Low-dimensional genotype-fitness mapping across divergent environments suggests a limiting functions model of fitness

Olivia M. Ghosh[1,2,4],* Grant Kinsler[2,3], Benjamin H. Good[2,4,5], and Dmitri A. Petrov[2,5]†

[1] *Department of Physics, Stanford University, Stanford, CA 94305, USA*
[2] *Department of Biology, Stanford University, Stanford, CA 94305, USA*
[3] *Department of Bioengineering, University of Pennsylvania, Philadelphia, PA, USA*
[4] *Department of Applied Physics, Stanford University, Stanford, CA 94305, USA*
[5] *Chan Zuckerberg Biohub – San Francisco, San Francisco, CA 94158, USA*

(Dated: April 5, 2025)

A central goal in evolutionary biology is to be able to predict the effect of a genetic mutation on fitness. This is a major challenge because fitness depends both on phenotypic changes due to the mutation, and how these phenotypes map onto fitness in a particular environment. Genotype, phenotype, and environment spaces are all extremely complex, rendering bottom-up prediction unlikely. Here we show, using a large collection of adaptive yeast mutants, that fitness across a set of lab environments can be well-captured by top-down, low-dimensional linear models that generate abstract genotype-phenotype-fitness maps. We find that these maps are low-dimensional not only in the environment where the adaptive mutants evolved, but also in more divergent environments. We further find that the genotype-phenotype-fitness spaces implied by these maps overlap only partially across environments. We argue that these patterns are consistent with a "limiting functions" model of fitness, whereby only a small number of limiting functions can be modified to affect fitness in any given environment. The pleiotropic side-effects on non-limiting functions are effectively hidden from natural selection locally, but can be revealed globally. These results combine to emphasize the importance of environmental context in genotype-phenotype-fitness mapping, and have implications for the predictability and trajectory of evolution in complex environments.

A fundamental challenge in evolutionary biology is developing a predictive framework that links genetic variation to phenotypic traits and, ultimately, to evolutionary fitness (19, 58). One major obstacle is the sheer scale of genotype space, which is nearly infinite, with the total number of possible genotypes reaching hyperastronomical numbers (33). These genetic variants modify various low-level molecular phenotypes, which in turn affect additional higher-level phenotypes, eventually causing changes to cellular phenotypes such as maximum growth rate or survival under stress (17, 21, 38). Building bottom-up, causal chains across these phenotypic levels is extremely difficult given the vast network of interactions at every level of biological organization. When we further consider that phenotypes depend dynamically on each other, it is likely that phenotype space is just as complex as genotype space, if not more so. Furthermore, when considering evolution, our primary interest is fitness, and how genetic changes translate into phenotypes that influence fitness within the evolution environment (15). Hence, our mapping must also account for environmental and ecological variables, such as chemical and physical conditions, shifts in competition, and other ecological interactions, which are themselves combinatorially vast and dynamic. The environment also modulates how genotypes map onto phenotypes, and how phenotypes map onto fitness (4, 13, 46). From this angle, the problem of building a predictive model of evolution appears impossible due to biology's curse of dimensionality.

But this high-dimensional picture is at odds with the fact that evolution can be predictable, at both the phenotypic and genotypic levels (40). There are many empirical examples of phenotypic convergence and parallelism throughout the tree of life, such as the repeated evolution of wings, eyes, hydrodynamic and crab-like body shapes, and modes of metabolism in plants (31, 43, 44, 45, 59). Such parallelism is not only limited to the highest phenotypic levels of biological organization. For instance, many studies have demonstrated parallelism at the genetic level in microbial evolution experiments(30, 53, 55, 60). The same genes are repeatedly involved in independent bouts of adaptation in nature as well (51). Widespread parallelism and convergence that span biological scales and transcend large phylogenetic distances suggest that there exist low-dimensional structures in the genotype-phenotype-fitness map, because it is both navigable and has a finite number of redundant solutions.

Another piece of evidence that this map's dimensionality must collapse at some level is the fact that adaptive evolution often efficiently finds mutations that confer large fitness benefits in experimental evolution (56) and in nature (7, 11). Under a model of random pleiotropy, as used in Fisher's model of phenotypic evolution (16, 36, 37), large-effect pleiotropic mutations are likely to disrupt many more functions than improve them. This is known as the cost of complexity (12, 37). The solution to this tension between apparent high-dimensionality and evident low-dimensionality must involve a reduction in the number of pleiotropic effects of adaptive mutations. In other words, adaptive mutants must be modular at some level. The empirical

* Correspondence: omghosh@stanford.edu
† Correspondence: dpetrov@stanford.edu

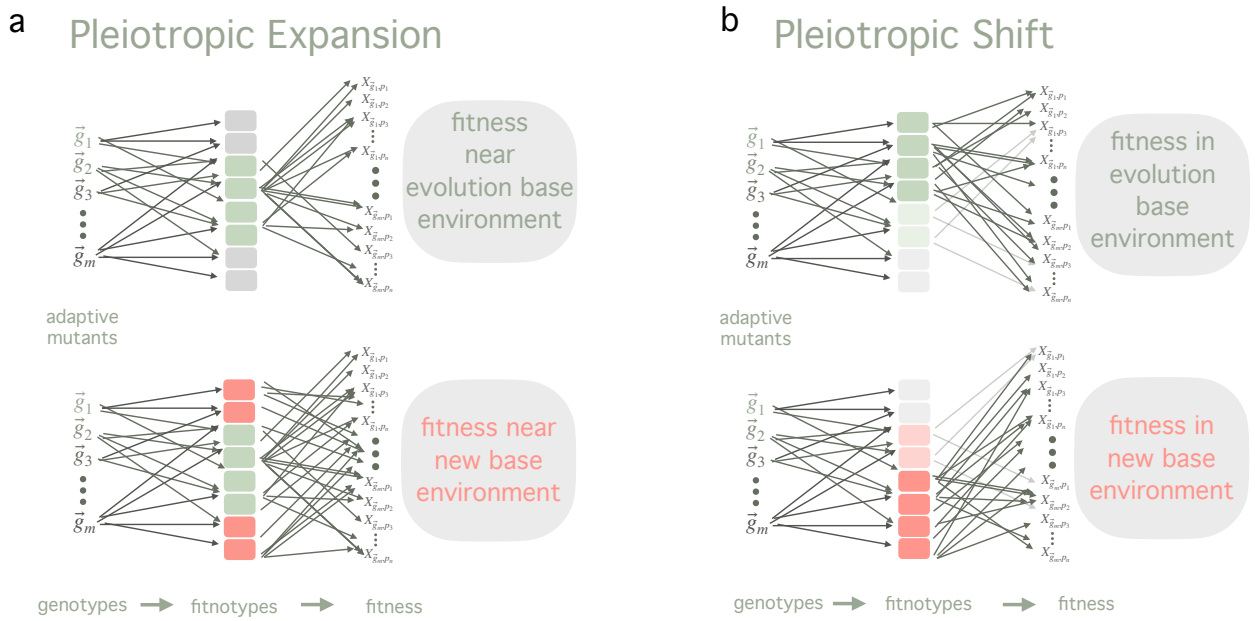

FIG. 1: **Two models for the nature of pleiotropy in adaptation** Left column represents genotypes (here, adaptive mutants). Right column represents fitness in a collection of environments. Middle column represents inferred structure of latent phenotype space from these left and right columns, or "fitnotypes." Colors correspond to environmental "weights." (a) Pleiotropic expansion would predict that adaptive mutants are generally pleiotropic and exhibit many diverse phenotypic effects in most environments, but are constrained to appear modular near their evolution environment, with a small number of fitnotypes determining fitness. (b) The pleiotropic shift model posits that adaptive mutants are still pleiotropic, but the dimensionality of their behavior is constrained in any environment, because the fitness function is dominated by just a few relevant, limiting challenges. These shift in importance in different base environments, and can partially overlap.

evidence, however, suggests that large-effect adaptive mutations are often pleiotropic and affect many distinct phenotypes (8, 10, 47, 48, 57). Indeed, many are in transcription factors affecting thousands of genes, or located in signaling pathways, known to be connected to many functions in the cell (9, 23, 24, 50, 51, 56). Therefore, large-effect adaptive mutants are unlikely to be strictly modular, with few phenotypic effects (3). There must instead be some alternative path towards effective dimensionality reduction in the mapping between genotypes, phenotypes, and fitness.

Consistent with this picture of emergent low-dimensionality, recent studies succeeded in building low-dimensional genotype-phenotype-fitness maps using top-down approaches. These approaches forego measurement of any specific phenotypes, and take advantage of advances in both the identification of genetic changes that drive fitness differences, and the measurement of fitness effects in high-throughput manner (2, 20, 22, 39, 42, 61). These approaches circumvent the difficulty of defining which phenotypes *should* be measured and instead attempt to infer the underlying phenotypic structure based on variation in fitness due to different genotypes, across different environments. Pan et al. used fitness data from human cell lines with genetic perturbations to infer the presence of latent biological functions that may be important to cancer (35, 39). Petti et al. studied a yeast cross in various lab environments and found that a low-dimensional linear model captured most of the relevant fitness variation, attributing it to variation in a few key biological "core processes" (34, 42). In both cases, the actual phenotypic traits or biological processes underlying fitness remained anonymous. Yet, by focusing only on traits relevant to fitness, these studies were able to infer structural properties of the genotype-phenotype-fitness landscape.

