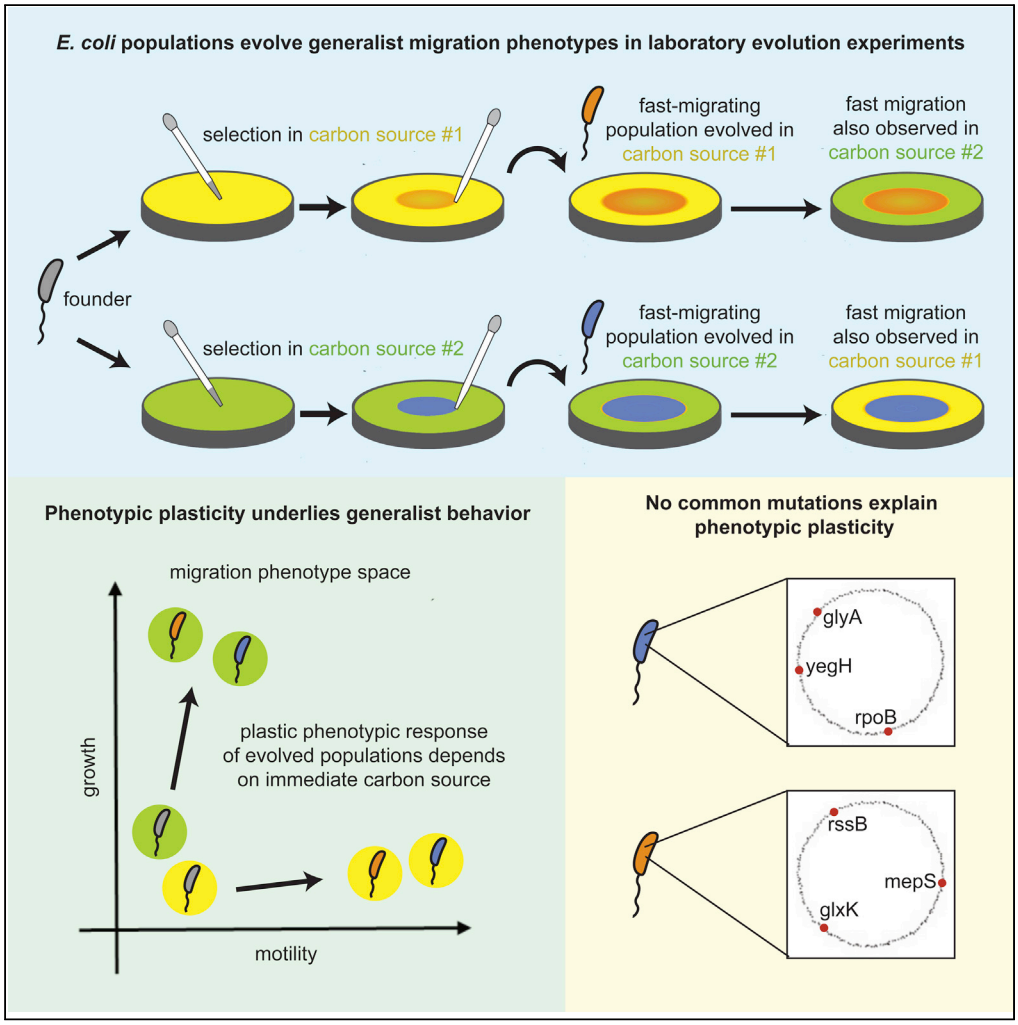
iScience

CellPress
OPEN ACCESS

Article

Evolution of Generalists by Phenotypic Plasticity



David T. Fraebel, Karna Gowda, Madhav Mani, Seppe Kuehn

seppe.kuehn@gmail.com

HIGHLIGHTS

Bacteria evolve fast migration in porous media by adapting growth and motility

Selection in one nutrient condition results in fast migration in other conditions

Fast migrating generalists evolve by plasticity in growth and motility

Selection in homogeneous environments can evolve generalist phenotypes

Fraebel et al., iScience 23, 101678
November 20, 2020
https://doi.org/10.1016/j.isci.2020.101678

iScience

CellPress
OPEN ACCESS

Article

Evolution of Generalists by Phenotypic Plasticity

David T. Fraebel,[1,2] Karna Gowda,[3,4] Madhav Mani,[5,6,7] and Seppe Kuehn[3,4,8,*]

SUMMARY

Adapting organisms face a tension between specializing their phenotypes for certain ecological tasks and developing generalist strategies that permit persistence in multiple environmental conditions. Understanding when and how generalists or specialists evolve is an important question in evolutionary dynamics. Here, we study the evolution of bacterial range expansions by selecting *Escherichia coli* for faster migration through porous media containing one of four different sugars supporting growth and chemotaxis. We find that selection in any one sugar drives the evolution of faster migration in all sugars. Measurements of growth and motility of all evolved lineages in all nutrient conditions reveal that the ubiquitous evolution of fast migration arises via phenotypic plasticity. Phenotypic plasticity permits evolved strains to exploit distinct strategies to achieve fast migration in each environment, irrespective of the environment in which they were evolved. Therefore, selection in a homogeneous environment drives phenotypic plasticity that improves performance in other environments.

INTRODUCTION

Organisms in nature often encounter varied environments throughout their lifetimes, each with its own demands on the phenotype. Therefore, the ability of a single genotype to thrive under different environmental conditions can be essential for a lineage's chance of long-term persistence. Organisms that can thrive in varied environments are called generalists. It is thought that generalists evolve in fluctuating environments, while fixed environments select for specialists (Kassen, 2002). In its native environment, a specialist is expected to have enhanced fitness relative to a generalist but reduced fitness in other environments (Elena and Lenski, 2003).

However, the results of experimental evolution studies show remarkably diverse outcomes. While trade-offs are sometimes observed (Cooper and Lenski, 2000), they are far from universal: studies with a constant selection environment (which are expected to produce specialists) often produce a mixture of specialists and generalists across replicate lineages (Velicer and Lenski, 1999). Additionally, trade-offs observed in experimental evolution are sometimes asymmetric: studies employing multiple selection conditions may produce specialists in one condition and generalists in another (Lee et al., 2009; Travisano, 1997). Other studies find that a majority of evolved lines are generalists, even with selection performed in a constant environment (Ostrowski et al., 2005; Fong et al., 2005).

A key question that emerges from these studies is understanding the limits and mechanisms of the evolution of phenotypic generality. Experiments suggest that generalists will only show enhanced fitness in environments that are sufficiently similar to their selection condition (Travisano et al., 1995; Travisano and Lenski, 1996). While this notion is intuitive, there are distinct avenues of phenotypic adaptation that could give rise to generalist phenotypes. An experimentally supported interrogation of the conditions under which generalists evolve, the environmental limits of their generality, as well as its underlying genotypic and phenotypic mechanisms, would shed light on how populations deploy phenotypic variation during adaptation (Pigliucci, 2001).

To address this, we selected *Escherichia coli* for faster migration through porous media with one of four different sugars as a carbon source and chemoattractant. We observed the evolution of generalists: selection for fast migration in any one sugar resulted in fast migration in all sugars. Because migration depends on both growth and motility, we measured these phenotypes for strains from each evolutionary history in all

[1]Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[2]Center for the Physics of Living Cells, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA

[3]Center for the Physics of Evolving Systems, University of Chicago, Chicago, IL 60637, USA

[4]Department of Ecology and Evolution, University of Chicago, Chicago, IL 60637, USA

[5]Department of Molecular Biosciences, Northwestern University, Evanston, IL 60208, USA

[6]NSF-Simons Center for Quantitative Biology, Northwestern University, Evanston, IL 60208, USA

[7]Department of Engineering Sciences and Applied Mathematics, Northwestern University, Evanston, IL 60208, USA

[8]Lead Contact

*Correspondence: seppe.kuehn@gmail.com

https://doi.org/10.1016/j.isci.2020.101678

four nutrient conditions to investigate the evolution of migration rate generality. We found that the environment determined the phenotype more than the evolutionary history: strains from all selection conditions exhibited a distinct growth and motility phenotype for each assay condition irrespective of their evolutionary history. That is, all strains observed in a particular environment showed similar adaptation of growth and motility regardless of their selection condition. Because the evolved strains in this study exhibit different phenotypes depending on their environment and because those phenotypes are associated with faster migration, we conclude that generalists evolved in our experiment via phenotypic plasticity (Scheiner, 1993).

RESULTS

Selection Enhances Migration of Bacterial Populations through Soft Agar

E. coli inoculated into low-viscosity agar depletes nutrients locally, as cells swim and divide in the porous, three-dimensional environment. This depletion establishes a nutrient gradient that drives chemotaxis outward and subsequent growth of the population (Adler, 1966; Wolfe and Berg, 1989). The result is a macroscopic colony that expands radially from the site of inoculation at a speed determined by growth, motility, and chemotaxis of its constituent cells (Croze et al., 2011). We performed time-lapse imaging on these colonies and observed an initial growth phase followed by radial expansion at a constant rate (Figure S1).

We performed experimental evolution by selecting *E. coli* for faster migration through soft agar. From a single ancestral strain (MG1655-motile, Coli Genetic Stock Center #8237), we performed selection for faster migration in M63 minimal medium with 0.2% w/v agar and one of four different sugars as carbon sources at 1 mM concentration: mannose, melibiose, *N*-acetylglucosamine (NAG), and galactose. After allowing a population to expand for 24 hr, we selected a small population of cells from the outermost edge of the colony and used these cells to inoculate a new low-viscosity agar plate (Figure 1A). This process of expansion and selection was repeated for ten rounds. In each condition, the sugar served as the sole energy source and chemoattractant for the expanding population. We chose these four sugars because we believed that the diverse set of genetic architectures involved in their chemotaxis, import, and metabolism could lead to diverse opportunities for genomic evolution and resultant phenotypic adaptation in each condition. All four sugars traverse the outer membrane through *OmpF*, but mannose and NAG use phosphotransferase systems to cross the inner membrane while melibiose and galactose rely on cation symporters *MelB* and *GalP* (Travisano and Lenski, 1996; Yazyu et al., 1984; Henderson et al., 1977). Once inside the cell, catabolism of NAG and galactose is regulated by repression from *NagC* and *GalR/GalS* (Plumbridge, 1991; Weickert and Adhya, 1993). Alternatively, melibiose catabolism is regulated by activation from *MelR* (Bourgerie et al., 1997) and mannose catabolism does not have a specific inducer, requiring only the global regulator cyclic AMP receptor protein to signal glucose starvation (Fraser and Yamazaki, 1980).

Across all four conditions, we observed a dramatic enhancement of migration rates due to selection. Migration rates of all lineages in melibiose and NAG more than doubled after just one round of selection (Figures 1C and 1D), while similar improvement in mannose and galactose was achieved after two rounds of selection (Figures 1B and 1E). The reason for the small drop in migration rates between rounds 1 and 2 in mannose is not known but might be due to the different growth history of the cells used at the beginning or rounds 1 and 2 (liquid culture and agar plates, respectively). In all four conditions, migration rates continued to improve from round 2 to round 10, albeit modestly compared to the initial increase. In round 10, migration rates had increased almost threefold in mannose and NAG, 2.5-fold in melibiose, and nearly 6-fold in galactose.

Selection for Fast Migration in One Sugar Results in Fast Migration in all Sugars

We asked whether migration rate evolution was specific to each selection condition or whether strains evolved in one sugar were adapted for fast migration in other sugars as well. Therefore, we measured the migration rates of evolved strains in all nutrient conditions. Specifically, the migration rates were measured for four independent replicate lineages from each selection condition (a total of 16 strains) in each of the four different assay conditions (mannose, melibiose, NAG, and galactose). Surprisingly, nearly all strains exhibited enhanced migration across all conditions relative to the founding strain in that condition (Figure 2). For example, a strain evolved for fast migration in NAG (squares, Figure 2) also exhibited faster migration than the ancestral strain in all three of the other sugars. Migration rates of non-native strains (that is, populations assayed in a condition different from their selection condition) were nearly always much greater than the ancestor's migration rate, often comparable to or exceeding the natively

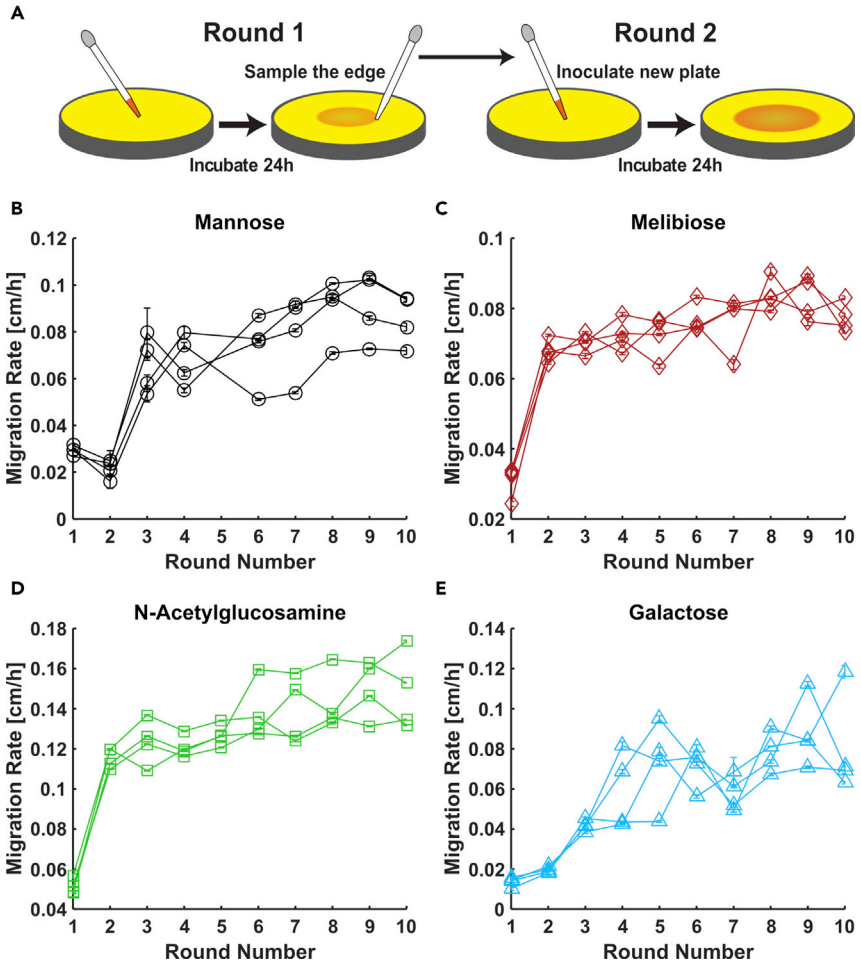


Figure 1. Repeated Selection Enhances *E. Coli* Migration through Soft Agar in Four Nutrient Conditions

(A) Schematic of migration selection procedure. Motile *E. coli* is inoculated into the bulk of a soft agar plate containing growth medium. Nutrient consumption and chemotaxis drive the growing population to expand radially across the plate at a constant migration rate. After a fixed interval, cells are sampled from the edge of the expanding colony and used to inoculate a new plate.

(B–E) Migration rates as a function of round of selection for experiments conducted in 0.2% w/v agar plates containing M63 minimal medium with one of four different carbon sources at 1 mM concentration: (B) mannose, (C) melibiose, (D) *N*-acetylglucosamine, and (E) galactose. Four replicate experiments were carried out to 10 rounds in each condition. Migration rates are measured by time-lapse imaging of expanding colonies followed by linear fit of colony radius versus time; error bars are 95% confidence intervals on fitted rates. No rates are reported for round 5 in mannose due to failure of the imaging device. See Transparent Methods for details of selection procedure and migration rate measurement. See also Figure S1.

evolved populations' migration rates. The non-native populations exhibit fast migration, despite having no evolutionary history in that nutrient condition. The objective of this study therefore became to understand how selection in one nutrient environment generically gives rise to fast migration in all nutrient environments.

