bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

**Authors:** Harry Kemble[1,2], Catherine Eisenhauer[1], Alejandro Couce[1,3], Audrey Chapron[1], Mélanie Magnan[1], Gregory Gautier[4,5], Hervé Le Nagard[1], Philippe Nghe[2], Olivier Tenaillon[1]*.

**Affiliations:**

[1]Infection, Antimicrobials, Modelling, Evolution, INSERM, Unité Mixte de Recherche 1137, Université Paris Diderot, Université Paris Nord, 75018 Paris, France.

[2]École Supérieure de Physique et de Chimie Industrielles de la Ville de Paris (ESPCI Paris), PSL Research University, UMR CNRS-ESPCI CBI 8231, 10 Rue Vauquelin, 75231 Paris Cedex 05, France.

[3]Department of Life Sciences, Imperial College, London, SW7 2AZ, UK.

[4]Centre de Recherche sur l'Inflammation, INSERM, UMRS 1149, 75018 Paris, France.

[5]Laboratoire d'Excellence INFLAMEX, Université Paris Diderot, Sorbonne Paris Cité, 75018 Paris, France.

*Correspondence to: olivier.tenaillon@inserm.fr.

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

**Abstract:** Our ability to predict the impact of mutations on traits relevant for disease and evolution remains severely limited by the dependence of their effects on the genetic background and environment. Even when molecular interactions between genes are known, it is unclear how these translate to organism-level interactions between alleles. We therefore characterized the interplay of genetic and environmental dependencies in determining fitness by quantifying ~4,000 fitness interactions between expression variants of two metabolic genes, in different environments. We detect a remarkable variety of environment-dependent interactions, and demonstrate they can be quantitatively explained by a mechanistic model accounting for catabolic flux, metabolite toxicity and expression costs. Complex fitness interactions between mutations can therefore be predicted simply from their simultaneous impact on a few connected molecular phenotypes.

Despite its centrality to medical and evolutionary genetics, our ability to predict the impact of mutations on even the apparently simplest of organismal traits (*1–8*), let alone complex ones (*9*), remains minimal. Three of the main factors proposed to account for this "missing heritability" (*9*) are: the large number of possible alleles at any locus, each having a potentially different impact on a gene's function; interaction between alleles at different loci (intergenic epistasis), such that their combined effect is not simply the sum of their individual effects; and interaction between genotype and environment, such that different genotypes respond to the environment in different ways (*1–9*). A promising inroad is the increasingly refined characterization of molecular interaction networks enabled by –omics approaches (*10*). Metabolic networks are the best-characterized of these, and are of great practical interest for medicine and engineering, but even for metabolic genes it remains unclear how

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

functional interactions at the molecular level translate to allelic interactions at the level of integrated traits relevant for disease, industry and adaptation (*11*).

We therefore developed an experimental system with which to systematically quantify the fitness interactions occurring between many alleles of two metabolic genes from the same pathway. Further, the design enabled us to probe the dependence of these interactions on environmentally modulated gene expression, a common non-genetic mechanism for the modification of physiological traits (*5*, *12*).

Our system was composed of the genes (*araA* and *araB*) encoding the enzymes responsible for the first two steps of the well-studied *Escherichia coli* L-arabinose-utilization pathway (*13*): L-arabinose isomerase (AraA) and L-ribulokinase (AraB), who together transform the sugar, L-arabinose, into the intermediate, L-ribulose-5-phosphate (Fig. 1A). L-ribulose-5-phosphate enters the pentose phosphate pathway (PPP) of central metabolism via further enzymatic reactions, ultimately supporting cell growth, but like many intermediates (*14*, *15*), its accumulation is toxic, retarding growth (*16*). Environmental modulation of gene expression was achieved by placing each of the two genes under an independent, *trans*-regulated chemically-inducible promoter.

For each promoter, 36 single-base variants were constructed, along with the initial "wildtype" sequence, and combined with all variants of the other promoter (Fig. 1B). The organismal phenotype, competitive fitness, was then measured for the entire set of 1,369 genotypes under three different inducer concentration combinations (Figs. 1C-D). Fitness was measured by tagging the mutant library with unique DNA barcodes (tens to thousands per genotype) (Figs. S1-2), culturing the pooled library for ~30 mean generations, and tracking barcode frequencies over time with Next-Generation Sequencing (Fig. S3). The barcodes act as internal replicates for every genotype, enabling precise fitness estimates at high-throughput

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

(log relative fitness, $F^{rel}$, median standard deviation of 0.0011 for single mutants and 0.0047 for double mutants; Fig. S4).