The first such approach applied to adaptation was taken by Kinsler et al., where they examined adaptive yeast mutants in a series of environments that varied in their similarity to the evolution condition (22). By introducing environmental perturbations, they analyzed patterns of fitness covariance among mutants and constructed a model in which genetic changes affected a set of latent phenotypes. They termed these quantities "fitnotypes," because they are fundamentally different from a measurable, traditional phenotype. Indeed, they are only detectable as latent variables that contribute to fitness variation, while phenotypic changes that do not affect fitness are hidden. These fitnotypes, in turn, determined fitness based on environment-specific weightings. If adaptive mutants are functionally identical because they evolved toward the same fitness optimum, they should exhibit congruent behavior in fitness space regardless of the number of environments tested, and hence appeear 1-dimensional. Conversely, functionally distinct mutants could reveal hidden phenotypic complexity when tested in new environments. Their model allowed them to infer the dimensionality of fitnotype space and predict relative fitness in novel environments. Kinsler et al. found that in environments similar to the evolution condition (i.e., subtle environmental perturbations), the mutants appeared low-dimensional (22). However, in more extreme environmental shifts, fitnotypes that had minimal contributions to fitness in the original

evolution condition grew to have substantial influence on fitness. This led them to reject a model of strict modularity and instead propose a model of fitness-relevant modularity, where adaptive mutants are globally pleiotropic - influencing many traits, but locally modular - influencing only a small number of traits that contribute to fitness in the evolution condition.

One explanation for this observation is the "pleiotropic expansion" model (Fig. 1a). In this framework, mutations selected in one environment are initially constrained to a low-dimensional phenotypic space. However, when placed in new environments, their pleiotropic effects become unconstrained, leading to an expansion in the number of phenotypes that influence fitness. If true, this suggests that evolution dynamically constrains mutants to explore the "flattest" paths in phenotype space within a given environment, eliminating the costs associated with affecting too many phenotypes. Yet, in novel environments, no such constraints exist, allowing the full phenotypic diversity (and likely the associated costs) of a mutation to be exposed.

An alternative explanation is the "pleiotropic shift" model (Fig. 1b), in which adaptive mutants always affect many phenotypes, but only a small subset of these phenotypes are relevant in any given environment. As environments change, the relative importance of different phenotypes shifts, reweighting their contributions to fitness. Under this model, genotype-phenotype-fitness maps are consistently low-dimensional, both in the original evolution environment and in other distant environments. Here, environmental constraints always impose low-dimensional structure, as fitness is always determined by a small number of dominant functional challenges in each environment. This model bears resemblance to previous work on linear pathway models, and is reminiscent of Liebeg's law of the minimum, which states that growth is limited by the first essential nutrient or metabolite to run out (6, 25, 29, 52, 54). In analogy to this work on limiting nutrients or limiting metabolites, we conceptualize this pleiotropic shift model as a "limiting functions model." Here, the set of functions that a cell must perform in order to thrive is large, but at any given time there will be a small number of functions that limit fitness. Thus, only a small number of "limiting functions" can be profitably modified in adaptation. Meanwhile the pleiotropic side-effects on non-limiting functions are effectively hidden from natural selection. Just as nutrients can shift from being in excess to being limited as the environment changes, so too can functions can shift between being non-limiting and limiting in new environmental contexts. To distinguish between these two models, we applied a set of environmental perturbations to three different base environments (one of which was the evolution environment) to construct distinct genotype-fitnotype-fitness maps, localized in environment space. Our findings ultimately reject the pleiotropic expansion model and strongly support the pleiotropic shift model. We interpret these results as evidence in favor of a "limiting functions" model of fitness. Environments impose a low-dimensional constraint on evolution, as each environment presents a limited number of dominant challenges that shape fitness outcomes. These results highlight the fundamental role of environmental context in shaping the genotype-phenotype-fitness landscape, providing new insights into how adaptive evolution navigates complex and fluctuating environments.

I. RESULTS

A. Hierarchical fitness assay of adaptive mutants generates local environmental neighborhoods in fitness space

In order to infer the phenotypic landscape of a set of adaptive mutants around three different base environments, we designed a fitness assay with hierarchical structure in both the genotype dimension and the environment dimension. To generate our dataset, we measured fitness across nearly 60 different lab environments for a library of DNA barcoded *Saccharomyces cerevisiae* mutants from previous evolution experiments (Fig. 2a) (1, 23, 27).

To measure the relative fitness of all mutants in the library for a single environmental condition, we utilize previously established methods that leverage DNA barcoding and high-resolution lineage tracking. We compete a pool of barcoded strains against a highly-abundant reference strain, the wild type (WT) ancestor, which takes up 95% of the culture. We quantify the frequencies of barcodes over time via amplicon sequencing. Based on these frequency trajectories, we can infer the relative fitness of thousands of strains (see Methods for more details) (5, 22, 56).

The mutations that differentiate the barcoded strains from the WT arose in several evolution experiments via serial dilution. One study, Levy et al, generated an initial set of adaptive mutants from a single WT ancestor that evolved in glucose-limited media with 48 hour dilution cycles ($T_{\text{transfer}} = 48$ hr) over 160 generations (27). Each mutant typically harbors one putatively causal genetic variant that improves fitness relative to the ancestor in this environment. Most of these mutations were either auto-diploidization events, or single nucleotide changes in genes in the Ras/PKA and TOR/Sch9 pathways, both nutrient-sensing pathways (56). In subsequent studies, four of the Ras/PKA mutants and one TOR mutant were re-barcoded and further evolved in two environments: the original 48 condition (hereafter, "2 Day") and a 24 hour condition (hereafter, "1 Day")(1, 23). Each of these genetic backgrounds and evolution conditions is represented as a row in the heatmap of Fig. 2(b). The cells complete fermentation of the glucose at around 20 hours and switch to non-fermentable carbon sources through respiration afterwards, so the 1 Day condition greatly limits the time the cells spend in respiration, amounting to a significant environmental shift despite unchanged media (28).

