Supplementary Information for

On the Deformability of Empirical Fitness Landscapes by Microbial Evolution

Djordje Bajić*, Jean C.C. Vila*, Zachary D. Blount, Alvaro Sanchez

Corresponding authors: Djordje Bajić (djordje.bajic@yale.edu) and Alvaro Sanchez (alvaro.sanchez@yale.edu)

This PDF file includes:

www.pnas.org/cgi/doi/10.1073/pnas.1808485115

Supporting Materials and Methods

Reconstruction of a prokaryotic genotype space.

All *in silico* explorations of genotype (1) space in this work took as a reference the *E. coli* model iJO1366 and consisted of both gene additions and deletions. Gene deletions were performed by constraining both upper and lower bounds of the reaction to zero. Gene additions were performed from a set of all known prokaryotic reactions. We used the BiGG database (1) to compile a dataset of all known reactions found across prokaryotic species. Conflicts in reaction directionality were resolved as follows i) if a reactions is found in the well benchmark *E. coli* iJO1366 model, use the properties given by this model, ii) if a reaction conflicts in directionality, only accept directions found across all models (e.g. if there is one model where a given reaction is irreversible, we set it as irreversible). We used this dataset to create a "universal" metabolic model that included all reactions found in *E. coli* iJO1366 as well as a set of all potential novel reactions. We removed reactions that would lead to erroneous energy-generating cycles using the GlobalFit algorithm (2). The algorithm was constrained to conserve reactions present in the original *E. coli* model. Removing any futile cycles from this "universal" model ensures that there will not be any futile cycles in any subset. The resulting network contains 4999 metabolic reactions and 585 nutrient uptake or sink reactions, of which 2758 and 255 were not found in the original *E. coli* model.

In silico simulation of growth through metabolic modeling.

Dynamic Flux Balance Analysis simulations were performed using the COMETS package ("Computation of Microbial Ecosystems in Time and Space", (3)) and the gurobi optimizer software. For computationally intensive simulations, we used the High Performance Facility at Yale University. For standard (non dynamic) FBA simulations, we used the COBRApy python package (4). Both Dynamic and Standard FBA optimizations were done using the parsimonious algorithm, in which a first optimization is done to maximize biomass yield, and a second one fixes this yield and minimizes total fluxes throughout the network (5). Unless otherwise stated, the default V_{max} was set in dynamic FBA simulations to 10 $mmol \times gr^{-1} \times hr^{-1}$ for all uptake reactions. Inorganic ions and gases where kept at high concentrations and where kept undepleted throughout the simulation (i.e lower bound : -1000 $mmol \times gr^{-1} \times hr^{-1}$, amount of metabolite: 1000 mmol). This was done to constrain our analysis to situations where growth is limited only by uptake of carbon sources. The unbounded nutrients are: ca2_e, cbl1_e, cl_e, co2_e, cobalt2_e, cu2_e, fe2_e, fe3_e, h_e, h2o_e, k_e, mg2_e, mn2_e, mobd_e, na1_e, nh4_e, ni2_e, pi_e, sel_e, slnt_e, so4_e, tungs_e, zn2_e. For the citrate simulation to avoid oxygen, nitrogen or proton limitation uptake was unconstrained by setting the vmax to 1000 $mmol \times gr^{-1} \times hr^{-1}$. Analysis of results was performed using GNU R language (6).

Fitness, environmental effects and deformability measurements.

To measure fitness, we use here (in both experiments and simulations) the Malthusian fitness measure that allows for a quantitative comparison across environments (7). Fitness of mutant M relative to ancestor A is therefore given as $F_M = \log([X'_M/X_M]/[X'_A/X_A])$, where X

and X' represent initial and final densities. For a pair of mutations, deformability can be then measured as $\delta_{ij} = F_{ij}^{(i)} + F_i^{(A)} - (F_{ij}^{(j)} + F_j^{(A)})$, where $F_x^{(y)}$ represents the fitness of genotype *x* in competition with genotype *y*. To compute environmental effects of mutations, the difference in secretion profile of mutants (as shown in Fig. 1C and Fig. S1) is computed for a given released molecule as sign(D) * (log(D)+1) where D is the amount released by the mutant minus that secreted by Ancestral *E. coli*. This log-modulus transformation (8) is applied to help visualization of the generally small differences in released amount, which can be either positive or negative. To measure environmental effect of a mutation (as used in Fig. S2 and Fig. S5), we use the Euclidean distance in the profile of released metabolic byproducts between a mutant and the *E. coli* ancestor using standard Flux Balance Analysis (4).

Simulation of the fitness landscape of aerobic growth on citrate by *E.coli*.

Starting with *E.coli* model iJO1366 we constructed metabolic models of the four mutants necessary to predict the fitness landscape involved in the evolution of aerobic citrate utilization in the Ara3 population of the LTEE. Unlike the LTEE ancestral strain REL606 (and *E. coli* generally), which possess the necessary genes for citrate utilization but do not express them in aerobic conditions, iJO1366 is able utilize both citrate and succinate if these reactions are unbounded (as FBA optimizes precisely regulation). Thus, the ancestral phenotype was recreated by knocking out three reactions CITt7pp (*citT*), SUCCt2_2pp (*dctA*) and SUCCt2_3pp (*dcuA* or *dcuB*). The reactions encoded by the first two genes (*citT* and *dctA*) are known to be involved in the evolution of aerobic growth on citrate in the LTEE whereas *dcuA* and *dcuB* are involved in dicarboxylate uptake in anaerobic conditions and are inactive in aerobiosis (9). This triple knockout represents the pre-citrate *E. coli* ancestor strain. The addition of CITt7pp simulates the promoter capture and consequent aerobic expression of CitT. Similarly, the addition of SUCCt2_2pp is equivalent to the first mutation (aerobic expression of *dctA*). We used dynamic FBA to predict the fitness landscape of these two mutations, calibrating the simulations to reflect the the experimental conditions. This involved i) setting the *in silico* media to reflect DM25 minimal glucose media (0.139mM glucose, 1.7mM citrate). Aerobic condition was simulated by keeping oxygen (o2_e) undepleted. ii) using published parameters pertaining to the physiology of *E. coli* (3) and iii) estimating the initial biomasses of each mutant prior to competition. Initial biomass for citrate simulations was determined using initial plate counts from pairwise competitions experiments (see also Fig. S4). We assume that average cell dry mass is $3.9 * 10^{-13}$ g which is the empirically measured cell dry mass of REL606 the ancestral strain used in the LTEE (10).

