Science Advances
AAAS

advances.sciencemag.org/cgi/content/full/6/23/eabb2236/DC1

Supplementary Materials for

Flux, toxicity, and expression costs generate complex genetic interactions in a metabolic pathway

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

*Corresponding author. Email: kembleharry@gmail.com (H.K.); philippe.nghe@espci.fr (P.N.); olivier.tenaillon@inserm.fr (O.T.)

Published 3 June 2020, *Sci. Adv.* **6**, eabb2236 (2020)
DOI: 10.1126/sciadv.abb2236

The PDF file includes:

Other Supplementary Material for this manuscript includes the following:

(available at advances.sciencemag.org/cgi/content/full/6/23/eabb2236/DC1)

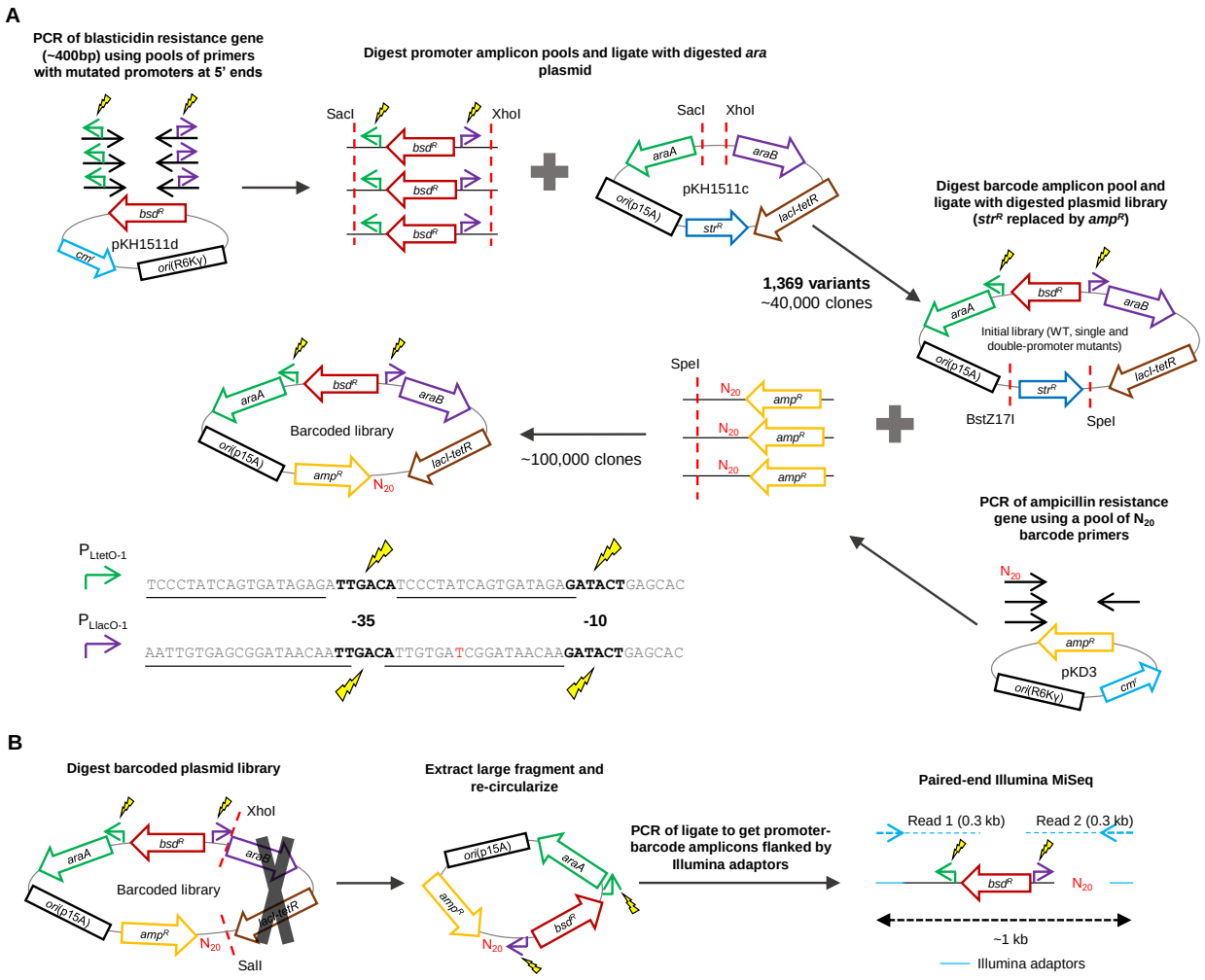

Fig. S1. Construction and characterisation of barcoded promoter-mutant plasmid library. (**A**) A blasticidin-resistance cassette (*bsd*^{R}) was amplified from pkH1511d using pools of primers carrying variants of the entire $P_{LtetO-1}$ (green arrow) and $P_{LlacO-1}$ (purple arrow) promoters at their 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 (*41*). 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 (*41*), 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.

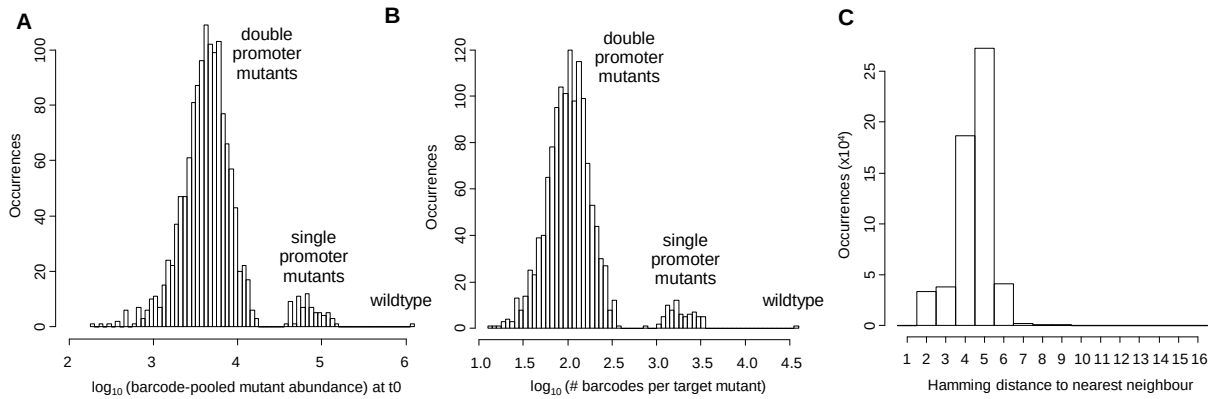

Fig. S2. Sequencing coverage and quality of barcoded mutant library. Data from t_0 of the preliminary competition experiment. (**A**) The total coverage (after pooling barcode counts) of each genotype is on the order of 10^3 for double mutants, 10^5 for single mutants and 10^6 for the "wildtype". These 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 process, in which immediately neighbouring sequences were assumed to be the result of PCR and sequencing errors.

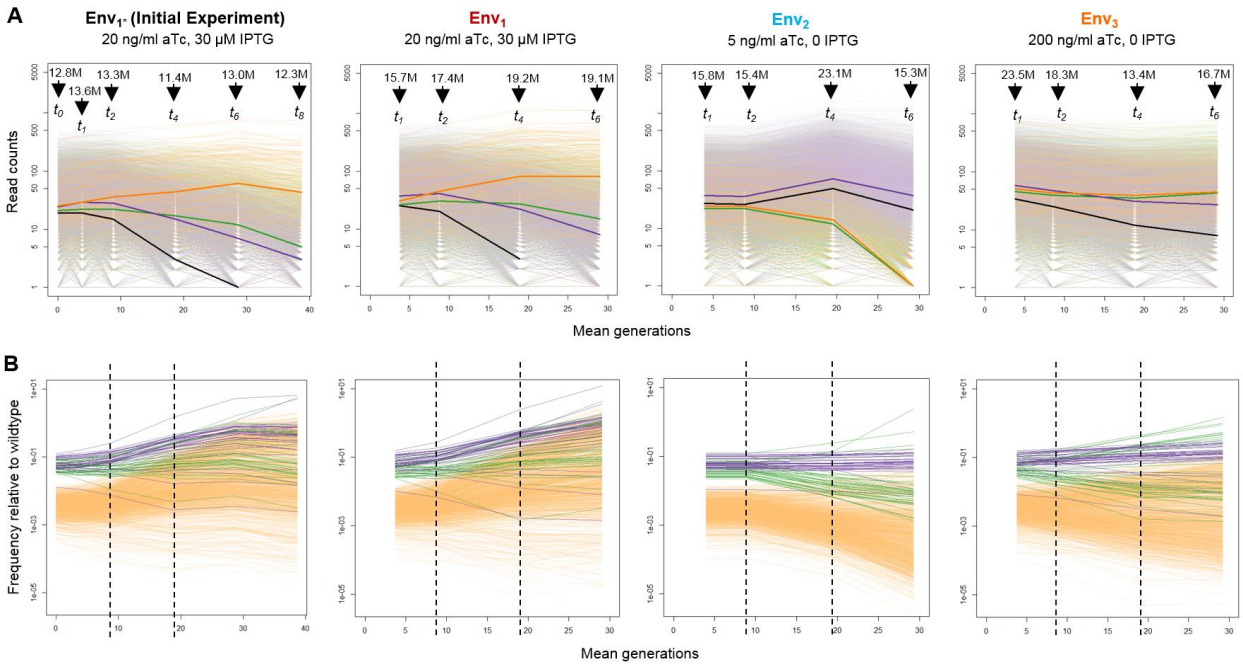

Fig. S3. Mutant dynamics during pooled competition assays under different inducer concentrations. (**A**) Example trajectories are shown for all barcodes associated to the wildtype (black), a single $P_{LtetO-1}$-*araA* mutant (green), a single $P_{LlacO-1}$-*araB* mutant (purple) and the resulting double 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 (see Fig. S4).