We generated a hierarchically-structured dataset

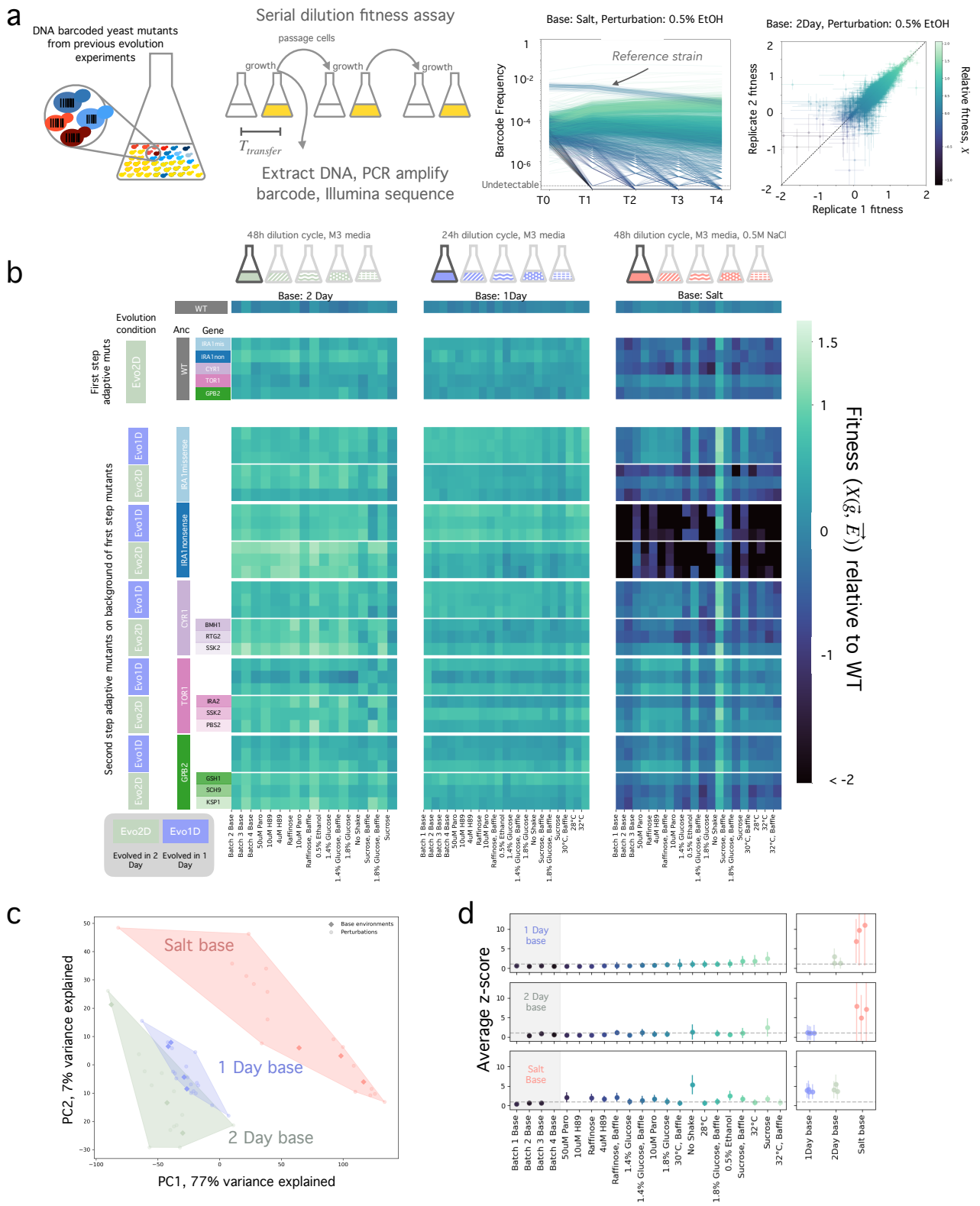

FIG. 2: **Fitness assay across hierarchical environment set enables comparisons across scales**(a) Schematic of the fitness assay. Yeast cells with DNA barcodes representing roughly 4000 different strains carrying adaptive mutations, from previous evolution experiments, are pooled together with the WT ancestor, which takes up 95% of the pool. The pool is serially diluted every 24 or 48 hours (T_{transfer}) into fresh media. At each transfer, the remaining cells are frozen for subsequent DNA extraction, PCR amplification, and Illumina sequencing of the barcode locus. The sequencing data allows for reconstruction of barcode-specific frequency trajectories through time, from which we can infer fitness with respect to the ancestor, in a highly replicable way (biological replicates shown in right panel). (b) Representative subset of genotype-by-environment fitness matrix is represented as a heatmap, with the fitness value $X(\vec{g}, \vec{E})$, mapped to a color. Columns are environment, and rows are mutants. The fitness assay was performed in a group of hierarchically structured environments, in which a set of identical environmental perturbations was applied to three distinct "base" environments ("2 Day" base, "1 Day" base, and "Salt" base). Genotypes comprised of adaptive mutations, either one step or two steps removed from the WT ancestor. Adaptive mutants were isolated from one evolution experiment in the 2 Day environment. 5 mutants were re-barcoded and re-evolved in the 1 Day and 2 Day environments. (c) Principal component analysis on the environmental covariance matrix of the fitness data. Each environment is projected onto the top two PCs, and the replicates of the unperturbed base environment are represented as saturated diamonds. (d) Average z-scores for each environment, calculated from each mutant's fitness mean and variance across the batch replicates of the unperturbed base environments. Colors are consistent across plots, where the order of the colors corresponds to the order of perturbations in the 1 Day base environment. Dashed gray line is a z-score of 1. Z-scores of within-base perturbations are measured with respect to focal base measurements (right panels). 1 Day and 2 Day appear similar, but both are distinct from Salt.

quantifying the fitness of our pool of adaptive mutants across three separate groups of environments. We chose three "base" environments around which to build our genotype-fitnotype-fitness maps: the two evolution conditions (2 Day, 1 Day) and one more extreme environment, "Salt," which has 0.5 M NaCl. Previous work showed that many of the single-step mutations exhibit strongly deleterious fitness effects in Salt (22). To each base environment, we applied a set of roughly 20 environmental perturbations (Fig. 2(b) columns represent different perturbations on three base environments). These changes included small alterations to incubation temperature (+/- 2°C), in glucose concentration (+/- 0.2%), or introducing sub-inhibitory amounts of drugs. Kinsler et al showed that such subtle environmental perturbations near the original evolution condition were enough to induce fitness variation containing signals of underlying latent phenotype space, so we quantified fitness across all ~ 60 environments (3 base environments $\times$ (17-20) perturbations). See supplemental information for more details about the environments.

Fig. 2b shows a representative subset of our data, where each row is a mutant and each column is an environment, grouped by base environment (either 2 Day, 1 Day, or Salt). The most obvious signal comes from the tradeoffs present in Salt conditions, where mutants that are adaptive in 1 Day or 2 Day are much more likely to be deleterious with respect to the ancestor in the Salt base. This genotype-by-environment fitness matrix will be the basis of our further analyses.

First, we characterized each environment as an m-dimensional fitness vector, where m is the total number of mutants. This mutant-embedding of environment space is helpful because it allows us to focus solely on environmental differences that matter to fitness for our focal set of adaptive mutations. We used principal component analysis (PCA) to project each environment vector into a two-dimensional space, onto the first two PCs. In Fig. 2c, each point is one environment (unique base-perturbation combination). We colored each environment according to its base, and found that each set of perturbations belonging to the same base remains mostly localized in this two-dimensional space. This is especially informative because the first two principal components explain nearly 82% of the variance in the data. 1 Day and 2 Day are more similar bases to each other, as was visually evident in the heatmap in Fig. 2b, and Salt is a more distinct base. Therefore, our set of environments provides a range of environmental scales across which we can compare fitnotype maps of adaptive mutants.

Due to the large number of environments in our assay, we performed our fitness measurement in batches of different perturbations. However, each batch contained one environment which was simply the unperturbed base, lending us internal batch replicates (diamonds in Fig. 2c). Differences between batch replicates reflect environmental factors that cannot be fully controlled experimentally, such as minor fluctuations in temperature and humidity, small inconsistencies in how the growth media were prepared, and biological noise. These batch replicates serve as a natural baseline of variation, which we can leverage to evaluate the magnitude of our environmental perturbations. To quantify these effects, we first measured the standard deviation across all batch replicates of the same base environment for a particular mutant. Then, for each mutant, we calculated a z-score in each perturbed environment by taking the absolute difference between its fitness in the perturbation with its mean fitness across the batch replicates of the corresponding base, scaled by inferred measurement error (see Methods). Finally, we took the average z-score across all mutants for that environment as a measure of the magnitude of the environmental perturbations, as "perceived" by the yeast cells and reflected in their fitness effects.

Fig. 2d demonstrates that most perturbations within a base environment typically result in average z-scores around 1 when compared to their batch replicates, indicating their very subtle effects on fitness. A notable exception is the "No Shake" perturbation in the Salt background, which differs significantly from the batch replicates of the Salt base. Whereas most mutants exhibit deleterious fitness effects relative to their ancestor in the Salt environments, in No Shake, most mutants appear to have a fitness advantage again. One explanation for this effect is that the ancestor (WT) is remarkably unfit in this Salt plus No Shake condition. Somehow, the adaptive mutations accumulated in 1 Day and 2 Day buffer the cells against the Salt when the flask is not shaking. This may hint at an interaction between osmotic pressure and oxygenation, and more generally suggests that environment-by-environment interactions may be present, and important, in determining the local phenotypic landscapes of adaptation.

B. Environmental perturbations generate base-dependent fitness effects, revealing ExE interactions

To further quantify these environment-by-environment interactions, we investigated the extent to which the same perturbation on different base environment backgrounds produced consistent changes in relative fitness. For example, does adding 0.5% extra glucose to M3 have a similar effect in a 24 hour dilution cycle and a 48 hour dilution cycle? And does adding salt disrupt the effect of additional glucose, if osmotic pressure exerts a more urgent challenge for the cells to respond to?

To isolate a perturbation's fitness effect on a particular mutant, we compared its relative fitness in the perturbed environment to its mean fitness across all batch replicates of the corresponding base, subtracting the base average from the perturbed measurement (henceforth, δX). Fig. 3b shows δX due to the same perturbation on different base environments, for one mutant with a mutation in the IRA1 gene. The x-axis represents δX on the Salt background, and the two y-axes are 1 Day and 2 Day δX. Each point is a comparison between bases for an individual perturbation, and