The overall distribution of fitness effects depended critically on the inducer environment, *ie*. the *trans*-regulatory input (Fig. 2A; Fig. S5A; Data S1). The proportion of beneficial effects varied from 88% in $Env_1$ (median $F^{rel}$ = 0.12) to 51% in $Env_3$ (median $F^{rel}$ = -0.03) and 12% in $Env_2$ (median $F^{rel}$ = -0.12). Further, the correlation of fitness effects between environments ranged from strongly positive ($Env_1$-$Env_3$, Pearson's r = 0.74, $p < 2.2x10^{-16}$) to weakly negative ($Env_1$-$Env_2$, Pearson's r = -0.11, $p = 1x10^{-4}$) (Fig. S5B), demonstrating that fitness in one environment can be an extremely poor predictor of fitness in other environments due simply to expression differences. At the level of individual alleles, all but one had changing patterns of effects across environments (Fig. 2B). In some environments, they were universally beneficial or deleterious across genetic backgrounds, and in others they switched between being beneficial and deleterious depending on the allele at the second promoter. This pervasive and inconsistent variability poses a clear challenge for the prediction of mutation effects.

To further characterize how the effects of mutations in one gene depended on the allele present at the other gene, we computed epistasis (*17*) for all mutation pairs in each environment. Epistasis evaluates quantitatively and qualitatively how the log fitness of a double mutant deviates from the sum of that of the constituent single mutants (Fig. 3A, Fig. S6A). Epistasis was found to be pervasive (89%, 39% and 81% of pairs in $Env_{1-3}$, respectively), heterogeneous and environment-dependent. A trend of antagonism reported for several other systems (*18*) was recovered between pairs of individually beneficial (negative epistasis in 89%, 72% and 100% in $Env_{1-3}$, respectively) and individually deleterious (positive epistasis 100% (1/1), 97% and 98%, respectively) mutations, while interactions between a beneficial and a deleterious mutation could be mostly positive or mostly negative,

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

depending on the environment and on which gene carried the beneficial/deleterious mutation. This epistatic diversity extended to individual mutation pairs, with more than 20% interacting both positively and negatively across environments (Figs. S6B-C). Notably, sign epistasis, an extreme interaction which occurs when the sign of a mutation effect changes in the presence of a second mutation (Fig. 1D), represented 31% of significant interactions in $Env_1$, 17% in $Env_2$ and 34% in $Env_3$.

Confronted with such a variety of interactions, we asked whether they might be understood simply in terms of the quantitative fitness effects of the interacting mutations, as has been found for some other mutation sets (*19*). We found that the effects of individual mutations were weakly predictive of the type and value of epistasis they exhibited with mutations at the second promoter (Fig. 3A scatterplots). In all environments, there was a significantly negative correlation between the sum of individual fitness effects and the value of epistasis (Pearson's r = -0.36, -0.37, -0.51 in $Env_{1\text{-}3}$, respectively; $p < 2.2\text{x}10^{-16}$ for all), a trend of diminishing returns that appears common across experimental systems (*19–22*) (Fig. S7A). However, when the two genes were considered separately, the relationship between individual fitness effects and epistasis was found to be markedly different between *araA* and *araB*: the negative correlation was stronger for $P_{LtetO\text{-}1}$-*araA* mutations being added to existing $P_{LlacO\text{-}1}$-*araB* mutations than for the inverse case (Figs. S7B-C; Pearson's r = -0.67, -0.73, -0.63 in $Env_{1\text{-}3}$, $p < 2.2\text{x}10^{-16}$ for all, *vs*. 0.12, -0.20 and -0.34, $p < 1.6\text{x}10^{-5}$ for all), in which the correlation can even be positive, an extremely rare trend in existing studies (*19*). Moreover, we found that the average trend was in some cases strikingly non-monotonic (Figs. S7B-C), revealing that different alleles of a particular promoter can cause similar fitness changes on their own but interact very differently with alleles at the second promoter.

The relationship between individual mutation effects and epistasis was further complicated by the fact that it could be different for different alleles of the same promoter. For example,

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

in $Env_1$, the numerous beneficial $P_{LtetO-1}$-*araA* mutations caused the average negative trend with $P_{LlacO-1}$-*araB* background fitness, while the rare deleterious $P_{LtetO-1}$-*araA* mutations showed no such trend (Fig. 3B, top panel). For individual $P_{LlacO-1}$-*araB* mutations in $P_{LtetO-1}$-*araA* backgrounds, the relationship was consistently non-monotonic, but had a different average direction for individually beneficial or deleterious alleles (Fig. 3B, bottom panel). Moreover, the trend for a given allele could vary greatly with the environment (Figs. S7B-C). These results demonstrate that genes interacting simply through their common participation in a linear pathway can exhibit complex, allele- and environment-dependent trends of epistasis. The smooth patterns exemplified by Fig. 3B, however, suggest that they may in principle be understood from an underlying phenotypic mechanism.