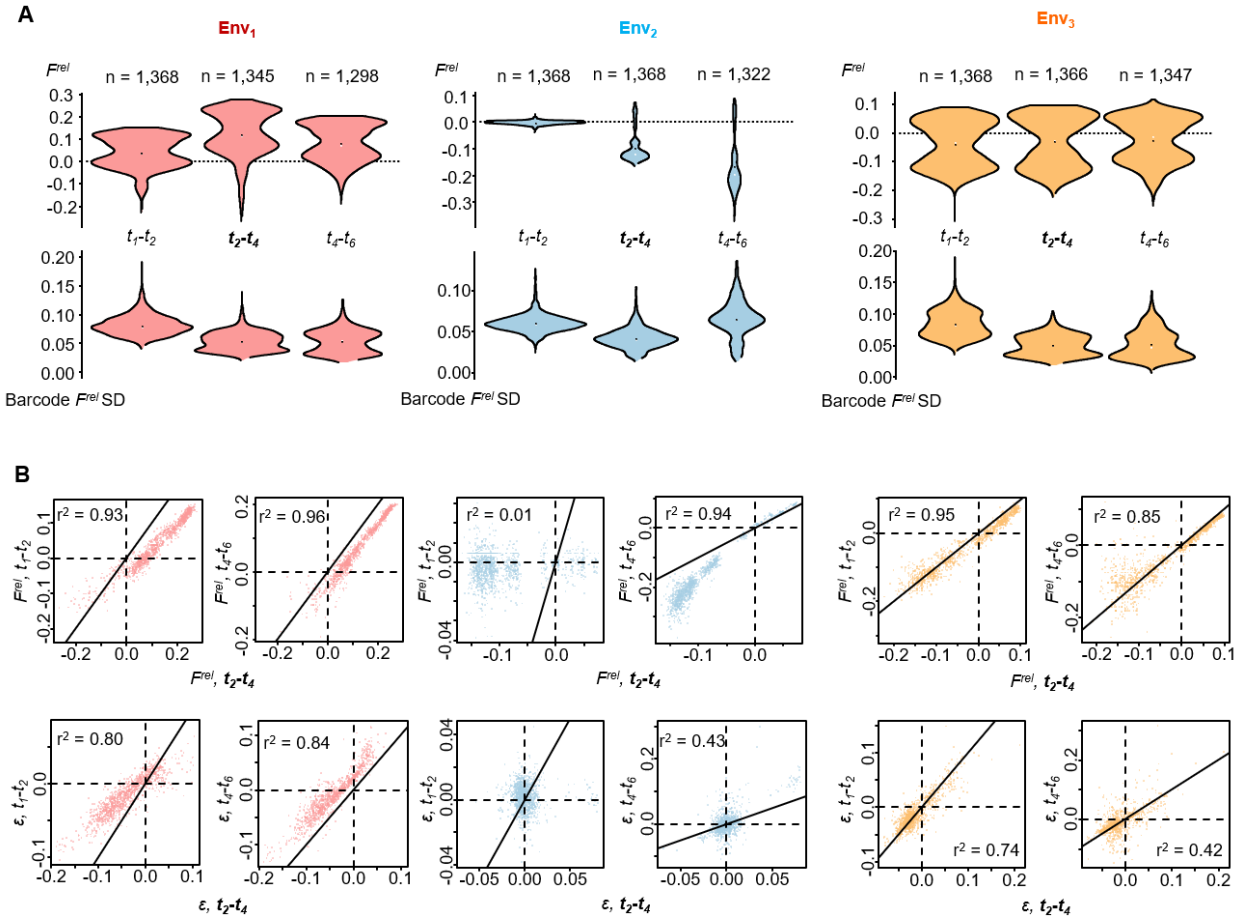

Fig. S4. Choice of time-window for fitness estimation. (**A**) Violin plots of quantifiable single- and double-mutation fitness effects (F^{rel}) (top row) and of the standard deviation (SD) in their associated barcode F^{rel} estimates (bottom row) obtained using different consecutive sampled timepoints, for environments Env_{1-3} (white point - median; black point – mean). The number of mutant genotypes for which F^{rel} was quantifiable is provided. Computation of mutant F^{rel} is described in Materials and Methods, and barcode F^{rel} standard deviation was necessarily computed using only barcodes still present in the later timepoint of each timepoint pair. The t_2-t_4 time-window (starting after >20 hours in final competition media) was selected as optimal for fitness estimation, due to the combination of strong selection and low intra-genotype variation observed across all environments. The extended negative F^{rel} tail seen in Env_2 at t_4-t_6 is associated with a large increase in intra-genotype variation, likely due to the increased sampling noise occurring when barcode abundances approach zero (see Fig. S5). As such, the beneficial tail is a more reliable estimator of selection strength. (**B**) Fitness (top row) and epistasis (bottom row) estimates from the chosen time-window, t_2-t_4, are plotted against the corresponding values from the t_1-t_2 and t_4-t_6 time-windows (diagonal line: y=x). Significant ($p < 0.05$) r^2 values are indicated. The one case where fitness and epistasis estimates are essentially uncorrelated to those from the chosen time-window is t_1-t_2 in Env_2 (low concentrations of both inducers), during which selection appears essentially absent.

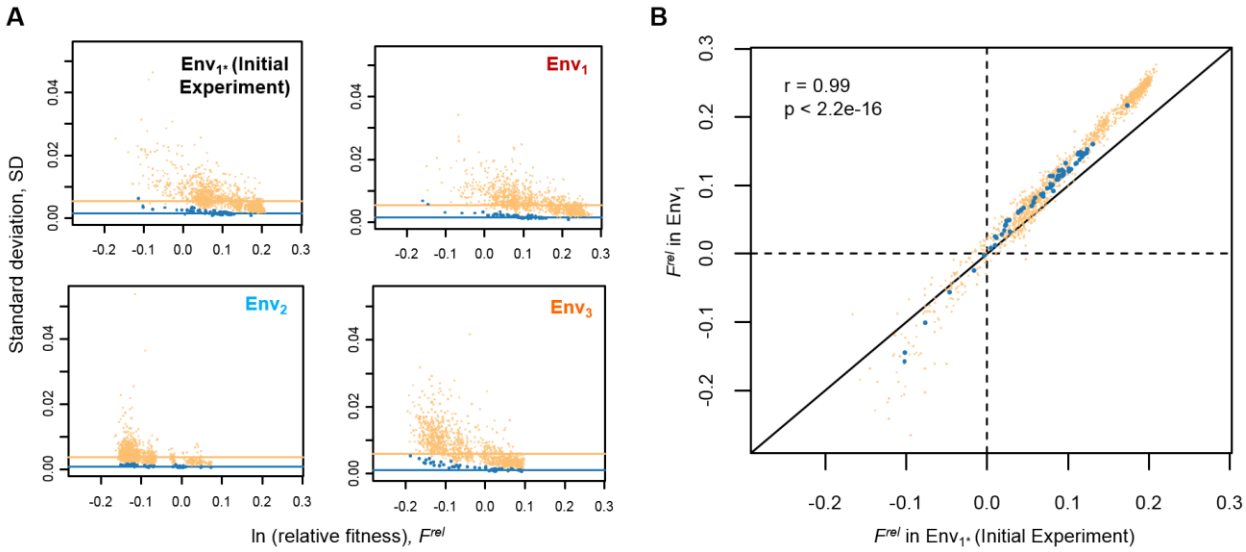

Fig. S5. Fitness measurement precision and reproducibility. (**A**) Fitness estimates are plotted against their corresponding bootstrap standard deviations (SD) for the different competition assays. Single mutants (blue) yield more precise estimates as they are associated to more barcodes than double mutants (orange). Precision is lower for less-fit genotypes due to their more rapidly decreasing abundances and so larger sampling 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.

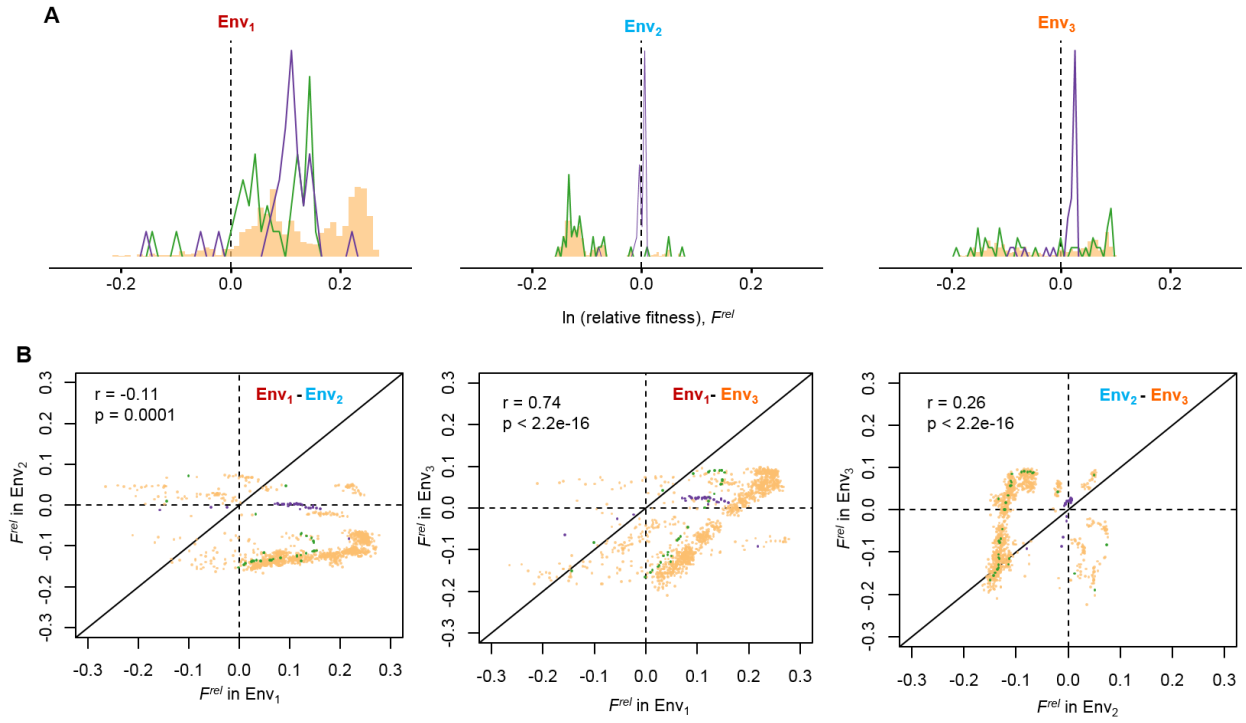

Fig. S6. Fitness effects of single and double mutations across environments. (**A**) Density distributions of fitness effects (F^{rel}) of single $P_{LtetO-1}$-*araA* mutations (green), single $P_{LlacO-1}$-*araB* mutations (purple) and double mutations (orange). (**B**) Correlations between mutant F^{rel} 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**.