To probe the limits of the observed phenotypic generality, we first measured the migration rates of the ancestor as well as two evolved strains from each selection condition in seven additional nutrient environments beyond the initial four: arabinose, dextrose, fructose, lactose, maltose, rhamnose, and sorbitol. Again, we found that nearly all the strains showed enhanced migration rates relative to the ancestral strain across all these conditions (Table S1). The evolved strains typically migrated 1.5- to 3-fold faster than the ancestor. We concluded that the nutrient generality of migration rate selection extends to

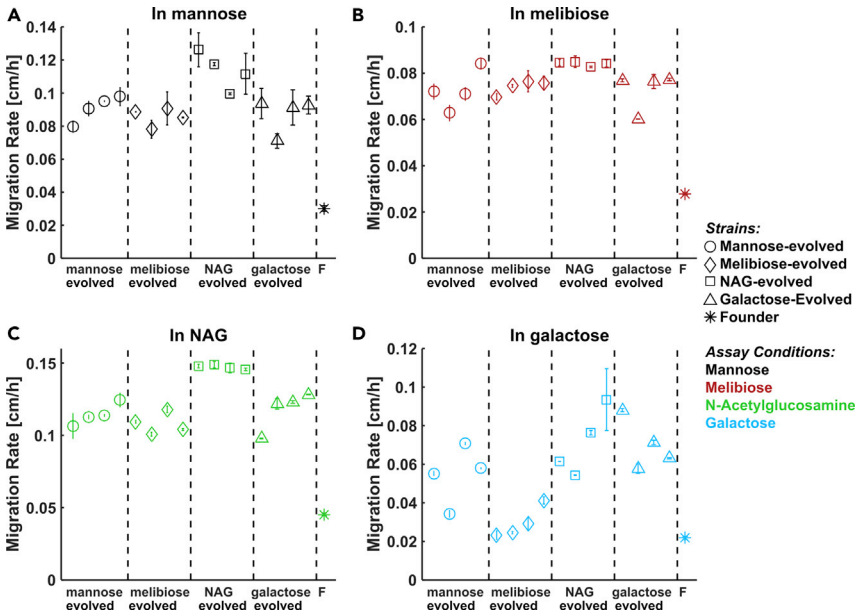


Figure 2. Nutrient Generality of Migration Rate Evolution

The 16 strains isolated after 10 rounds of selection (four from each nutrient condition, Figures 1B–1E) were assayed for enhanced migration rate in each of the four nutrient conditions used for the selection experiments: (A) mannose, (B) melibiose, (C) NAG and (D) galactose. Migration rates of evolved strains are presented as mean ± standard error of two replicate plates for each strain in each condition, except as noted below. Marker shapes denote selection condition; marker colors denote assay condition (legend). Migration rates of the founding strain (F) in each condition are presented as mean ± standard error of 3–4 replicate plates in each condition. For the following exceptions, error bars are standard error on migration rate from a single plate: The rightmost NAG-evolved strain and leftmost gal-evolved strain had only one usable plate in in NAG (C). For the rightmost two galactose-evolved strains in galactose (D), rates are reproduced from round 10 of the selection experiments (Figure 1E). See also Figures S2 and S3 and Table S1.

many different carbon sources within the regime of M63 medium with 0.2% w/v agar and 1 mM sugar. In contrast, migration rates of all 16 evolved strains presented in Figure 2 did not exhibit fast migration in lysogeny broth (LB) rich medium and instead exhibited a drop in migration rate relative to the ancestor (Figure S2). For comparison, we also measured three independently evolved strains isolated after 10 rounds of selection in LB and found that they migrate around 50% faster than the ancestor (Figure S3). So, while repeated selection still enhances migration in LB, the generality of the strains evolved in minimal medium does not extend to this rich medium where chemotaxis and growth are driven by amino acids (Adler, 1966).

What Makes a Generalist?

We next set out to understand how phenotypes evolved under selection for fast migration in one sugar give rise to fast migration in other sugars. Population-level migration through soft agar depends on both growth and motility of individual cells. Therefore, selection for faster migration can be driven by enhancements to growth rate, chemotactic response, and undirected motility (Croze et al., 2011; Fraebel et al., 2017). For example, increases in running speed or tumble frequency are known to drive faster migration through soft agar (Ni et al., 2017). We therefore asked what phenotypic changes permit a lineage selected on one sugar to migrate more rapidly than the ancestral population in another sugar (Figure 2).

To understand why strains from distinct evolutionary histories (selection conditions) show enhanced migration across different environments (assay conditions), we considered the growth and motility of each evolved strain in each assay condition. For the present discussion, we considered growth and motility as abstract traits, but our hypotheses do not depend on the number or identity of the traits considered. In this framework, we proposed that there are three main possibilities for how the generality of fast migration could evolve (Figure 3).

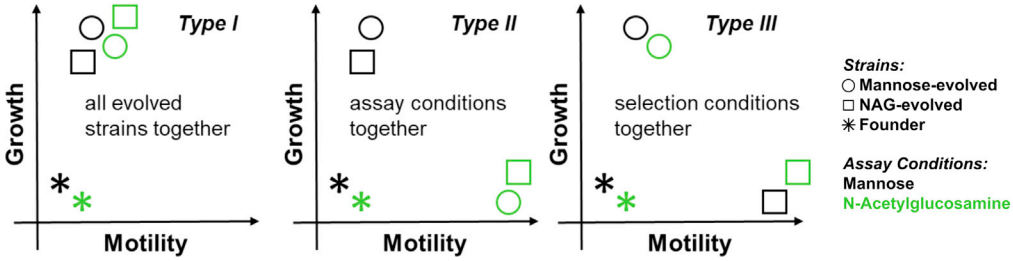


Figure 3. Hypothesized Modes of Evolution of Migration Rate Generality

We propose three distinct possibilities for the underlying phenotypic basis of the nutrient generality of migration rate adaptation presented in Figure 2. Type I (universality): All evolved strains show the same adaptation from founder across all assay conditions. There is no separation of evolved phenotypes in the two-dimensional phenotypic space of motility and growth. Type II (plasticity): Evolved strains exhibit a flexible adapted state. Evolved strains show a characteristic adaptation for each assay condition, regardless of their evolutionary history. Phenotypes separate by assay condition (marker color). Type III (degeneracy): Evolved strains display a characteristic phenotype for each evolutionary history, regardless of assay condition. Phenotypes separate by selection condition (marker shape).

Type I: Universal Adaptation

Neither the selection condition nor the assay condition has an impact on the evolved phenotype. Instead, there is a single phenotype conferring fast migration through 0.2% w/v agar and 1 mM sugar, irrespective of any differences in import, metabolism, and chemotactic affinity between different sugars. Evolved strains could achieve this phenotype across all selection conditions. Generality would then be achieved as long as evolved strains exhibit the adapted phenotype across different assay conditions. For example, the evolved strains could have achieved a growth rate adaptation that does not depend on the specific sugar, allowing them to grow rapidly (and thereby exhibit fast migration) in all 1 mM sugar assay conditions, or they could exhibit a change in run-tumble statistics which confer a migration rate advantage in soft agar. However, "type I" generality need not be this simple. The ideal phenotype could be a particular combination of growth and motility enhancements, corresponding to a distinct direction away from founder in the two-dimensional space of phenotypes. The key feature of "type I" generality is that there is no separation of evolved phenotypes by either selection condition or assay condition. Instead there is a single, universal phenotype that is achieved by all evolved strains in all conditions (Figure 3).

Type II: Phenotypic Plasticity

In each sugar, there exists a distinct growth/motility phenotype conferring fast migration. The adaptive value of growth or motility can easily depend on the assay condition. For example, in a condition that supports slow growth of the ancestral strain, increases in growth may confer a greater advantage compared to motility. These differences could also arise at the molecular level due to differences in import, metabolism, and chemotactic affinity for each sugar. In this scenario, generality would emerge if evolved strains exhibit the different adapted phenotypes for each assay condition, irrespective of their selection condition. This would mean that selection acts to put evolved populations in a plastic, adaptive state. Once in this state, cells could adapt their phenotype to the particular balance of growth and motility needed to enhance migration in different environments. We refer this mechanism for evolving generalists as "plasticity" because it requires that the same genotype (evolved strain) and exhibits distinct phenotypes (growth/motility) in different nutrient conditions. In this situation, a strain's phenotype is determined more by its assay condition than selection condition. Graphically, this would mean that the evolved phenotypes separate by assay condition (Figure 3). For example, if evolved strains across different evolutionary histories showed enhanced growth in mannose but enhanced motility in NAG, we would conclude that the nutrient generality of migration rate evolution was achieved through phenotypic plasticity.

Type III: Degeneracy

There is a degenerate set of distinct phenotypes which all confer faster migration on the evolved strains in all assay conditions. Suppose that there is a distinct method of adaptation associated with each selection condition and evolved populations exhibit these different adaptations in all assay conditions. Generality would be possible in this case as long as different growth and motility phenotypes could produce similar migration rates in a particular environment. In this sense, the evolved phenotypes would be degenerate at

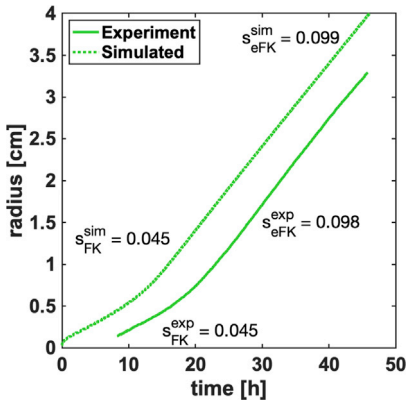


Figure 4. Theory-Experiment Comparison of Migrating Front Displacement in Time

Simulated and measured front position as a function of time for the founding strain in *N*-acetylglucosamine. The labeled rates, s_{FK} for the early times, and enhanced migration rates, s_{eFK} for the later times, are labeled for simulation (*sim*) and experiment (*exp*), respectively, in units of cm/h. For details of the simulation and parameter values see Transparent Methods. See also Figure S4.

the level of migration rate. Graphically, this would correspond to the case where evolved phenotypes separate by selection condition (Figure 3). For example, if mannose-evolved strains enhance growth in all environments, while strains evolved in NAG enhance motility across all environments, we would classify the observed migration rate generality as "type III".

Quantitative Relationship between Phenotypes and Migration Rates

Selection in our experiment is performed on migration rate. Therefore, to test which of these three hypotheses accounts for the evolution of generalists we observe in Figure 2 we must first relate phenotypes (growth and motility) quantitatively to the migration rate of the population. To accomplish this, we use an established reaction-diffusion formalism (Cremer et al., 2019) to describe the spatiotemporal dynamics of the chemoattractant and nutrient source and the bacterial population. The model takes the following form:

$$\frac{\partial \rho}{\partial t} = D_b \nabla^2 \rho - \nabla(v\rho) + \frac{k_g \rho c}{K_g + c}, \qquad \text{(Equation 1)}$$

$$\frac{\partial c}{\partial t} = D_c \nabla^2 c - \frac{k_g \rho c}{Y(K_g + c)}, \qquad \text{(Equation 2)}$$

$$\text{with } v = \chi \log\left(\frac{1 + c/A}{1 + c/B}\right), \qquad \text{(Equation 3)}$$

where ρ is a field describing the bacterial density in space and time and c describes the nutrient/chemoattractant field. D_b and D_c are diffusion constants for bacteria and nutrients, respectively, and k_g is the maximum growth rate with affinity K_g and biomass yield Y. A and B are constants that determine the range of attractant concentrations over which cells respond chemotactically. χ controls the strength of the chemotactic response and can be decomposed further $\chi = fD_b$, where D_b is the diffusion constant and f is a factor that depends on the bias of cells swimming up a gradient of attractant (Croze et al., 2011).

In the limit of a homogeneous field of nutrients ($\nabla c = 0$), the second term in Equation 1 tends to zero and we recover the familiar Fisher-Kolmogorov (FK) model. In this limit, the migration rate of the population rate is given by $s_{FK} \propto D_b k_g$. At long times, when the population has consumed available nutrients at the center of the plate, scaling arguments show that the chemotactic front moves with an enhanced migration rate $s_{eFK} \propto \chi k_g = fD_b k_g$ (Cremer et al., 2019), which is larger than s_{FK} since f is typically greater than 1. In either case, faster growth drives faster migration by more rapidly creating a larger population to drive colony expansion.

Simulations of the model in Equations (1), (2), and (3) show that expanding colonies transition from migrating at a rate s_{FK} at early times to s_{eFK} at late times (Figure 4). Migration at a rate s_{FK} occurs when the initial population is diffusing but has not yet consumed enough nutrients to create a sufficiently large gradient for chemotaxis to be important. Once nutrients have been consumed locally, which takes a time approximately $t_{xo} \sim Yc_0/k_g$, with c_0 the initial nutrient supply, the population transitions to migrating at a rate

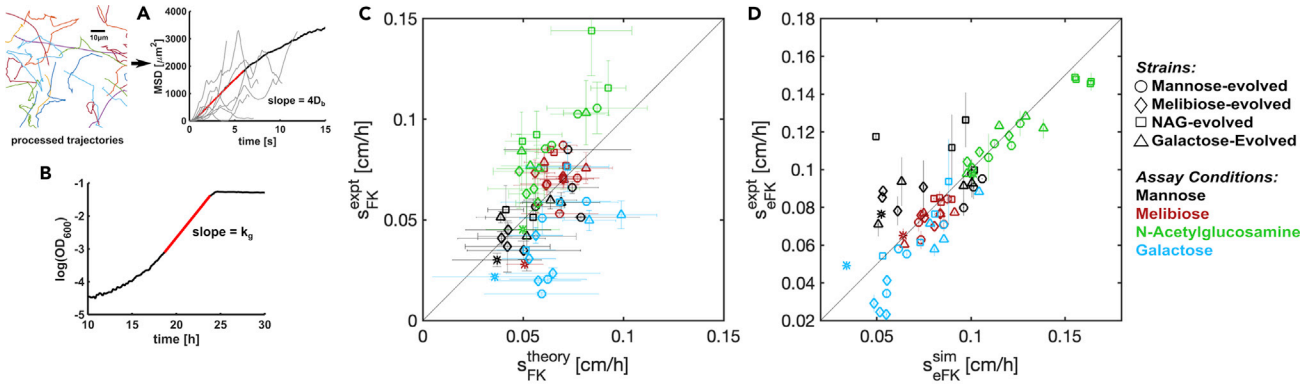


Figure 5. Quantified Phenotypes Predict Front Migration Rates of Founder and Evolved Strains.

(A) Scheme for measuring bacterial diffusion constants. For each strain in each condition, thousands of individuals were recorded swimming (7856 ± 5623, mean ± standard deviation trajectories per experiment). Videos were automatically processed into trajectories, and squared displacement was calculated for each cell. Examples of single-cell squared displacement traces given in light gray. Population-level mean squared displacement (MSD, black line) was calculated by averaging over single-cell traces at each frame (see Figure S5). D_b was inferred from a linear fit to the MSD vs time (red line), assuming cells were diffusing in two dimensions.

(B) Maximum growth rates were measured by continuously measuring the optical density of well-mixed liquid cultures. k_g was fitted from the slope of log(OD_{600}) versus time and averaged over replicate wells. See Transparent Methods for details of both experiments.

(C) Measured s_{FK} migration rates for all strains in all conditions plotted against the theoretical expectation of $s_{FK} = \sqrt{2D_b k_g}$. The black line corresponds to perfect agreement between measurement and prediction. Error bars in measured rate are standard errors on slopes fit to front position in time. Errors on predicted rates are computed via error propagation from the uncertainty in D_b and k_g from the regressions in A and B. Pearson's correlation 0.51, p value 2×10^{-5}.

(D) Simulated and measured s_{eFK} for all stains in all conditions. Simulations are of the model in Equations 1–3 as described in Transparent Methods. Errors on predicted rates are smaller than the size of the markers. Pearson's correlation 0.83, p value 8×10^{-19}.

See also Figure S4.

s_{eFK} (Figure S4). We observe this crossover in our experiments and simulations of the model presented above (Figure 4).

Therefore, a population could evolve faster migration by increasing D_b, k_g, or f or any combination of these phenotypic parameters. In some conditions (e.g., galactose) where the founder initially grows slowly, the crossover time (t_{xo}) may approach the duration of one round of selection (24 hr). In this case, an adaptation that increases the growth rate and brings $t_{xo} <$ 24 hr will result in substantially faster migration simply because the faster growth allows for an expanding population to access the migration rate $s_{eFK} > s_{FK}$ earlier in time. In this case, large increases in the migration rate, of order f, could be observed even for only modest increases in the growth rate (k_g) and no evolutionary modification of motility (D_b and χ).

Quantifying the Impact of Evolved Phenotypes on Migration Rate

We measured phenotypic parameters for the founder and evolved strains in order to determine which of the three hypotheses in Figure 3 best explained the evolution of generalists in our experiment. We measured the growth and diffusion of the ancestor, as well as the 16 evolved strains, presented in Figure 2 in each of the four environments. These experiments were conducted in liquid minimal medium identical to the plates used for migration assays but without agar. Motility was measured by performing high-throughput single-cell tracking on populations of cells with phase contrast microscopy (Perlova et al., 2019), and diffusion constants (D_b) were inferred from the slope of the mean squared displacement across individual trajectories (Figure 5A). We note that measured diffusion constants in liquid are within approximately 10% of the diffusion constant in 0.2% w/v agar where our selection experiment took place (Croze et al., 2011). Growth was measured by monitoring the optical density of well-mixed liquid cultures, and maximum growth rates (k_g) were fitted during the exponential phase of the growth curve (Figure 5B). We assume a fixed and small K_g in all conditions (see Transparent Methods).

