcoRV extremity of pSkunk-bkb and a downstream extension overlapping the upstream extremity of aKH150312a. PCR-amplification; overlaps, $P_{LlacO-1}$ and BBa_B1002 all introduced on primers	*E. coli* K12 MG1655 genomic DNA	oKH150312c, oKH150312e	-
aKH151120a	pKH1503a^{evo} backbone, containing *ori*s, *aadA1* Str/Sp-resistance gene and *lacI-tetR* ($P_{LlacO-1}$-*araBA* removed), with a downstream extension containing an NcoI site. PCR-amplification; extension introduced on reverse primer	pKH1503a^{evo}	oKH150312a, oKH151120a	SphI, NcoI
aKH151120b	*araB* coding region followed by BBa_B1004 artificial terminator (BioBrick Foundation), with an upstream extension containing SacI and XhoI restriction sites and a downstream extension containing an SphI restriction site. PCR-amplification; extensions and BBa_B1004 introduced on primers	pKH1503a^{evo}	oKH151120b, oKH151120c	SacI, SphI
aKH151120c	*araA* coding region followed by BBa_B1002 artificial terminator (BioBrick Foundation), with an upstream extension containing a SacI restriction site and a downstream extension containing an NcoI restriction site. PCR-amplification; extensions introduced on primers	pKH1503a^{evo}	oKH151120d, oKH151120e	SacI, NcoI
pSW23T::*attP*-bkb	Linearised pSW23T::*attP*. Double-digest of pSW23T::*attP* at Multiple Cloning Site	pSW23T::*attP* (*38*)	-	SpeI, SacII
aKH151126a	*bsd* Bsd-resistance cassette (*inc*. T1 terminator), with an upstream extension overlapping the SacII extremity of pSW23T::*attP*-bkb and a downstream extension overlapping the SpeI extremity of pSW23T::*attP*-bkb. PCR-amplification; overlaps introduced on primers	pBSK-BSD1	oKH151126a, oKH151203a	-

Table S2.

DNA fragments used for cloning in this study.