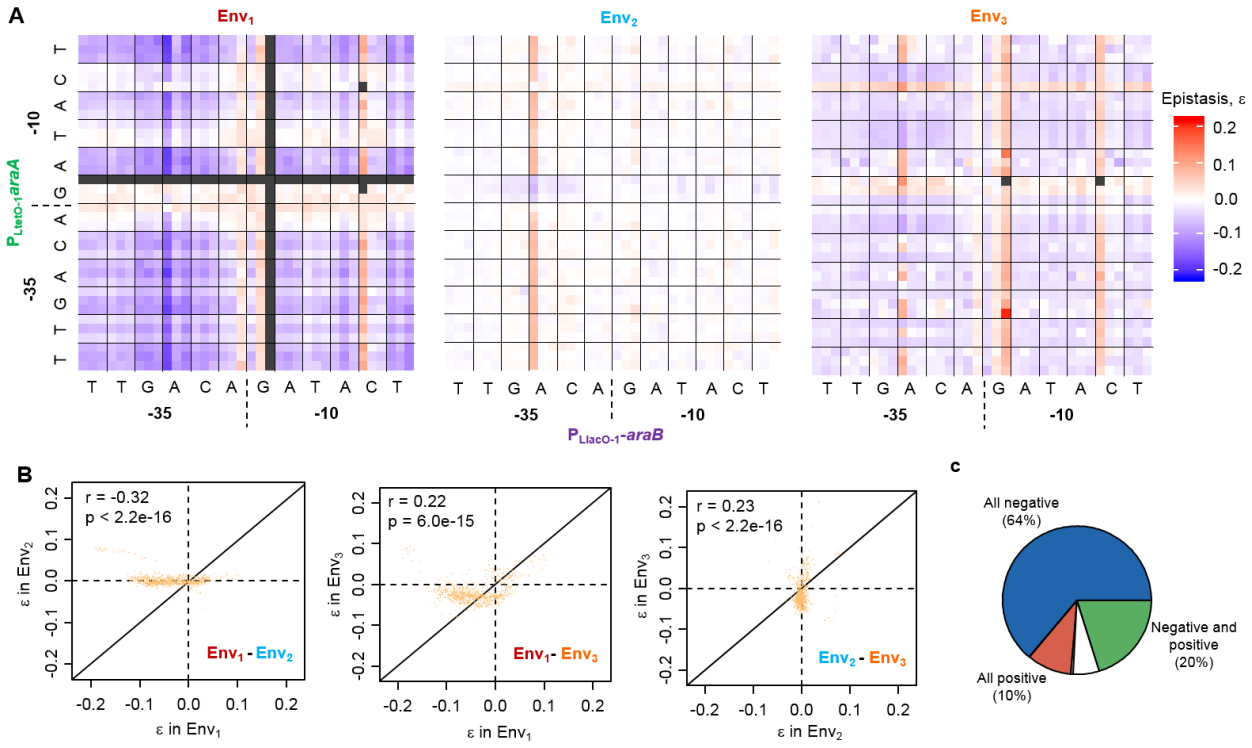

Fig. S7. Epistasis across environments. (**A**) Genotype-epistasis maps. "-35" and "-10" denote the RNA polymerase-binding hexamers. Letters show the wildtype base at each position. The three mutants at each position are ordered alphabetically, as in Fig. 2A. Grey denotes incomputable epistasis coefficients. (**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.

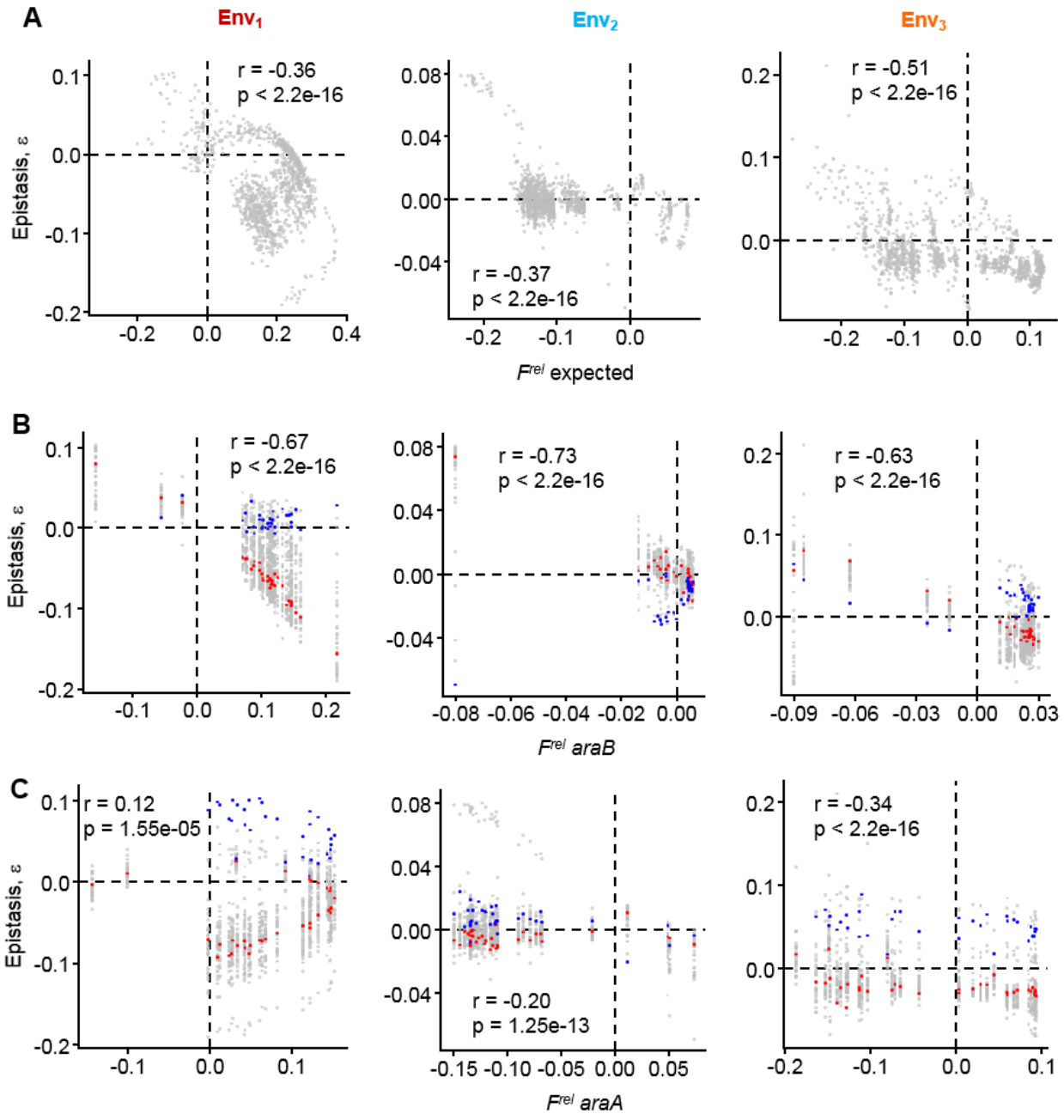

Fig. S8. Correlations between individual fitness effects and epistasis. (**A**) In all environments, the sum of the fitness effects of two individual mutations (F^{rel} expected) correlates negatively with the epistasis they experience when combined, a trend of diminishing returns and losses (Pearson's 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).

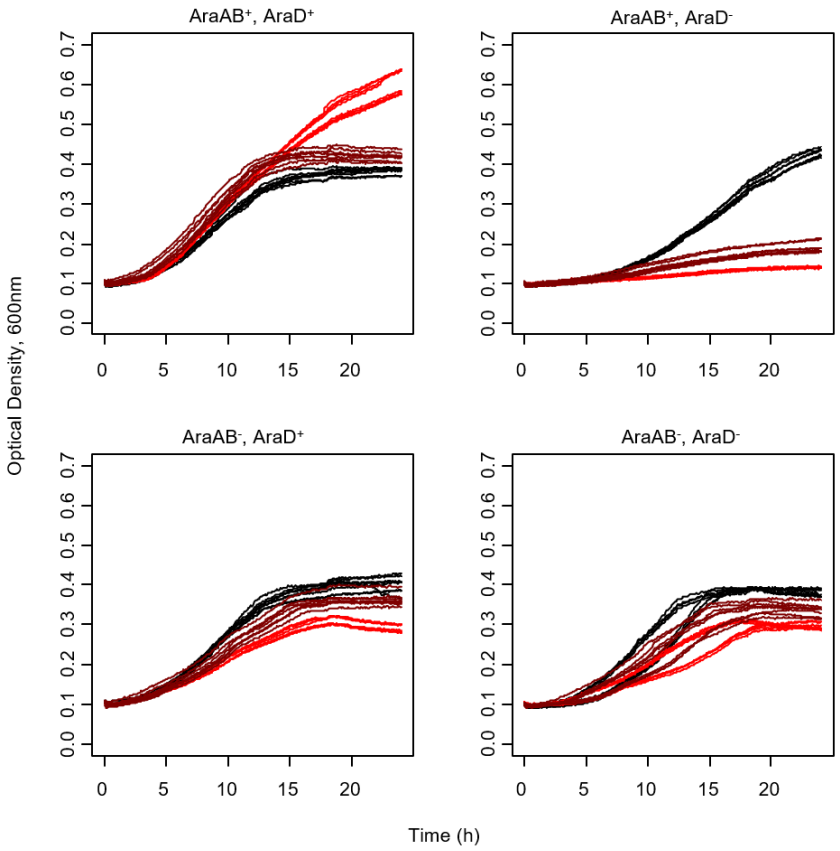

Fig. S9. Toxicity of the metabolic intermediate, L-ribulose-5-phosphate. As shown in Fig. 1A, L-ribulose-5-phosphate lies immediately between the AraB and AraD enzymes in the arabinose pathway. Each plot shows replicate growth curves of a different strain in M9 plus 0.5% casamino acids and inducers (Env_1 concentrations), with 0.4% arabinose (red), 0.004% arabinose (brown) or no arabinose (black). Top-left: A strain with *araA* and *araB* on the plasmid under $P_{LtetO-1}$ and $P_{LlacO-1}$ inducible promoters and a functional *araD* gene shows improved growth in the presence of arabinose. Top-right: A strain with *araA* and *araB* on the plasmid under $P_{LtetO-1}$ and $P_{LlacO-1}$ inducible promoters and a deleted *araD* gene has its growth drastically inhibited by the presence of arabinose. Bottom row: For strains lacking induction of *araA* and *araB* (plasmid as for $AraAB^+$ but with $P_{LtetO-1}$ and $P_{LlacO-1}$ promoters removed), the presence or absence of *araD* has no significant effect on growth, confirming that it is the processing by AraA and AraB of L-arabinose into L-ribulose-5-phosphate, and its consequent accumulation, that inhibits growth when AraD is absent.