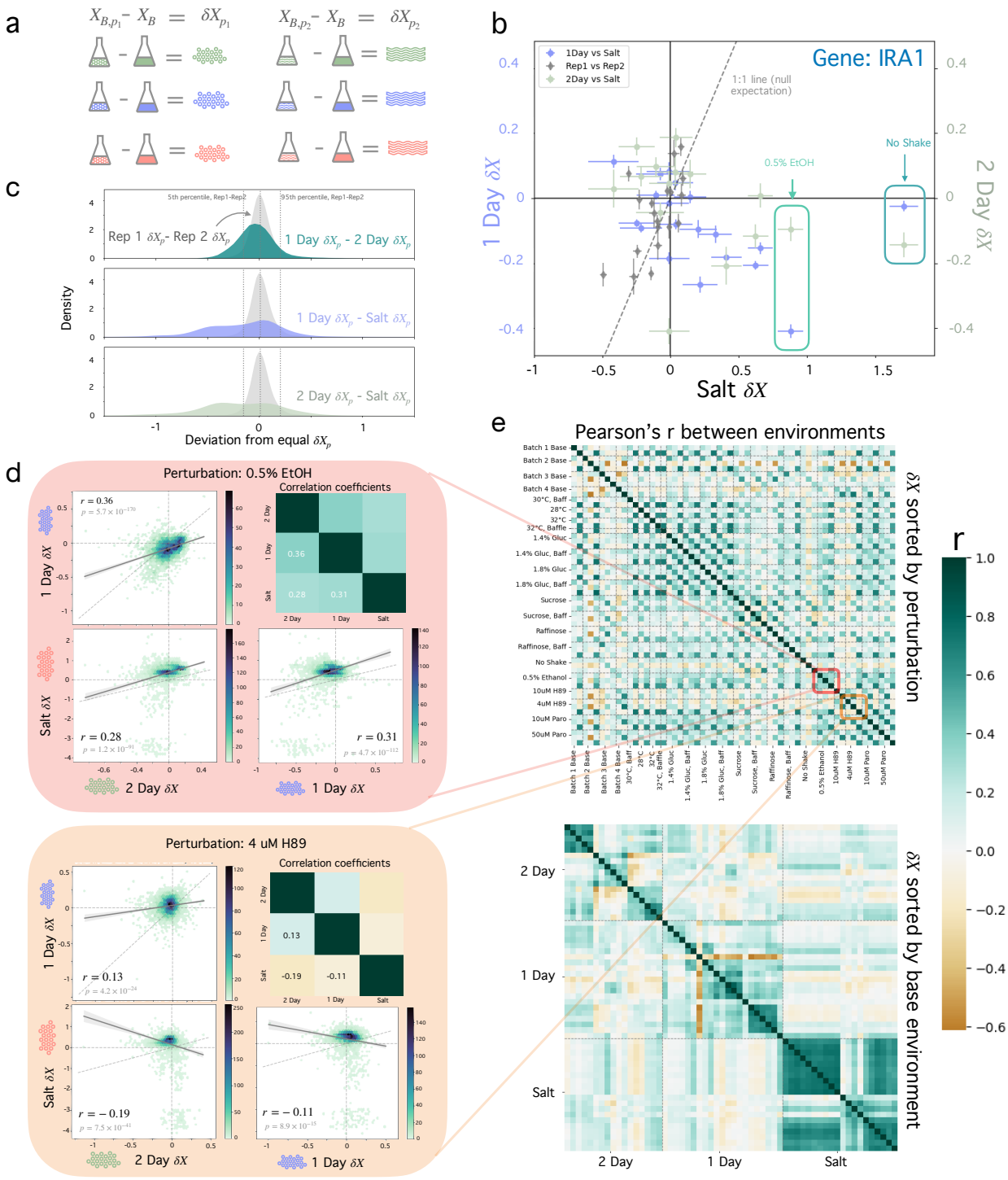

FIG. 3: **Widespread environment by environment interactions**(a) Schematic of delta fitness, using visual language from Fig. 2. To get the δX effect, we subtract each mutant's average base fitness from fitness in the perturbation environment. (b) ExE interactions for particular mutant (mutation in IRA1 gene). Each point is a perturbation, with 2Day δX vs Salt δX in green, and 1Day δX vs Salt δX in blue. Gray points represent δX between two biological replicates of the same environment, as a null expectation for deviations from 1-1 line. (c) Distribution of deviation from additivity, or the deviation from a 1-1 line in the plot in panel b, aggregated across all mutants and all perturbations. Different colors are different base-environment comparisons, and gray is a replicate-replicate comparison. Vertical lines are 5th, 50th, and 95th percentiles from rep-rep comparison, respectively. (d) 2 dimensional histograms of δX due to focal perturbation on two different base environments. Color of pixel corresponds to number of mutants in the bin. Top right shows correlation coefficient for each environmental comparison (e) Correlation matrix between environmental perturbations, clustered by perturbation (then batch) (top) and base environment (bottom). Block diagonal form is more apparent on the right, suggesting that the base environment is important for determining δX.

each perturbation is represented twice: once comparing δX in Salt and 1 Day (in blue), and once comparing δX in Salt and 2 Day (in green). The No Shake perturbation (right-most box) has little effect on top of the 1 Day and 2 Day bases, but has a highly beneficial effect on a Salt background, suggesting an interaction between salinity and oxygenation. The addition of 0.5% ethanol is also beneficial on a Salt background, but deleterious on a 1 Day background. Neither relationship shows significant correlation (1 Day-Salt p=0.07,

2 Day-Salt p=0.15, null hypothesis of no correlation).

Technical noise in the fitness measurement process could also be a source of uncorrelated δX, so we compared these base-to-base correlations with replicate-replicate δX correlations. We took two biological replicates from the same condition, and reasoned that any deviation from the 1:1 line could only be driven by noise, and could serve as a baseline for establishing signal of ExE interactions. We show these replicate-replicate correlations in Fig. 3b as gray points (Pearson's r=0.58, p=0.009), and find that for this mutant in IRA1, many perturbations far exceed the spread in replicate-replicate comparisons, confirming that apparent non-additivity of perturbations is driven by more than just measurement noise.

We next aggregated the deviation from additivity for all mutants and all environments and compared them to this null expectation. In Fig. 3c, we show the distribution of deviations from the 1:1 line for two biological replicates in gray, which is narrow and centered around 0, compared to deviations across different base comparisons. In contrast to the replicate-replicate deviations, the distributions corresponding to δX across bases are substantially wider, and the comparisons with Salt seem to have a bimodal distribution. We find that, in aggregate across all mutants, 30-70% of measurements showed δX deviations different than zero (32% for 1Day-2Day, 65% for 2Day-Salt, 69% for 1Day-Salt; FDR 10%). Hence, we see strong evidence for individual mutants often behaving unpredictably upon the "same" environmental perturbations.

We next wondered how widespread these ExE interactions were, and if they were dominated by particular perturbations. We compared fitness changes for each pair of base environments under the same perturbation, for all mutants. Despite the non-additivity of ethanol for the mutant in panel (b), panel (d) (top) shows that there are many mutants whose δX effects with ethanol are fairly predictable across base environments. Other perturbations generate fitness effects that are less correlated across background environments. For instance, adding 4 μM H89, an inhibitor of PKA, induces the opposite fitness behavior on background environments with and without 0.5M NaCl (Fig. 3d, bottom).

Because individual environmental perturbations vary in their degree of environment-by-environment interactions, we quantified patterns across the whole space of environments in our assay by calculating the Pearson correlation coefficient, r, of δX between two environments. If the effect of an environmental perturbation does not depend on its base environment, we would expect to see high correlations between δX vectors within-perturbation, and low or uncorrelated δX vectors between different perturbations. If instead the base environment plays a dominant role in determining the effect of a perturbation, then even though the effect of perturbations has effectively been isolated by subtracting out the base fitness in δX, we might expect to see large magnitudes of r for environments within the same base. We can visualize these two scenarios with a heat map of the correlation matrix between environments. In the base-independent scenario, a correlation matrix organized by perturbation would show strong block diagonal form, whereas a matrix organized by base environment would not. We organized the correlation matrix with the two separate partitions of the data (perturbation partitions and base partitions), and found the opposite result. There appears to be little structure in the matrix organized by perturbation, but when we cluster environments by base, we find a much stronger signal of block-diagonality, suggesting strong ExE interactions. Using a tool from network theory called the "modularity" score, Q, we quantified this block diagonality in each clustering strategy. Typically used to quantify the modularity of a network, Q can intuitively be understood as a measure from 0 to 1 of how much more density falls within a partition than would be expected of a random matrix, where the row sums are conserved. For the perturbation-clustered matrix (upper), Q=0.08, and for the base-clustered matrix (lower), Q=0.35. Therefore, we can conclude that across our collection of mutants, we see evidence for widespread ExE interactions. This suggests that our collection of mutants has distinct phenotypic effects in each of the base environments.

C. Emergence of GxExE interactions in simple fitness landscapes

These environment-by-environment interactions naturally emerge in a simple model of a fitness landscape. To see this, we assume that fitness is mediated through a low-dimensional space of coarse-grained environmental variables, ψ_k, where $k = 1, 2, ..., K$. Cells are not infinitely sensitive to the underlying physical and chemical properties of the environment, so it is reasonable to assume fitness might arise from this low-dimensional space. Likewise, we also see evidence for low-dimensionality in empirical work across many biological systems (22, 39, 42, 49). Therefore, we write the fitness of mutant m as a function of both its genotype, $\vec{g_m}$, and the coarse-grained environmental variables $\vec{\psi}$.

$$X(\vec{g}_m, \vec{E}) = X(\vec{g}_m, \vec{\psi}(\vec{E})) \tag{1}$$

We model the fitness effect of an environmental perturbation by decomposing the environment into two parts: the base environment, E_B, and the change in the environment due to the perturbation, δE_p. If the perturbation is small, we expand the fitness function to write the change in fitness due to an environmental perturbation as

$$\begin{aligned}\delta X_p(\vec{g}_i|\vec{E}_B) &= X(\vec{g}_i, \vec{\psi}(\vec{E}_B + \delta\vec{E}_p)) - X(\vec{g}_i, \vec{\psi}(\vec{E}_B)) \\ &\approx \sum_k \frac{\partial X}{\partial \psi_k}\bigg|_{\vec{\psi}=\vec{\psi}(E_B)} \cdot (\vec{\psi}(\vec{E}_B + \delta\vec{E}_p) - \vec{\psi}(\vec{E}_B))\end{aligned} \tag{2}$$

This equation is a generalized version of the standard linear fitness models proposed in quantitative genetics literature (26) , as well as an extension of the

model used in Kinsler et al. to assign fitness to adaptive mutants via a low-dimensional latent phenotype space (22). We can cast this equation for δX_P into more standard notation, where the partial derivative plays the role of a "trait" or a "phenotype," and the $\vec{\psi}(\vec{E}_B + \delta\vec{E}_p) - \vec{\psi}(\vec{E}_B)$ is a selection gradient. To simplify our notation, we can write

$$\delta X_p(\vec{g}_i|\vec{E}_B) \approx \sum_k \phi_{i,B,k}\beta_{k,B,p} \tag{3}$$

where $\phi_{i,B,k}$ is an effective phenotype that depends on the genotype and base environment, and $\beta_{k,B,p}$ is the selection gradient that depends on the base environment and perturbation.

Equation (2) shows that we should generically expect gxExE effects via the δX_p metric, unless the base environments are so close to each other that the $\vec{\phi}_{i,B}$ is actually base-independent, and the weights $\vec{\beta}_{B,p}$ are the same across bases. Thus, the observations in Fig. 3 of ExE interactions are consistent with this simple model, since we chose our base environments to be far from one another. We can infer that they are distant enough to induce widespread gxExE interactions. Even so, if the environmental *perturbations* are small, the linear approximation in Eq. (3) is still a good one. We thus cull the Salt + No Shake perturbation (which was not subtle according to Fig. 2d), and henceforth use a linear fitness model as in Eg. (3) to probe the quantitative properties of environmental dependence of phenotypic landscapes.

D. Environmental shifts away from evolution condition do not induce pleiotropic expansion

Having established that environment-by-environment interactions are prevalent in our data, we next sought to distinguish whether adaptive mutants fit the pleiotropic expansion or pleiotropic shift models presented in Fig. 1, using a linear model of fitness as in Eq. (3). Kinsler et al. proposed that adaptive mutations could have large fitness effects *and* be pleiotropic because they are modular near their evolution condition(22). In other words, a set of adaptive mutations might look the same in environments that are very similar to their evolution condition because they affect fitnotypes in the same way. But there are no guarantees for low-dimensional behavior amongst this set of adaptive mutants when their fitness is measured in new environments that are dissimilar from the evolution condition. Under this model, we would expect the fitness of a set of adaptive mutants to be well-approximated by a low-dimensional model in the evolution condition (and subtle perturbations). But as the environment shifts, new traits may become important for survival, so the previously hidden phenotypic effects may reveal themselves. Does this manifestation of hidden pleiotropy tend to drive higher-dimensional behavior of adaptive mutants in foreign environments?