To this end, we constructed a quantitative model of the metabolic pathway, where fitness results from a balance between the benefit of flux (*23*) and the costs of intermediate toxicity (*14*, *24*, *25*) and AraA and AraB protein expression (*26–28*). Log fitness was computed as $F = \left(\omega + u\varphi - \frac{v}{1/\eta - \varphi}\right)(1 - \theta_A A - \theta_B B)$, with $\omega$ a basal growth rate, $u$ and $v$ terms describing the catabolic benefit and toxicity cost of pathway flux ($\varphi$), $A$ and $B$ the cellular activity of the two enzymes, and θ the cost of enzyme expression. Flux depended on AraA and AraB activities as $\varphi = \frac{1}{1/A + 1/B + \eta}$ (*25*, *29*).

Each promoter mutation was then characterized as a change in the activity (*via* expression) of AraA or AraB. Because most mutations lay outside of the repressor binding sites governing promoter inducibility (Fig. 1B), the fold-change in activity caused by each mutation was kept constant across inducer environments. Parameters describing the fitness function, wildtype activities in the 3 environments and expression effects of individual mutations were then optimized to fit the observed data (Data S2; Fig. S8A).

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The fitted model is in excellent agreement with our data, yielding $r^2$ values of 0.98 between experimental and simulated fitness effects and 0.82 between experimental and simulated epistasis coefficients (Fig. 4A-B; Fig. S8B-C; see Fig. S9 for more minimal models). Notably, the model is capable of recapitulating the diverse and complex trends of epistasis seen in the data. In particular, we find that the non-monotonic relationships between single-mutant fitness and the fitness impact of alleles at the second promoter are well explained by the single mutants lying at two sides of a phenotypic optimum (Fig. 4B). Such overshooting, which is also the cause of sign epistasis (Fig. 4C) (*30*), is relatively common in our dataset, mostly because L-ribulose-5-phosphate toxicity results in an optimum in the flux-fitness relationship (*24*, *25*) (Fig. S10). Two alleles of the same gene may thus result in similar fitness changes individually but cause substantially different expression levels and fluxes, resulting in different responses to mutations at the second gene. This is principally due to enzymes possessing different degrees of flux control on each side of the optimum, with lower levels of one resulting in the second having less control.

The model reveals how the biology underlying a linear pathway can result in heterogeneous, environmentally dependent intergenic interactions. When fitness depends solely on flux (*23*, *25*), the nature of epistasis should be guaranteed by pathway topology alone (*25*). Under the slightly more complex selection pressure resulting from metabolite toxicity (*24*, *25*) and gene expression costs, however, interactions can be both positive and negative. We find that epistatic categories form several localized zones over the fitness landscape, their size and position dependent on the wildtype phenotype, controlled here by the environment (Fig. 4C; Fig. S11). Encouragingly, these zones are generally large and orderly enough to be predictable, but only through knowledge of the underlying landscape and the position of the relevant genotypes within it.

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

The importance of this knowledge becomes immediately apparent when considering the existence of a disease threshold (Fig. 4D). The two alleles shown can lead to disease, but only when they co-occur, and only in one particular environment. The model thus provides a mechanism by which potential physiological defects can be manifested, aggravated or alleviated by particular combinations of alleles and environments (*1–7*, *9*). Insight into intergenic fitness landscapes for other biological systems, and for genes connected by more complex topologies, will be indispensable for progress across medicine, bio-engineering and evolution.

**References and Notes:**

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

2. J. L. Badano, N. Katsanis, Beyond Mendel: an evolving view of human genetic disease transmission. *Nature Reviews Genetics*. **3**, 779–789 (2002).

3. K. M. Dipple, E. R. B. McCabe, Modifier Genes Convert “Simple” Mendelian Disorders to Complex Traits. *Molecular Genetics and Metabolism*. **71**, 43–50 (2000).

4. B. Lanpher, N. Brunetti-Pierri, B. Lee, Inborn errors of metabolism: the flux from Mendelian to complex diseases. *Nature Reviews Genetics*. **7**, 449–459 (2006).

5. 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. *Human Genetics*. **132**, 1077–1130 (2013).

6. J. Vockley, P. Rinaldo, M. J. Bennett, D. Matern, G. D. Vladutiu, Synergistic Heterozygosity: Disease Resulting from Multiple Partial Defects in One or More Metabolic Pathways. *Molecular Genetics and Metabolism*. **71**, 10–18 (2000).

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

8. J. Hou *et al.*, The hidden complexity of Mendelian traits across yeast natural populations (2016), doi:10.1101/039693.

9. T. A. Manolio *et al.*, Finding the missing heritability of complex diseases. *Nature*. **461**, 747–753 (2009).

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

11. L. Xu, B. Barker, Z. Gu, Dynamic epistasis for different alleles of the same gene. *Proceedings of the National Academy of Sciences*. **109**, 10420–10425 (2012).