E. coli Long-Term Evolution Experiment.

Briefly, twelve populations of *E. coli* B were founded in 1988 from clone REL606. The populations were initially identical, save for half having a mutation that permitted growth on arabinose. (See below.) These have since been evolved in DM25 minimal glucose medium under conditions of daily, 100-fold serial transfer, and incubation at 37°C with 120 rpm orbital shaking. Samples of each population are frozen every 500 generations [38]. DM25 is Davis-Mingioli broth supplemented with 25 mg/L glucose. (Per liter: 7g potassium

phosphate dibasic trihydrate, 2g potassium phosphate monobasic anhydrous, 1g ammonium sulfate, 0.5g sodium citrate, 1mL 10% magnesium sulfate, and 1mL 0.2% thiamine.)

Isolation and Preparation of Test Strains.

ZDB89 is a Cit^{++} clone isolated from the Ara-3 population sample frozen for generation 35,000 during the LTEE. Cit^{-} revertants arise spontaneously from Cit^{+} and Cit^{++} clones due to recombination-mediated collapse of the tandem cit amplification to the ancestral genotype at that locus. We isolated a Cit^{-} revertant, ZDB757, by first passaging ZDB89 in a glucose-only medium for five days. This passaging does not constitute a selection, but nonetheless enriches for Cit^{-} revertants by eliminating the selective penalty for losing the ability to grow on citrate. Passage cultures were spread on LB plates, and Cit^{-} mutants screened for by patching colonies to LB and Minimal Citrate (MC) plates to identify clones that no longer grew on citrate. The Cit^{-} phenotype was confirmed by streaking on Christensen's Citrate Agar. Recombineering with the pKO3 suicide plasmid (11) was used to delete the *dctA* gene from ZDB89 and ZDB757, producing the Cit^{+} *dctA*$^{-}$ and Cit^{-} *dctA*$^{-}$ constructs, ZDB912 and ZDB904, respectively. To permit differentiation of competitors during fitness assays, we isolated Ara^{+} revertants of each of the aforementioned clones and constructs. Briefly, Ara^{-} strains lack the ability to use arabinose, and form red colonies on Tetrazolium Arabinose (TA) plates, while Ara^{+} revertants are mutants with restored ability to grow on arabinose, and form white colonies on TA. The ancestral strain of the Ara-3 population and its descendants are Ara^{-}. We isolated Ara^{+} revertants by plating clone or construct cultures on Minimal Arabinose (MA) plates. Revertants were competed against their Ara^{-} parents to verify marker state neutrality. Clones, constructs, and revertants are listed in Supplementary Table 1. Derivation of constructs and revertants are shown in Fig. S9.

Experimental fitness Assays.

Fitness was assayed in pairwise competitions. Competitors with opposite Ara marker states were inoculated from frozen stocks into 10 mL LB broth, and incubated overnight at 37°C with 120 rpm orbital shaking to permit revival and elimination of traces of glycerol cryoprotectant. To precondition the competitors, each competitor revival culture was then diluted 100-fold in 0.85% saline, and 100 L of the diluted culture used to inoculate 9.9 mL DM25 with ten-fold replication. These culture were grown for 24 hours at 37°C with 120 rpm orbital shaking, after which they were transferred via 100-fold dilution into 9.9 mL volumes of fresh DM25, and grown for another 24 hours under the same conditions. Ten competition cultures were prepared for each competitor pairing by inoculating each 9.9 mL DM25 with 50 L of each preconditioned competitor. A single replicate preconditioning culture of each competitor for each competition was inoculated so that each competition was inoculated from a single preconditioned culture of the competitors. Upon inoculation with the competitors, 100 L of a 100-fold dilution of each was spread on TA to permit enumeration of the initial frequency of each competitor. 100 L of a 1000-fold dilution was also plated for each competition including at least one Cit^{+} or Cit^{++} competitor. Colonies were counted following 48 hours of plate incubation at 37°C. Following 24 hours incubation under the same conditions used for preconditioning, 100 L of 10,000-fold

dilutions of each competition were plated on TA to permit final enumeration of the competitors. 100,000-fold dilutions were also plated for competitions including at least one Cit^{+} or Cit^{++} competitor.

Exploration of deformability in the local mutational neighborhood of *E. coli*.

To systematically analyze the local mutational neighborhood of *E.coli* we construct a set of metabolic models consisting of every viable single and double mutation, considering both additions and deletions from our universal reaction set and using as a reference the *E. coli* iJO1366 model (4389 and 9636050 genotypes, respectively). We removed from this analysis and all subsequent analysis essential genes, sink reactions, diffusion reactions, as well as those genes leading to artifacts (H_2 or CO_2 limitation). We used dynamic flux balance analysis to simulate competition assays of each mutant with its immediate ancestor. We chose to perform our in-silico competitions in anaerobic conditions because in under aerobic conditions FBA incorrectly predicts complete oxidation of glucose at saturating level. The simulations started with an initial glucose concentration of 0.0001mM and assayed co-culture growth during 10hr, a period during which glucose was never exhausted, i.e. growth remained exponential. All simulations were done with the mutant starting at low frequency (1%, 10^{-10} gr. dry cell weight, for 9.9×10^{-9} gr. for the ancestor) in anaerobic glucose minimal media (unless otherwise stated, see detailed parameters in supplement and supplementary tables ST2).

Simulation of long-range fitness landscape deformation.

In order to explore the long-range effects of landscape deformation, we started from a one-step mutation from ancestor *E. coli* model and performed random walks in genotype space by sampling 1024 mutations (without replacement) among both deletions and additions. In addition to previously mentioned artifacts we excluded from this analysis reactions that had led to CO2 or H2 limitation in a unique pair, even if they did not have this effect with other reactions . The pairs are: SHSL2 and SHSL2r ,DHORD_NAD and DHORDi, ENO and HADPCOADH, LEUTA and LLEUDr, P5CRx and PRO1y, in BIGG database notation (1). To prevent irreversible loss of viability, the sampling procedure also ignored all reactions that were essential in a minimal model capable of growing on glucose in anaerobic conditions. The minimal model was built by sequentially removing reactions while possible, following (12)). At regular intervals along the random walk, fitness was measured as before in competition with the mutant and the ancestor (wild-type) using dynamic flux balance analysis (COMETS (3)). To determine the growth rates of genotype in ancestral vs mutant environments we repeated this procedure except at each step, growth rate was measured in the environment provided by the mutant and the ancestor using standard flux balance analysis (COBRApy (4)). These environments were simulated by setting uptake rates for each secreted metabolites to the excretion rate of the respective ancestor.

Computation of null models for growth in the absence of epistasis.