Using measured k_g and D_b parameters for all 16 strains in each of 4 conditions, we predicted the front migration rate early in the colony expansion process by computing s_{FK} and comparing this value to the measured early time expansion rates for 63 of 68 strain/condition combinations and found good

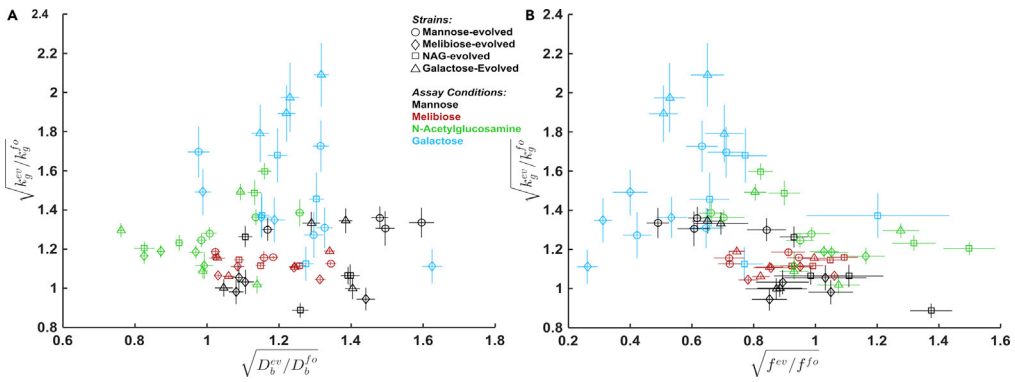


Figure 6. Phenotypic Adaptation Suggests Plasticity

(A) Motility and growth adaptation in liquid media of all 16 evolved strains presented in Figure 2 in four different assay conditions. The founder has a diffusion constant of 52 ± 1.3, 63 ± 1.2, 58 ± 1.1, and 56 ± 1.7 $\mu m^2 s^{-1}$ in mannose, melibiose, NAG, and galactose, respectively, fitted slope (MSD vs time) ± standard error after combining data from two independent experiments (Figure S6). The founder has a maximum growth rate of 0.18 ± 0.05, 0.29 ± 0.01, 0.30 ± 0.05, and 0.16 ± 0.08 hr^{-1} in mannose, melibiose, NAG, and galactose, respectively; mean ± standard deviation of ten replicate wells spread over two independent plates. We present adaptation of diffusion constant as $\sqrt{D_b^{ev}/D_b^{fo}}$ and adaptation of maximum growth rate as $\sqrt{k_g^{ev}/k_g^{fo}}$ since these quantities contribute linearly to fractional enhancement of migration rate. Error bars are standard error calculated using error propagation.

(B) Shows changes in the chemotactic coefficient plotted versus changes in growth rate.

f^{ev} and f^{fo} denote the founder and evolved coefficients which are computed from the observed late-time migration rates and measured diffusion constants and growth rates as $f = s_{eFK}^2/(D_b k_g)$. The founder has an f^{fo} of 17 ± 1.6, 6.5 ± 0.4, 15.3 ± 1.3, and 7.5 ± 1.2 for mannose, melibiose, NAG, and galactose, respectively. Error bars are from error propagation. See Figure S6 for a plot of $\sqrt{D_b^{ev}/D_b^{fo}}$ vs $\sqrt{f^{ev}/f^{fo}}$.

See also Figures S5–S8.

agreement (Figure 5C). For 5 condition/strain combinations, s_{FK} could not reliably be determined from imaging data (see Transparent Methods). Therefore, our measurements of D_b and k_g accurately account for changes in migration rates by the evolved strains early in the colony expansion process.

Accounting for the late-time migration rates in founding and evolved strains, (s_{eFK}) requires measuring χ and therefore f. Measuring f in high throughput, at the single-cell level, is challenging because it requires applying controlled attractant gradients while measuring single-cell motility. We therefore inferred this parameter from our data. To accomplish this, we note that $f \propto s_{eFK}^2/(D_b k_g)$, where s_{eFK} is the measured migration rate late in the colony expansion process (Figure 4). To measure f for the founding strain, we measured colony expansions for 48 hr to ensure the colony had enough time to consume the initial supply of nutrients. We measured the migration rate via regression of front position in time (Figure 4) and computed $s_{eFK}^2/(D_b^{fo} k_g^{fo})$. For the evolved strains, we measured s_{eFK} as the migration rate during the last 5 hr of expansion (during an 24 hr expansion) and inferred f using independent measurements of D_b and k_g on evolved strains.

With these inferred values of f for all strains, we paramterized the model in Equations 1, 2, and 3 for all strains in all media conditions (see Transparent Methods). We simulated front migration for all strains and measured s_{eFK}^{sim} for all strains. Figure 5D compares this simulated migration rate with our experimental observations and again shows good agreement.

In summary, high-throughput single-cell tracking allowed us to quantify the bacterial diffusion constant D_b in all 68 distinct strain/condition combinations shown in Figure 2. We combined these results with measurements of growth rate k_g and an established reaction-diffusion model to infer the impact of selection on the chemotaxis coefficient $\chi = fD_b$. We then were able to directly compare measured front migration rates with model predictions and observed good agreement (Figures 5C and 5D).

Measured Phenotypes Suggest Generality Evolved by Plasticity

Having quantified how the evolved phenotypes impact migration rate, we could then ask which hypothesis for generality (Figure 3) is correct. Migration rate can be increased in the evolved strains by changes in f, D_b,

Source of Variance	F	p value
Selection condition (α_i)	1.01	0.42
Assay condition (β_j)	12.80	7.61×10^{-6}
Strain (γ_k)	1.65	0.12
Selection condition x assay condition ($\alpha\beta_{ij}$)	0.60	0.79

Table 1. Assay Condition Drives Differences in Diffusion Constant Adaptation

Summary statistics for ANOVA on the model presented in Equation 4 using $\sqrt{D_b^{ev}/D_b^{fo}}$ as the response variable. The F-statistic describes the ratio of between-group variability to within-group variability. Here, we only found a significant ($p < 0.05$) F-statistic for assay condition.

or k_g. The contribution of any adaptation in these parameters to the evolved strains increase in migration rate relative to the founder is given by: f^{ev}/f^{fo}, ${D_b}^{ev}/{D_b}^{fo}$, and ${k_g}^{ev}/{k_g}^{fo}$. Therefore, to distinguish between the three types of generality presented in Figure 3, we examine the distributions of these quantities. Figures 6A and 6B show these quantities for all strains in all conditions.

A quick inspection of the evolved phenotypes suggested phenotypic plasticity. The data seemed to roughly separate by assay condition (Figures 6A and 6B). For example, strains measured in mannose typically exhibited a larger enhancement to diffusion but a more modest enhancement to growth when compared to strains measured in galactose. In addition, the increases in f appeared to concentrate in strains assayed in NAG, whereas strains assayed in galactose appeared to have reduced chemotactic capabilities irrespective of their selection condition.

We utilized a statistical approach to evaluate the significance of any separation of evolved phenotypes by selection condition and/or assay condition. Specifically, we performed analysis of variance (ANOVA) on the following linear mixed-effects model:

$$y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + (\alpha\beta)_{ij} + \varepsilon_{ijk} \quad \text{(Equation 4)}$$

This model decomposes a response variable y_{ijk} (${k_g}^{ev}/{k_g}^{fo}$, ${D_b}^{ev}/{D_b}^{fo}$, or f^{ev}/f^{fo}) in terms of deviations from the global mean μ attributable to different groupings of the data. α_i is a fixed effect due to selection condition, with i indexing the four selection conditions. β_j is a fixed effect due to assay condition, with j indexing the four assay conditions. γ_k is a random effect for each of the 16 evolved strains, with k indexing the evolved strain. $(\alpha\beta)_{ij}$ is an interaction term included to allow for unique effects due to particular combinations of selection and assay conditions. For example, if some strains grew particularly well in the condition they were selected for compared to the other strains, this would manifest itself as a significant interaction term. Lastly, ε_{ijk} is a noise term in the form of normally distributed random disturbances.

ANOVA analysis allowed us to determine which groupings, if any, displayed significant differences in the response variable. Performing ANOVA on ${D_b}^{ev}/{D_b}^{fo}$, we found that assay condition is the only grouping with a significant ($p < 0.05$) F-statistic (Table 1). That is, we rejected the null hypothesis that all the β_j are zero. We concluded that there is a significant difference between the assay conditions in the adaptation of undirected motility. To investigate which assay conditions show significant departures from the global mean and in which direction, we performed *post-hoc* testing via a non-parametric bootstrapping approach. This revealed that strains measured in mannose were significantly above the global mean in ${D_b}^{ev}/{D_b}^{fo}$, while strains measured in NAG were significantly below the mean, and strains in melibiose and galactose showed no significant change in ${D_b}^{ev}/{D_b}^{fo}$ relative to the global mean (Figure S7).

Bacterial diffusion constants are approximately related to the run speed and duration in the following way: $D_b \approx {v_r}^2 \tau_r$. Therefore, changes in diffusion constants can be achieved through changes in run duration and/or speed. We asked whether changes in ${D_b}^{ev}/{D_b}^{fo}$ were achieved with a consistent microscopic strategy within each assay condition. Therefore, we investigated correlations between changes in run statistics and changes in diffusion constant across all evolved strains within each assay condition (Figure S8). We found that ${D_b}^{ev}/{D_b}^{fo}$ was significantly correlated with a particular microscopic strategy for strains measured in each assay condition, regardless of their evolutionary history. For example, in mannose, changes in D_b

Source of Variance	F	p value
Selection condition (α_i)	7.00	0.01
Assay condition (β_j)	31.86	3.10×10^{-10}
Strain (γ_k)	1.15	0.36
Selection condition x assay condition ($\alpha\beta_{ij}$)	4.68	3.78×10^{-4}

Table 2. Assay Condition is the Main Driver of Differences in Growth Adaptation

Summary statistics for ANOVA on the model presented in Equation 4 using $\sqrt{k_g^{ev}/k_g^{fo}}$ as the response variable. The F-statistic describes the ratio of between-group variability to within-group variability. Here, we found significant F-statistics for selection condition, assay condition, and the interaction term between them. The largest F-statistic was associated with assay condition, indicating this predictor is the dominant source of variability in $\sqrt{k_g^{ev}/k_g^{fo}}$. See also Tables S2–S4.

were correlated with changes in τ_r but not v_r. Conversely, in NAG, changes in diffusion constant were correlated with changes in both v_r and τ_r.

Similarly, performing ANOVA with $\sqrt{k_g^{ev}/k_g^{fo}}$ as the response variable shows that, by far, the most significant effect is due to assay condition (largest F-statistic, Table 2). *Post-hoc* testing revealed that strains measured in galactose have above-average growth rate adaptation, while strains measured in mannose and melibiose were below average, and strains measured in NAG showed no significant change relative to the global mean (Table S2). We also found smaller but still significant ($p < 0.05$) effects due to selection condition and the interaction term (Table S3). *Post-hoc* testing on the α_i revealed only one significant coefficient: melibiose-evolved strains exhibited a below-average growth rate enhancement. For the interaction term, *post-hoc* showed only one significant term: an increased growth rate adaptation for strains evolved in galactose when measured in galactose (Table S4). We concluded that the assay condition is the strongest determinant of growth rates in evolved strains.

Lastly, ANOVA with f^{ev}/f^{fo} as the response variable again showed that the most significant effect was due to assay condition (largest F-statistic, Table 3). *Post-hoc* testing revealed here that strains measured in NAG exhibited above-average chemotactic adaptation, while strains measured in galactose were below average, and strains measured in mannose and melibiose showed no significant change relative to the global mean (Table S5). We also found a smaller but still significant ($p < 0.05$) effect due to selection condition. *Post-hoc* testing on the α_i revealed that strains selected in NAG were above average in their chemotactic adaptation, while strains selected in mannose were below average, and strains selected in melibiose and galactose showed no significant deviation from the global mean (Table S6).

These ANOVA analyses provided clear evidence that the evolved phenotypes presented in Figures 6A and 6B separate by assay condition. We concluded that, within each of the four environments, evolved strains exhibited similar adaptation in growth and motility, regardless of their evolutionary history. The result is consistent with "type II" generality (i.e., plasticity). Evolved strains across all selection conditions appeared to have evolved phenotypic plasticity that allowed them to enhance population-level migration by adapting cellular phenotypes to meet the unique demands of each nutrient condition.

Mutations Present in Evolved Strains Cannot Predict Phenotypes in Any Condition

We next asked whether specific mutations present in strains evolved for faster migration could explain the migration, growth, and motility phenotypes we observed. We performed whole genome sequencing on all 16 evolved strains and identified *de novo* mutations relative to the common ancestor present at a frequency of at least 20% (Deatherage and Barrick, 2014) (Tables S7 and S8). We identified 33 unique mutations occurring in diverse genes and pathways including biosynthesis of essential molecules, stress response, and metabolite import. We observed no single mutation common to all evolved strains, nor were any mutations common to all 4 evolved strains from any particular selection condition. However, 24 of the 33 mutations were present in more than one strain. By grouping together mutations occurring in the same gene, as well as those in adjacent intergenic regions, we identified 21 unique genes or flanking regions where mutations occurred, which we term mutational targets. Twelve of those targets were affected in multiple strains, and of those, 8 were affected across strains with different selection conditions. Unlike previous studies (Fraebel et al., 2017), we did not observe mutations with an obvious interpretation in terms of their

Source of Variance	F	p value
Selection condition (α_i)	7.66	4.02×10^{-3}
Assay condition (β_j)	20.04	8.27×10^{-8}
Strain (γ_k)	1.47	0.18
Selection condition x assay condition ($\alpha\beta_{ij}$)	2.16	0.05

Table 3. Assay Condition is the Main Driver of Differences in Apparent Chemotactic Adaptation
Summary statistics for ANOVA on the model presented in Equation 4 using f^{ev}/f^{fo} as the response variable. The F-statistic describes the ratio of between-group variability to within-group variability. Here, we found significant F-statistics for selection condition and assay condition. The largest F-statistic was associated with assay condition, indicating this predictor is the dominant source of variability in f^{ev}/f^{fo}. See also Tables S5 and S6.

impact on evolved phenotypes. Therefore, we took a statistical approach to interpreting the sequencing data.

Having grouped mutations by their target, we created a mutation candidacy matrix (X) which describes the presence (1) or absence (0) of each observed mutation in each strain. We then asked whether the presence or absence of these mutations could predict the migration phenotypic parameters of each strain in each condition. To do this, we performed linear regression using mutation candidacy as the predictor variables and the adaptation in either growth rate (k_g), diffusion constant (D_b), or the chemotactic coefficient (f) as the dependent variable, e.g., $\sqrt{f^{ev}/f^{fo}} = \eta_0 + X\vec{\eta} + \vec{\varepsilon}$, where η_0 is an intercept, $\vec{\eta}$ is a vector of regression coefficients expressing the impact of each mutation on the response variable, and $\vec{\varepsilon}$ is a noise term with zero mean and variance σ^2. We performed L_1-regularized regression (least absolute shrinkage and selection operator [LASSO]) to avoid overfitting (Yi and Xu, 2008; Li and Sillanpää, 2012), using leave-one-out cross-validation to determine the best value of the regularization hyperparameter. For several assay conditions, LASSO regression selected a model with only an intercept (e.g., giving $\sqrt{f^{ev}/f^{fo}} = \eta_0 + \vec{\varepsilon}$), indicating that no mutation predicted the migration parameter across the set of strains. In cases where a model with non-zero $\vec{\eta}$ was selected (i.e., some mutations were identified as having predictive power), the improvement in the mean squared error (estimated using cross-validation) relative to a model with only an intercept was small, and the statistical significance of a non-zero $\vec{\eta}$ was unclear.

These observations led us to undertake a numerical investigation using surrogate data to assess the performance of our LASSO regressions (see Transparent Methods for details). The purpose of this investigation was to determine whether regressions that selected a non-zero $\vec{\eta}$ had robust out-of-sample predictive power. If not, then we would consider it unlikely that any mutation in these regressions truly predicted the migration phenotypic parameters. Briefly, we constructed surrogate data sets where the coefficients $\vec{\eta}$ were specified, and the signal-to-noise ratio was effectively modulated by specifying the number of non-zero coefficients in $\vec{\eta}$ and the random noise strength σ (Figures S9B and S9C). Each data set comprised a training set and an out-of-sample test set, where the predictors in both training and test sets matched the candidacy matrix (X) in the number of observations, number of variables, and first-order statistics (presence frequency of each mutation). On each of 3×10^5 instances of surrogate data sets, we performed LASSO regressions as described above and evaluated performance by computing the coefficient of determination R^2 using the out-of-sample data.

We developed a test statistic M to relate these surrogate data regressions to the real regressions on migration parameters (see Transparent Methods for details) since the true coefficient vectors $\vec{\eta}$ and noise strength σ for the regressions on real data are unknown. This statistic measured the improvement in model fits (mean squared error estimated using cross-validation) for the model selected by LASSO relative to a model with only an intercept (η_0). In the surrogate data, we observed that M correlated positively with the out-of-sample predictive power of a model inferred by LASSO, giving us a clear relationship between the test statistic and regression performance (Figures 7, S9D, and S9E). We then computed our test statistic for the real data regressions for adaptation in growth rate, diffusion constants, or chemotactic coefficients from the mutation candidacy matrix and observed small values of the test statistic M (Figure 7), indicating low predictive power is likely (out-of-sample $R^2 \approx 0$). We concluded that, for our data, mutations are highly unlikely to predict the measured phenotypic adaptation. In sum, the presence of "individual" mutations has no predictive power of evolved phenotypes.