12. M.-J. Favé *et al.*, Gene-by-environment interactions in urban populations modulate risk phenotypes. *Nature Communications*. **9** (2018), doi:10.1038/s41467-018-03202-2.

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

14. 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 Computational Biology*. **13**, e1005371 (2017).

15. P. J. O'Brien, R. Bruce, *Endogenous Toxins: Targets for Disease Treatment and Prevention* (Wiley-VCH, Hoboken, NJ, USA, 2010), vol. 1.

16. E. Englesberg *et al.*, L-arabinose-sensitive, L-ribulose 5-phosphate 4-epimerase-deficient mutants of Escherichia coli. *Journal of Bacteriology*. **84**, 137–146 (1962).

17. J. B. Wolf, E. D. Brodie III, M. J. Wade, Eds., in *Epistasis and the Evolutionary Process* (Oxford University Press, NY, USA, 2000), p. 10.

18. J. A. G. M. de Visser, T. F. Cooper, S. F. Elena, The causes of epistasis. *Proceedings of the Royal Society B: Biological Sciences*. **278**, 3617–3624 (2011).

19. D. Berger, E. Postma, Biased Estimates of Diminishing-Returns Epistasis? Empirical Evidence Revisited. *Genetics*. **198**, 1417–1420 (2014).

20. 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).

21. H.-H. Chou, H.-C. Chiu, N. F. Delaney, D. Segre, C. J. Marx, Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. *Science*. **332**, 1190–1192 (2011).

22. C. Li, W. Qian, C. J. Maclean, J. Zhang, The fitness landscape of a tRNA gene. *Science*. **352**, 837–840 (2016).

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

24. A. G. Clark, Mutation-selection balance and Metabolic Control Theory. *Genetics*. **129**, 909–923 (1991).

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

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

27. M. Kafri, E. Metzl-Raz, G. Jona, N. Barkai, The Cost of Protein Production. *Cell Reports*. **14**, 22–31 (2016).

28. 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 Genetics*. **10**, e1004149 (2014).

29. H. Kacser, J. A. Burns, The Molecular Basis Of Dominance. *Genetics*. **97**, 639–666 (1981).

30. 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).

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

32. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature 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. *Archives of Biochemistry and Biophysics*. **150**, 774–781 (1972).

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

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

36. G. Fritz *et al.*, 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 Ptac and Ptrc Promoters by Natural and Synthetic Inducers. *Biotechnology Progress*. **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. *Proceedings of the National Academy of Sciences*. **97**, 6640–6645 (2000).

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

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

40. G. A. Wray, The evolutionary significance of cis-regulatory mutations. *Nature Reviews Genetics*. **8**, 206–216 (2007).

41. R. Lutz, Independent and tight regulation of transcriptional units in Escherichia coli via the LacR/O, the TetR/O and AraC/I1-I2 regulatory elements. *Nucleic Acids Research*. **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 Genetics*. **6**, e1001042 (2010).

43. J. B. Kinney, A. Murugan, C. G. Callan, E. C. Cox, Using deep sequencing to characterize the biophysical mechanism of a transcriptional regulatory sequence. *Proceedings of the National Academy of Sciences*. **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 Computational Biology*. **8**, e1002811 (2012).

45. L. Bintu *et al.*, Transcriptional regulation by the numbers: models. *Current Opinion in Genetics & Development*. **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** (2017), doi:10.7554/eLife.25192.

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. *Biochimica et Biophysica Acta*. **1219**, 653–659 (1994).

48. K. S. Sarkisyan *et al.*, Local fitness landscape of the green fluorescent protein. *Nature*. **533**, 397–401 (2016).

49. S. F. Levy *et al.*, 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 Engineering Design and Selection*. **20**, 315–318 (2007).

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

52. P. D. Schloss *et al.*, Introducing mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities. *Applied and Environmental Microbiology*. **75**, 7537–7541 (2009).

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

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

54. R. T. Hietpas, J. D. Jensen, D. N. A. Bolon, Experimental illumination of a fitness landscape. *Proceedings of the National Academy of Sciences*. **108**, 7896–7901 (2011).

55. T. N. M. Nguyen, Q. G. Phan, L. P. Duong, K. P. Bertrand, Effects of carriage and expression of the Tn10 tetracycline-resistance operon on the fitness of Escherichia coli K12. *Molecular Biology and Evolution*. **6**, 213–225 (1989).

56. A. L. Koch, The protein burden of lac operon products. *Journal of Molecular Evolution*. **19**, 455–462 (1983).

57. M. S. Bienick *et al.*, The Interrelationship between Promoter Strength, Gene Expression, and Growth Rate. *PLoS ONE*. **9**, e109105 (2014).