We built a null model for the expected growth rates of mutants, under the assumption that the effect of each mutation on the growth rate is independent. We denote the growth rate

of the ancestral genotype M by g_M, and that of each single mutant i by $g_i=w_i g_M$, where i can represent any of all N possible mutations (in our case N=4181). Here we introduced the parameter $w_i=g_i/g_M$ representing the relative effect of mutation i on the growth rate. If two mutations i and j do not interact with one another, their effects on growth rate are multiplicative: $w_{ij}=w_i w_j$. For a mutant that contains Q mutations relative to the ancestor M, we can thus calculate its growth rate relative to the ancestor as

$$g_Q = \left(\prod_{i=1}^{P} w_i \right) g_M .$$

Long range deformation in an adaptive trajectory.

To provide a mechanistic example of an adaptive long-range deformation with epistasis, we sequentially removed reactions that i) had a detrimental effect, and ii) affected lactate secretion. We reached a 6 step mutant genotype that was unable to use lactate. The subsequent addition of 4 of these removed reactions to this mutant led first to the secretion of lactate (upon addition of LDH, lactate dehydrogenase), and then to the consumption of this lactate only after 3 additional mutational steps (ACKr - acetate kinase, PFL - pyruvate-formate lyase, ATPs - cytosolic ATP synthase). For each mutation along the trajectory (LDH -ATPs-PFL-ACKr,) we measured fitness as before, by simulating the competition of each mutant with its immediate ancestor. The effect of lactate cross-feeding on fitness was assessed by repeating this analysis, albeit with lactate removed from the environment at each step (Fig. S8).

Supporting figures

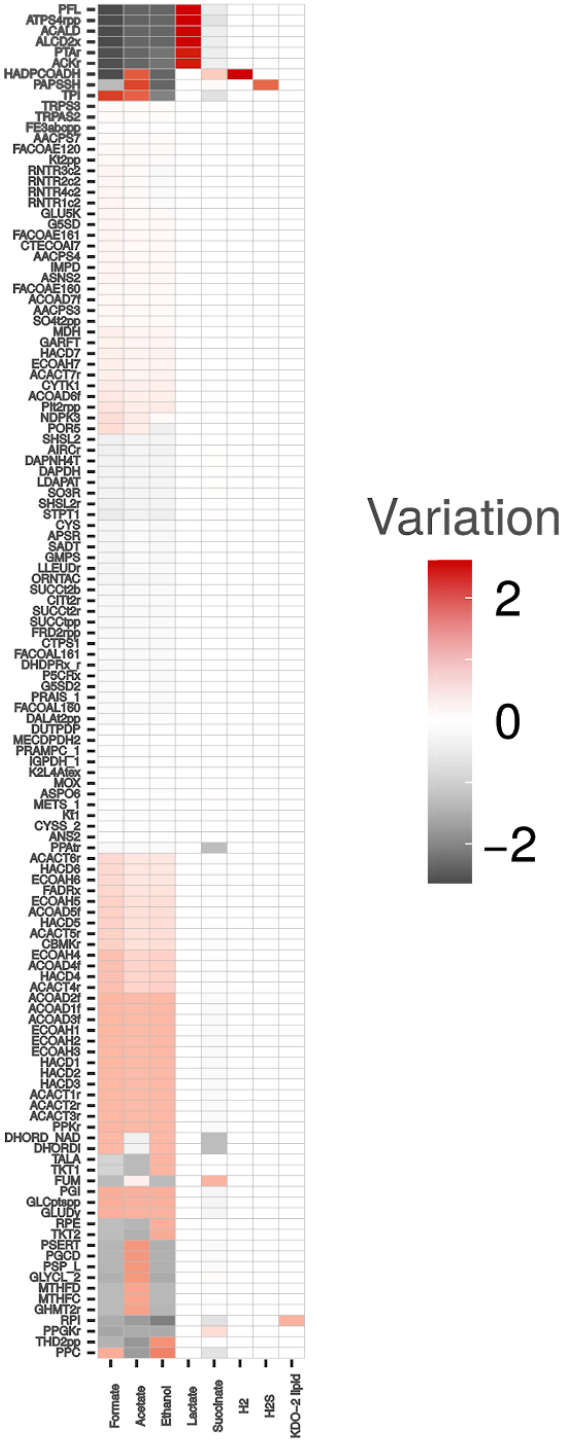

Fig. S1. Variation in the secretion profile of single mutants (see fig. 1C main text)., full set of mutants with environmental effect.

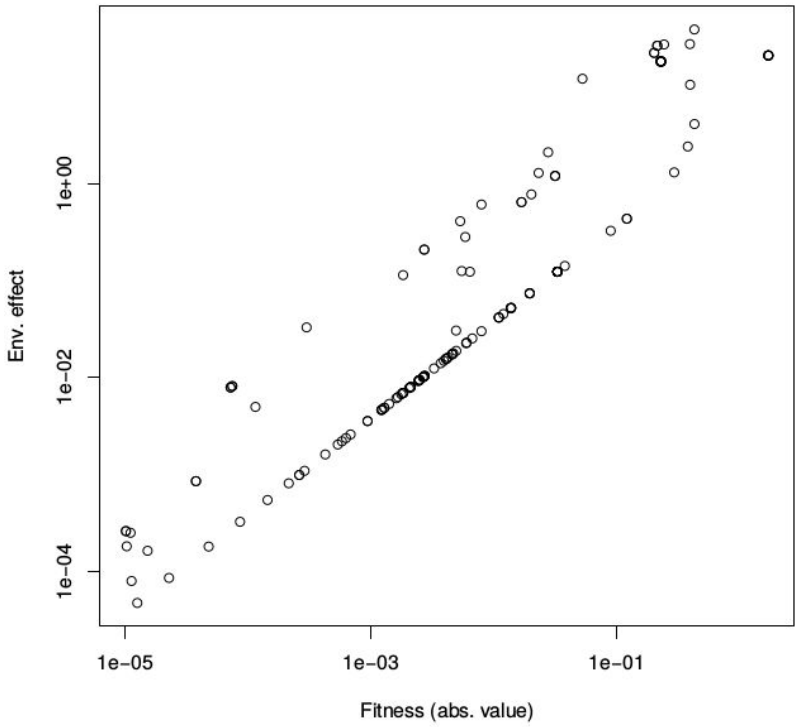

Fig. S2. Fitness and environmental effects are correlated. We plot the environmental effects of each single mutant (calculated as discussed in the methods) as a function of the fitness effect of that mutation. Both metrics are strongly correlated (Pearson's $\rho = 0.61$, $P<10^{-6}$).