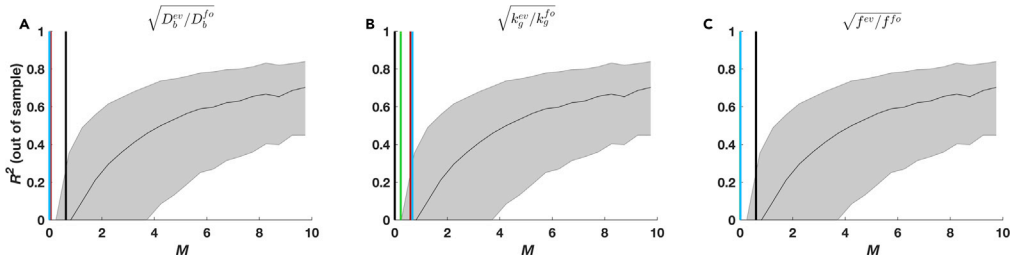


Figure 7. Numerical Investigation of LASSO Regression Demonstrates Mutations Have Limited Predictive Power for Migration Phenotypes

(A–C) Relationship between the test statistic M and out-of-sample R^2 obtained from surrogate data regressions (see Transparent Methods), with M values from regressions on real migration phenotypic parameters overlaid. M and R^2 values from all surrogate data simulations are combined, binned by M, and the quartiles of R^2 within the bins are shown as a function of M. The same quartile curves are shown in A–C. Vertical lines indicate the M values resulting from LASSO regressions on real data for the quantities (A) $\sqrt{D_b^{ev}/D_b^{fo}}$, (B) $\sqrt{k_g^{ev}/k_g^{fo}}$, and (C) $\sqrt{f^{ev}/f^{fo}}$, , where the colors indicate the assay condition (black: mannose, red: melibiose, green: N-acetylglucosamine, blue: galactose).
See also Figure S9, Tables S7 and S8.

DISCUSSION

The central finding of this study is the emergence of generalists when bacterial populations are selected for faster migration through a porous minimal medium environment. We found that distinct phenotypic strategies gave rise to fast migration, and, remarkably, these strategies were determined more by the immediate nutrient condition than by the evolutionary history of the strains being measured. We concluded that repeated selection in any condition drove fast migration in all conditions via distinct, plastic phenotypic responses to each nutrient condition. As a result, the fast migration in all conditions emerged via plasticity, which is a byproduct of selection for fast migration in any one condition. Therefore, in this case, plastic phenotypes evolved even in homogeneous selection conditions.

It is not yet clear why selection in a single environment results in plastic phenotypic responses to other environments that increase migration rate. Perhaps the simplest explanation is that similar physiological changes occur during evolution in all four nutrient conditions. However, the molecular mechanisms that give rise to the observed generality via phenotypic plasticity are not yet clear. Our statistical analysis of the genetic variation observed in evolved strains shows that there is no simple genetic basis for this plastic adaptive response. Given what is known about sugar uptake and metabolism in *E. coli*, some of our observed mutations could be targeting the uptake or metabolism of multiple sugars. For example, *EnvZ* regulates expression of *OmpF*, the outer membrane porin responsible for import of all four sugars (Mizuno and Mizushima, 1987; Travisano and Lenski, 1996). *NagA* is essential for metabolism of NAG but also has a role in regulating expression of the *nag* regulon, including *nagC* (Plumbridge, 1991). There is evidence that *NagC* is capable of repressing the mannose phosphotransferase system (PTS) system (Plumbridge, 1991) and the galactose transporter (El Qaidi et al., 2009). Therefore, the mutations we observed in *nagA* for two galactose-evolved strains could have an impact in mannose, NAG, and galactose environments.

The mechanistic basis for how metabolism is coupled to motility remains unclear, and elucidating this fully will be necessary to understand why evolved strains in some conditions exhibit large changes in both motility and growth (e.g., galactose). The fact that the evolved strains did not exhibit fast migration in a rich medium, where amino acids are responsible for growth and chemotaxis, is consistent with the idea that the plastic generalist adaptation observed in this study was specific to sugars. Therefore, the limits on generalist evolution due to plasticity in this system would appear to be defined by the chemical identity of the nutrient/chemoattractant. Given that we observed mutations in regulatory elements (e.g., *rssB, wzzE, envZ*), it is possible that the plasticity we observed in the evolved strains might be best understood at the regulatory level. In this case, the plastic response would be selected in a homogeneous environment, due to regulatory variation that results in changes in gene expression impacting motility and metabolism in other environments. Therefore, the shared molecular-level features of the evolved strains which were responsible for the phenotypic plasticity outcomes in different nutrient conditions may be explained by gene expression measurements (Huang et al., 2005).

More broadly, the smooth phenotypic variation we observed across strains and environments (Figure 6) is consistent with a complex genetic basis where the plastic response of a given genotype depends on genetic variants at a large number of loci. One possible mechanism for genetic complexity in evolved plastic responses would be phenotypic impacts from a large number of weak regulatory linkages (Gerhart and Kirschner, 2007), which is supported by genetic screens on motility (Girgis et al., 2007). The apparent smooth phenotypic variation is in contrast to phenotypic switches, where plasticity drives discrete phenotypic variation (Balaban, 2004). It would be interesting to design future experiments to determine when and why selection favors continuous or discrete plastic responses. Perhaps one way to approach this problem is to compare the present results to selection in fluctuating environments. Fluctuating selection is believed to favor the evolution of cooperative (non-linear) responses in biological systems, for example, in proteins (Stiffler et al., 2015), and one might speculate that the same could be true for phenotypic plasticity.

Since the seminal work of Baldwin and later Waddington (Baldwin, 1896; Waddington, 1953), a main focus of work on phenotypic plasticity has been to understand the relationship between plastic responses to environmental change and subsequent genetic adaptation (Ghalambor et al., 2015; Schaum and Collins, 2014; Ho and Zhang, 2018). Our results suggest another possible role of phenotypic plasticity in evolution: namely, that selection in one environment can potentiate new physiological responses to other environments resulting in the evolution of generalists even in homogeneous environmental conditions. Future work should focus on understanding why selection in one environment can elicit new phenotypic responses to other environments and the scope of phenotypic generality that can evolve by this mechanism.

Limitations of the Study

The central limitation of the present study is that it relies on a statistical argument to support the claim that generalists evolve via phenotypic plasticity. As a result, the physiological basis of the observed plastic responses is not yet known. Future work should examine the impact of selection for fast migration on gene expression profiles across all evolved lineages to discern conserved patterns of expression which could then inform more detailed mechanistic interrogation of the evolution of plasticity.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Seppe Kuehn (seppe.kuehn@gmail.com).

Materials Availability

All strains generated in this study are available upon request. No reagents were generated.

Data and Code Availability

Details of the detected mutations are shown in Table S8. Parameter values for the simulations are given in the Supplementary Information. Source code for simulations is available upon request. The accession number for the DNA sequencing data reported in this paper is NCBI: PRJNA667609.

METHODS

All methods can be found in the accompanying Transparent Methods supplemental file.

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.isci.2020.101678.

ACKNOWLEDGMENTS

D.F. and S.K. acknowledge support from the National Science Foundation Physics Frontiers Center Program (PHY 0822613 and PHY 1430124). K.G. acknowledges support from the James S. McDonnell Foundation Postdoctoral Fellowship Award (#220020499). M.M. acknowledges the Simons Foundation (597491). M.M. is a Simons Foundation Investigator. We would like to thank Jun Song for advice in selecting the form of the ANOVA model, Tatyana Perlova for providing technical support and software for the single-cell tracking assays, and Elizabeth Ujhelyi for assistance with sequencing.

AUTHOR CONTRIBUTIONS

Conceptualization, D.T.F. and S.K.; Methodology, D.T.F, K.G., M.M., and S.K.; Formal Analysis, D.T.F., K.G., and S.K.; Experiments, D.T.F.; Investigation, D.T.F., K.G., and S.K., LASSO regression, K.G.; Reaction-diffusion simulations, D.T.F. and S.K.; Writing, D.T.F, K.G., and S.K.; Supervision, M.M. and S.K.; Funding acquisition, M.M. and S.K.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: July 1, 2020
Revised: September 22, 2020
Accepted: October 9, 2020
Published: November 20, 2020

REFERENCES

Adler, J. (1966). Chemotaxis in bacteria. Science *153*, 708–716.

Balaban, N.Q. (2004). Bacterial persistence as a phenotypic switch. Science *305*, 1622–1625.

Baldwin, J.M. (1896). A new factor in evolution. Am. Nat. *30*, 441–451.

Bourgerie, S.J., Michán, C.M., Thomas, M.S., Busby, S.J., and Hyde, E.I. (1997). DNA binding and DNA bending by the MelR transcription activator protein from Escherichia coli. Nucleic Acids Res. *25*, 1685–1693.

Cooper, V., and Lenski, R.E. (2000). The population genetics of ecological specialization in evolving Escherichia coli populations. Nature *407*, 734–736.

Cremer, J., Honda, T., Tang, Y., Wong-Ng, J., Vergassola, M., and Hwa, T. (2019). Chemotaxis as a navigation strategy to boost range expansion. Nature *575*, 658–663.

Croze, O.A., Ferguson, G.P., Cates, M.E., and Poon, W.C. (2011). Migration of chemotactic bacteria in soft agar: role of gel concentration. Biophys. J. *101*, 525–534.

Deatherage, D.E., and Barrick, J.E. (2014). Identification of mutations in laboratory-evolved microbes from next-generation sequencing data using breseq. Methods Mol. Biol. *1151*, 165–188.

El Qaidi, S., Allemand, F., Oberto, J., and Plumbridge, J. (2009). Repression ofgalP, the galactose transporter inEscherichia coli, requires the specific regulator ofN-acetylglucosamine metabolism. Mol. Microbiol. *71*, 146–157.

Elena, S.F., and Lenski, R.E. (2003). Microbial genetics: evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. Nat. Rev. Genet. *4*, 457–469.

Fong, S.S., Joyce, A.R., and Palsson, B.Ø. (2005). Parallel adaptive evolution cultures of Escherichia coli lead to convergent growth phenotypes with different gene expression states. Genome Res. *15*, 1365–1372.

Fraebel, D.T., Mickalide, H., Schnitkey, D., Merritt, J., Kuhlman, T.E., and Kuehn, S. (2017). Environment determines evolutionary trajectory in a constrained phenotypic space. eLife *6*, e24669.

Fraser, A.D., and Yamazaki, H. (1980). Mannose utilization in *Escherichia coli* requires cyclic AMP but not an exogenous inducer. Can. J. Microbiol. *26*, 1508–1511.

Gerhart, J., and Kirschner, M. (2007). The theory of facilitated variation. Proc. Natl. Acad. Sci. U S A *104*, 8582–8589.

Ghalambor, C.K., Hoke, K.L., Ruell, E.W., Fischer, E.K., Reznick, D.N., and Hughes, K.A. (2015). Non-adaptive plasticity potentiates rapid adaptive evolution of gene expression in nature. Nature *525*, 372–375.

Girgis, H.S., Liu, Y., Ryu, W.S., and Tavazoie, S. (2007). A Comprehensive Genetic Characterization of Bacterial Motility. PLoS Genet. *3*, e154.

Henderson, P.J., Giddens, R.A., and Jones-Mortimer, M.C. (1977). Transport of galactose, glucose and their molecular analogues by Escherichia coli K12. Biochem. J. *162*, 309–320.

Ho, W.-C., and Zhang, J. (2018). Evolutionary adaptations to new environments generally reverse plastic phenotypic changes. Nat. Commun. *9*, 1–11.

Huang, S., Eichler, G., Bar-Yam, Y., and Ingber, D.E. (2005). Cell fates as high-dimensional attractor states of a complex gene regulatory network. Phys. Rev. Lett. *94*, 12.

Kassen, R. (2002). The experimental evolution of specialists, generalists, and the maintenance of diversity. J. Evol. Biol. *15*, 173–190.

Lee, M.-C., Chou, H.-H., and Marx, C.J. (2009). Asymmetric, bimodal trade-offs during adaptation of Methylobacterium to distinct growth substrates. Evolution *63*, 2816–2830.

Li, Z., and Sillanpää, M.J. (2012). Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. Theor. Appl. Genet. *125*, 419–435.

Mizuno, T., and Mizushima, S. (1987). Isolation and characterization of deletion mutants of ompR and envZ, regulatory genes for expression of the outer membrane proteins OmpC and OmpF in Escherichia coli. J. Biochem. *101*, 387–396.

Ni, B., Ghosh, B., Paldy, F.S., Colin, R., Heimerl, T., and Sourjik, V. (2017). Evolutionary remodeling of bacterial motility checkpoint control. Cell Rep. *18*, 866–877.

Ostrowski, E.A., Rozen, D.E., and Lenski, R.E. (2005). Pleiotropic effects of beneficial mutations in Escherichia coli. Evolution *59*, 2343–2352.

Perlova, T., Gruebele, M., and Chemla, Y.R. (2019). Blue light is a universal signal for Escherichia coli chemoreceptors. J. Bacteriol. *201*, e00762–18.

Pigliucci, M. (2001). Phenotypic Plasticity: Beyond Nature and Nurture (Johns Hopkins University Press).

Plumbridge, J.A. (1991). Repression and induction of the nag regulon of Escherichia coll K-12: the roles of nagC and nagA in maintenance of the uninduced state. Mol. Microbiol. *5*, 2053–2062.

Schaum, C.E., and Collins, S. (2014). Plasticity predicts evolution in a marine alga. Proc. Biol. Sci. *281*, 20141486.

Scheiner, S.M. (1993). Genetics and evolution of phenotypic plasticity. Annu. Rev. Ecol. Syst. *24*, 35–68.

Stiffler, M.A., Hekstra, D.R., and Ranganathan, R. (2015). Evolvability as a function of purifying selection in TEM-1 -lactamase. Cell *160*, 882–892.

Travisano, M. (1997). Long-term experimental evolution in Escherichia coli. VI. Environmental constraints on adaptation and divergence. Genetics *146*, 471–479.

Travisano, M., and Lenski, R.E. (1996). Long-term experimental evolution in Escherichia coli. IV. Targets of selection and the specificity of adaptation. Genetics *143*, 15–26.

Travisano, M., Vasi, F., and Lenski, R. (1995). Long-term experimental evolution in Escherichia coli. III. Variation among replicate populations in correlated responses to novel environments. Evolution *49*, 189–200.

Velicer, G.J., and Lenski, R.E. (1999). Evolutionary trade-offs under conditions of resource abundance and scarcity: experiments with bacteria. Ecology *80*, 1168–1179.

Waddington, C.H. (1953). Genetic assimilation of an acquired character. Evolution *7*, 118–126.

Weickert, M.J., and Adhya, S. (1993). Control of transcription of gal repressor and isorepressor genes in Escherichia coli. J. Bacteriol. *175*, 251–258.

Wolfe, A.J., and Berg, H.C. (1989). Migration of bacteria in semisolid agar. Proc. Natl. Acad. Sci. U S A *86*, 6973–6977.

Yazyu, H., Shiota-Niiya, S., Shimamoto, T., Kanazawa, H., Futai, M., and Tsuchiya, T. (1984). Nucleotide sequence of the melB gene and characteristics of deduced amino acid sequence of the melibiose carrier in Escherichia coli. J. Biol. Chem. *259*, 4320–4326.

Yi, N., and Xu, S. (2008). Bayesian LASSO for quantitative trait loci mapping. Genetics *179*, 1045–1055.

iScience, Volume *23*

Supplemental Information

Evolution of Generalists by Phenotypic Plasticity

David T. Fraebel, Karna Gowda, Madhav Mani, and Seppe Kuehn

Supplemental Figures and Tables

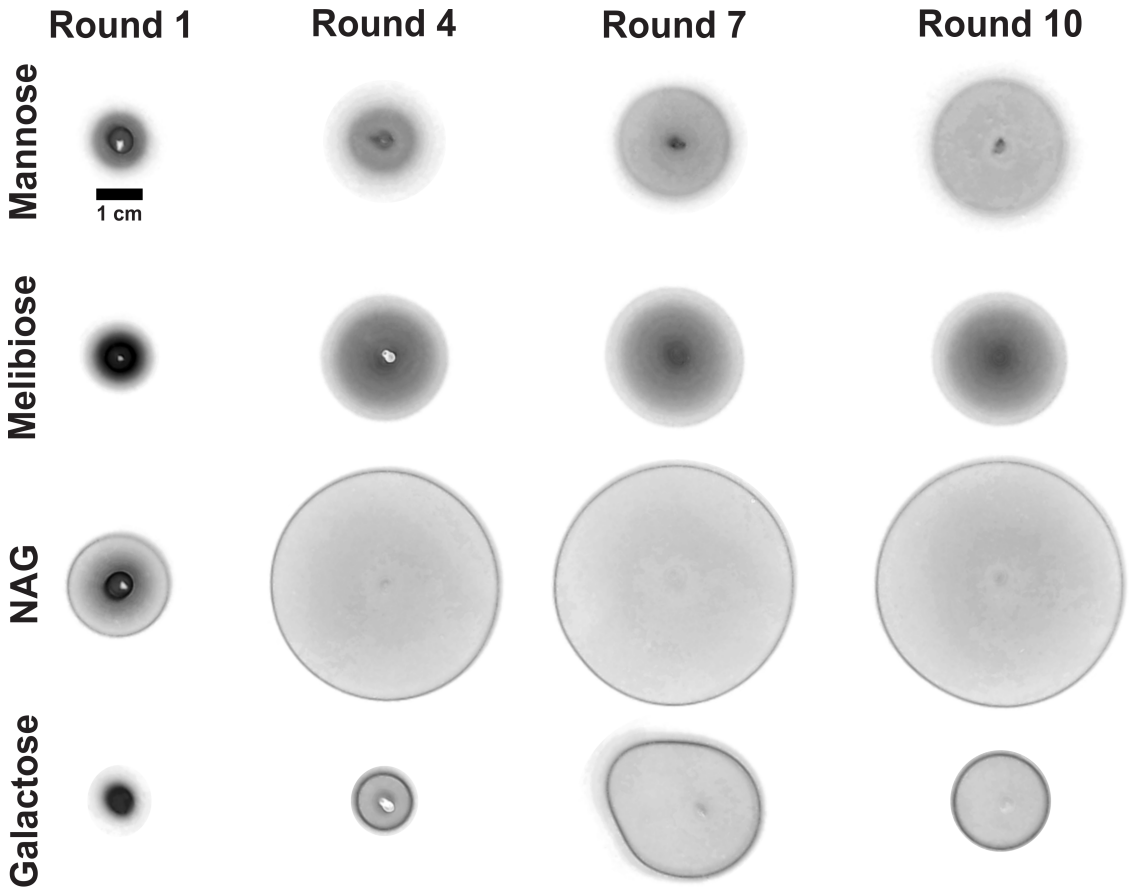


Figure S1: ***E. coli*** **colonies throughout selection experiment (related to Figure 1).** Example images of expanded colonies after 24 hours of migration at rounds 1, 4, 7 and 10 of the selection experiments described in the main text. 1 cm scale bar applies to all images. Darker regions correspond to higher cell density. Grayscale images were background subtracted, inverted, and had their dynamic range adjusted for better contrast. The asymmetry observed in galactose round 7 is due to inhomogeneity in the soft agar plate.