58. S. Brroks, A. Gelman, G. L. Jones, X.-L. Meng, *Handbook of Markov Chain Monte Carlo* (Chapman and Hall/CRC, Boca Raton, Fl, USA, 2011), *Handbooks of Modern Statistical Methods*.

59. E. Firnberg, M. Ostermeier, PFunkel: Efficient, Expansive, User-Defined Mutagenesis. *PLoS ONE*. **7**, e52031 (2012).

60. M. Gossen, H. Bujard, Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences*. **89**, 5547–5551 (1992).

61. A. Pant *et al.*, Effect of LexA on Chromosomal Integration of CTXϕ in Vibrio cholerae. *Journal of Bacteriology*. **198**, 268–275 (2016).

**Acknowledgments:** We thank A. Birgy, A. Decrulle, I. Matic, M. Deyell, A. Soler and D. Mazel for providing genetic material and technical advice, A. Baron, J. Chatel, A. Bridier-Nahmias and the CRI cytometry facility for technical assistance, and L.-M. Chevin, B. Gaut, E. Denamur and C. Landry for critical reading of the manuscript. MiSeq sequencing was performed using equipment provided by the Genetics Department of Bichat-Claude Bernard Hospital. **Funding:** This work was supported by the European Research Council under the European Union's Seventh Framework Programme (ERC grant 310944 to O.T.). H.K. was supported by the Ecole Doctorale Frontières du Vivant (FdV) – Programme Bettencourt.

**Author contributions:** H.E.K., P.N. and O.T. conceived the idea for the experiment; H.E.K. and O.T. designed the experiments; H.E.K., C.E., A.C., A.E.C., M.A.M., G.G. and H.L.N. performed the experiments; H.E.K., P.N. and O.T. performed the analyses; H.E.K., P.N. and

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

O.T. wrote the paper. **Competing interests:** The authors declare no competing interests.

**Data and materials availability:** All genotype fitness estimates, along with their bootstrap 95% CIs and the number of replicates used to compute them, are provided in Supplementary Table 6. Raw and processed sequencing data has been submitted to GEO (accession number GSE115725). Custom code used in this study is available from the authors upon request.

**Supplementary Materials:**

Materials and Methods

Figures S1-S11

Tables S1-S5

References (*31-61*)

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

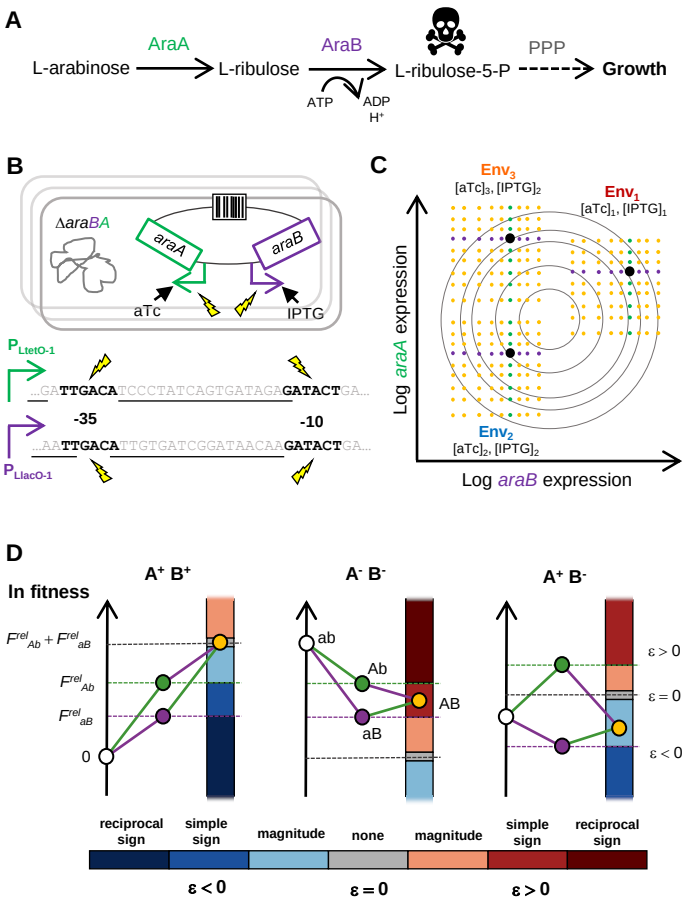


**Fig. 1.** Quantitative mapping of fitness interactions between expression variants of two metabolic genes in expression-modifying environments. (**A**) L-arabinose pathway of *E. coli*. (**B**) *araA* and *araB* were placed under the control of inducible promoters, making their expression sensitive to the concentration of their respective inducers, anhydrotetracycline (aTc) and isopropyl β-D-1-thiogalactopyranoside (IPTG). A barcoded library of mutant promoter combinations was constructed, with mutations targeting the -35 and -10 RNA-polymerase binding hexamers (black letters). Underlined bases are annotated repressor binding sites. (**C**) Competitive fitness was measured under different inducer concentrations defining three environments. $P_{LtetO-1}$ single mutants – green; $P_{LlacO-1}$ single mutants – purple; double mutants - orange. Contours are hypothetical fitness isoclines. (**D**) Epistasis was quantified for all mutant promoter pairs across environments. Epistasis can be categorized as either magnitude or sign type. Sign epistasis is further categorized as simple (effect of one