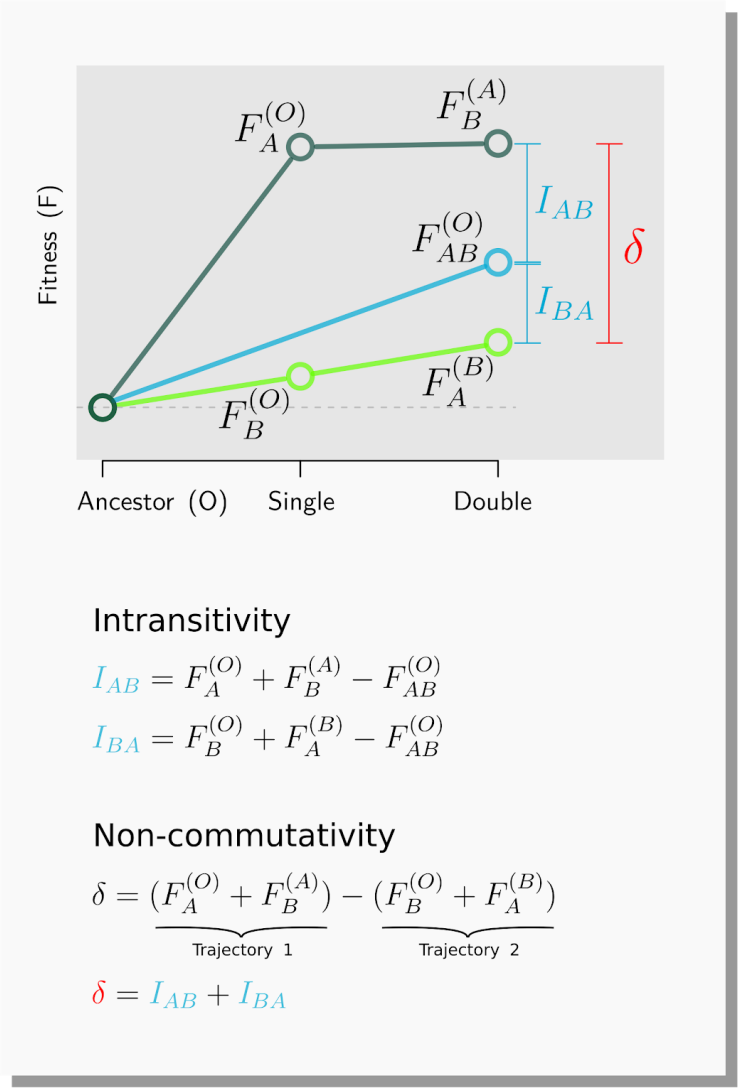

Fig. S3. Relation between non-commutativity (δ) and intransitivity (I) in a hypothetical two mutation genotype space. We denote the fitness of mutant X in competition at low frequency with mutant Y by $F_X^{(Y)}$. As shown in the diagram, noncommutativity (δ) is the sum of the intransitivities in both trajectories (I_{AB} and I_{BA}). However, to compute intransitivity, we need to perform a competition of the double mutant (AB) versus the original ancestor (O).

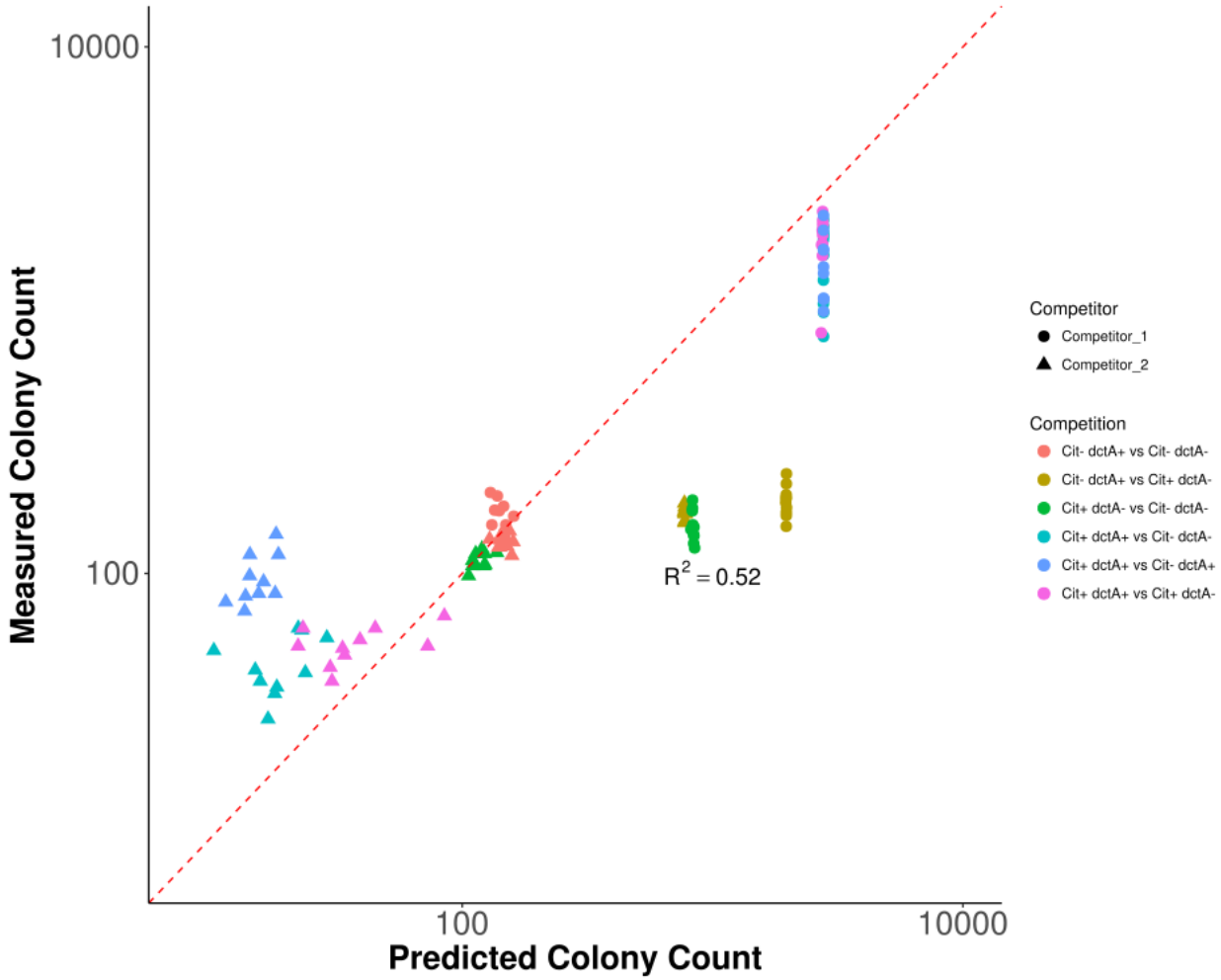

Fig. S4. Prediction of competition assay outcomes in the path to strong aerobic growth on citrate in *E. coli* was compared to the measured colony counts for each competition assay. All 120 competition assays were simulated using dynamic FBA (see Methods), and the experiments were performed as explained in the main text (Methods)