Primer name	Sequence (5' -> 3')
oKH150202d	ATGGCAGAAATTCGAAAGC
oKH150312a	GCGGCATGCATTTACGTTGA
oKH150312b	AGCGCGTCGGCCGGTCGAATGCATAAGCTTACTAACTAGTGAGAGCGTTCACCGACAAAC
oKH150312c	AGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCAATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTG AGCACACCCGTTTTTTTGGATGGAGTG
oKH150312e	TTTTGCACCATTCGATGGTGTCAACGTAAATGCATGCCGCGCGAAAAAACCCCGCCGAAGCGGGGTTTTTTGCGTTAGCGACGAAACC CGTAATAC
oKH150401c	ATTCATTAATGCAGCTGGC
oKH151120a	TTTTTCCATGGGATATCGTTAGCCCACCCAG
oKH151120b	TTTTTGAGCTCCACAGCTAACCTCGAGACCCGTTTTTTTTGGATGGAGTG
oKH151120c	TTTTTGCATGCCGCGCGGCAAAACCCCGCCGAAGCGGGGTTTTCGGCGTTATAGAGTCGCAACGGCCT
oKH151120d	TTTTTGAGCTCTGCGACTCTATAAGGACACG
oKH151120e	TTTTTCCATGGGCGAAAAAACCCCGCCGA
oKH151126a	GATAAGCTTGATATCGAATTCCTGCAGCCCGGGGGATCCACTAGTGCGGCCGCGTGAGCCAGTGTGACTCTAGT
oKH151203a	CGTTTTATTTGATGCCTCTAGCACGCGTACCATGGAGCTCCACCGCGGATAGGAACTTCACGCTAGGG
KO-araBA-fwd	ACTCTCTACTGTTTCTCCATACCCGTTTTTTTGGATGGAGTGAAACGATGGTGTAGGCTGGAGCTGCTTC
KO-araBA-rev	ATCAGGCGTTACATACCGGATGCGGCTACTTAGCGACGAAACCCGTAATACATATGAATATCCTCCTTAG
verif-araBA-fwd	TTGCATCAGACATTGCCGTC
verif-araBA-rev	GTTGGCTTCTAATACCTGGCG
KO-lacIZYA-fwd	GTATGGCATGATAGCGCCCGGAAGAGAGTCAATTCAGGGTGGTGAATGTGGTGTAGGCTGGAGCTGCTTC
KO-lacIZYA-rev	AGCGCAGCGTATCAGGCAATTTTTATAATTTAAACTGACGATTCAACTTTCATATGAATATCCTCCTTAG
verif-lacIZYA-fwd	GTGATGACTATCAACTGGCAC
verif-lacIZYA-rev	CTATTGCTGGCAAGCTGGTG
KO-fucK-fwd	TCCGGCTACCGGGCCTGAACAAGCAAGAGTGGTTAGCCGGATAAGCAATGGTGTAGGCTGGAGCTGCTTC
KO-fucK-rev	AAATTAACGGCGAAATTGTTTTCAGCATTTCACACTTCCTCTATAAATTCCATATGAATATCCTCCTTAG
verif-fucK-fwd	AACGCACCAACTCAACCTGG
verif-fucK-rev	TTGATGCGGATGATGTCAGG
oBarcodeBla-fwd	TTTTTACTAGTGGCGCGCCGTCGACTTNNNNNATNNNNNATNNNNNATNNNNNATCTTCAGATCCTCTACGCCGG
oBarcodeBla-rev	TACACTCCGCTAGCGCTGATGTCCGGCGGTGCCAGGTGGCACTTTTCGGG
oLinkBarcode-fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNCGTGTCCTTATAGAGTCGCAG
oLinkBarcode-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNGTCCGGCGTAGAGGATCTG
oBarcodeSeq-fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNGTGAACGCTCTCACTAGTGG
oBarcodeSeq-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNCAAGATCCGGCCACGATGC
c1 (*8*)	TTATACGCAAGGCGACAAGG
c2 (*8*)	GATCTTCCGTCACAGGTAGG

Table S3.

PCR Primers used in this study, excluding those used directly for promoter mutagenesis.