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

mutation changes sign in presence of the other) or reciprocal (effects of both mutations change sign in the presence of the other). Capitalized letters represent mutant alleles of $_{PLtetO-1}$-*araA* and $P_{LlacO-1}$-*araB*. Superscript plus and minus denote that individual alleles are beneficial or deleterious, respectively.

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

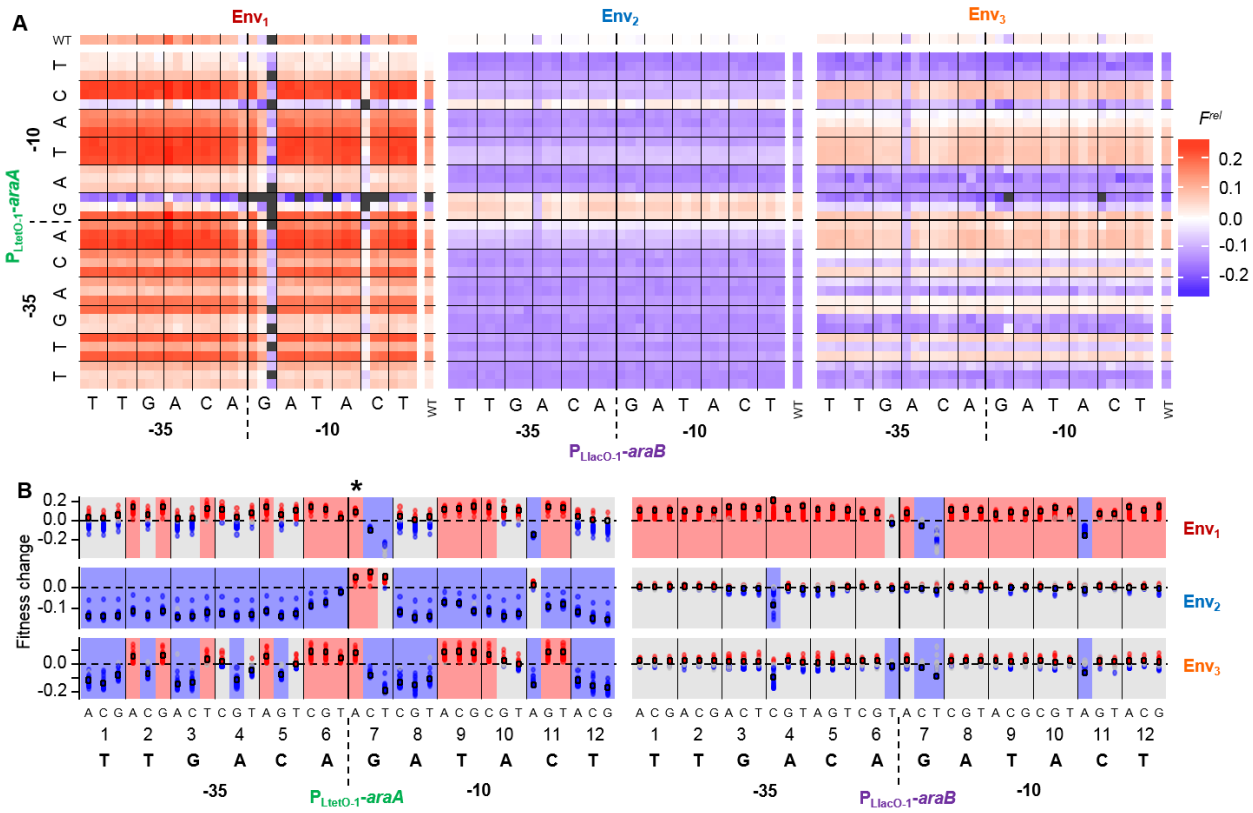


**Fig. 2.** Fitness effects of promoter mutations across backgrounds and environments. (**A**) Genotypes are colored according to the natural logarithm of their fitness relative to the wildtype ($F^{rel}$). Grey denotes unquantifiable fitness effects. Letters show wildtype bases, and the 3 mutations at each position are ordered alphabetically, as in **B**. 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 ($Env_1$); 5 ng/ml aTc and no IPTG ($Env_2$); 200 ng/ml aTc and no IPTG ($Env_3$). (**B**) Fitness changes when an allele of one promoter is added to alleles of the second promoter. Large points indicate the "background" promoter is wildtype. Red, blue and grey points indicate positive, negative and non-significant fitness changes, respectively. Red, blue and grey rectangles indicate, in that environment, an allele can be beneficial but never deleterious, deleterious but never beneficial, or both beneficial and deleterious. G7A of $P_{LtetO-}$

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

$_1$-*araA* (*) is the only allele conferring a qualitatively consistent fitness effect (beneficial) across all backgrounds and environments.