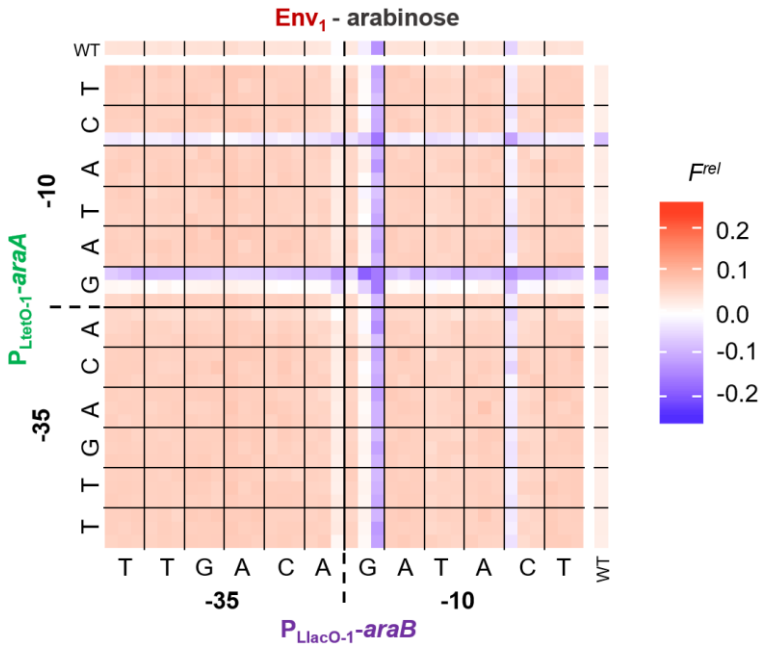

Fig. S10. Genotype-fitness map in absence of arabinose, under Env$_1$ inducer concentrations. Genotypes are colored according to the natural logarithm of their fitness relative to the wildtype (F^{rel}). "-35" and "-10" denote the RNA polymerase-binding hexamers. Letters show wildtype bases, and the 3 mutations at each position are ordered alphabetically, as in Fig. 2. Single promoter mutants make up the right-most column (*araA*) and top row (*araB*). Inducer concentrations were 20 ng/ml aTc and 30 µM IPTG, as for Env$_1$.

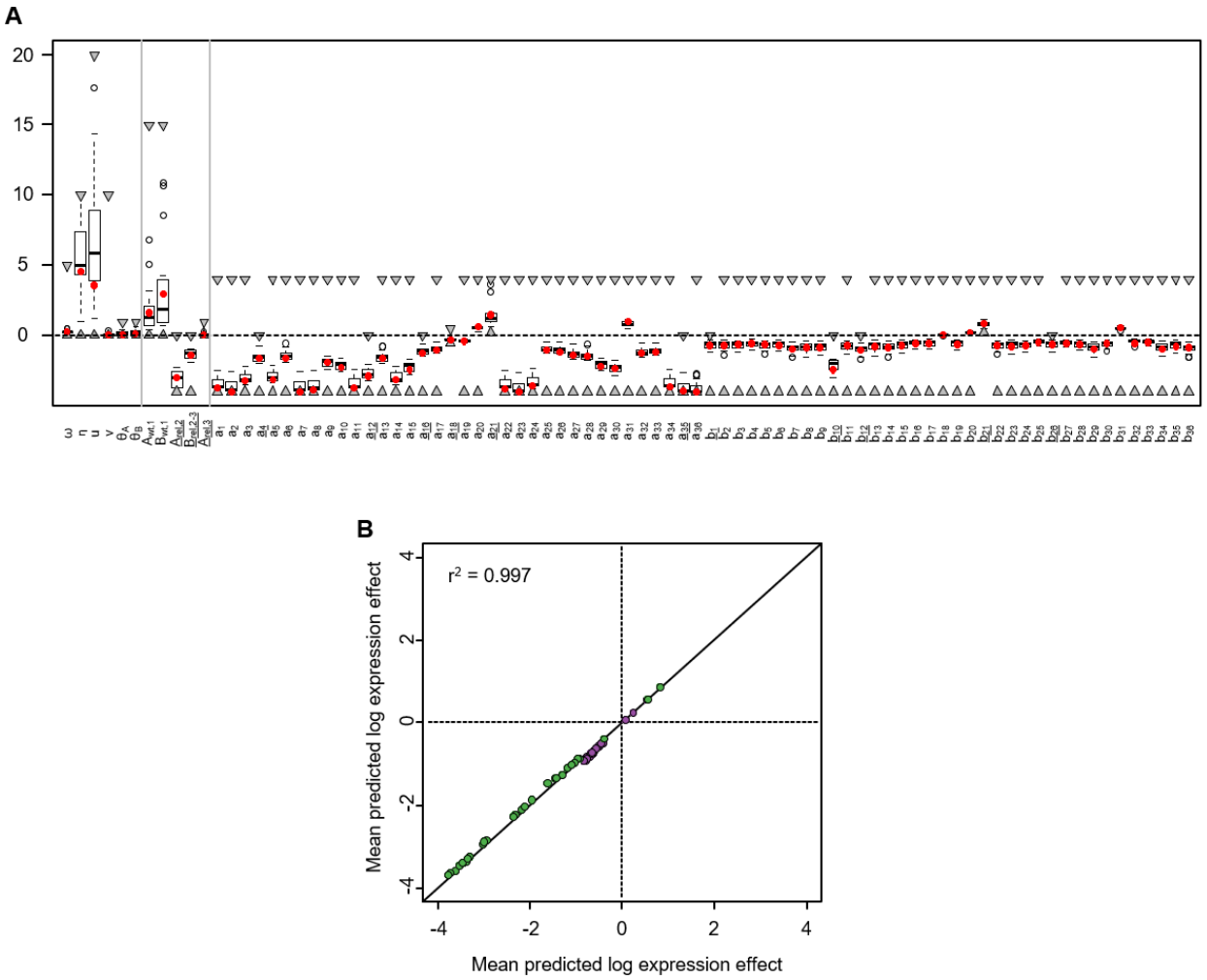

Fig. S11. Robustness of parameter estimates. (**A**) Boxplots show parameter estimate distributions from the best 2.5% of Markov chains (n = 800 chains). Red points show parameter estimates from the best chain. Triangles show bounds of the uniform prior distributions. Parameter descriptions are given in Data S2. Vertical grey 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. (**B**) The best 2.5% of Markov chains were randomly split into two sub-sample halves, and the mean mutational expression effect predictions of each half are plotted against each other. Only those mutations with fully "naïve" expression effect prior bounds (-4,4) are included in this analysis (n = 61: 30/36 *araA* (green) and 31/36 *araB* (purple) promoter mutations). 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.

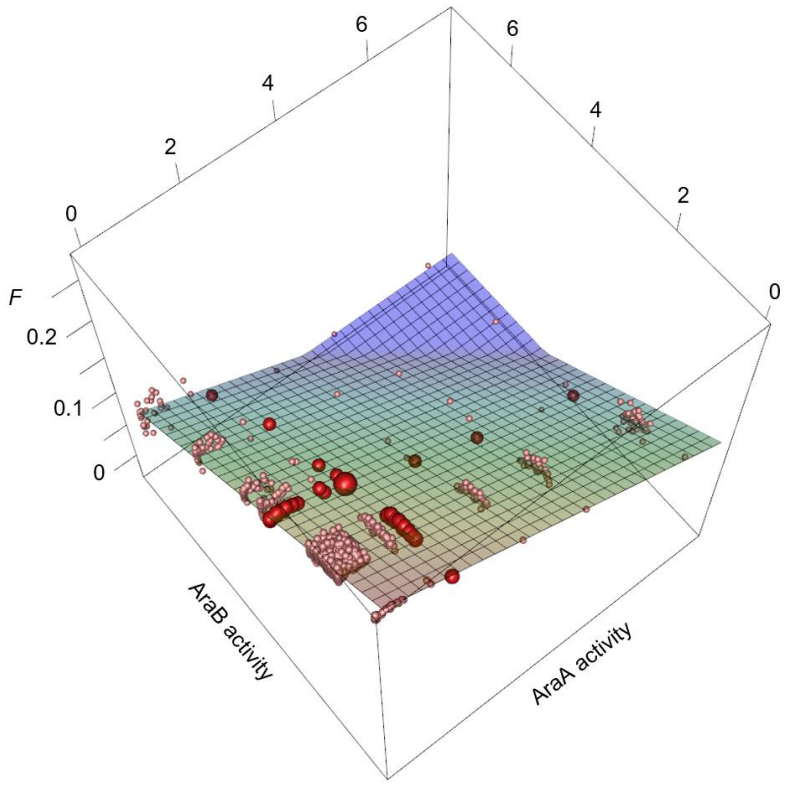

Fig. S12. Fitted activity-fitness model in absence of arabinose, under Env_1 inducer concentrations. Spheres are positioned according to predicted activity levels and observed F^{rel}. Largest sphere is wildtype, intermediate-sized spheres are single mutants, small pale spheres are double mutants.