To test this, we first focused on the adaptive mutants that arose on the WT background in the 2 Day base environment. We partitioned our environments by base and asked whether the same set of mutants yielded the same low-dimensional behavior around each base environment. Specifically, we wondered whether the 2 Day base would look different from the other two bases because it has the special distinction of being the evolution environment for this set of mutants. Under the "pleiotropic expansion" model, we expect the dimensionality of the mutant set to appear higher around Salt and 1 Day than 2 Day.

To infer the dimensionality of the latent space, we did singular value decomposition (SVD) and quantified the variance explained by each dimension. By definition, the first k SVD components generate the k-dimensional linear model with the lowest reconstruction error, and as k approaches the rank of the input matrix, the variance explained approaches 100% (14).

We compared the fraction of variance explained by each component across the three different base environments. Due to different levels of noise and measurement resolution, we have different limits of detection for each base. We compared the number of components that fall above the variance explained by the limit of detection. We obtained this limit by running SVD on 100 folds of a noise-only matrix, filled with 0-centered Gaussian random variables, with standard deviation equal to inferred standard error due to measurement noise in the fitness assay. As shown in Fig. 4b, the δX of the mutants that evolved in the 2 Day base are well-approximated with a 3 dimensional model in environments belonging to the 2 Day base. When the same set of mutants is measured in the same set of environmental perturbations, on the 1 Day base, the same number of dimensions fits. And the most distant environment, Salt, shows 4 dimensions. So for this particular set of mutants, we find that dimensionality in the evolution condition base, 2 Day, does not substantially differ from the non-evolution condition bases.

But this set of mutants is effectively one data point. To fill out this picture, we leverage the second-step mutants from subsequent evolution experiments, represented as lower rows in Fig. 2b. These mutations arose in both the 2 Day base and 1 Day base, so we now have two evolution conditions to compare, and to lend more evidence for or against the pleiotropic expansion hypothesis. These sets of mutants serve as biological replicates – they represent multiple instances of diversification from different genetic backgrounds, in two different environments. For each set of mutants, we can quantify their dimensionality in their evolution environment, and the other two bases. We visualize these comparisons as a scatter plot, with dimensionality in evolution condition on the x-axis, and dimensionality in distant environmental base on the y-axis (Fig. 4b). This panel shows our expectation under the pleiotropic expansion model, whereby dimensionality in the non-evolution base is systematically higher than the evolution base, for the same set of mutants. Here, we show two different quantifications of dimen-

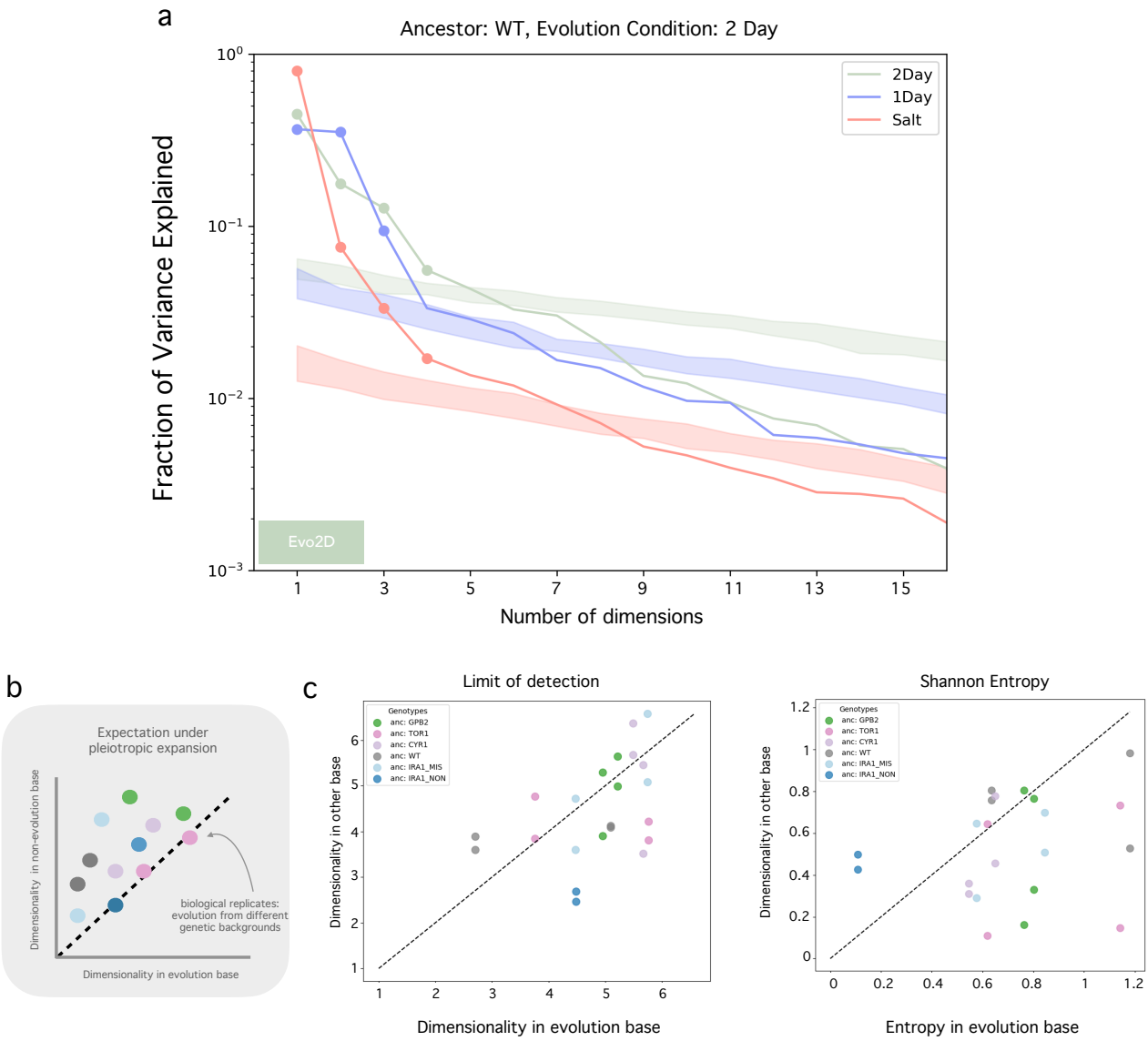

FIG. 4: **Pleiotropic expansion model is not supported by dimensionality reduction across all bases** (a) Fraction of variance in δX explained by component for each of the three environmental bases. Points are colored if they fall above limit of detection. Limit of detection was obtained by doing SVD on 1000-folds of noise-only matrices, using inferred measurement error as the standard deviation for the random variable in each fold. These mutants evolved in 2Day base, and have 4 dimensions above the limit of detection in 2Day, 4 in Salt, and 3 in 1Day. (b) If pleiotropic expansion were true, we should expect systematically higher dimensionality in a non-evolution condition base than an evolution condition, amongst a set of biological replicates (ie independent evolution experiments). (c) Using two metrics for dimensionality (number of dimensions that fall above the limit of detection, and Shannon entropy), we find no evidence for pleiotropic expansion. If anything, the dimensionality in the evolution base tends to be slightly higher.

sionality. In the left panel, we count the number of components that fall above the limit of detection. In the right panel, we quantify the Shannon entropy of the distribution of variance across the different components that fall above the limit of detection, which takes into consideration the "steepness" of the scree plot.

Overall, we find no evidence for the pleiotropic expansion hypothesis. In general, there is no clear relationship between dimensionality and whether a group of mutants evolved to the base or not. In fact, the scatter plots indicate that the evolution condition appears to be higher-dimensional more often than the other bases, but more replicates would be necessary to confidently make such a claim. We can claim, however, that a similar number of dimensions is necessary to build fitnotype maps for each set of mutants, regardless of where they evolved.

E. Pleiotropic reweighting of important latent dimensions across environmental bases

Having discovered a similar number of independent latent phenotypes to be important in each base environment, we next wondered the extent to which different fitnotype spaces overlap. To quantify the overlap in fitnotypes across different base environments, we assess the accuracy of a model trained on δX in one base, predicting δX in another base. It is difficult in general to compare latent spaces, but the accuracy of prediction is invariant to the choice of basis in the matrix factorization, so it is a useful metric for this task. If the fitnotype spaces fully overlap, we will find