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

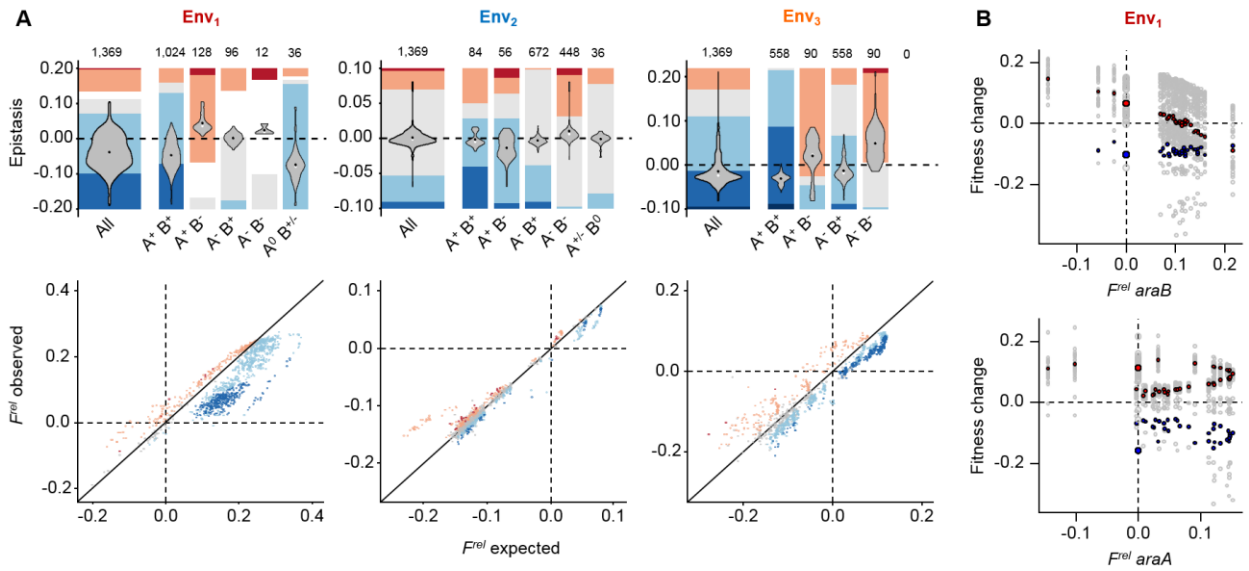


**Fig. 3.** Strength, types and trends of epistasis across environments. (**A**) Violins show epistasis for different kinds of mutation pairs (white point - median; black point - mean). Mutation pairs may contain mutations that are individually both beneficial ($A^+ B^+$), both deleterious ($A^- B^-$) or mixed ($A^+ B^-$ and $A^- B^+$), or one of which confers an undetectable effect ($A^0 B^{+/-}$ and $A^{+/-} B^0$). The number of each such pair is provided. Stacked bars show fractions of different epistasis types (colors as Fig. 1D, with white where epistasis could not be computed). Scatterplots show fitness of double mutants against that expected if mutation effects combined additively. Points colored as in Fig. 1D. (**B**) Relationship between background fitness and the fitness change induced by mutations in the second promoter, in $Env_1$. Top: *araA* promoter mutations added to existing *araB* promoter mutations; bottom: inverse case. Colored points highlight particular alleles. Top: $P_{LtetO-1}$-*araA* alleles T2C (red) and G7C (blue). Bottom: $P_{LlacO-1}$-*araB* alleles T1A (red) and C11A (blue). Large points show effects in the wildtype background.