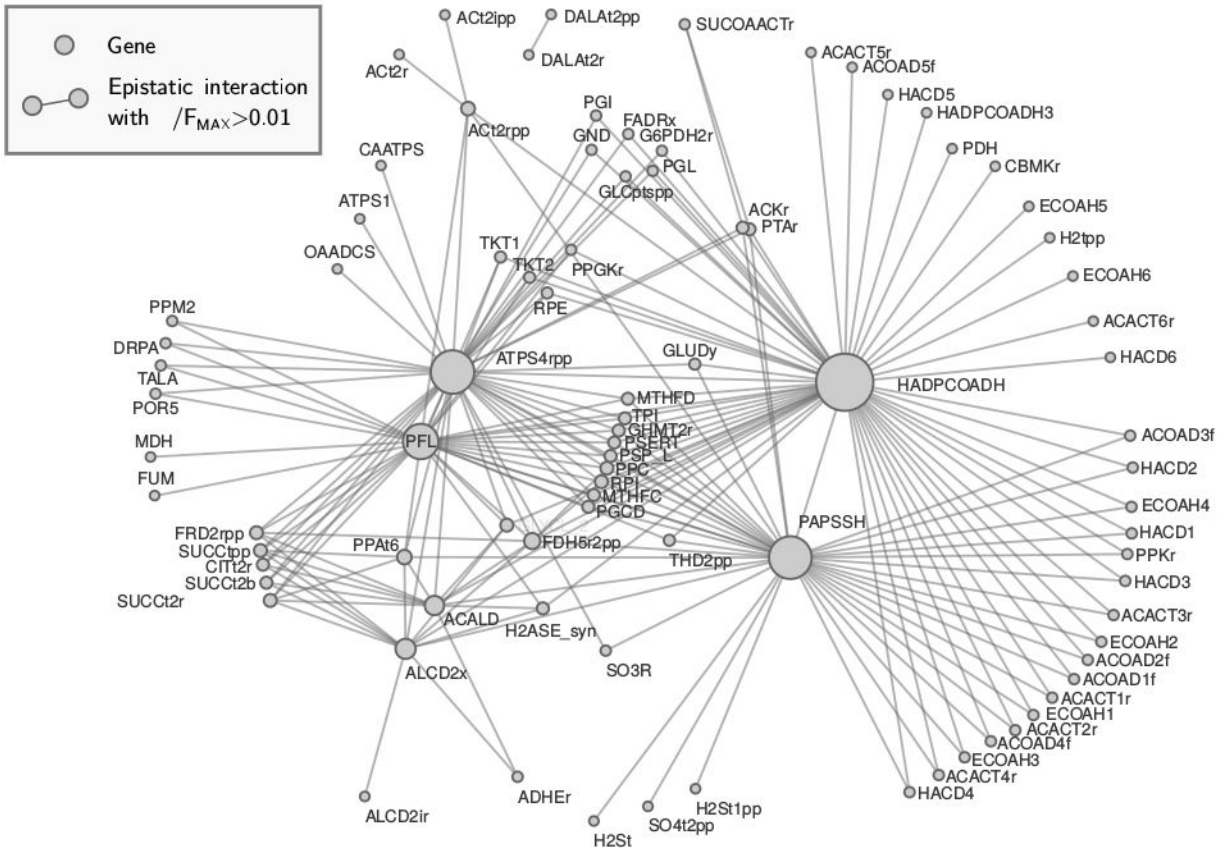

Fig. S5. Same network shown in Fig. 3B showing all reaction labels.

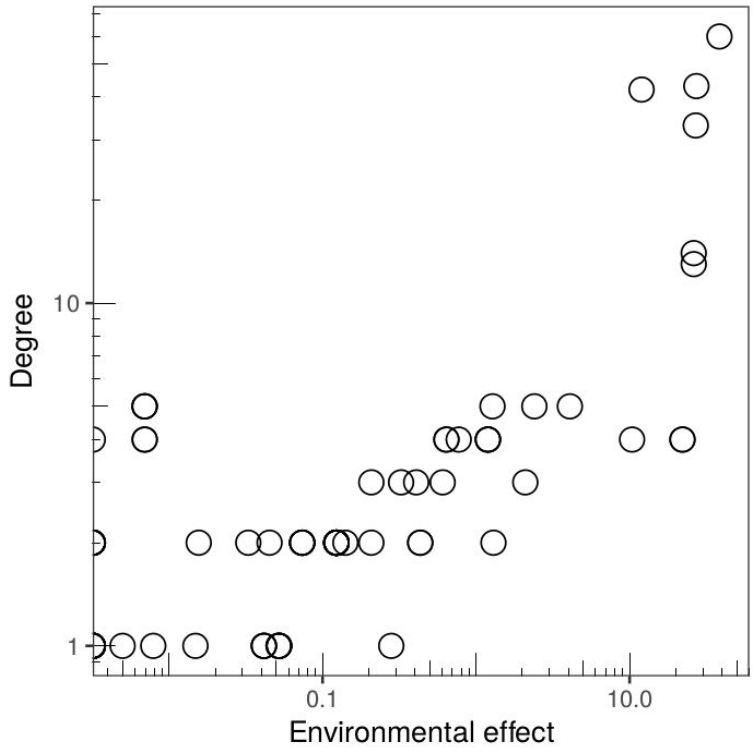

Fig. S6. Environmental effect and deformability are correlated in the local genotype space of *E. coli*. Here, the degree (i.e. number of interactions) in the network presented in figure 3C (main text) is used as a proxy for deformability.