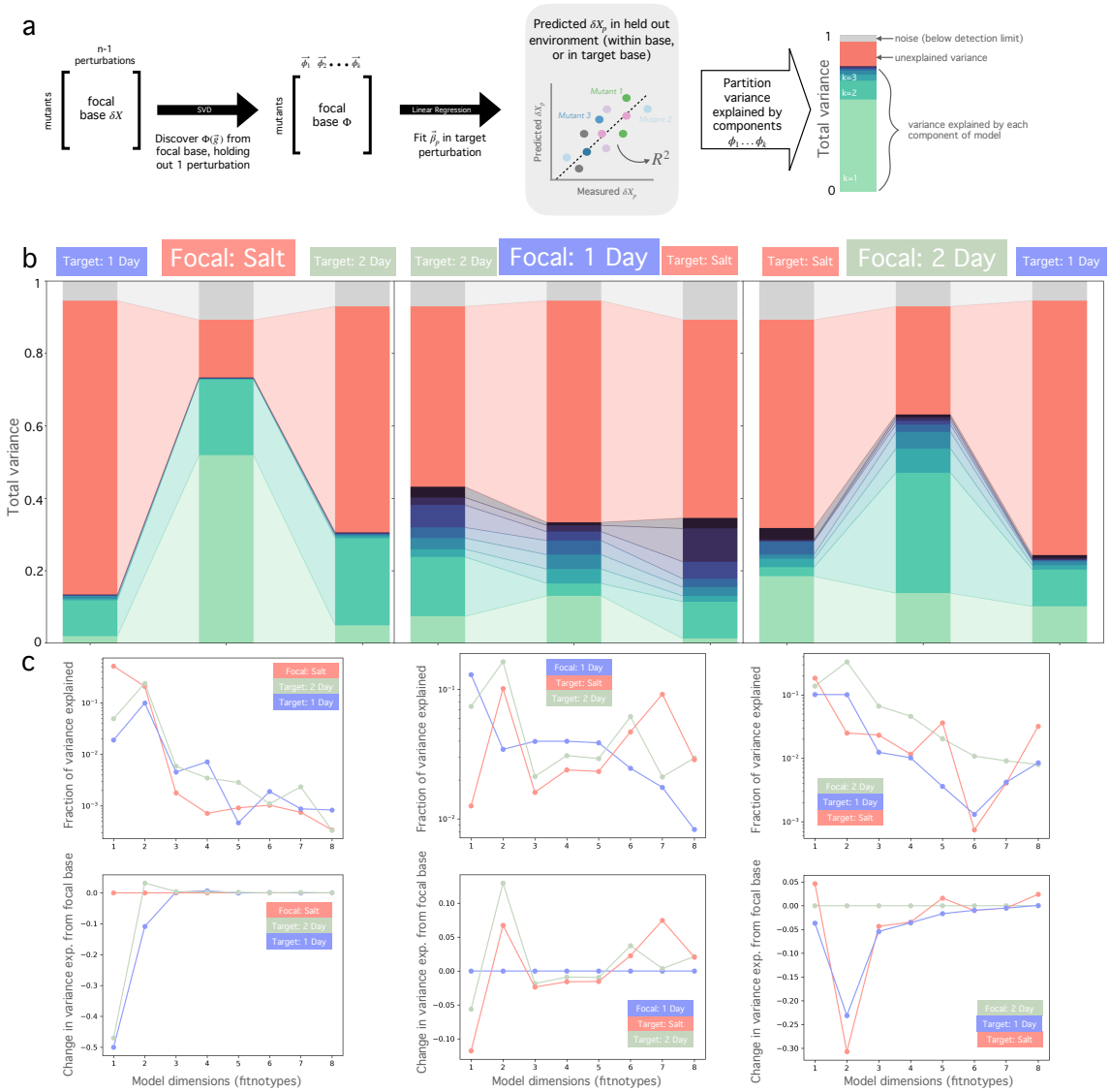

FIG. 5: **Fitness prediction within and across base environments imply a pleiotropic shift** (a) Schematic for prediction approach, and explanation of visualization. Prediction is quantified via coefficient of determination. Total variance across all held-out perturbations is partitioned into variation outside our limit of detection (in gray), unexplainable variation due to non-linear effects or heterogeneity in perturbation effects (in salmon), and explained variation in blue and green tones. Explained variation is further partitioned into the fraction explained by each component (because they are orthogonal), which allows us to quantify their "weight" in determining fitness. (b) Model performance for focal base and target bases, across all three focal bases. Focal base is always in the center bar, and results from target bases flank. Coefficient of determination is aggregated across all held-out perturbations. (c) Top row shows fraction of variance explained by model dimension for each focal base and target bases. Bottom row is difference in fraction variance explained by model dimension between target base and focal base.

equivalent predictability for environments within base and outside the base. In contrast, if fitnotypes are orthogonal, prediction accuracy will drop precipitously across base environments. To test this in practice, we use SVD to "discover" the fitnotypes in one base (focal base) from δX in each perturbation, inferring the $\vec{\phi}$ vectors from Eq. (3) for a set of mutants (Fig. 5a). Then, we use these $\vec{\phi}$ vectors as features in a linear regression model, and fit $\vec{\beta}$ parameters for a held-out test perturbation. We are essentially asking: if fitnotypes are entirely overlapping across base environments, and we allow our model to freely reweigh these fitnotypes to best predict fitness in a new base, how close can we get to within-base predictability? We quantify predictive power for the test perturbation across the whole group of genotypes, and generate a coefficient of determination, or R^2, to assess the model's success. We partition the explained variance by component to understand how much each fitnotype contributes to our ability to predict fitness in a particular base, and represent these fractions as blue-green tones in the stacked bar plot representing total variance (Fig. 5a, right). We chose to use an 8 component model for the full set of genotypes in our dataset to give the model extra flexibility, but clearly in some cases the effective dimensionality is much lower. The gray fraction of the bar plot corresponds to the unexplained variation in an ideal 8-dimensional model, obtained through SVD on the base itself (with no environments held out). We represent unexplained variation as a salmon-colored fraction. This represents fitnotypic variation that we cannot predict with our model, either due to noise, orthogonal variation, or non-linear behaviors. By comparing R^2 between held-out test perturbations from

the focal base and test perturbations from target bases, we can begin to understand how the inferred fitnotypes $\vec{\phi}$ depend on the base environment, and the structure of their overlap across bases.

Fig. 5b shows two key results from this quantification of fitnotype space for the complete set of mutants. First, by focusing on the salmon-colored fraction in the flanking plots, we see that fitnotypes discovered in Salt and 2 Day have substantial predictive power for held-out perturbations within their own set of base environments. But this predictability drops significantly in the other bases, suggesting that there is considerable fitnotypic variation that only appears in these other bases. The middle plot shows that discovering fitnotypes in the 1 Day base has limited predictability across all three bases, and in fact is an even better model for the 2 Day base than itself. This heterogeneity emphasizes how much local phenotypic landscapes are shaped by the environment. Our second key finding concerns the weights of each component of our model, which we denote in different shades of green and blue in the bar plots. In Fig. 5c, the top row shows the fraction of variance explained in each base, by each fitnotype. The bottom row shows the change in fraction of variance explained by each fitnotype from the base environment. We see qualitatively different patterns in each base environment. In the Salt base, we see two dominant fitnotypes that explain nearly 75% of the variation. However, this first fitnotype is effectively unimportant for making predictions in the two other bases. In contrast, the 1 Day base has a much more similar distribution of predictive power across the fitnotypes, and this is conserved in the target bases, although there is some significant re-weighting of the first, second, and seventh components. The 2 Day base is somewhere in the middle. There is a substantial gap in predictability between within-base and out-of-base predictions, and there is also evidence for fitnotype reweighting.

These findings, a gap in prediction power and a re-weighting of fitnotypic importance, combine to strongly support the pleiotropic shift hypothesis. We see direct evidence that there are hidden phenotypes that remain hidden in certain bases, but are revealed in others. And of the overlapping phenotypic space, what contributes most heavily to local fitness often shifts. It is worth noting that the inputs to our model for each base environment are, on the surface, identical. The same genotypes and the same physical and chemical changes to the base environment make up the rows and columns of the input matrix. Therefore, the heterogeneity in fitnotype space across base environments is entirely attributable to how the base environment constrains the fitness function.

II. DISCUSSION

In this study, we aimed to understand how genotype-phenotype-fitness maps shift across environments. Building such maps has been a difficult task for several reasons. Genotype space is vast, it is unclear which phenotypes are important to fitness, and environments can modulate both the genotype-to-phenotype mapping, and phenotype-to-fitness mapping. Previous work tackled the initial question of mapping phenotype space for a group of adaptive mutants near their evolution environment, using a top-down linear model of fitness variation across subtly perturbed environments (22). This technique yields a map of abstract "phenotypes" that are revealed through their effect on fitness that Kinsler et al termed "fitnotypes."

They found evidence that locally insignificant fitnotypes can become important for determining fitness in environments far from the evolution condition. They did not, however, systematically build fitnotype spaces around different distant environments. In generating full genotype-fitnotype-fitness maps around distinct base environments, which are themselves distant from each other, we were able to explore the nature of local and global pleiotropy for adaptive mutants and distinguish between two models of how pleiotropy might manifest.

One hypothesis, the pleiotropic expansion hypothesis (Fig. 1a), is that the dimensionality of a model capturing the fitness variation of adaptive mutants is small near their evolution condition, but grows dramatically around distant bases. We did not see evidence supporting this model: we reliably found instead that the dimensionality of our models were consistently low across bases, regardless of whether they evolved in the base environment or not. This allowed us to reject the "pleiotropic expansion" model.

We instead found strong evidence for the alternative, the pleiotropic shift model (Fig. 1b). In this model, genetic mutations do generate a pleiotropic collection of phenotypic effects, but the dimensionality of the phenotypic variation that affects fitness remains limited even in environments that are distant from the evolution condition. This pattern is consistent with a limiting functions model of fitness. In any particular environment, there is a small number of key limiting functions that can be altered to affect fitness. These are the functions that determine the cells' ability to surmount limiting challenges and thrive. Fitness variation is thus dominated by this small set of functions, suggesting that phenotypic diversity in an adapting population is constantly being projected down onto a few key dimensions. The choice of these dimensions changes as the environment shifts.

How might such a pleiotropic architecture of limiting functions arise? Similar questions have been explored extensively in the context of nutrient availability (25, 29, 54). If we think of cellular growth as a function of available nutrients, many different nutrients may be required. However, at any given time, growth will be constrained by just one of these—the limiting nutrient—while the others remain in excess. This principle is well understood in both ecology and chemistry; for example, in a chemical reaction, the limiting reagent determines the reaction's progress. Fitness is a more complex function than a simple chemical

reaction, but at its core, it is governed by a series of biochemical processes that drive growth. Differences in these processes create differences between cellular performance, manifesting as fitness advantages or defects. It is therefore reasonable to extend the concept of limiting nutrients or reactants to the level of fitness. Importantly, the key constraint on fitness may not always be a nutrient deficiency. It could be the need to withstand extreme heat, buffer against extreme pH, or overcome some other environmental challenge. Ultimately, a cell "perceives" its environment as a set of challenges, but in any given moment, closely related cells will be limited by only one or a few of these challenges. After all, optimizing growth rate is irrelevant if the cell is dying from osmotic stress.