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

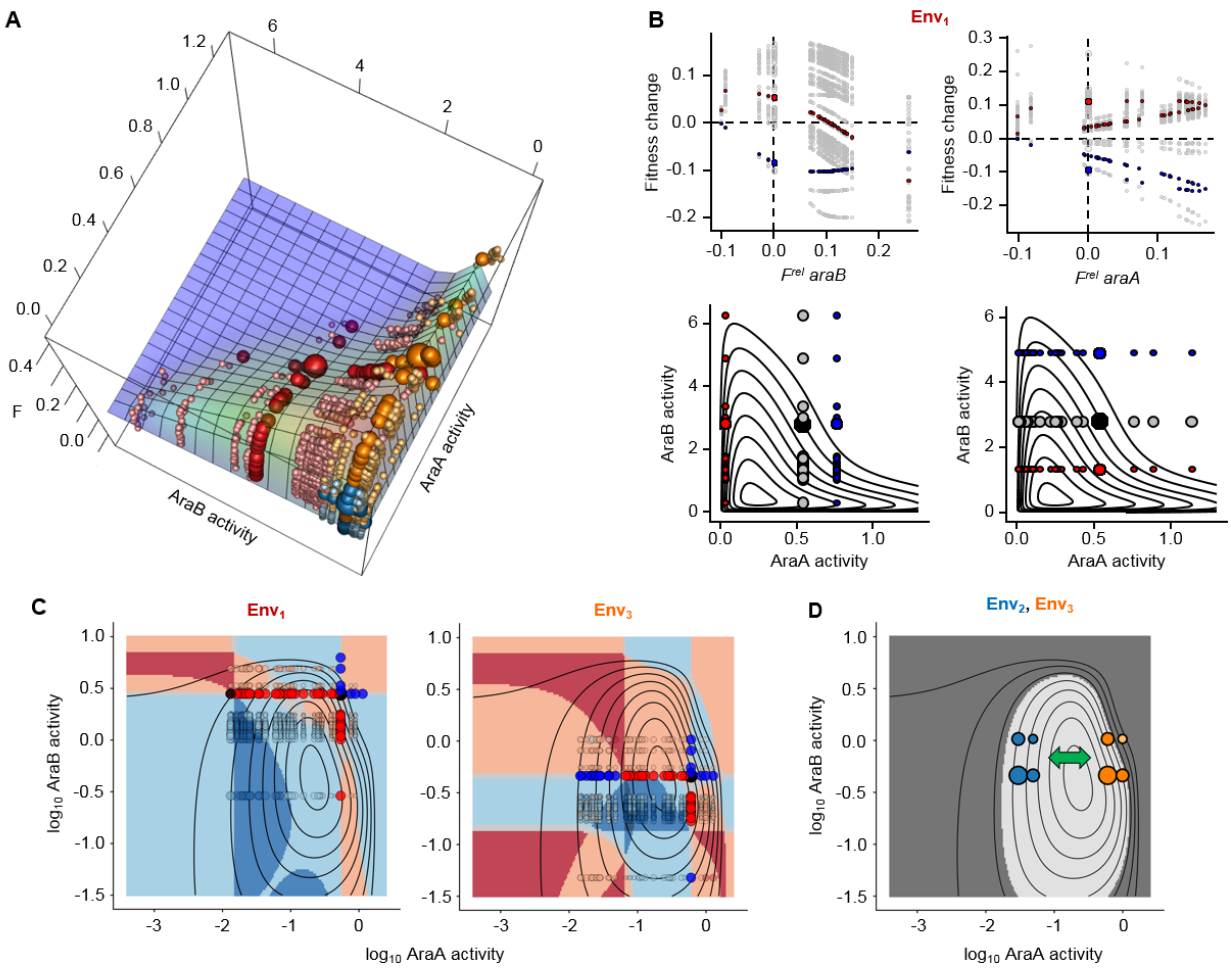


**Fig. 4.** Mechanistic basis of heterogeneous, environmentally dependent epistasis. (**A**) Fitted activity-fitness model. Spheres are positioned according to predicted activity levels and observed $F^{rel}$ ($Env_{1\text{-}3}$ – red, blue, orange). Three largest spheres are wildtype, intermediate-sized spheres are single mutants, small pale spheres are double mutants. (**B**) Upper plots recapitulate Fig. 3B. Lower plots show highlighted genotypes within fitness landscape (black point is wildtype; other large points are single mutants, grey for the gene considered as carrying the "background" alleles). (**C**) Fitness surface on log activity scale, colored by predicted intergenic epistasis type (colors as Fig. 1D; determined as non-significant (grey) if magnitude < 0.005). Large black point is wildtype. Smaller, opaque blue, red and black points are single mutants, colored by observed $F^{rel}$ (deleterious, beneficial and neutral, respectively). Transparent points are double mutants, colored by observed epistasis type and

bioRxiv preprint doi: https://doi.org/10.1101/362327; this version posted July 27, 2018. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-NC-ND 4.0 International license.

sized by epistasis strength. (**D**) Dark grey marks area below a hypothetical disease threshold (40% of maximum fitness). Points are four genotypes in $Env_2$ (blue) and $Env_3$ (orange): wildtype (largest), C11A of $P_{LtetO-1}$-*araA* and G7T of $P_{LlacO-1}$-*araB* (intermediate size), and the resulting double-mutant (smallest). Green arrow represents a change in activity levels caused by non-genetic factors like ageing or environment. A disease state results here from one combination of alleles and environment (pale orange).