Strain name	Description/Usage	Genotype	Engineering method / Supplier	Antibiotic / supplements used for selection / adaptation
K12 MG1655	"Wildtype" laboratory strain	F$^-$ λ$^-$ *ilvG*$^-$ *rfb*-50 *rph*-1	A. Couce; Coli Genetic Stock Centre #6300	-
PIR1	*pir*-expressing strain for cloning and maintenance of *pir*-dependent plasmids (thymidine auxotroph)	F$^-$ Δ*lac*169 *rpo*S(am) *rob*A1 *cre*C510 *hsd*R514 *end*A *rec*A1 *uidA*(Δ*MluI*)::*pir-116*	A. Soler and D. Mazel	Erm + dT
DH5α	Standard strain for plasmid cloning and maintenance	F$^-$ λ$^-$ Φ80*lac*ZΔM15 Δ(*lac*ZYA-*arg*F) U169 *rec*A1 *end*A1 *hsd*R17 (rK$^-$, mK$^+$) *pho*A *sup*E44 *thi*-1 *gyr*A96 *rel*A1	Lab stock	-
DH5α Δ*araBA*::*cat*	Intermediate for construction of DH5α Δ*araBA*	DH5α Δ*araBA*::*cat*	Datsenko-Wanner (pKD46) (*8*)	Cm
DH5α Δ*araBA*	Preliminary tests; used as alternative to DH5α in this study	DH5α Δ*araBA*::*FRT*	Datsenko-Wanner (pCP20) (*8*)	-
MG1655 Δ*araBA*::*cat*	Intermediate for construction of MG1655 Δ*araBA*	MG1655 Δ*araBA*::*cat*	Datsenko-Wanner (pKD46) (*8*)	Cm
MG1655 Δ*araBA*	Preliminary tests; intermediate for construction of MG1655 Δ*araBA* Δ*lacIZYA*::*cat* and MG1655 Δ*araBA* D-ara$^{+/evo}$	MG1655 Δ*araBA*::*FRT*	Datsenko-Wanner (pCP20) (*8*)	-
MG1655 Δ*araBA* Δ*lacIZYA*::*cat*	Preliminary tests	MG1655 Δ*araBA*::*FRT* Δ*lacIZYA*::*cat*	Datsenko-Wanner (pKD46) (*8*)	Cm
MG1655 Δ*araBA* D-ara$^{+/evo}$	MG1655 Δ*araBA* derivative able to metabolise D-arabinose using genes of the *fuc* operon, due to a *fucR* mutation rendering the operon D-arabinose-inducible. Further adapted to D-arabinose for ~ 60 generations, and a single colony isolated. Intermediate for construction of MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK*::*cat*	MG1655 Δ*araBA*::*FRT* *fucR*$^{D\text{-}ara}$ D-ara^{evo}	Incubated in M9 + D-arabinose until visible growth (6 days). Then, serially transferred in M9 + D-arabinose for ~ 60 generations before isolation of a single colony (see refs. (*3*–*5*))	D-arabinose
MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK*::*cat*	Intermediate for construction of MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK*	MG1655 Δ*araBA*::*FRT* *fucR*$^{D\text{-}ara}$ D-ara^{evo} Δ*fucK*::*cat*	Datsenko-Wanner (pKD46) (*8*)	Cm
MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK*	Intermediate for construction of MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat*	MG1655 Δ*araBA*::*FRT* *fucR*$^{D\text{-}ara}$ D-ara^{evo} Δ*fucK*::*FRT*	Datsenko-Wanner (pCP20) (*8*)	-
MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat*	Intermediate for construction of MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat [pKH1503a]*	MG1655 Δ*araBA*::*FRT* *fucR*$^{D\text{-}ara}$ D-ara^{evo} Δ*fucK*::*FRT* Δ*lacIZYA*::*cat*	Datsenko-Wanner (pKD46) (*8*)	Cm
MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* *[pKH1503a]*	Intermediate for construction of MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo} *[pKH1503a]*	MG1655 Δ*araBA*::*FRT* *fucR*$^{D\text{-}ara}$ D-ara^{evo} Δ*fucK*::*FRT* Δ*lacIZYA*::*cat* *[pKH1503a]*	Plasmid transformation (electroporation)	Str
MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo} *[pKH1503a^{evo}]*	MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* *[pKH1503a]* derivative adapted to alternating D- and L-arabinose in presence of 10μM IPTG for ~45 generations, and a single large colony isolated. Evolved plasmid (pKH1503a^{evo}) used as template for further plasmid constructs; intermediate for construction of MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo}	MG1655 Δ*araBA*::*FRT* *fucR*$^{D\text{-}ara}$ D-ara^{evo} Δ*fucK*::*FRT* Δ*lacIZYA*::*cat* D/L-ara^{evo} *[pKH1503a^{evo}]*	Incubated in M9 + 10μM IPTG + D-arabinose until visible growth (2 weeks). Then, serially transferred in M9 + 10μM IPTG + alternating D- and L-arabinose for ~45 generations before isolation of a single large colony	Alternating D- and L-arabinose (+ IPTG + Str)
MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo}	Final engineered/adapted plasmidless host strain for barcoded promoter-mutant plasmid library; able to utilize L-arabinose in presence of plasmid-expressed AraA and AraB, and D-arabinose in presence of plasmid-expressed AraB	MG1655 Δ*araBA*::*FRT* *fucR*$^{D\text{-}ara}$ D-ara^{evo} Δ*fucK*::*FRT* Δ*lacIZYA*::*cat* D/L-ara^{evo}	Plasmid curing	Ribitol (*9*) (+ IPTG + Cm)

Table S4.

E. coli strains used in this study. Cm: chloramphenicol (10 μg/ml); dT: thymidine (30 μg/ml); Erm: erythromycin (20 μg/ml); Str: streptomycin (50 μg/ml); IPTG: isopropyl β-D-1-

thiogalactopyranoside. For adaptation, D- and L-arabinose were present at 0.3% and 0.2% w/v, respectively.