We are presented with a range of possibilities for how environments might vary under this model. Two distant environments might present entirely orthogonal selection pressures to a population. In a simple case, one environment might have an abundance of a fermentable carbon source for the yeast, so optimizing growth rate or lowering lag phase for fermentation metabolism the best function to optimize in a fitness competition. But a different environment might only have non-fermentable carbon sources, and so changes to fermentation metabolism will have little effect on fitness in this second environment. Any pleiotropic effects from the first environment that bear on respiration metabolism will be revealed, and differentiate between the various mutants. Imagine another scenario, in which both of these media are placed in an incubator at 37°C. Now, while the base media are still "orthogonal" metabolically, the yeast are most concerned with surviving the heat shock, and so the limiting function is heat survival in both environments. Any variation in that phenotype will dominate the relative fitness between mutants in both environments, so we will infer these underlying fitnotypes to be fully overlapping. Conceptualizing environments in this manner presents an intriguing possibility for understanding environmental similarity, which is a slippery concept. The notion of overlap in limiting functions could be a useful organism-specific metric for parameterizing environmental distance, from the perspective of competitive fitness.

One natural question to emerge is why there seems to be overlap in the phenotypes affected in our pool of mutants, and the limiting functions that are important in our distant, non-evolution base environments. A priori, it is not clear that variation in the hidden phenotypic effects from a group of mutants will necessarily have non-zero projections onto an arbitrary environment's limiting functions. One possibility is that single bouts of evolution can exhaustively explore all phenotypic diversity, due to the tight integration of cellular and genetic networks, such that any possible important phenotype is already present in an adapting population of sufficient size. Or, perhaps we are biased by our genotype set. Most of the mutations in our collection are either autodiploidization events, which affect the entire genome, or in signaling pathway, which are known to be especially pleiotropic (51, 56). These mutants are also atypical in that they harbor large beneficial fitness effects, despite their genetic proximity. What might we find by applying similar environmental perturbations and measuring the relative fitness of a different set of genotypes? In particular, it would be fascinating to compare our results to a set of genotypes that have adapted for much longer, either in experimental evolution with Lenski's LTEE *E. coli* lines for example, or as a result of evolution in natural settings by looking at a large collection of natural isolates that represent the extent of standing genetic diversity within a species (18, 32, 41).

Still another possibility for why we tend to detect limiting functions across the three base environments is that maybe the bases are not in fact sufficiently distant. All three are lab environments with glucose-limited media, so while we do observe many differences in fitness, perhaps we are constraining the environment space more than we thought. Or perhaps the very fact that yeast cells can thrive in all these environments itself necessarily acts as a built-in constraint on fitness assays and growth, meaning that the range of possible environments will always be limited. So it is interesting to speculate: what we will learn when we apply many more perturbations to a very large number of base environments? Perhaps we will discover that there are endless ways to project phenotypic diversity onto fitness, and that at a certain point the collection of genotypes will become the limiting variable. Or perhaps we will discover that environment space, constrained by survivability, is actually small, and consistently mappable. These extensions of environment space provide a promising avenue for future work.

The limiting functions model naturally predicts that the fitness effects of environmental perturbations are often heavily dependent on the base environment. If base environments present distinct major challenges, the projection of fitness onto these key dimensions will be different, even as a result of the same underlying physical and chemical changes. Thus, idiosyncrasy in δX effects across bases is to be expected. However, this idiosyncrasy does not mean that prediction writ large is hopeless. The fact that dimensionality remains constrained locally in any base environment raises the possibility of practical applications. If a series of exploratory fitness measurements in carefully chosen environments is sufficient to infer relevant fitnotypes, perhaps the problem of fitness prediction will turn into the problem of fitness extrapolation.

Consider the implications of this model in an applied setting: cancer. We can predict that driver mutations will tend to have a similar, low-dimensional set of fitnotypic effects in the untreated host, evolving in response to the limiting challenges presented by the tumor's need to thrive and spread in tissue. These same mutations should simultaneously carry along hidden phenotypic effects that are likely to make response to treatment unpredictable, as treatments naturally generate a whole new set of limiting challenges. Different

treatments should also present a set of disparate challenges to tumors, but our model would predict that the space of possible outcomes associated with a particular treatment is finite, and perhaps even learnable. In other words, extrapolating fitness to new environments might be feasible because the number of parameters to "learn" in any new environment is relatively low, and therefore the phenotypic diversity that matters to fitness, ergo health outcomes, is limited.

More broadly, the limiting functions model offers insight into how evolution might proceed over successive environmental epochs. First, a clonal population accumulates multiple adaptive mutations that compete with each other for fixation. All of these adaptive mutations confer a fitness benefit relative to the ancestor, and tend to affect similar phenotypes that matter to fitness in their environment, but they harbor a reservoir of hidden phenotypic diversity. When the environment shifts and a new epoch begins, a different set of phenotypes will be important for fitness, reshuffling the relative fitness of each mutant. New groupings of similar mutants emerge, but the total number of clusters is roughly conserved. The pool of phenotypic diversity amongst a group of adaptive mutations is large when integrated across multiple environments, but is never so large in a single epoch that mutants behave in completely idiosyncratic ways. Environments present limiting challenges that impose low-dimensional structure on phenotype space, and dominate the fitness response. This low-dimensionality provides some hope that evolutionary prediction is possible, and brings us one step closer to the ambitious goal of building a universally relevant genotype-phenotype-fitness map.

III. METHODS

A. Experimental procedure for barcoded fitness assay

1. Experimental overview of fitness assay

We performed fitness assays as previously described ((22), (28), (56)). We pool all barcoded strains and compete them together against a reference strain (in this case the wild type ancestor, WT, represented at 95% of the pool at the beginning of the competition) by growing them for a predetermined transfer duration (T_{transfer}) in 100 mL of glucose-limited media. We grew the cells in incubators set a temperatures determined by environmental prescription, for either 24 hours or 48 hours per cycle. All environments except "No Shake" were rotated at 223 rpm. At the end of each cycle, after the cells have grown in direct competition, we transfer 400 uL of saturated media into fresh media, or roughly 5×10^7 cells, preferentially sampling lineages that have grown to higher frequency over the course of the cycle due to increased fitness relative to the mean population fitness. We dilute the cells at a 1:250 ratio, corresponding to roughly 8 generations of cell division over the course of the growth cycle, and we perform 4 total dilution cycles. At each transfer, we save the rest of the untransferred cells in sorbitol for DNA extraction and PCR amplification of the barcode region. We sequence 5 timepoints to get estimates for the frequency of each barcoded lineage, and use the log-fold change in frequency to infer the relative fitness of each strain with respect to the WT ancestor, correcting for the changing population mean fitness. With these methods, in each environment we were able to infer the relative fitness of thousands of strains. We perform fitness assays in multiple environments (Supplemental information).

2. Yeast strains

All strains in this study are of genetic background Matα, ura3Δ0::Gal-Cre-KanMX-1/2URA3-loxP-Barcode-1/2URA3-HygMX-lox66/71. Mutant strains come from the evolution experiments described in (27), (1),(23)).

3. Growth environments

All of our growth environments are based on M3 minimal, glucose-limited media. Our environments are defined by the media in the flask, additional components added to the media, the shape of the flask, the shaking status, and the temperature of the incubator. Due the scale of our experiment, we grouped our environmental perturbations into 4 batches, each with 3 base conditions and between 4 and 5 perturbations. This meant a total of 12-15 environments per batch. We had two biological replicates per environment. The first batch was temperature perturbations, the second batch was glucose gradients, the third batch was additional carbon sources, and the fourth batch was drugs and ethanol. See Supplemental Information for a full list of environments.

4. DNA Extraction

We extracted genomic DNA according to a similar procedure in (22). Briefly, we thawed 400 μL of cells for each sample. We spun down the cells at 3500 rpm for 3 minutes, and discarded the supernatant. We washed the cells by adding 400 μL of sterile water and spun them down at 3500 rpm for 3 minutes, and discarded the supernatant. We resuspended the pellet in 400 μL of extraction buffer, and incubated at 37°C for 30 minutes. Then, we added 40 μL of 0.5M EDTA, and vortexed. Then, we added 40 μL of 10% SDS, and vortexed. We next added 56 μL of proteinase K, and vortexed very briefly. We incubated at 37°C for 30 minutes, and put the samples on ice for 5 minutes. We then added 200 μL of 5M potassium acetate, and manually shook the tubes. We incubated on ice for 30 minutes. Then, we spun the tubes for 10 minutes at maximum speed (13000 rpm), and transferred supernatant to a new tube with 750 μL isopropnaol, and rested on ice for 5 minutes. We spun it down for 10 minutes at max speed and discarded supernatant. We eashed twice with 750 μL of 70% ethanol, vortexing

very briefly and spinning at maximum speed ofr 2 minutes, then discarding liquid. We resuspended in 50 μL Tris pH 7.5, and left it on the bench overnight if the pellet was not fully resuspended. We next added 1 μL of 20 mg/mL RNase A and incubated at 65°C for 30 minutes. To digest the ancestor (so that we do not sequence our high-abundance reference strain), we added 5.5 μL of Cutsmart buffer and 1μL of ApaL1 restriction enzyme to each tube. We incubated the tubes at 37°C overnight, and quantified with Qubit.

5. PCR amplification of barcode locus

To amplify the barcode region for sequencing, we perform two steps of PCR. For the first step, our goal is to tag individual molecules with UMIs and include first step primers to allow for combinatorial indexing. We divided each sample into 8 separate PCR tubes in order to eliminate the influence of PCR jackpotting. For a single set of 8 reactions, we used 200 μL HotStartTaq Polymerase master mix, 8 μL each of forward and reverse primers (10 μM), 118 μL nuclease-free water, 16 μL of 50 mM MgCl2, and 50 μL of genomic DNA. We ran 3 cycles of the following program.