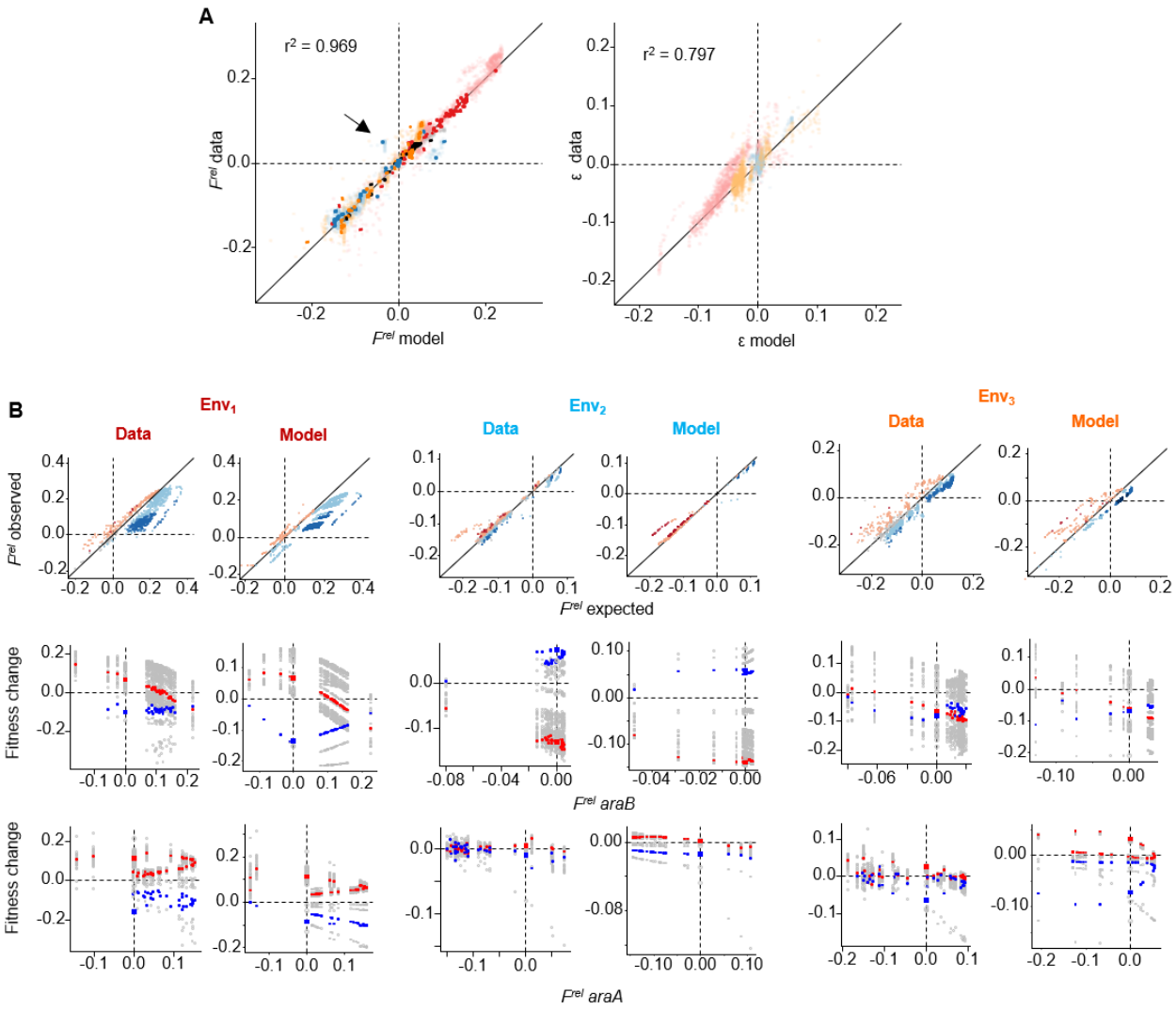

Fig. S13. Descriptive power of the flux-toxicity-expression burden model. (**A**) Correlations between observed values and those predicted by the model. Left – fitness (n = 5,447 mutant measurements); right – epistasis (n = 5,109 single-single-double mutant triplet measurement sets); p < 2.2e-16 for both. Opaque points are single-mutants. Points are coloured by environment, as in Fig. 4A, with black indicating the no-arabinose control environment. 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)). (**B**) 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.

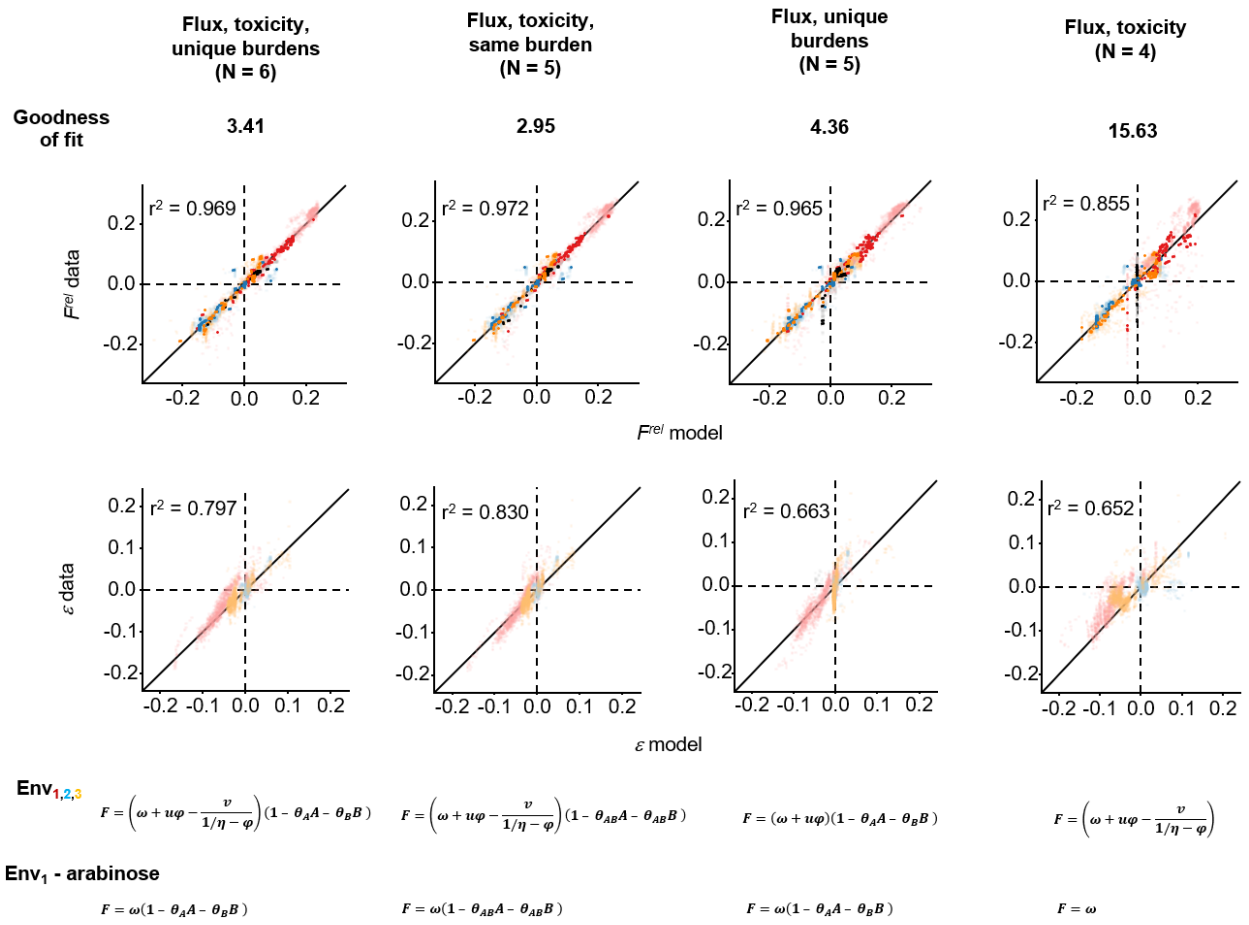

Fig. S14. Goodness-of-fit comparison of different phenotype-fitness models. Correlations between observed values and those predicted by different model variations. Top row – fitness (n = 5,447 mutant measurements); bottom row – epistasis (n = 5,109 single-single-double mutant triplet measurement sets); $p < 2.2e\text{-}16$ for all. Opaque points are single-mutants. Points are coloured by environment, as in Fig. 4A, with black indicating the no-arabinose control environment. 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 = 10,556). 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. For each model, the functions describing fitness in $Env_{1\text{-}3}$ and the no-arabinose control environment are provided.

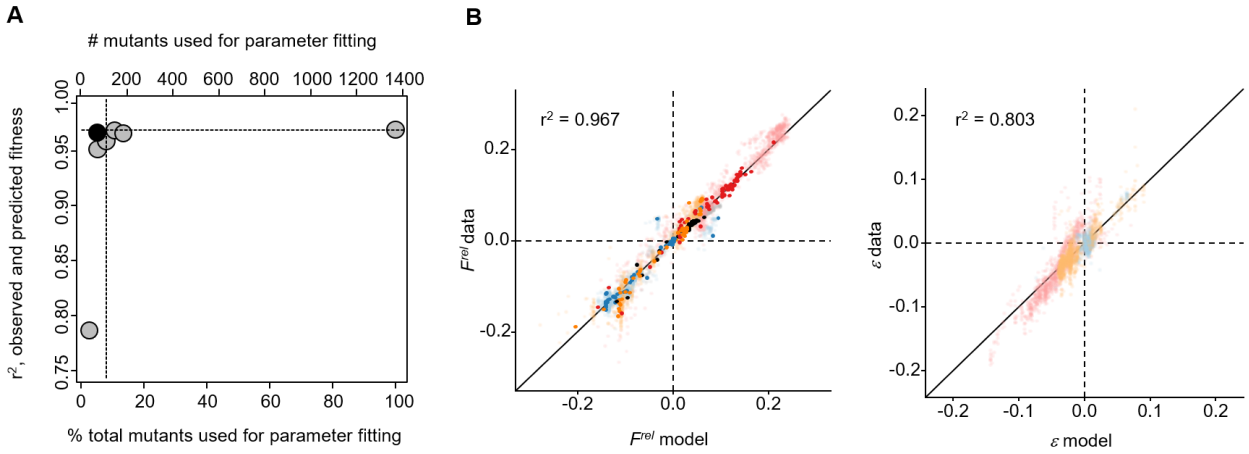

Fig. S15. Predictive power of the flux-toxicity-expression burden model. (**A**) Quality of the fit (r^2 between predicted and observed F^{rel} of all mutants across all environments) obtained from model parameterization on varying fitness data subsets. Grey points: varying fractions of mutants were sampled randomly, but evenly such that for a given sub-sample each allele was represented an equal number of times, and their observed F^{rel} across all environments was used for model fitting. Black point: the observed F^{rel} of the set of 72 single-mutants across all environments was used for model fitting. Vertical dashed line marks just 5% of all mutants, and horizontal dashed line marks the quality of fit obtained from parametrization on the complete set of mutants. (**B**) Correlations between observed values and those predicted by the flux-toxicity-expression burden model parameterized on only single-mutants (black point in **A**). Left – fitness (n = 5,447 mutant measurements); right – epistasis (n = 5,109 single-single-double mutant triplet measurement sets); $p < 2.2e\text{-}16$ for both. Opaque points are single-mutants. Points are coloured by environment, as in Fig. 4A, with black indicating the no-arabinose control environment.