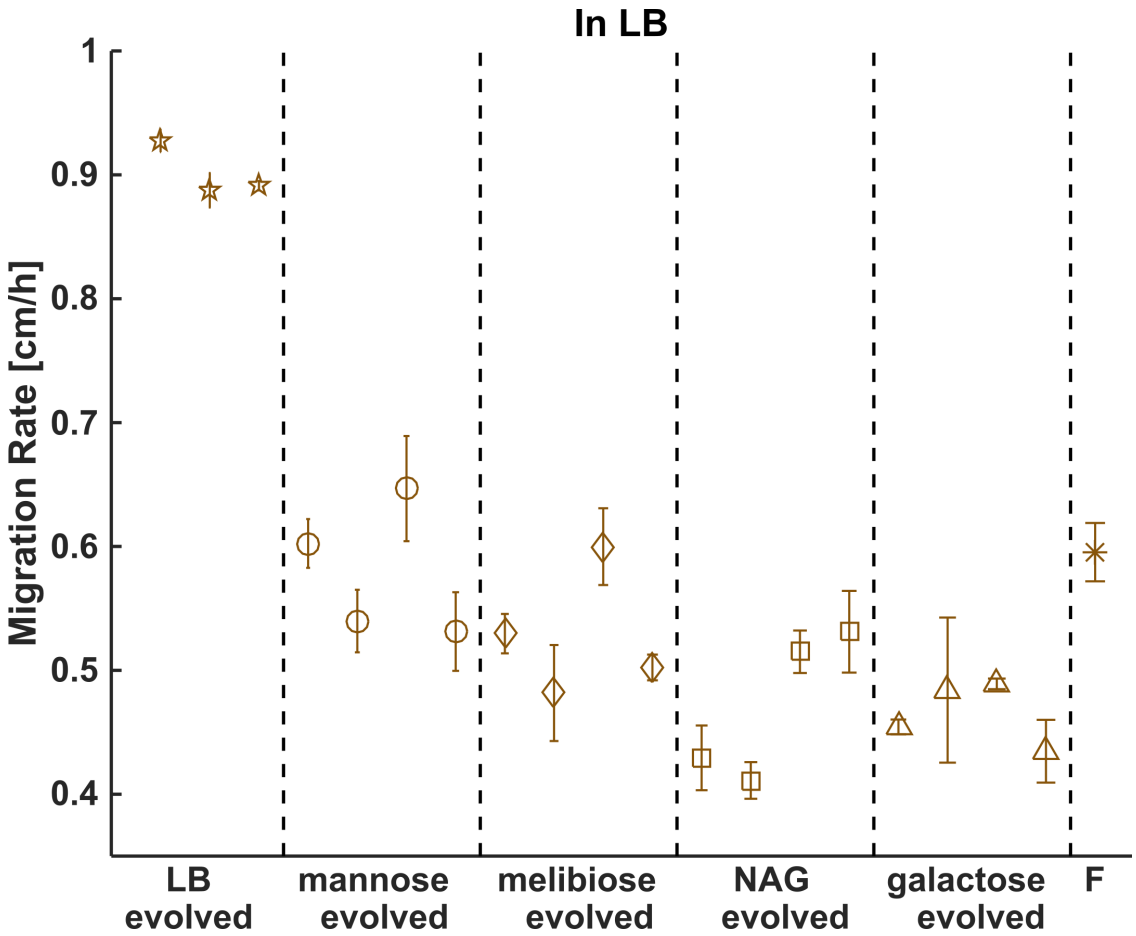


Figure S2: **Nutrient generality does not extend to rich medium (related to Figure 2).** The 16 strains isolated after 10 rounds of selection (four from each nutrient condition, main text Figure 1) were assayed for enhanced migration rate in LB rich medium. Migration rates of these strains are presented as mean ± standard error of two replicate plates. For comparison, we also measured migration rates of the founding strain (F) as well as three strains isolated after 10 rounds of selection in LB (Figure S3). Migration rates of these strains are presented as mean ± standard error of four replicate plates. Migration rate assays were conducted as described in main text methods, with the following exceptions: We used 10 cm petri dishes containing LB with 0.2 % w/v agar. Seed cultures were grown overnight in 5 mL LB and time-lapse imaging was performed every two minutes for eight hours. Only the first five hours were analyzed, since the LB-evolved strains reach the boundary of a 10 cm plate around this time. During image analysis, front location was determined by locating peaks in the radial density profiles.

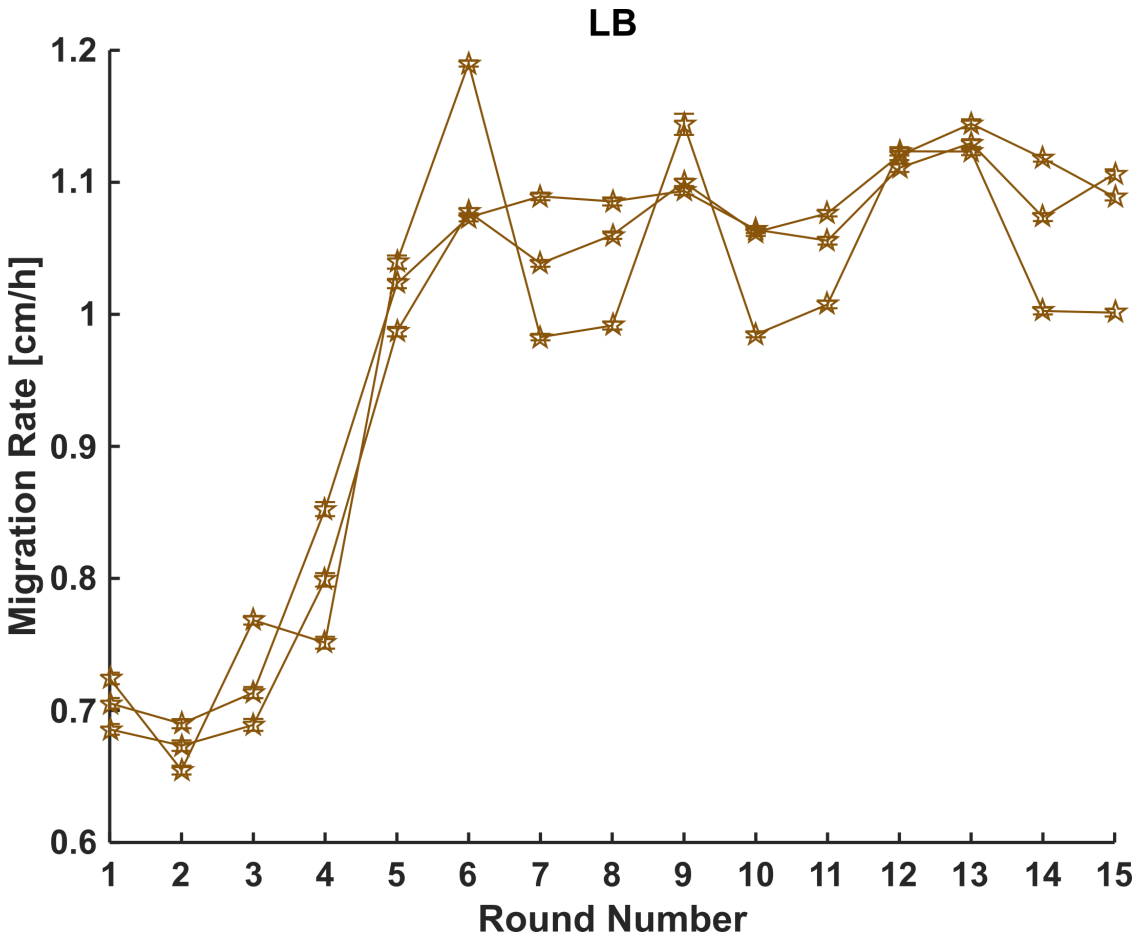


Figure S3: **Repeated selection enhances *E. coli* migration through soft agar in rich medium (related to Figures 1 and 2).** Migration rates as a function of round of selection for three replicate experiments performed in lysogeny broth (LB) rich medium. Selection and migration rate measurement were performed as described in main text methods, with the following exceptions: We used 15 cm petri dishes (containing LB with 0.2 % w/v agar) and selection was performed every 8 hours due to the fast migration rates in this condition. Seed cultures were grown overnight in 5 mL LB and time-lapse imaging was performed every minute. After imaging, four 50 µL samples were removed from the outermost edge of the expanding colony. During image analysis, front location was determined by locating peaks in the radial density profiles. These experiments were carried out to 15 rounds, however, the strains isolated after only 10 rounds were used for the comparison with 10-round minimal medium evolved strains described in the main text and presented below.

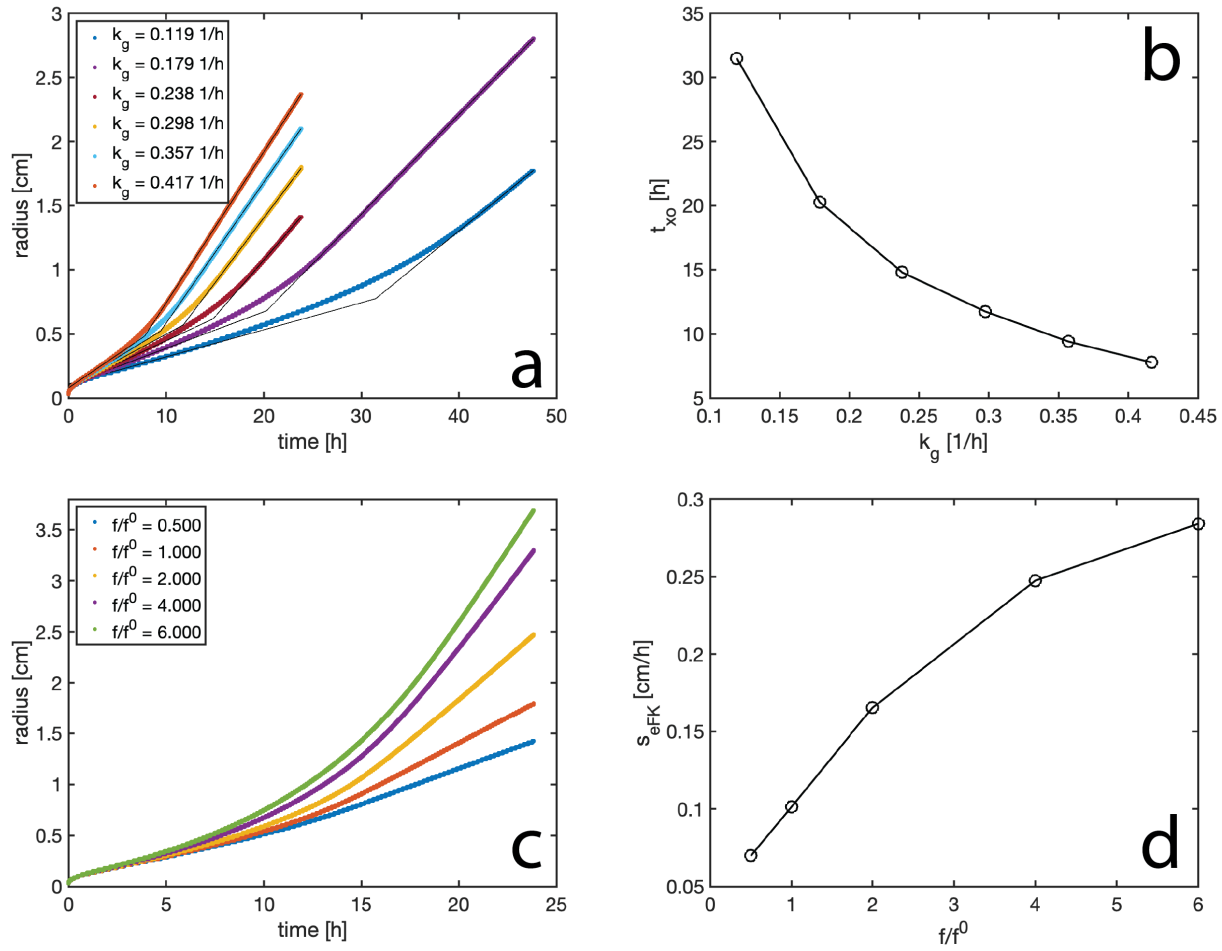


Figure S4: **Scaling of migration rates with phenotypic parameters (related to Figure 4).** Numerical simulations of the reaction-diffusion model from the main text support scaling arguments. All simulations were performed using phenotypic parameters for the founder strain in NAG as an example (see Methods). (a) Simulations for varying growth rates (k_g) showing colony radius versus time. Black lines are linear fits to early (s_{FK}) and late (s_{eFK}) times. The intersection of these two fits is used to measure the cross-over time (t_{xo}) from Fisher migration rates ($\sqrt{D_b k_g}$) to enhanced Fisher wave migration rates ($\sqrt{f D_b k_g}$). (b) Shows t_{xo} as a function of k_g, a fit to the data (not shown) gives $t_{xo} \propto k_g^{-1.1\pm0.05}$ as expected. (c) Same as in (a) except here the chemotactic coefficient f is varied by the factor shown in the legend relative to the starting value for founder in NAG (f^{fo} =15.3). Note the early migration rates are nearly constant (s_{FK}). (d) Shows the late time migration rate determined via linear regression as a function of f. A fit to the data (not shown) gives $s_{eFK} \propto f^{0.51}$ as expected from theory (Cremer et al., 2019, Main Text).

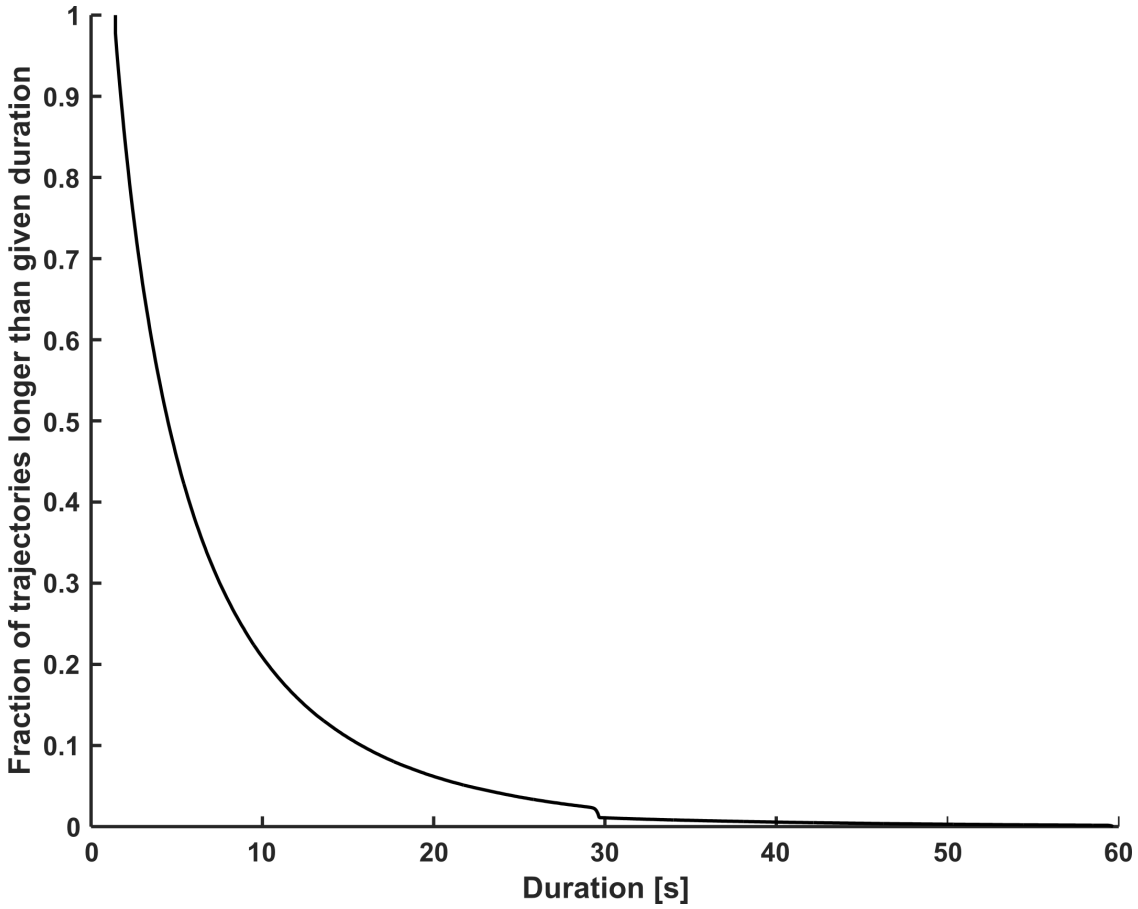


Figure S5: **Most trajectories do not extend beyond time interval used for fitting MSD (related to Figures 4 and 5).** Complementary cumulative distribution function of trajectory duration observed in all single-cell tracking experiments. To measure diffusion constant we fit mean-squared displacement over the interval from 1 to 6 seconds into the MSD trace (Figure 4a, main text).