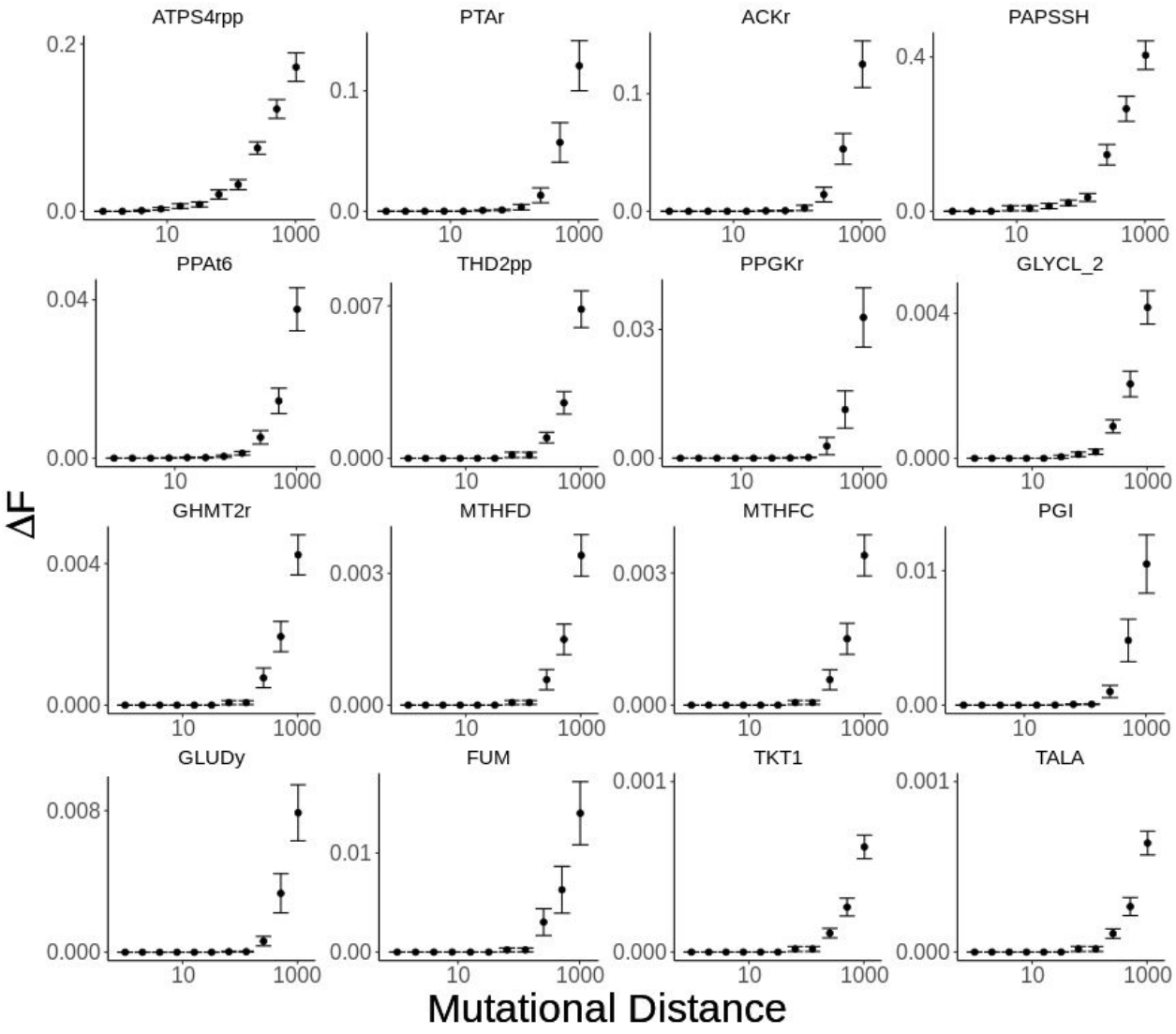

Fig. S7. Additional examples of long range deformation similar to figure 4C (main text) using the 16 non-essential mutations M with largest environmental effect (shown as subpanel titles) other than ACKr (which is shown in Fig. 4C in the main text). We show average fitness differences (ΔF) in competition with A vs. in competition with M. ΔF always increases with mutational distance. Error bars represent SEM (N=100).