Primer name	Sequence (5' -> 3')
oPtetLib-fwd-1	TTTTTGAGCTCGTGCTC**AGTATC**TCTATCACTGATAGGGA**TGTCA*N***TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-2	TTTTTGAGCTCGTGCTC**AGTATC**TCTATCACTGATAGGGA**TGTC*N*A**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-3	TTTTTGAGCTCGTGCTC**AGTATC**TCTATCACTGATAGGGA**TGT*N*AA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-4	TTTTTGAGCTCGTGCTC**AGTATC**TCTATCACTGATAGGGA**TG*N*CAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-5	TTTTTGAGCTCGTGCTC**AGTATC**TCTATCACTGATAGGGA**T*N*TCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-6	TTTTTGAGCTCGTGCTC**AGTATC**TCTATCACTGATAGGGA***N*GTCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-7	TTTTTGAGCTCGTGCTC**AGTAT*N***TCTATCACTGATAGGGA**TGTCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-8	TTTTTGAGCTCGTGCTC**AGTA*N*C**TCTATCACTGATAGGGA**TGTCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-9	TTTTTGAGCTCGTGCTC**AGT*N*TC**TCTATCACTGATAGGGA**TGTCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-10	TTTTTGAGCTCGTGCTC**AG*N*ATC**TCTATCACTGATAGGGA**TGTCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-11	TTTTTGAGCTCGTGCTC**A*N*TATC**TCTATCACTGATAGGGA**TGTCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPtetLib-fwd-12	TTTTTGAGCTCGTGCTC***N*GTATC**TCTATCACTGATAGGGA**TGTCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGT GACTCTAGTAG
oPlacLib-rev-1	TTTTTCTCGAGGTGCTC**AGTATC**TTGTTATCCGATCACAA**TGTCA*N***TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-2	TTTTTCTCGAGGTGCTC**AGTATC**TTGTTATCCGATCACAA**TGTC*N*A**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-3	TTTTTCTCGAGGTGCTC**AGTATC**TTGTTATCCGATCACAA**TGT*N*AA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-4	TTTTTCTCGAGGTGCTC**AGTATC**TTGTTATCCGATCACAA**TG*N*CAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-5	TTTTTCTCGAGGTGCTC**AGTATC**TTGTTATCCGATCACAA**T*N*TCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-6	TTTTTCTCGAGGTGCTC**AGTATC**TTGTTATCCGATCACAA***N*GTCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-7	TTTTTCTCGAGGTGCTC**AGTAT*N***TTGTTATCCGATCACAA**TGTCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-8	TTTTTCTCGAGGTGCTC**AGTA*N*C**TTGTTATCCGATCACAA**TGTCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-9	TTTTTCTCGAGGTGCTC**AGT*N*TC**TTGTTATCCGATCACAA**TGTCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-10	TTTTTCTCGAGGTGCTC**AG*N*ATC**TTGTTATCCGATCACAA**TGTCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-11	TTTTTCTCGAGGTGCTC**A*N*TATC**TTGTTATCCGATCACAA**TGTCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG
oPlacLib-rev-12	TTTTTCTCGAGGTGCTC***N*GTATC**TTGTTATCCGATCACAA**TGTCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG

Table S5.

Forward and reverse primer sets for promoter mutagenesis. -35 and -10 RNA polymerase-binding hexamers are in bold. N (italicised) denotes a mix of all 4 bases.

Data S1.

Mutant fitness estimates with their 95% bootstrap confidence intervals and the number of barcodes used for their estimation. Genotype nomenclature is [$P_{LtetO-1}$-*araA* mutation].[$P_{LlacO-1}$-*araB* mutation].

Data S2.

Parameter estimates for complete phenotype-fitness model. Prior bounds are provided (bold indicates bounds guided by expression measurements), along with the upper, lower and median estimates from the best 2.5% of Markov chains, and the estimates from the single best chain.