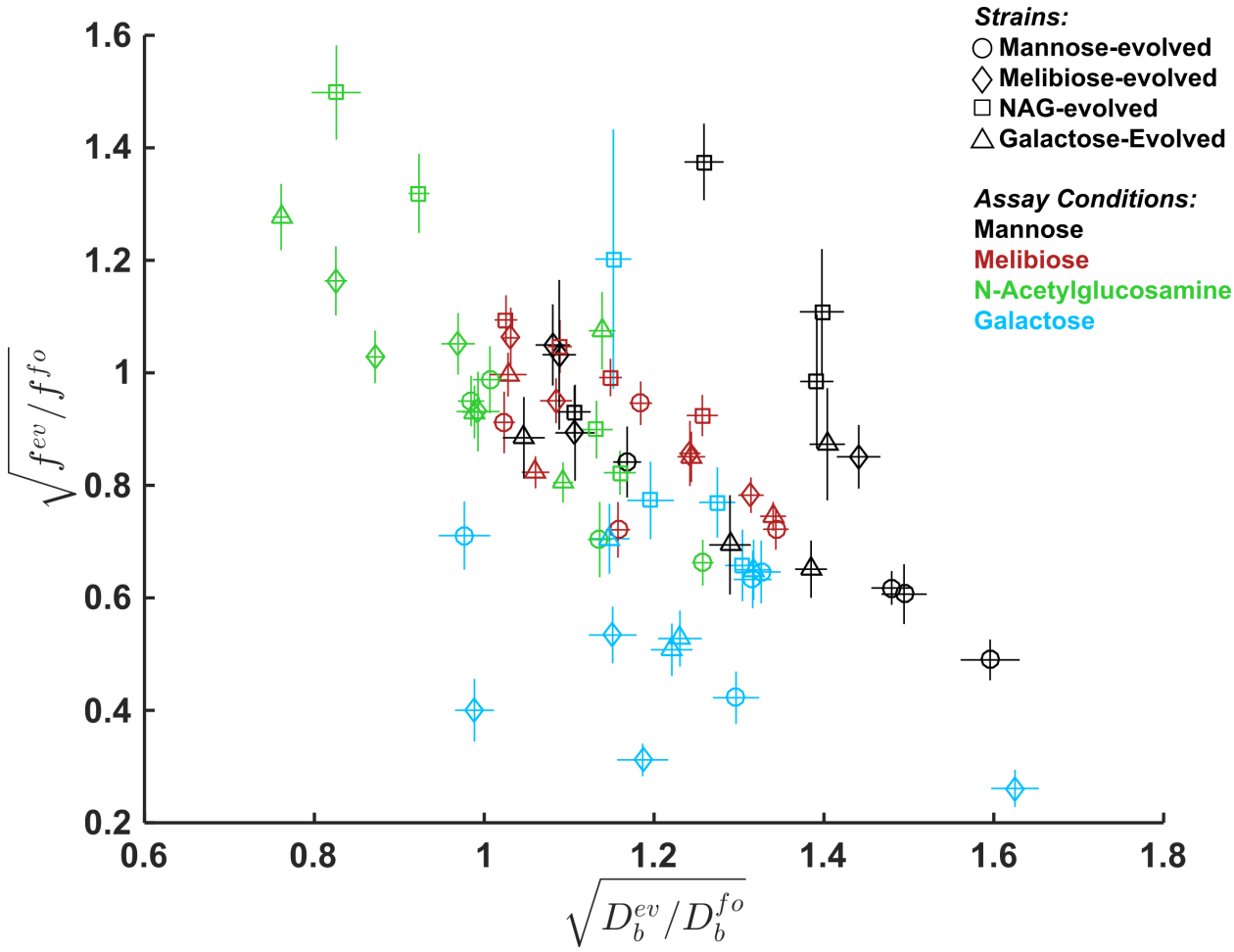


Figure S6: **Adaptation of f and D_b (related to Figure 6).** Plot identical to the two panels of Figure 6 in the main text except that it shows D_b versus f. For each strain in each condition, we compute the apparent contribution of chemotaxis to migration rate, $f = s_{eFK}^2/(D_b k_g)$. As with diffusion constant and growth rate (main text Figure 4), we calculate the contribution to migration rate enhancement attributable to evolution in this quantity, $\sqrt{f^{ev}/f^{fo}}$. Error bars are standard error calculated using error propagation.

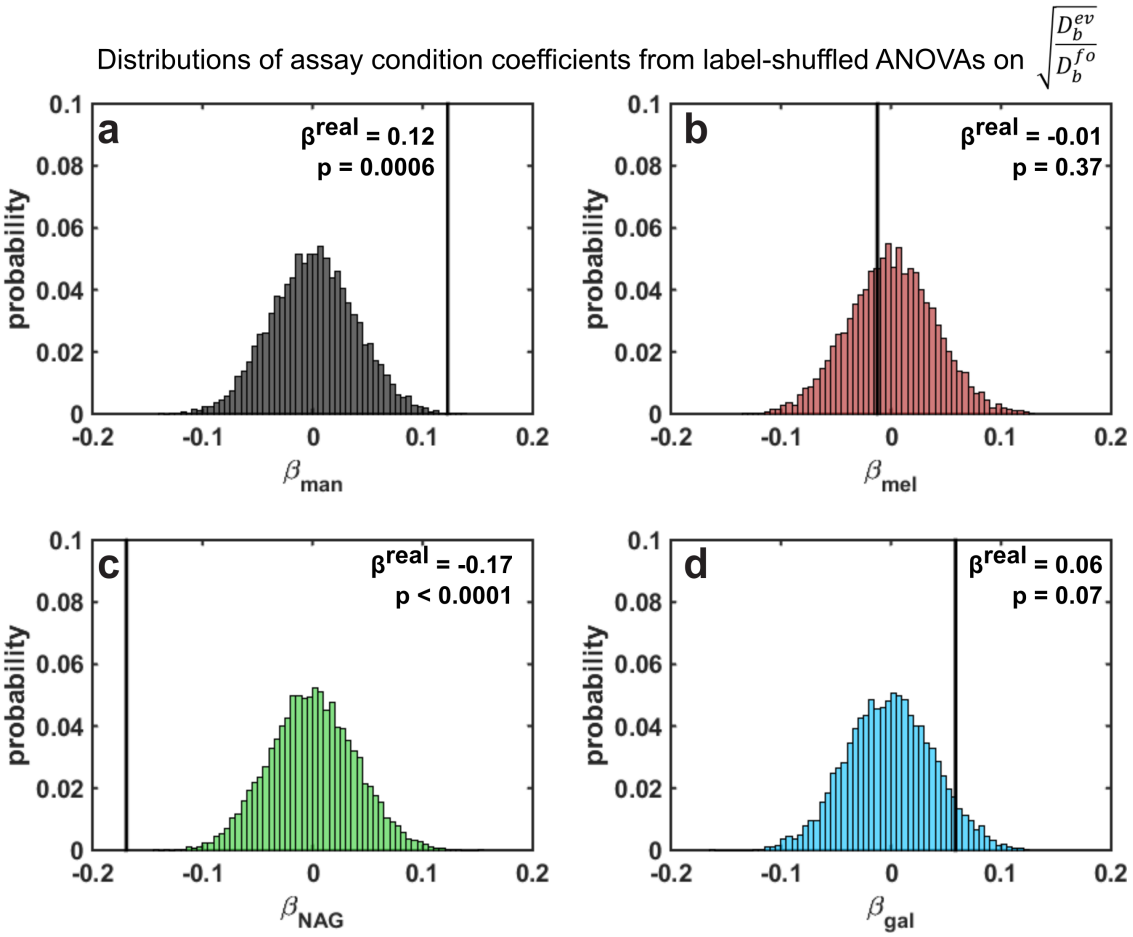


Figure S7: **Two assay conditions have significant effects on diffusion adaptation (related to Table 1).** Results of bootstrapping approach used to determine which assay conditions show significant ($p < 0.05$) departures from the global mean in $\sqrt{D_b^{ev}/D_b^{fo}}$, and in which direction. ANOVA coefficients from the properly-labeled data set are presented as β^{real} (vertical black lines). p-values are computed by determining the fraction of corresponding coefficients from label-shuffled ANOVAs that are higher or lower than β^{real}, depending on its sign. $p < 0.0001$ indicates none of the 10 000 ANOVAs on label-shuffled data had a more negative coefficient than β_{NAG}^{real} See main text methods for details.

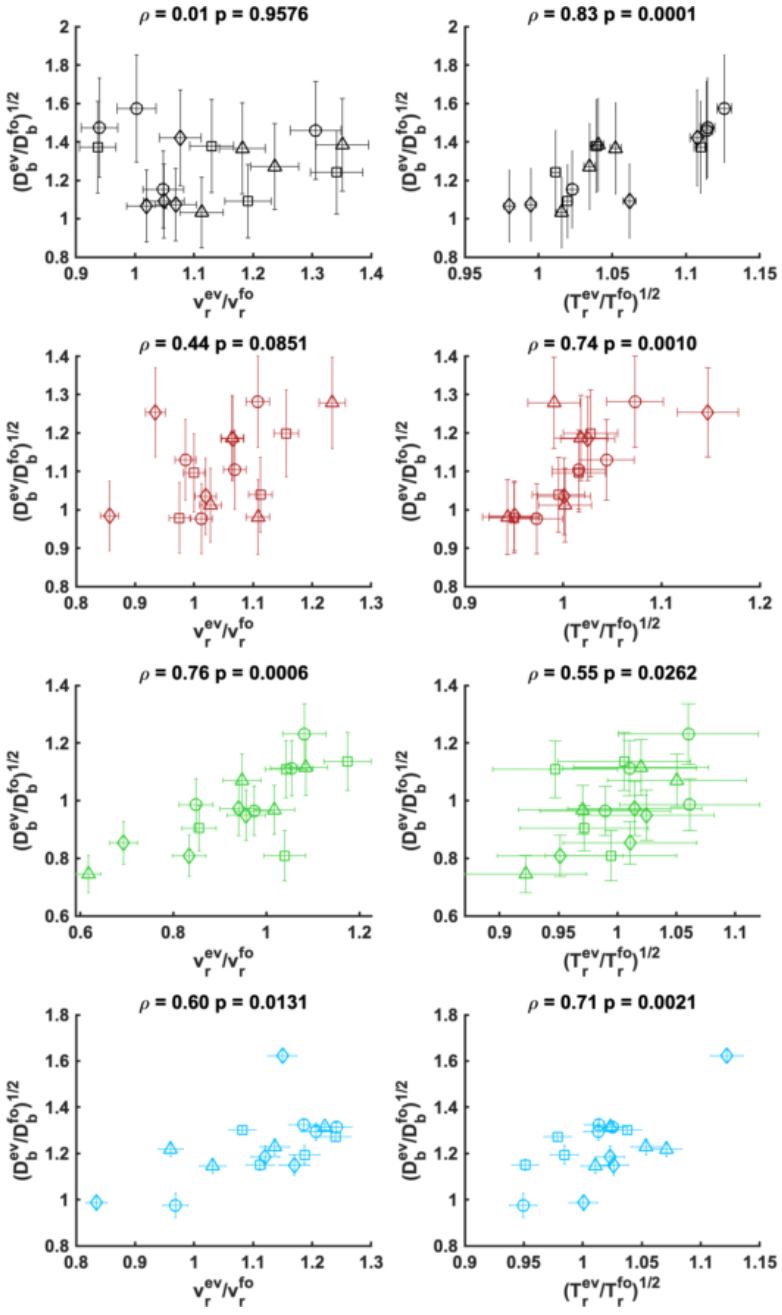


Figure S8: **Plasticity in D_b extends to run-tumble statistics (related to Table 1).** For each single-cell tracking experiment presented in main text Figure 4c, we detected runs and tumbles using the Hidden Markov Model classifier in Pytaxis (Perlova et al., 2019). For each detected run, we compute its duration and average speed. We average over all runs in an experiment to obtain mean run speed v_r and duration T_r of each strain in each condition. The founder has a run speed of $16.1 \pm 0.5, 17.4 \pm 0.3, 16.5 \pm 0.7$ and 16.8 ± 0.4 μm s^{-1} and a run duration of $0.29 \pm 0.002, 0.32 \pm 0.02, 0.31 \pm 0.03$ and 0.31 ± 0.007 seconds in mannose, melibiose, NAG and galactose respectively, mean $\pm$ standard deviation of two replicate experiments. Since $D_b \approx v_r^2 T_r$ to examine the contribution of v_r and T_r to the migration rate enhancement we correlated $\sqrt{D_b^{ev}/D_b^{fo}}$ with v_r^{ev}/v_r^{fo} and $\sqrt{T_r^{ev}/T_r^{fo}}$. Error bars are from error propagation. For each panel, we obtain a Pearson correlation coefficient and associated p-value which is shown in the title.

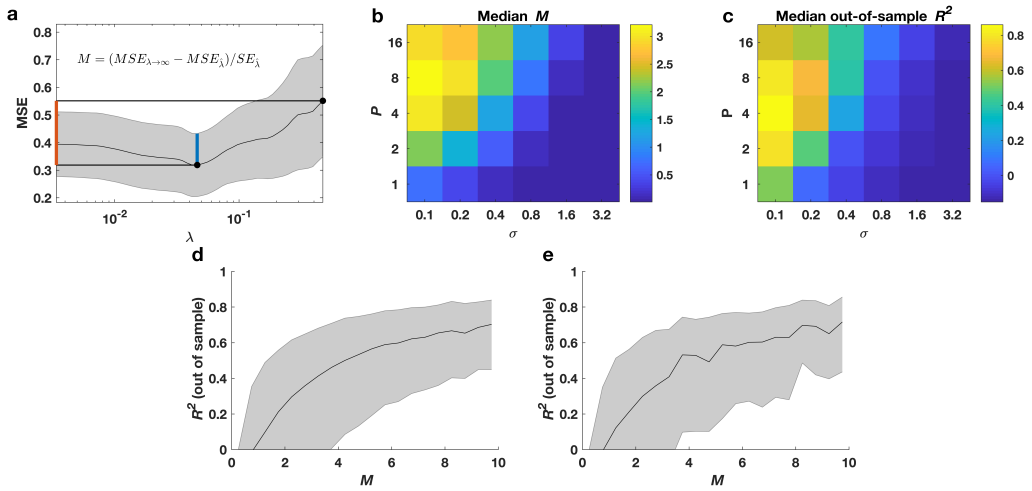


Figure S9: **Numerical investigation of LASSO regression demonstrates a correlation between a test statistic and regression performance (related to Figure 7).** (a) Schematic demonstrating the M statistic. M measures the improvement of the model that minimizes cross-validation MSE relative to the trivial model with only an intercept term (red line). This statistic is scaled by the estimated standard error of the MSE at its minimum (blue line) to incorporate uncertainty in the minimum MSE estimate. (b-c) Results of surrogate data simulations at different values of P (the number of true nonzero regression coefficients) and σ (the standard deviation of the noise term). Median values of M across 10^4 simulations per (P, σ) combination show that high M is achieved in a high signal-to-noise regime, i.e., when σ is sufficiently small and $P > 1$. Median out-of-sample R^2 values are also largest in this high signal-to-noise regime. (d) $M - R^2$ relationship from surrogate data. M and R^2 values from all 3×10^5 surrogate data simulations are combined, binned by M, and the quartiles of R^2 within the bins are shown as a function of M. The same quartile curves are shown in all three panels. (e) shows the same plot as (d) except using only 3×10^4 surrogate data simulations.).