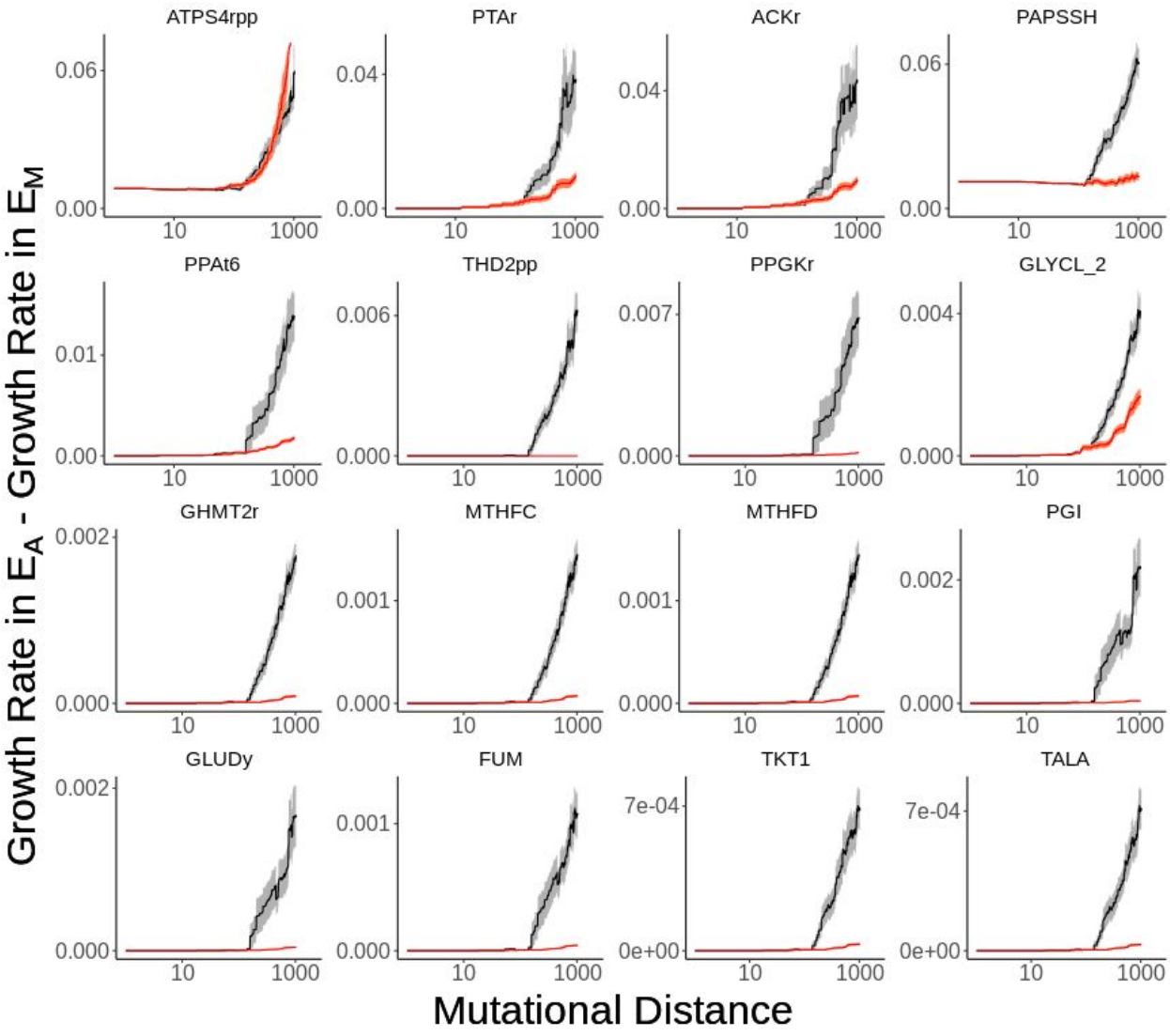

Fig. S8. Average difference (absolute value) in growth rate between environments E_M and E_A (in grams of dry cell weight × hr^{-1}) at varying genotype distances (gray line, shading represents SEM; N=100). Additional examples are shown, similar to figure 4D (main text), using the environments generated by the 16 non-essential mutations with largest environmental effect. Average difference in growth across environments always increases with mutational distance. Gray shading is SEM (N=100).

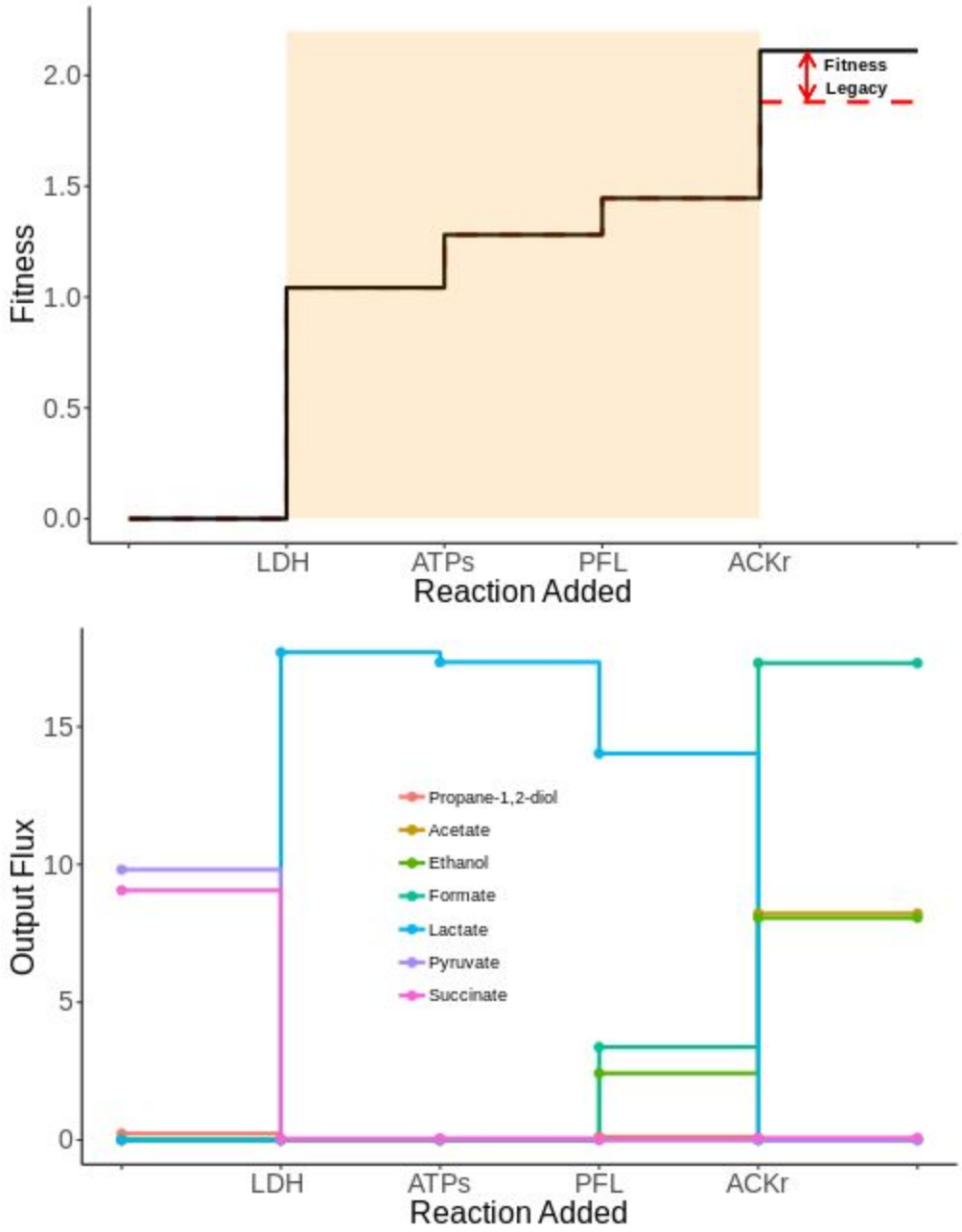

Fig. S9. Breakdown of the example of long range effects in an adaptive trajectory given in Fig. 4. In the top panel we show the incremental fitness increase of each mutant as predicted by competition with it's immediate ancestor. The dotted red line shows the fitness predicted when excreted lactate is removed from the environment. Whilst lactate production only requires a single mutation, this environmental change does not affect the fitness of immediate descendants and instead leaves a 'legacy' (shaded region) that persists and requires multiple interacting mutations to be 'felt'. In the bottom panel we show the FBA predicted output flux of secondary metabolites when glucose is in excess (i.e uptake rate = 10 mmol $\times$ gr^{-1} $\times$ hr^{-1}).