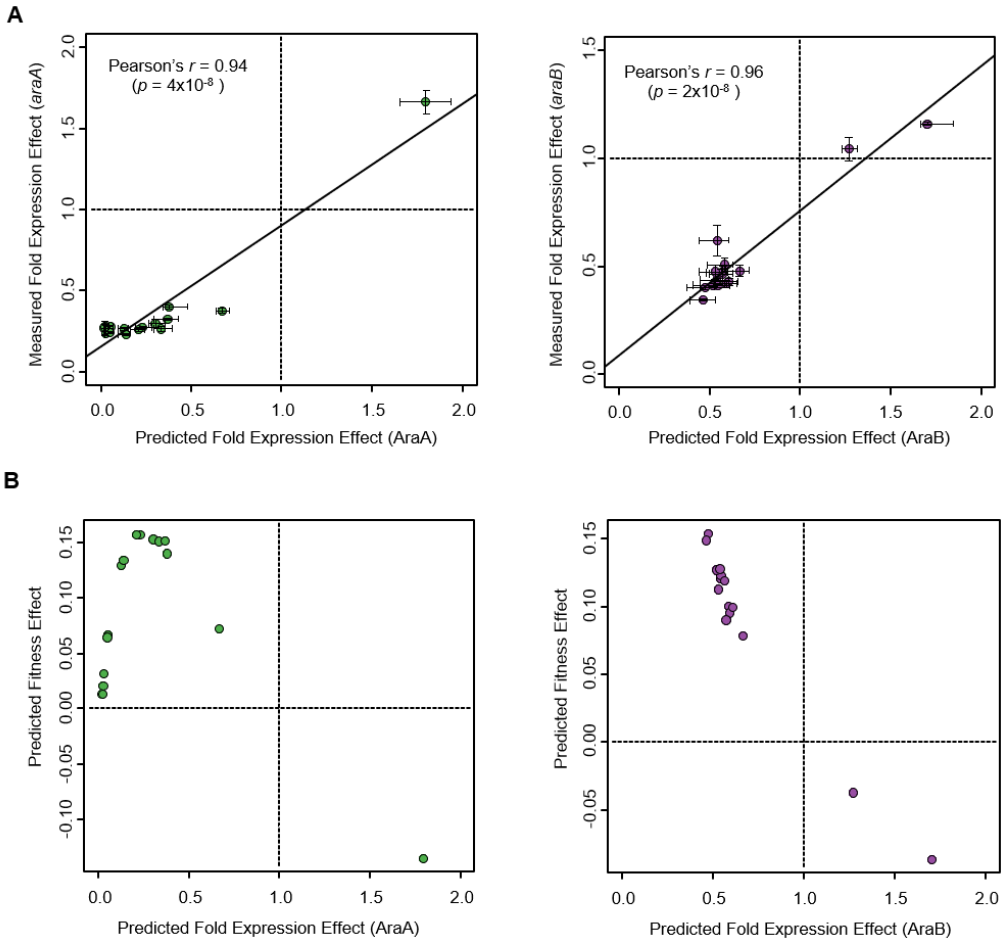

Fig. S16. Validation of inferred mutational expression effects by single-mutant expression measurements. (**A**) Correlations between predicted and observed fold expression effects for a randomly isolated subset of single *araA* (left) and *araB* (right) promoter mutants. Mutational expression effects were measured by RT-qPCR of "wildtype" and single-mutant clones grown in M9 plus 0.5% casamino acids and Env_1 inducer concentrations after random isolation from 2 single-mutant libraries and Sanger-sequencing. Only those isolated mutations with fully "naïve" expression effect prior bounds are included in this analysis (n = 16/36 *araA* and 15/36 *araB* single promoter mutants). Measured expression effects shown are the median of 5 replicates performed on 2 different days (cultures), with normalization to "wildtype" expression performed separately on each day (error-bars: standard error). The "wildtype" expression reference was estimated based on the mean of 4 independently isolated clones. Predicted expression effects shown are the median of the best 2.5% of Markov chains (error-bars: interquartile range). Diagonal line: linear regression. (**B**) Predicted relationship between expression-effect and fitness-effect in Env_1 for the set of single-mutants shown in **A**. Predicted expression effects shown are the median of the best 2.5% of Markov chains. The non-monotonic relationship for *araA* in particular highlights the non-triviality of the agreement shown in **A**.

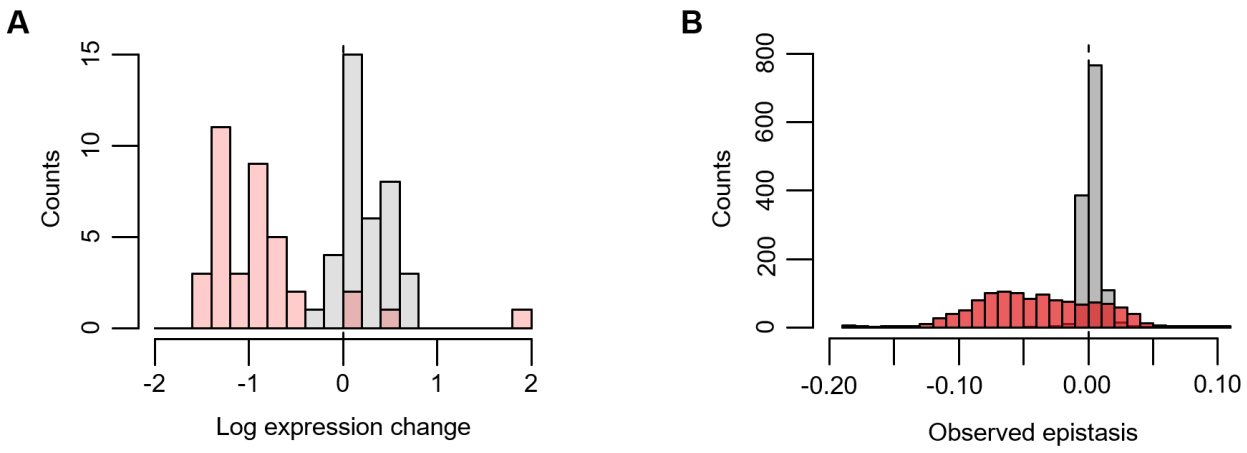

Fig. S17. Lack of direct interactions between promoters at the level of expression. (**A**) Distribution of log expression effects for the set of randomly isolated single-mutants (n = 37/72). As expected, the change in expression from the aggregated non-mutated promoters (grey) is centred close to zero, and the measured expression range is narrow compared to that for the aggregated mutated promoters (pink). (**B**) Distribution of epistasis observed for Env_1 (red; n = 1,223) and the no-arabinose control environment containing Env_1 inducer concentrations (grey; n = 1,296). As expected, epistasis vanishes in the absence of arabinose, supporting the conclusion that fitness interactions result from the shape of the phenotype-fitness landscape rather than from direct interactions between promoter mutations at the level of expression.

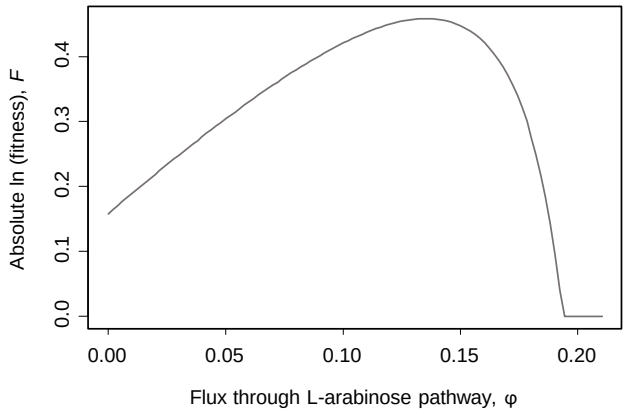

Fig. S18. Flux-fitness relationship predicted by model. The fitted model results in the existence of a particular flux that is optimal for fitness (*18*, *21*). As the flux exceeds this optimum, the rapid accumulation of the toxic intermediate, L-ribulose-5-phosphate, causes a steep fitness decline.

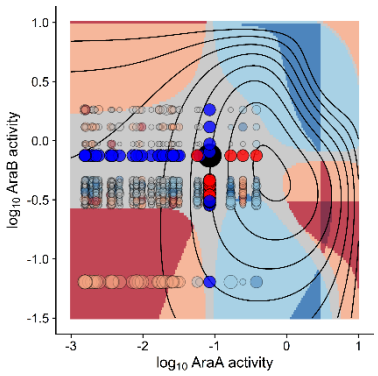

Fig. S19. Fitness surface coloured by predicted epistasis category in Env_2 (as for Fig. 4C). The vast majority of interactions in this environment are predicted, and observed, to be weak (see blue points in Fig. S13A, right panel).

Table S1. Plasmids used in this study. Amp: ampicillin (100 µg/ml); Bsd: blasticidin; Cm: chloramphenicol (10 µg/ml); Kan: kanamycin (50 µg/ml); Spec: spectinomycin (50 µg/ml); Str: streptomycin (50 µg/ml).