Strain	Arabinose	Dextrose	Fructose	Lactose	Maltose	Rhamnose	Sorbitol
mannose 10A	2.8 ± 0.18	2.1 ± 0.14	1.7 ± 0.02	1.6 ± 0.11	2.0 ± 0.10	1.7 ± 0.30	2.0 ± 0.03
mannose 10B	2.3 ± 0.15	2.2 ± 0.03	1.8 ± 0.02	1.6 ± 0.06	2.1 ± 0.003	1.2 ± 0.12	1.9 ± 0.08
melibiose 10A	2.1 ± 0.18	2.1 ± 0.21	1.9 ± 0.20	1.5 ± 0.05	1.7 ± 0.01	1.5 ± 0.31	1.9 ± 0.05
melibiose 10B	2.4 ± 0.39	1.9 ± 0.01	2.2 ± 0.005	1.5 ± 0.04	1.7 ± 0.01	2.3 ± 0.29	1.9 ± 0.13
NAG 10A	3.5 ± 0.33	3.0 ± 0.09	2.7 ± 0.18	1.7 ± 0.14	2.1 ± 0.06	2.3 ± 0.08	3.1 ± 0.002
NAG 10B	3.7 ± 0.51	3.0 ± 0.27	3.0 ± 0.32	1.9 ± 0.01	2.0 ± 0.03	2.6 ± 0.04	3.0 ± 0.34
galactose 10A	2.6 ± 0.44	2.5 ± 0.001	2.6 ± 0.12	1.6 ± 0.02	1.8 ± 0.03	2.4 ± 0.41	2.4 ± 0.05
galactose 10B	1.9 ± 0.22	1.9 ± 0.04	2.0 ± 0.12	1.2 ± 0.04	1.5 ± 0.05	2.0 ± 0.04	1.8 ± 0.01

Table S1: **Nutrient generality extends to a variety of other sugars (related to Figure 2).** We assayed the migration rates of the ancestor as well as two evolved strains isolated after 10 rounds from each selection condition in a variety of sugars. Rates are presented as fold change compared to the founder's migration rate in the same condition, mean $\pm$ standard deviation of two replicate plates for each strain in each condition. The founder has a migration rate of 0.026 ± 0.001, 0.062 ± 0.005, 0.034 ± 0.004, 0.069 ± 0.002, 0.069 ± 0.005, 0.027 ± 0.002, 0.047 ± 0.004 cm h^{-1} in arabinose, dextrose, fructose, lactose, maltose, rhamnose and sorbitol, respectively, mean $\pm$ standard deviation of two replicate plates in each condition.

Assay Coefficient	$(\beta)^{real}$	p-value
AssayCond=man	-0.13	0.0038
AssayCond=mel	-0.15	0.0011
AssayCond=nag	0.0009	0.49
AssayCond=gal	0.27	$p < 0.0001$

Table S2: **Three assay conditions have significant effects on growth rate adaptation (related to Table 2).** Results of bootstrapping approach used to determine which assay conditions show significant ($p < 0.05$) departures from the global mean in $\sqrt{k_g^{ev}/k_g^{fo}}$, and in which direction. p-values are computed by determining the fraction of corresponding coefficients from label-shuffled ANOVAs that are higher or lower than β^{real}, depending on its sign. $p < 0.0001$ indicates none of the 10 000 ANOVAs on label-shuffled data had a more negative coefficient than β_{gal}^{real}. See main text methods for details.

Selection Coefficient	$(\alpha)^{real}$	p-value
SelectionCond=man	0.05	0.16
SelectionCond=mel	-0.13	0.0044
SelectionCond=nag	-0.02	0.36
SelectionCond=gal	0.09	0.05

Table S3: **Melibiose-evolved strains show below-average growth rate enhancement (related to Table 2).** Results of bootstrapping approach used to determine which selection conditions show significant ($p < 0.05$) departures from the global mean in $\sqrt{k_g^{ev}/k_g^{fo}}$, and in which direction. p-values are computed by determining the fraction of corresponding coefficients from label-shuffled ANOVAs that are higher or lower than α^{real}, depending on its sign. See main text methods for details.

Interaction Coefficient	$(\alpha\beta)^{real}$	p-value
SelectionCond=man x AssayCond=man	0.13	0.09
SelectionCond=man x AssayCond=mel	-0.02	0.42
SelectionCond=man x AssayCond=nag	-0.01	0.48
SelectionCond=man x AssayCond=gal	-0.10	0.15
SelectionCond=mel x AssayCond=man	-0.01	0.47
SelectionCond=mel x AssayCond=mel	0.08	0.19
SelectionCond=mel x AssayCond=nag	0.02	0.40
SelectionCond=mel x AssayCond=gal	-0.09	0.17
SelectionCond=nag x AssayCond=man	-0.05	0.32
SelectionCond=nag x AssayCond=mel	0.03	0.35
SelectionCond=nag x AssayCond=nag	-0.13	0.09
SelectionCond=nag x AssayCond=gal	-0.11	0.11
SelectionCond=gal x AssayCond=man	-0.07	0.25
SelectionCond=gal x AssayCond=mel	-0.09	0.17
SelectionCond=gal x AssayCond=nag	-0.14	0.06
SelectionCond=gal x AssayCond=gal	0.30	0.0013

Table S4: **Galactose-evolved strains have a 'home field advantage' in growth rate adaptation (related to Table 2).** Results of bootstrapping approach used to determine which selection x assay ($(\alpha\beta)_{ij}$) interaction terms show significant ($p < 0.05$) departures from the global mean in $\sqrt{k_g^{ev}/k_g^{fo}}$, and in which direction. ANOVA coefficients from the properly-labeled data set are presented as $(\alpha\beta)^{real}$. p-values are computed by determining the fraction of corresponding coefficients from label-shuffled ANOVAs that are higher or lower than $(\alpha\beta)^{real}$, depending on its sign. See main text methods for details.

Assay Coefficient	$(\beta)^{real}$	p-value
AssayCond=man	0.02	0.33
AssayCond=mel	0.06	0.16
AssayCond=nag	0.16	0.0012
AssayCond=gal	-0.24	$p < 0.0001$

Table S5: **Two assay conditions have significant effects on apparent chemotactic adaptation (related to Table 3).** Results of bootstrapping approach used to determine which assay conditions show significant ($p < 0.05$) departures from the global mean in $\sqrt{f^{ev}/f^{fo}}$, and in which direction. p-values are computed by determining the fraction of corresponding coefficients from label-shuffled ANOVAs that are higher or lower than β^{real}, depending on its sign. $p < 0.0001$ indicates none of the 10 000 ANOVAs on label-shuffled data had a more negative coefficient than β_{gal}^{real}. See main text methods for details.

Selection Coefficient	$(\alpha)^{real}$	p-value
SelectionCond=man	-0.12	0.01
SelectionCond=mel	-0.02	0.33
SelectionCond=nag	0.18	0.0005
SelectionCond=gal	-0.03	0.28

Table S6: **Two selection conditions have significant effects on apparent chemotactic adaptation (related to Table 3).** Results of bootstrapping approach used to determine which selection conditions show significant ($p < 0.05$) departures from the global mean in $\sqrt{f^{ev}/f^{fo}}$, and in which direction. p-values are computed by determining the fraction of corresponding coefficients from label-shuffled ANOVAs that are higher or lower than α^{real}, depending on its sign. See main text methods for details.

Targets	Mutations observed	Present in these strains
glyA	H165H	man10A, man10C, man10D, nag10A, nag10C, nag10D, gal10B
yegH	R335R	man10B, mel10B, gal10A
rpoB[12], rpoC[3]	P552L[1], I569L[2], +9bp[3]	man10D[1], nag10A[2], nag10C[3]
rssB	A280T	mel10A, mel10D, gal10B, gal10C, gal10D
mepS[1245] lpxT→ / →mepS[3]	IS5(–)+4bp[1], E171*[2] IS1(+)+8bp[34], IS1(+)+9bp[5]	mel10A[1], mel10B[2], nag10B[3] nag10C[4], nag10D[5]
glxK	P210L	mel10B, mel10C, mel10D
yeaR	Δ1::IS186(–)+6bp::Δ1	mel10B, gal10B
frdA	G393V	mel10B, gal10B
envZ	Δ36bp[1], Δ1bp[2]	nag10A[1], nag10D[2]
rph[13] pyrE← / ←rph[2]	Δ1bp[1], A→G[2] +A[3]	nag10B[1], gal10A[2], gal10D[3]
nagA	Δ1bp[13], A→C[2] Δ10bp[4]	gal10A[12], gal10B[34]
metK→ / →galP	C→A[1], G→T[2]	gal10C[1], gal10D[2]
yghG	E30*	mel10C
mokB← / →trg	G→T	mel10D
wzzE	IS1(–)+9bp	nag10B
yffR→ / →yffS	C→A	nag10D
osmC	D90N	gal10B
yncE→ / ←ansP	C→A	gal10C
yggI	G159G	gal10C
ligB	A142V	gal10D
nrfG→ / →gltP	G→T	gal10D

Table S7: **The set of mutations shared between strains with different evolutionary histories. Related to Figure 7.** Novel mutations present at frequencies of 20 % or greater in the 16 evolved strains presented in main text Figure 2. Whole-genome sequencing and analysis was performed as described in main text methods with an average coverage of 59.8 ± 11.2 (mean $\pm$ standard deviation across strains). We group mutations by target since some genes exhibit different mutations across strains and since some strains have mutations in intergenic regions adjacent to genes affected in other strains. For these cases, superscripts indicate which strains had which mutations, and whether they occurred in the coding region or the intergenic space (that is, superscripts are specific to each row of the table where they appear). This target-level grouping was used for the candidacy matrix used in the LASSO regressions. Strains are designated by their selection condition (mannose, melibiose, N-acetylglucosamine, galactose), rounds of selection (10 for these strains) and replicate (A,B,C,D, since four independent lineages were sequenced from each selection condition). Notation convention for mutations is described in the *breseq* documentation (`http://barricklab.org/twiki/pub/Lab/ToolsBacterialGenomeResequencing/documentation`).

evidence	*position*	*mutation*	*frequency*	*annotation*	*gene*	*# reads*
man10A (35.8x cvg)						
RA	2,685,013	G→A	36.30%	H165H (CAC→CAT)	glyA ←	44
man10B (55.4x cvg)						
RA	2,138,906	G→A	23.80%	R335R (CGG→CGA)	yegH →	38
man10C (50.9x cvg)						
RA	2,685,013	G→A	42.20%	H165H (CAC→CAT)	glyA ←	57
man10D (52.1x cvg)						
RA	2,685,013	G→A	100%	H165H (CAC→CAT)	glyA ←	50
RA	4,182,899	C→T	100%	P552L (CCG→CTG)	rpoB →	61
mel10A (52.2x cvg)						
RA	1,291,079	G→A	23.60%	A280T (GCG→ACG)	rssB →	29
JC JC	2,270,300	IS5 (–) +4 bp	26.20%	coding (322□325/567 nt)	mepS →	7//9
mel10B (52.1x cvg)						
RA	542,516	C→T	37.20%	P210L (CCG→CTG)	glxK →	35
JC JC	1,879,829	Δ1 :: IS186 (–) +6 bp :: Δ1	40.00%	coding (115□120/360 nt)	yeaR ←	6//26
RA	2,138,906	G→A	29.30%	R335R (CGG→CGA)	yegH →	41
RA	2,270,489	G→T	40.00%	E171* (GAA→TAA)	mepS →	35
RA	4,381,141	C→A	43.00%	G393V (GGT→GTT)	frdA ←	58
mel10C (57.9x cvg)						
RA	542,516	C→T	46.90%	P210L (CCG→CTG)	glxK →	34
RA	3,113,390	C→A	25.40%	E30* (GAA→TAA)	yghG ←	63
mel10D (71.1x cvg)						
RA	542,516	C→T	28.40%	P210L (CCG→CTG)	glxK →	39
RA	1,291,079	G→A	100%	A280T (GCG→ACG)	rssB →	48
RA	1,492,441	G→T	100%	intergenic (□312/□29)	mokB ← / → trg	41
NAG10A (64.3x cvg)						
RA	2,685,013	G→A	100%	H165H (CAC→CAT)	glyA ←	65
JC	3,535,134	Δ36 bp	79.40%	coding (700□735/1353 nt)	envZ ←	44
RA	4,182,949	A→C	100%	I569L (ATC→CTC)	rpoB →	98
NAG10B (54.7x cvg)						
JC JC	2,269,892	IS1 (+) +8 bp	38.20%	intergenic (+325/□80)	lpxT → / → mepS	20//6
RA	3,815,949	Δ1 bp	100%	pseudogene (601/669 nt)	rph ←	52
JC JC	3,969,335	IS1 (–) +9 bp	75.50%	coding (305□313/1047 nt)	wzzE →	33//47
NAG10C (75.8x cvg)						
JC JC	2,270,177	IS1 (+) +8 bp	44.80%	coding (199□206/567 nt)	mepS →	29//23
RA	2,685,013	G→A	100%	H165H (CAC→CAT)	glyA ←	77
JC	4,186,107	+9 bp	23.4%	coding (758/4224 nt)	rpoC →	15
NAG10D (64.2x cvg)						
JC JC	2,270,208	IS1 (+) +9 bp	47.10%	coding (230□238/567 nt)	mepS →	27//21
RA	2,564,500	C→A	23.10%	intergenic (+128/□23)	yffR → / → yffS	78
RA	2,685,013	G→A	100%	H165H (CAC→CAT)	glyA ←	68
RA	3,535,204	Δ1 bp	20.40%	coding (665/1353 nt)	envZ ←	54
gal10A (54.6x cvg)						
RA	702,332	Δ1 bp	100%	coding (420/1149 nt)	nagA ←	20
RA	702,339	A→C	100%	V138G (GTA→GGA)	nagA ←	25
RA	2,138,906	G→A	100%	R335R (CGG→CGA)	yegH →	43
RA	3,815,813	A→G	77.10%	intergenic (□45/+21)	pyrE ← / ← rph	35
gal10B (73.2x cvg)						
RA	702,330	Δ1 bp	48.30%	coding (422/1149 nt)	nagA ←	29
JC	702,404	Δ10 bp	16.90%	coding (339□348/1149 nt)	nagA ←	9
RA	1,291,079	G→A	25.10%	A280T (GCG→ACG)	rssB →	44
RA	1,556,892	G→A	46.30%	D90N (GAT→AAT)	osmC →	54
JC JC	1,879,829	Δ1 :: IS186 (–) +6 bp :: Δ1	39.40%	coding (115□120/360 nt)	yeaR ←	20//21
RA	2,685,013	G→A	51.90%	H165H (CAC→CAT)	glyA ←	81
RA	4,381,141	C→A	25.50%	G393V (GGT→GTT)	frdA ←	88
gal10C (78.4x cvg)						
RA	1,291,079	G→A	27.10%	A280T (GCG→ACG)	rssB →	45
RA	1,524,403	C→A	44.70%	intergenic (+35/+78)	yncE → / ← ansP	65
RA	3,088,201	C→A	53.10%	intergenic (+341/□83)	metK → / → galP	32
RA	3,090,231	T→C	39.80%	G159G (GGT→GGC)	yggI →	110
gal10D (63.4x cvg)						
RA	1,291,079	G→A	68.40%	A280T (GCG→ACG)	rssB →	41
RA	3,088,203	G→T	100%	intergenic (+343/□81)	metK → / → galP	34
RA	3,815,899	+A	59.20%	pseudogene (651/669 nt)	rph ←	76
RA	3,820,746	G→A	51.50%	A142V (GCA→GTA)	ligB ←	66
RA	4,294,169	G→T	22.40%	intergenic (+30/□312)	nrfG → / → gltP	49

Table S8: **All mutations present in evolved strains. Related to Figure 7.** Novel mutations present at frequencies of 20 % or greater in the 16 evolved strains presented in main text Figure 2. Whole-genome sequencing and analysis was performed as described in main text methods with an average coverage of 59.8 ± 11.2 (mean $\pm$ standard deviation across strains). Strains are designated by their selection condition (mannose, melibiose, N-acetylglucosamine, galactose), rounds of selection (10 for these strains) and replicate (A,B,C,D, since four independent lineages were sequenced from each selection condition). Notation convention for mutations is described in the *breseq* documentation (`http://barricklab.org/twiki/pub/Lab/ToolsBacterialGenomeResequencing/documentation`).

Transparent Methods

Migration rate assay and selection experiment

Plates were prepared by autoclaving agar into laboratory-grade water, cooling to 55 °C, then adding sterile stock solutions of M63 media components and the desired carbon source on a heated stir plate. The media had a final agar concentration of 0.2 % w/v and a final sugar concentration of 1 mM. 22 mL of media was added to a 10 cm petri dish and allowed to gel before being wrapped with parafilm and stored at 4 °C until use. Plates were thermalized for 24 h before use in the 30 °C environmental chamber where all migration experiments were conducted (Darwin Chambers).