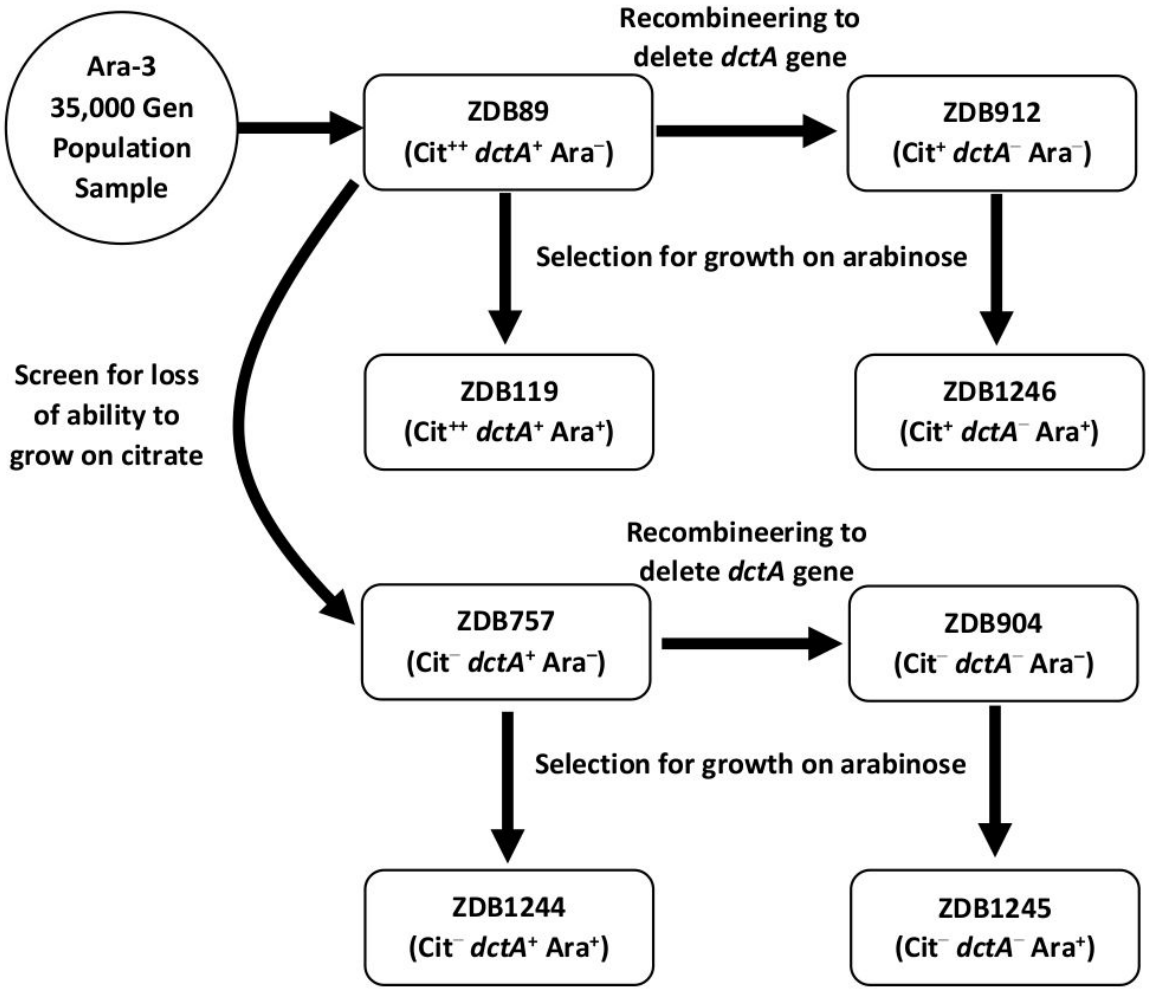

Fig. S10. Derivation of Ara-3 strains used in competition experiments.

Table ST1. Ara-3 strains used in competition experiments.

Clone	**Phenotype**	***citT* Duplication**	***dctA* Mutation**	**Ara marker**
ZDB89	**Cit++**	+	+	–
ZDB119	**Cit++**	+	+	+
ZDB757	**Cit–**	–	+	–
ZDB904	**Cit–**	–	–	–
ZDB912	**Cit+**	+	–	–
ZDB1244	**Cit–**	–	+	+
ZDB1245	**Cit–**	–	–	+
ZDB1246	**Cit+**	+	–	+

Table ST2. Parameters used in dFBA simulations

Parameters	**Default Parameters (Fig 1,3 and 4)**	**Citrate Simulations (Fig2)**
Default V_{max} (mmol × gr.$^{-1}$ × hr^{-1})	10	10
Default uptake Km (uM)	0.01	0.01
Death rate (%)	0	0.01
Time Step (hr)	0.1	0.1
Cycles	100	240
Space Width (cm)	0.02	0.02
Simulation layout	Single cell	Single cell

References

1. King ZA, et al. (2016) BiGG Models: A platform for integrating, standardizing and sharing genome-scale models. *Nucleic Acids Res* 44(D1):D515–22.

2. Fritzemeier CJ, Hartleb D, Szappanos B, Papp B, Lercher MJ (2017) Erroneous energy-generating cycles in published genome scale metabolic networks: Identification and removal. *PLoS Comput Biol* 13(4):e1005494.

3. Harcombe WR, et al. (2014) Metabolic resource allocation in individual microbes determines ecosystem interactions and spatial dynamics. *Cell Rep* 7(4):1104–1115.

4. Ebrahim A, Lerman JA, Palsson BO, Hyduke DR (2013) COBRApy: COnstraints-Based Reconstruction and Analysis for Python. *BMC Syst Biol* 7:74.

5. Lewis NE, et al. (2010) Omic data from evolved E. coli are consistent with computed optimal growth from genome-scale models. *Mol Syst Biol* 6:390.

6. R Core Team (2017) *R: A language and environment for statistical computing* (R Foundation for Statistical Computing) Available at: https://www.R-project.org/.

7. Wagner GP (2010) The measurement theory of fitness. *Evolution* 64(5):1358–1376.

8. John JA, Draper NR (1980) An Alternative Family of Transformations. *J R Stat Soc Ser C Appl Stat* 29(2):190–197.

9. Six S, Andrews SC, Unden G, Guest JR (1994) Escherichia coli possesses two homologous anaerobic C4-dicarboxylate membrane transporters (DcuA and DcuB) distinct from the aerobic dicarboxylate transport system (Dct). *J Bacteriol* 176(21):6470–6478.

10. Gallet R, et al. (2017) The evolution of bacterial cell size: the internal diffusion-constraint hypothesis. *ISME J* 11(7):1559–1568.

11. Link AJ, Phillips D, Church GM (1997) Methods for generating precise deletions and insertions in the genome of wild-type Escherichia coli: application to open reading frame characterization. *J Bacteriol* 179(20):6228–6237.

12. Pál C, et al. (2006) Chance and necessity in the evolution of minimal metabolic networks. *Nature* 440(7084):667–670.