Plasmid name	Description	DNA fragments used for construction (this study)	Construction method / Supplier	Antibiotic used for selection	Accidental mutations / Sequence conflicts
pKD3 (*38*)	PCR template plasmid for Datsenko-Wanner (*38*) 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	-
pKD4 (*38*)	PCR template plasmid for Datsenko-Wanner (*38*) gene deletion, containing a *kan*^{R} Kan-resistance cassette flanked by *FRT* sites and an R6Kγ *pir*-dependent *ori*	-	Lab stocks	Kan	-
pKD46 (*38*)	Plasmid with L-arabinose-inducible λ Red expression cassette for Datsenko-Wanner (*38*) recombineering; temperature-sensitive *ori* (repA101ts) for easy curing	-	Lab stocks	Amp	-
pCP20 (*38*)	Plasmid with yeast *FLP* recombinase expression cassette for Datsenko-Wanner (*38*) resistance-gene excision; temperature-sensitive *ori* (repA101ts) for easy curing	-	Lab stocks	Amp	-
pSkunk3-BLA (*61*)	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 (*41*)	*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 (*62*)); 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 (*41*) and *lacI* and *tetR* repressor genes under constitutive promoters (*41*)	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 (*41*)). 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* (*63*)	*ori*$V_{R6K\gamma}$ (*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 S2. DNA fragments used for cloning in this study.

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 (*61*)	-	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 (*41*)	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* (*63*)	-	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 S3. PCR Primers used in this study, excluding those used directly for promoter mutagenesis. -35 and -10 RNA polymerase-binding hexamers are in bold.

Primer name	Sequence (5' -> 3')
oKH150202d	ATGGCAGAAATTCGAAAGC
oKH150312a	GCGGCATGCATTTACGTTGA
oKH150312b	AGCGCGTCGGCCGGTCGAATGCATAAGCTTACTAACTAGTGAGAGCGTTCACCGACAAAC
oKH150312c	AGCCAGAAAACCGAATTTTGCTGGGTGGGCTAACGATATCAATTGTGAGCGGATAACAATTGACATTGTGAGCGGATAACAAGATACTG AGCACACCCGTTTTTTTGGATGGAGTG
oKH150312e	TTTTGCACCATTCGATGGTGTCAACGTAAATGCATGCCGCGCGAAAAAACCCCGCCGAAGCGGGGTTTTTTGCGTTAGCGACGAAACC CGTAATAC
oKH150401c	ATTCATTAATGCAGCTGGC
oKH151120a	TTTTTCCATGGGATATCGTTAGCCCACCCAG
oKH151120b	TTTTTGAGCTCCACAGCTAACCTCGAGACCCGTTTTTTTGGATGGAGTG
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
KO-araD-rev	GTTTGATTGGCTGTGGTTTTATACAGTCATTACTGCCCGTAATATGCCTTCATATGAATATCCTCCTTAG
verif-araD-rev	CGCGACAGTAAAGGCCATAC
oBarcodeBla-fwd	TTTTTACTAGTGGCGCGCCGTCGACTTNNNNNATNNNNNATNNNNNATNNNNNATCTTCAGATCCTCTACGCCGG
oBarcodeBla-rev	TACACTCCGCTAGCGCTGATGTCCGGCGGTGCCAGGTGGCACTTTTCGGG
oLinkBarcode-fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNCGTGTCCTTATAGAGTCGCAG
oLinkBarcode-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNGTCCGGCGTAGAGGATCTG
oBarcodeSeq-fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGNNNNNNGTGAACGCTCTCACTAGTGG
oBarcodeSeq-rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGNNNNNNCAAGATCCGGCCACGATGC
oPtetWT-fwd	TTTTTGAGCTCGTGCTC**AGTATC**TCTATCACTGATAGGGA**TGTCAA**TCTCTATCACTGATAGGGAGGCGCGCCGTGAGCCAGTGTGACT CTAGTAG
oPlacWT-rev	TTTTTCTCGAGGTGCTC**AGTATC**TTGTTATCCGATCACAA**TGTCAA**TTGTTATCCGCTCACAATTATAGGAACTTCACGCTAGGG

oQPCR-araA-fwd	TCGGTTTCTCCGTCAATA
oQPCR-araA-rev	ATGGTGTAGCAGCTTTCG
oQPCR-araB-fwd	GAAAGTGCACGCAGACATCC
oQPCR-araB-rev	ATAGTGTTGTTCGGCGCTCA
c1 (*38*)	TTATACGCAAGGCGACAAGG
c2 (*38*)	GATCTTCCGTCACAGGTAGG
k1 (*38*)	CAGTCATAGCCGAATAGCCT
k2 (*38*)	CGGTGCCCTGAATGAACTGC

Table S4. *E. coli* strains used in this study. Cm: chloramphenicol (10 μg/ml); dT: thymidine (30 μg/ml); Erm: erythromycin (20 μg/ml); IPTG: isopropyl β-D-1-thiogalactopyranoside; Kan: kanamycin (50 μg/ml); Str: streptomycin (50 μg/ml). For adaptation, D- and L-arabinose were present at 0.3% and 0.2% w/v, respectively.

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*Mlu*I)::*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) (*38*)	Cm
DH5α Δ*araBA*	Preliminary tests; used as alternative to DH5α in this study	DH5α Δ*araBA*::*FRT*	Datsenko-Wanner (pCP20) (*38*)	-
MG1655 Δ*araBA*::*cat*	Intermediate for construction of MG1655 Δ*araBA*	MG1655 Δ*araBA*::*cat*	Datsenko-Wanner (pKD46) (*38*)	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) (*38*)	-
MG1655 Δ*araBA* Δ*lacIZYA*::*cat*	Preliminary tests	MG1655 Δ*araBA*::*FRT* Δ*lacIZYA*::*cat*	Datsenko-Wanner (pKD46) (*38*)	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. (*33*–*35*))	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) (*38*)	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) (*38*)	-
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) (*38*)	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 (*39*) (+ IPTG + Cm)
MG1655 Δ*araBA* D-ara$^{+/evo}$ Δ*fucK* Δ*lacIZYA*::*cat* D/L-ara^{evo} Δ*araD*::*kan*^{R}	Δ*araD* derivative of final plasmid library host strain; the toxic intermediate, L-ribulose-5-phosphate, accumulates in the presence of L-arabinose and plasmid-expressed AraA and AraB	MG1655 Δ*araBAD*::*kan*^{R} *fucR*$^{D\text{-}ara}$ D-ara^{evo} Δ*fucK*::*FRT* Δ*lacIZYA*::*cat* D/L-ara^{evo}	Datsenko-Wanner (pKD46) (*38*)	Kan

Table S5. Forward and reverse primer sets for promoter mutagenesis. -35 and -10 RNA polymerase-binding hexamers are in bold. N (italicised) denotes a uniform mix of all 4 bases.

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*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

Data S1. Mutant fitness estimates with their 95% bootstrap confidence intervals and the number of barcodes used for their estimation. Genotype nomenclature is [$P_{LtetO\text{-}1}$-*araA* mutation].[$P_{LlacO\text{-}1}$-*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.

REFERENCES AND NOTES

1. C. R. Scriver, P. J. Waters, Monogenic traits are not simple: Lessons from phenylketonuria. *Trends Genet.* **15**, 267–272 (1999).

2. J. L. Badano, N. Katsanis, Beyond Mendel: An evolving view of human genetic disease transmission. *Nat. Rev. Genet.* **3**, 779–789 (2002).

3. D. N. Cooper, M. Krawczak, C. Polychronakos, C. Tyler-Smith, H. Kehrer-Sawatzki, Where genotype is not predictive of phenotype: Towards an understanding of the molecular basis of reduced penetrance in human inherited disease. *Hum. Genet.* **132**, 1077–1130 (2013).

4. C. A. Argmann, S. M. Houten, J. Zhu, E. E. Schadt, A next generation multiscale view of inborn errors of metabolism. *Cell Metab.* **23**, 13–26 (2016).

5. T. A. Manolio, F. S. Collins, N. J. Cox, D. B. Goldstein, L. A. Hindorff, D. J. Hunter, M. I. McCarthy, E. M. Ramos, L. R. Cardon, A. Chakravarti, J. H. Cho, A. E. Guttmacher, A. Kong, L. Kruglyak, E. Mardis, C. N. Rotimi, M. Slatkin, D. Valle, A. S. Whittemore, M. Boehnke, A. G. Clark, E. E. Eichler, G. Gibson, J. L. Haines, T. F. C. Mackay, S. A. McCarroll, P. M. Visscher, Finding the missing heritability of complex diseases. *Nature* **461**, 747–753 (2009).

6. E. J. O'Brien, J. M. Monk, B. O. Palsson, Using Genome-scale Models to Predict Biological Capabilities. *Cell* **161**, 971–987 (2015).

7. L. Xu, B. Barker, Z. Gu, Dynamic epistasis for different alleles of the same gene. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 10420–10425 (2012).

8. H. Kemble, P. Nghe, O. Tenaillon, Recent insights into the genotype–phenotype relationship from massively parallel genetic assays. *Evol Appl.* **12**, 1721–1742 (2019).

9. M.-J. Favé, F. C. Lamaze, D. Soave, A. Hodgkinson, H. Gauvin, V. Bruat, J.-C. Grenier, E. Gbeha, K. Skead, A. Smargiassi, M. Johnson, Y. Idaghdour, P. Awadalla, Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nat. Commun.* **9**, 827 (2018).

10. R. Schleif, Regulation of the L-arabinose operon of *Escherichia coli*. *Trends Genet.* **16**, 559–565 (2000).

11. J. Ewald, M. Bartl, T. Dandekar, C. Kaleta, Optimality principles reveal a complex interplay of intermediate toxicity and kinetic efficiency in the regulation of prokaryotic metabolism. *PLOS Comput. Biol.* **13**, e1005371 (2017).

12. E. Englesberg, R. L. Anderson, R. Weinberg, N. Lee, P. Hoffee, G. Huttenhauer, H. Boyer, L-arabinose-sensitive, L-ribulose 5-phosphate 4-epimerase-deficient mutants of *Escherichia coli*. *J. Bacteriol.* **84**, 137–146 (1962).