All migration assays were initiated by growing 5 mL cultures of *E. coli* (strain MG1655-motile, Coli Genetic Stock Center #8237) overnight to saturation from frozen stock in 5 mL of liquid M63 with 30 mM of sugar matching the plate to be used. 10 µL of saturated culture was injected into the center of a soft agar plate. Time-lapse imaging was performed for 24 h at 5 min intervals on the expanding colonies using webcams (Logitech HD Pro Webcam C920) in a dark box with pulsed illumination provided by warm white LED strips (LEDMO SMD2835) around the periphery of each plate. Automated image analysis was used to extract migration rates as described below.

For selection experiments, eight 20 µL samples were removed from the outermost edge of the expanding colony after imaging. The sample was briefly vortexed and 10 µL was immediately injected into a fresh plate from the same batch to initiate the next round of selection. The remainder of the sample was preserved at -80 °C on 25 % glycerol. This process of imaging, sampling and inoculation was repeated until ten rounds of selection had been completed, at which point a final sample was taken and preserved.

Image analysis

Webcam-acquired images were processed by custom written MATLAB code. First, images were background subtracted by an image created from six early time points before growth had occurred. The colony's center was determined by applying Canny edge detection and a circular Hough transform to an image near the end of the experiment. Next, radial intensity profiles were constructed for each image along a line outwards from the colony's center. Local cell density is monotonic with pixel intensity. Location of the colony's edge was determined by applying a threshold to the intensity profiles, and s_{eFK} for evolved strains was determined by linear regression of the edge's position versus time during the final 5 hours of expansion. Calibration was performed by imaging a test target to determine the number of pixels per centimeter. To measure s_{FK} on evolved strains colony expansion data was inspected manually and a linear regression was performed on the front position in time as early as possible in the colony expansion process. In five cases, the cross over from s_{FK} to s_{eFK} was not obvious and therefore s_{FK} was not determined (strain:conditions - gal 10D:gal, gal 10C:gal, nag 10D:man,gal, nag10C:man). Due to errors in determining s_{FK} in evolved strains, estimating f via $f = (s_{eFK}/s_{FK})^2$, was prone to large uncertainties and therefore this approach was not used.

To reliably determine s_{FK} and s_{eFK} for the founding strain a 48 h colony expansion was recorded and the early and late migration rates were determined as descibed above. An example is shown in Figure 4.

Numerical simulations

Simulations of the model in Equations (1)-(3) were carried out using custom written MATLAB scripts with an approach identical to that described in a previous publication (Fraebel et al., 2017). Briefly, we assumed azimuthal symmetry and derivatives were approximated by central differences on a 1-D lattice. The rate of migration was determined by locating the bacterial front in space using a threshold on density. Linear regressions were then used to determine the migration rates. The scaling arguments in the main text regarding the migration rates s_{FK} and s_{eFK} are supported in the SI (Figure S5).

The model contains the following parameters for each strain: k_g, K_g, Y, f, D_b, ($\chi = fD_b$), A and B. K_g is small for *E. coli* consuming sugars and was fixed at 0.3 µM for all sugars and assumed not to evolve (value taken from a previous work, Fraebel et al., 2017). Optical density measurements in well-mixed conditions were used to measure k_g. Y was measured for the founder in all four conditions by serial dilution and plate enumeration of a stationary-phase culture grown in liquid media. We assume that Y is unchanged by

selection. D_b was measured via single cell tracking and f was inferred from the late time migration rates as described in the text. A and B were initially inferred by straddling the peak chemotactic response reported for each sugar in (Adler et al., 1973) by an order of magnitude in each direction. A was subsequently tuned to achieve migration rates from simulation which matched the founding strain (Figure 4) and then fixed for all other strains in the same nutrient condition.

Growth rate measurement

Seed cultures were grown from frozen stocks in 5 mL of M63 with 1 mM sugar for 36 hours to saturation. Cultures were then diluted $1:1000$ into a 48-well plate containing 1 mL of fresh media in each well. Optical density was measured in a plate reader (Tecan Infinite 200 or BMG Labtech CLARIOstar) by 600 nm absorbance every 10 minutes with 200RPM shaking between measurements. Maximum growth rates (k_g) were acquired by linear fit of the linear portion $log(OD)$ versus time (exponential growth phase) just before the roll-off to stationary phase. Fitting intervals were determined manually and were typically two to five hours in duration.

Single-cell tracking

Glass slides and cover slips were cleaned by sonication in acetone followed by 1 M KOH, passivated with 2 mg mL^{-1} bovine serum albumin and rinsed with laboratory-grade water before use. Cultures were grown from frozen stocks in 5 mL of M63 with 1 mM sugar to early-mid exponential phase, $OD600 \approx 0.14$. 5 µL of culture was added to the passivated region of a slide, covered with the passivated side of a cover slip, and the chamber was sealed with Devcon 5 Minute Epoxy. Videos of swimming cells were acquired for 30 or 60 seconds at 12 frames per second with a Point Grey model FL3-U3-32S2M-CS camera and a phase contrast microscope (Omano OM900-T inverted) at 10x magnification. Illumination was provided by a high-brightness white LED (LED Supply 07040-PW740-L) to avoid the 60 Hz flickering that observed with the stock halogen lamp. Experiments were performed in a 30 °C environmental chamber (Darwin Chambers). For each evolved strain in each condition, 5 videos were acquired from different sites on the glass slide. For the founder, this process was repeated for two independent replicate slides in each condition.

Videos were processed into trajectories and run-tumble classified using PyTaxis (`https://github.com/tatyana-perlova/pytaxis`) (Perlova et al., 2019). Briefly, this software segments videos to obtain cellular coordinates, links coordinates from each frame into trajectories and filters out trajectories of cells stuck to the glass slide as well as trajectories fewer than 20 frames. Subsequently, filtered trajectories were analyzed using custom MATLAB code. First, we calculated each cell's squared displacement at each frame. Next, we averaged over cells to obtain a mean squared displacement versus time trace for each strain in each condition. A five-second interval near the beginning of the trace was observed to be linear across datasets, indicating diffusive motion of the cells. The slope of this region was used to compute the diffusion constant (D_b) after dividing a prefactor of 4, since swimming cells were confined to two dimensions in this experiment. After this interval, the traces become sub-linear due to the presence of cells with lower diffusivity making up these very long trajectories (Figure 6a). Computing the distribution of trajectory lengths confirmed that such cells constitute a minority of the total trajectories (Figure S5).

Analysis of variance and post-hoc testing

We performed 3-way ANOVA on the model presented in Equation (4) using the MATLAB function anovan. We used either $\sqrt{k_g^{ev}/k_g^{fo}}$, $\sqrt{D_b^{ev}/D_b^{fo}}$ or $\sqrt{f^{ev}/f^{fo}}$ as the response variable, and selection condition, assay condition and strain as predictor variables. Selection condition, assay condition and the interaction term between those two predictors were all treated as fixed effects. Strain was treated as a random effect nested into selection condition. This serves as a strain-specific noise term which accounts for the lineage to lineage stochastic variation in phenotypes within one selection condition.

ANOVA results were used to determine which predictors have significant differences in the response variable between their levels. For predictors that showed a significant F-statistic, we further investigated

which coefficients showed significant departures from the global mean using a non-parametric bootstrapping approach. To do this, we first randomly permuted the response variable to create a randomly-labeled dataset. We then performed the ANOVA on the randomly-labeled data and computed the coefficients for the predictor being studied. We repeated the process of random permutation and ANOVA 10^4 times to construct a distribution of ANOVA coefficients from randomly-labeled datasets. Finally, we computed a p-value by considering the fraction of coefficients from shuffled data larger/smaller than the coefficients from ANOVA on the properly-labeled data. The intent of this approach is to highlight the probability that the particular arrangement of selection conditions, assay conditions and strains we observed in the phenotypic space defined by $\sqrt{k_g^{ev}/k_g^{fo}}$, $\sqrt{D_b^{ev}/D_b^{fo}}$ and $\sqrt{f^{ev}/f^{fo}}$ (Figure 6) could arise due to random chance. If an ANOVA coefficient from our data is unlikely given a distribution of coefficients from ANOVAs on randomly-labeled data, we conclude that the associated group has a meaningful departure from the global mean in whichever response variable is being investigated.

Whole genome sequencing and analysis

Cultures of each evolved strain were grown from frozen stocks in 5 mL LB overnight to saturation. Genomic DNA was purified using the Qiagen DNeasy UltraClean Microbial Kit. Input DNA was quantified by Qubit and sequencing libraries were prepared using the NexteraXT kit from Illumina. Purified, amplified libraries were quantified by Qubit and Bionanalyzer and normalized with the bead-based method of the NexteraXT kit. Sequencing was performed on an Illumina MiSeq system following pooling, dilution and denaturation of the bead-normalized libraries according to MiSeq system specifications. Reads were demultiplexed and adapter trimmed with the onboard Illumina software. Sequence data for the ancestral strain was used from our previous work, where we sequenced the founder with an average depth of 553X by aggregating reads from four separate sequencing reactions (Fraebel et al., 2017).

Analysis was performed using the *breseq* computational pipeline in polymorphism mode. Breseq uses an empirical error model and a Bayesian variant caller to predict mutations. The software uses a threshold on the empirical error estimate (E-value) to call variants (Barrick and Lenski, 2009). The value for this threshold used here was 0.01, and we report all mutations present in evolved populations at a frequency of 0.2 or above (Barrick and Lenski, 2009). All other parameters were set to their default values. Reads were aligned to the MG1655 genome (INSDC U00096.3). We note that breseq is not well suited to predicting large structural variation. We excluded any mutations present in the founder from further analysis in order to consider novel genetic variation present in the evolved strains.

Numerical investigation of LASSO regression

We sought to understand whether or not the presence of mutations in specific targets could be associated with phenotype changes in the evolved strains relative to the ancestor. We investigated the ability of L_1-regularized regression (LASSO) (Hastie, Tibshirani, and Friedman, 2008; Hastie, Tibshirani, and Wainwright, 2016) to predict the values of phenotypic parameters from mutational target candidacy using a numerical analysis of surrogate data. As discussed in the main text, preliminary work showed that some regressions yielded models with only an intercept term. We therefore designed a computational approach to assess when LASSO regression could reliably detect true nonzero regression coefficients.

We generated surrogate data from a model of the form $\vec{Y} = \eta_0 + X\vec{\eta} + \vec{\varepsilon}$. We varied two control parameters to generate these data, P and σ: P is the number of randomly-selected elements in the regression coefficient vector $\vec{\eta}$ that are drawn from a standard normal distribution with mean zero and variance of unity, with all other elements set to zero; σ controls the magnitude of the noise term $\vec{\varepsilon}$, whose elements are drawn from a normal distribution with mean zero and variance σ^2. In order to simulate the structure of the true mutation candidacy matrix, a predictor matrix X is generated by randomly shuffling along columns of the candidacy matrix, which preserves the number of observations ($N = 16$), the number of mutational targets (21), and the frequency of each mutation. Without loss of generality, we set the intercept $\eta_0 = 0$. From $\vec{\eta}$, X, and $\vec{\varepsilon}$, we obtain the surrogate response vector $\vec{Y}$. To generate additional data to be reserved for out-of-sample testing, we repeat the shuffling procedure to generate a new predictor matrix X_{OOS} with $N = 16$ observations, and

sampled a new noise vector $\vec{\varepsilon}_{OOS}$ to obtain response values $\vec{Y}_{OOS} = X_{OOS}\vec{\eta} + \vec{\varepsilon}_{OOS}$, while keeping the $\vec{\eta}$ fixed.

For each surrogate data set, we used LASSO regression (MATLAB R2017b) to fit a linear model of the form $\vec{Y} = \eta_0 + X\vec{\eta} + \vec{\varepsilon}$. LASSO minimizes the following objective function:

$$\min_{\eta_0, \vec{\eta}} \left\{ ||Y - \eta_0 - X\vec{\eta}||_2 + \lambda ||\vec{\eta}||_1 \right\}. \tag{1}$$

The LASSO procedure generated a set of models over a range of values for the regularization hyperparameter λ. Leave-one-out cross-validation was used to estimate the model mean squared error (MSE) as a function of λ. Model selection was performed by choosing the value of $\lambda = \hat{\lambda}$ that minimized the cross-validated MSE. At $\lambda = \hat{\lambda}$, LASSO regression produces estimates for the model parameters: $\hat{\eta}_0$, $\hat{\vec{\eta}}$. Additionally, we inferred a model with only an intercept (i.e., $\vec{Y} = \eta_0$, which is the model resulting from the limit $\lambda \to \infty$). We then constructed a statistic, M which measures the improvement of the model MSE for $\hat{\lambda}$ relative to the $\lambda \to \infty$ limit:

$$M = \frac{(MSE_{\lambda \to \infty} - MSE_{\hat{\lambda}})}{SE_{\hat{\lambda}}},$$

where $MSE_{\lambda \to \infty}$ is the MSE value of the model with only an intercept, $MSE_{\hat{\lambda}}$ is the MSE of the model inferred at $\hat{\lambda}$, and $SE_{\hat{\lambda}}$ is the estimated standard error of $MSE_{\hat{\lambda}}$ determined by cross-validation (Figure S9a). Model evaluation was performed by computing the coefficient of determination R^2 of the selected model applied to the out-of-sample predictor X_{OOS} and response $\vec{Y}_{OOS}$:

$$R^2 = 1 - \frac{\sum_i (y^i_{OOS} - \hat{y}^i_{OOS})^2}{\sum_i (y^i_{OOS} - \bar{Y}_{OOS})^2},$$

where i is an index over the 16 data points in the out-of-sample data set, $\hat{y}^i_{OOS} = \hat{\eta}_0 + x^i_{OOS}\hat{\vec{\eta}}$ and $\bar{Y}_{OOS} = \frac{1}{N}\sum_i y^i_{OOS}$ and $N = 16$ is the number of data points. Numerical evaluation of predictive power using out of sample surrogate data allowed us to test the ability of the fit to generalize beyond the training data.

We performed the steps above for a grid of P and σ values ($P \in \{1, 2, 4, 8, 16\}$ and $\sigma \in \{0.1, 0.2, 0.4, 0.8, 1.6, 3.2\}$), generating 10^4 instances of surrogate data for each (P, σ) combination. We chose these values of σ to sample both high and low noise regimes for the surrogate data. The average magnitude of the P non-zero entries in $\vec{\eta}$ is $\sqrt{2/\pi} = 0.79$, so σ values range from much smaller (0.1) to much larger (3.2) than the regression coefficients. These two limits on σ characterize low and high noise regimes respectively.

The results of these simulations are shown in Figure S9b-c. For each (P, σ) combination the heat map shows median values of M across the 10^4 instances of surrogate data. The values of M are highest in a low noise regime, corresponding to $P > 1$ and σ sufficiently small. Median out-of-sample R^2 values are also highest (> 0.5) in this low noise regime, suggesting that high values of M indicate situations where the LASSO procedure yields good inferences of the regression coefficients. To make the relationship between M and out-of-sample R^2 explicit, we combined the values of M and R^2 obtained in all surrogate data sets from all models ($n = 3 \times 10^5$ in total) and binned values by M in intervals of 0.5. Within bins, we computed the quartiles (25th, 50th, 75th percentiles) of R^2, which we plotted in Figure S9d. We observed that the relationship between M and out-of-sample R^2 is indeed monotonically increasing, with an interquartile range that decreases as M increases. Thus, as M increases, it is increasingly likely that LASSO has correctly inferred the linear model. We verified that the relationship between M and out-of-sample R^2 is essentially unchanged when one-tenth of the data are used (Figure S9e).

In our data we do not know P (the true number of genes predicting the phenotype) or σ since both are inferred by regression. However, we can compute M during cross-validation. We therefore use the statistic M, computed not on surrogate data, but on the true data (phenotypes and mutations) and use the measured values of M to assess the predictive power of our model.

We next performed the LASSO procedure on the experimental data with the true mutational target candidacy matrix X for three different response variables, the quantities $\sqrt{D_b^{ev}/D_b^{fo}}$, $\sqrt{k_g^{ev}/k_g^{fo}}$, and $\sqrt{f^{ev}/f^{fo}}$. Separate regressions were performed for all evolved strains in each of four nutrient conditions. Using these

regressions performed on the true data we then computed our statistic M, and these values are shown as colored vertical lines in Figure 7. As in the surrogate data simulations, for each real data regression there are 16 observations, 21 mutational targets, and leave-one-out cross validation is used for determining $\hat{\lambda}$. In each case, the value of M obtained is near zero, and the numerical simulations indicate that an out-of-sample $R^2 \approx 0$ is likely. We conclude that there is little predictive power in the presence or absence of specific mutations in terms of the evolved phenotypes.

Supplemental References

Adler, J., Hazelbauer, G.L., and Dahl, M.M. (1973). Chemotaxis toward sugars in Escherichia coli. Journal of Bacteriology *115*.3, 824–847.

Barrick, J.E. and Lenski, R.E. (2009). Genome-wide Mutational Diversity in an Evolving Population of Escherichia coli. Cold Spring Harbor Symposia on Quantitative Biology *74*, 119–129.

Hastie, T., Tibshirani, R., and Friedman, J. (2008). The Elements of Statistical Learning (Springer).

Hastie, T., Tibshirani, R., and Wainwright, R. (2016). Statistical Learning with Sparsity (CRC Press).