13. J. B. Wolf, E. D. Brodie III, M. J. Wade, *Epistasis and the Evolutionary Process* (Oxford Univ. Press, 2000), pp. 20–38.

14. J. A. G. M. de Visser, T. F. Cooper, S. F. Elena, The causes of epistasis. *Proc. Biol. Sci.* **278**, 3617–3624 (2011).

15. D. Berger, E. Postma, Biased estimates of diminishing-returns epistasis? Empirical evidence revisited. *Genetics.* **198**, 1417–1420 (2014).

16. A. I. Khan, D. M. Dinh, D. Schneider, R. E. Lenski, T. F. Cooper, Negative epistasis between beneficial mutations in an evolving bacterial population. *Science* **332**, 1193–1196 (2011).

17. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segrè, C. J. Marx, Diminishing returns epistasis among beneficial mutations decelerates adaptation. *Science* **332**, 1190–1192 (2011).

18. E. Szathmáry, Do deleterious mutations act synergistically? Metabolic Control Theory provides a partial answer. *Genetics* **133**, 127–132 (1993).

19. H. Kacser, J. A. Burns, The molecular basis of dominance. *Genetics* **97**, 639–666 (1981).

20. D. E. Dykhuizen, A. M. Dean, D. L. Hartl, Metabolic flux and fitness. *Genetics* **115**, 25–31 (1987).

21. A. G. Clark, Mutation-selection balance and metabolic control theory. *Genetics* **129**, 909–923 (1991).

22. L. Perfeito, S. Ghozzi, J. Berg, K. Schnetz, M. Lässig, Nonlinear fitness landscape of a molecular pathway. *PLOS Genet.* **7**, e1002160 (2011).

23. H.-H. Chou, N. F. Delaney, J. A. Draghi, C. J. Marx, Mapping the fitness landscape of gene expression uncovers the cause of antagonism and sign epistasis between adaptive mutations. *PLOS Genet.* **10**, e1004149 (2014).

24. F. Blanquart, G. Achaz, T. Bataillon, O. Tenaillon, Properties of selected mutations and genotypic landscapes under Fisher's Geometric Model. *Evolution* **68**, 3537–3554 (2014).

25. T. N. Starr, J. W. Thornton, Epistasis in protein evolution. *Protein Sci.* **25**, 1204–1218 (2016).

26. J. Otwinowski, I. Nemenman, Genotype to Phenotype Mapping and the Fitness Landscape of the *E. coli lac* Promoter. *PLOS ONE.* **8**, e61570 (2013).

27. L. Keren, J. Hausser, M. Lotan-Pompan, I. Vainberg Slutskin, H. Alisar, S. Kaminski, A. Weinberger, U. Alon, R. Milo, E. Segal, Massively parallel interrogation of the effects of gene expression levels on fitness. *Cell* **166**, 1282–1294.e18 (2016).

28. B. D. Towbin, Y. Korem, A. Bren, S. Doron, R. Sorek, U. Alon, Optimality and sub-optimality in a bacterial growth law. *Nat. Commun.* **8**, 14123 (2017).

29. P. Nghe, M. Kogenaru, S. J. Tans, Sign epistasis caused by hierarchy within signalling cascades. *Nat. Commun.* **9**, 1451 (2018).

30. G. Diss, B. Lehner, The genetic landscape of a physical interaction. *eLife* **7**, e32472 (2018).

31. S. Rozen, H. Skaletsky, Primer3 on the WWW for General Users and for Biologist Programmers. *Methods Mol. Biol.* **132**, 365–386 (2000).

32. D. G. Gibson, L. Young, R.-Y. Chuang, J. C. Venter, C. A. Hutchison III, H. O. Smith, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods* **6**, 343–345 (2009).

33. D. J. Leblanc, R. P. Mortlock, The metabolism of D-arabinose: Alternate kinases for the phosphorylation of D-ribulose in *Escherichia coli* and *Aerobacter aerogenes*. *Arch. Biochem. Biophys.* **150**, 774–781 (1972).

34. D. J. LeBlanc, R. P. Mortlock, Metabolism of D-Arabinose: A New Pathway in *Escherichia coli*. *J. Bacteriol.* **106**, 90–96 (1971).

35. D. J. LeBlanc, R. P. Mortlock, Metabolism of D-Arabinose: Origin of a D- Ribulokinase Activity in *Escherichia coli*. *J. Bacteriol.* **106**, 82–89 (1971).

36. G. Fritz, J. A. Megerle, S. A. Westermayer, D. Brick, R. Heermann, K. Jung, J. O. Rädler, U. Gerland, Single Cell Kinetics of Phenotypic Switching in the Arabinose Utilization System of *E. coli*. *PLOS ONE*. **9**, e89532 (2014).

37. A. Khlebnikov, J. D. Keasling, Effect of *lacY* Expression on Homogeneity of Induction from the P_{tac} and P_{trc} Promoters by Natural and Synthetic Inducers. *Biotechnol. Prog.* **18**, 672–674 (2002).

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

39. L. Katz, Selection of *AraB* and *AraC* Mutants of *Escherichia coli* B/r by Resistance to Ribitol. *J. Bacteriol.* **102**, 593–595 (1970).

40. G. A. Wray, The evolutionary significance of *cis*-regulatory mutations. *Nat. Rev. Genet.* **8**, 206–216 (2007).

41. R. Lutz, H. Bujard, 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).

42. R. K. Shultzaberger, D. S. Malashock, J. F. Kirsch, M. B. Eisen, The Fitness Landscapes of *cis*-Acting Binding Sites in Different Promoter and Environmental Contexts. *PLOS Genet.* **6**, e1001042 (2010).

43. J. B. Kinney, A. Murugan, C. G. Callan Jr., E. C. Cox, Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proc. Natl. Acad. Sci. U.S.A.* **107**, 9158–9163 (2010).

44. R. C. Brewster, D. L. Jones, R. Phillips, Tuning Promoter Strength through RNA Polymerase Binding Site Design in *Escherichia coli*. *PLOS Comput. Biol.* **8**, e1002811 (2012).

45. L. Bintu, N. E. Buchler, H. G. Garcia, U. Gerland, T. Hwa, J. Kondev, R. Phillips, Transcriptional regulation by the numbers: Models. *Curr. Opin. Genet. Dev.* **15**, 116–124 (2005).

46. M. Lagator, T. Paixão, N. H. Barton, J. P. Bollback, C. C. Guet, On the mechanistic nature of epistasis in a canonical *cis*-regulatory element. *eLife* **6**, e25192 (2017).

47. M. Kimura, A. Takatsuki, I. Yamaguchi, Blasticidin S deaminase gene from *Aspergillus terreus* (*BSD*): A new drug resistance gene for transfection of mammalian cells. *Biochim. Biophys. Acta* **1219**, 653–659 (1994).

48. K. S. Sarkisyan, D. A. Bolotin, M. V. Meer, D. R. Usmanova, A. S. Mishin, G. V. Sharonov, D. N. Ivankov, N. G. Bozhanova, M. S. Baranov, O. Soylemez, N. S. Bogatyreva, P. K. Vlasov, E. S. Egorov, M. D. Logacheva, A. S. Kondrashov, D. M. Chudakov, E. V. Putintseva, I. Z. Mamedov, D. S. Tawfik, K. A. Lukyanov, F. A. Kondrashov, Local fitness landscape of the green fluorescent protein. *Nature* **533**, 397–401 (2016).

49. S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, G. Sherlock, Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature* **519**, 181–186 (2015).

50. M. Goldsmith, C. Kiss, A. R. M. Bradbury, D. S. Tawfik, Avoiding and controlling double transformation artifacts. *Protein Eng. Des. Sel.* **20**, 315–318 (2007).

51. C. Pusch, H. Schmitt, N. Blin, Increased cloning efficiency by cycle restriction−ligation (CRL). *Tech. Tips Online.* **2**, 35–37 (1997).

52. P. D. Schloss, S. L. Westcott, T. Ryabin, J. R. Hall, M. Hartmann, E. B. Hollister, R. A. Lesniewski, B. B. Oakley, D. H. Parks, C. J. Robinson, J. W. Sahl, B. Stres, G. G. Thallinger, D. J. Van Horn, C. F. Weber, Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Appl. Environ. Microbiol.* **75**, 7537–7541 (2009).

53. A. Meyerhans, J.-P. Vartanian, S. Wain-Hobson, DNA recombination during PCR. *Nucleic Acids Res.* **18**, 1687–1691 (1990).

54. R. T. Hietpas, J. D. Jensen, D. N. A. Bolon, Experimental illumination of a fitness landscape. *Proc. Natl. Acad. Sci.U.S.A.* **108**, 7896–7901 (2011).

55. E. Dekel, U. Alon, Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436**, 588–592 (2005).

56. T. N. Nguyen, Q. G. Phan, L. P. Duong, K. P. Bertrand, R. E. Lenski, Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of Escherichia coli K12. *Mol. Biol. Evol.* **6**, 213–225 (1989).

57. A. L. Koch, The protein burden of *lac* operon products. *J. Mol. Evol.* **19**, 455–462 (1983).

58. M. Kafri, E. Metzl-Raz, G. Jona, N. Barkai, The cost of protein production. *Cell Rep.* **14**, 22–31 (2016).

59. M. S. Bienick, K. W. Young, J. R. Klesmith, E. E. Detwiler, K. J. Tomek, T. A. Whitehead, The Interrelationship between promoter strength, gene expression, and growth rate. *PLOS ONE.* **9**, e109105 (2014).

60. S. Brroks, A. Gelman, G. L. Jones, X.-L. Meng, Handbook of Markov Chain Monte Carlo, in *Handbooks of Modern Statistical Methods*, G. Fitzmaurice, Ed. (Chapman and Hall/CRC, 2011).

61. E. Firnberg, M. Ostermeier, PFunkel: Efficient, expansive, user-defined mutagenesis. *PLOS ONE.* **7**, e52031 (2012).

62. M. Gossen, H. Bujard, Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. U.S.A.* **89**, 5547–5551 (1992).

63. A. Pant, D. Anbumani, S. Bag, O. Mehta, P. Kumar, S. Saxena, G. B. Nair, B. Das, Effect of LexA on Chromosomal Integration of CTXϕ in *Vibrio cholerae*. *J. Bacteriol.* **198**, 268–275 (2016).