1. 94°C 10 min
2. 94°C 3 min
3. 55°C 1 min
4. 68°C 1 min
5. Repeat steps 2-4 for a total of 3 cycles
6. 68°C 1 min
7. Hold at 4°C

We did a clean-up (standard DNA purification procedure) of the step 1 PCR product, and then performed step 2 PCR. We split our template DNA into 3 reactions, and for a set of 3 reactions, we used 45 μL template, 65.5 μL nuclease-free water, 30 μL Q5 buffer, 2.5 μL 10uM Forward and Reverse Nextra primers, 3 μL 10mM dNTPs, and 1.5 μL Q5 polymerase. We then ran the following program for the second step of PCR, which amplifies for 20 cycles.

1. 98°C 30 sec
2. 98°C 10 sec
3. 62°C 20 sec
4. 72°C 30 sec
5. Repeat steps 2-4 19 times
6. 72°C 3 min
7. Hold at 4°C

We did a sample clean-up using standard DNA purification techniques, and quantified with Qubit.

6. Sample pooling and amplicon sequencing, and data processing

We pooled our samples and sent them to Admera Health for quality control, bead cleanup, and sequencing. We sequenced on HiSeq and Novaseq S4. We processed our data in exactly the same way as (22) to obtain barcode counts and frequencies for each sample, from raw sequencing reads.

B. Analysis

1. Fitness inference

We adapted the fitness inference method introduced in Ascensao et al (5), and closely follow their derivation of maximum likelihood fitness estimate below. We want to infer the fitness of a barcoded yeast strain from the read count information. Read counts for a particular barcode are generateed through a sequence of noisy processes, such as stochastic growth, bottle-necking, DNA extraction, PCR amplification, and sequencing. Each of these processes is individually a counting process, so we can model the read count of barcode i at timepoint t as a negative binomial random variable, which essentially allows us to think about the actual value of the read count as being sample from an overdispersed Poisson distribution, where the variance is larger than the mean:

$$r_{i,t} \sim \mathrm{NB}(\mu_{i,t}, c_t) \tag{4}$$

where the expected value of the read count is

$$\langle r_{i,t} \rangle = \mu_{i,t} = R_t f_0 e^{(s-\bar{x}_t)t} \tag{5}$$

and the variance is

$$\mathrm{var}(r_{i,t}) = c_t \langle r_{i,t} \rangle \tag{6}$$

There are several parameters we need to infer before we can try to infer s. In order to calculate the expected value of the read count, $\langle r_{i,t} \rangle$, we must have an estimate of the mean fitness of the barcoded population with respect to the ancestor, $\bar{x}_t$, which we can infer from the trajectories of known neutral barcoded lineages. For a set of neutral barcodes, we have (by definition):

$$f_t^{(n)} = f_0^{(n)} e^{-\bar{x}_t t} \tag{7}$$

$$\bar{x}_t = -\frac{1}{t} \log \frac{f_t^{(n)}}{f_0^{(n)}} \tag{8}$$

We can actually calculate this, with the caveat that read counts are somewhat noisy estimates of the underlying frequency. Also, since we assume all the barcodes that identify neutral lineages should have identical fitnesses (s=0), we can take the sum of all neutral barcodes at a given timepoint, $R_t^{(n)}$, to calculate the mean fitness.

$$\bar{x}_t = -\frac{1}{t} \left[\log \frac{R_t^{(n)}}{R_t} - \log \frac{R_0^{(n)}}{R_0} \right] \tag{9}$$

where

$$R_t^{(n)} = \sum_{i \in \text{neutral}} r_{i,t} \tag{10}$$

Alternatively, we can calculate the mean fitness within time intervals, where we have

$$\bar{x}_{j,k} = \sum_{j}^{k} -\frac{1}{k-j}\left[\log\frac{R_k^{(n)}}{R_k} - \log\frac{R_j^{(n)}}{R_j}\right] \tag{11}$$

We can estimate the error in the mean fitness value by bootstrapping over neutral barcodes and recalculating the mean fitness. Alternatively, we can calculate the median absolute deviation between subsamples of neutral barcodes. To infer the variance of the negative binomial distribution, we can also leverage our neutral barcodes. We can think about all the potential sources of noise in the stages from cell number in the flask to read count. There is biological growth noise and bottlenecking noise, which should be correlated across timepoints (ie it accumulates), and there is uncorrelated technical noise from DNA extraction, 30 rounds of PCR, and sequencing. All of these processes are counting processes, so we expected the variance of the frequencies to be proportional to the mean. Note, we not are treating the total coverage at a particular timepoint as a random variable. So we have

$$\text{var}(f_{i,t}) \propto \frac{\langle r_{i,t}\rangle}{R_t} = c_t\frac{\langle r_{i,t}\rangle}{R_t} \tag{12}$$

Given that we have a variance that is strongly dependent on the mean, we apply a variance-stabilizing transformation. Let's define

$$\phi_{i,t} \equiv \sqrt{f_{i,t}} \tag{13}$$

Applying the square root transformation effectively decouples the variance from the mean. We can think about how the variance in $f_{i,t}$ is related to the variance in $\phi_{i,t}$. We can start with the Langevin equation for the change in frequency of a mutation due to drift.

$$f(t+\delta t) = f(t) + \sqrt{\frac{f(t)(1-f(t))\delta t}{N_e}}Z \tag{14}$$

Consider the change in ϕ over one time interval:

$$\begin{aligned}
\delta\phi &= \phi(t+\delta t) - \phi(t) \\
&= \sqrt{f(t+\delta t)} - \sqrt{f(t)} \\
&= \sqrt{f(t) + \sqrt{\frac{f(t)(1-f(t))\delta t}{N_e}}Z} - \sqrt{f(t)} \\
&= \sqrt{f(t) + \sqrt{\frac{f(t)\delta t}{N_e}}Z} - \sqrt{f(t)} \quad (f \ll 1,\ \sqrt{\tfrac{f(t)\delta t}{N_e}}Z \ll 1) \\
&= \sqrt{f(t)}\left(1 + \frac{1}{2}\sqrt{\frac{\delta t}{N_e}}Z\right) - \sqrt{f(t)} \\
&= \sqrt{\frac{f(t)\delta t}{4N_e}}Z
\end{aligned}$$

So the variance in ϕ due to genetic drift over a single cycle is $\frac{1}{4N_e}$. Between two timepoints, we expect neutral lineages to change due only to noise from genetic drift (which accumulates over time) and technical noise at each timepoint (which is independent at each timepoint). If there are enough read counts that the central limit theorem applies, we can add the variances:

$$\kappa_{t,t'} = \text{var}(\phi_{i,t} - \phi_{i,t'}) = \eta_t + \eta_{t'} + \frac{|t-t'|}{N_e} \tag{15}$$

We can measure κ for each time interval, and treat the η_t and N_e as unknown parameters that we can fit by minimizing the squared errors of the expected relationship. We can also do this with inverse variance weighting by bootstrapping over barcodes when estimating κ.

$$\vec{\eta}, N_e = \arg\min_{\hat{\eta}, N_e} \sum_{t,t'} \frac{\left(\eta_t + \eta_{t'} + \frac{|t-t'|}{4N_e} - \hat{\kappa}_{t,t'}\right)}{\text{var}(\hat{\kappa}_{t,t'})} \tag{16}$$

We only use timepoints and barcodes with read count exceeding $r_{i,t} = 50$.

We can related these quantities back to the variance in frequency, and obtain an estimate for the dispersion parameter, c_t.

$$c_t = (4\eta_t + \frac{1}{N_e})R_t \tag{17}$$

Having obtained our parameter estimates for intermediate parameters, we can finally infer the parameter value for fitness, s, using the likelihood function for a negative binomial random variable. The likelihood for a particular value of s and $f_{0,i}$ given a barcode read count of r_i, is

$$P(r_i|s, f_{0,i}) = \prod_t \frac{\Gamma(r_{t,i} + \frac{\mu_{t,i}}{c_t - 1})}{\Gamma(r_{t,i}+1)\Gamma(\frac{\mu_{t,i}}{c_t-1})} \frac{(c_t-1)^{r_{t,i}}}{c_t^{r_{t,i}+\frac{\mu_{t,i}}{c_t-1}}} \tag{18}$$

We can marginalize over values of f_0, i

$$\mathcal{L}(s|r) = \prod_i \int df_{0,i} P(r_i|s, f_{0,i}) \tag{19}$$

and find the maximum likelihood fitness,

$$\hat{s} = \arg\max_s \log \mathcal{L}(s|r) \tag{20}$$

We use a grid search to find the value of s. We can use the likelihood function to infer a standard error on our parameter estimation, approximated as the inverse of the square-root Fisher information,

$$\text{std } \hat{s} = \frac{1}{\sqrt{-\partial_s^2 \log \mathcal{L}(s|r)|_{\hat{s}}}} \tag{21}$$

To get a fitness estimate for a single environment, we infer fitness separately for each replicate, and then take a weighted average across our two biological replicates, weighted by the inverse of the inferred standard error.

2. Fitness processing

There were some technical problems in sequencing, resulting in several environments with poor data quality. We dropped these from further analysis. These included all temperature perturbations to the 2 Day base (all Batch 1 conditions), the Batch 4 Salt Base, and 10 μM H89 on the Salt base.

3. PCA on environment space

To assess environmental similarity, we did principal component analysis on our fitness data. First, we took fitness data for all environments and centered and scaled it using `sklearn.preprocessing.StandardScaler`. We then used `sklearn.decomposition.PCA` on the transpose of our standard fitness matrix (which normally has columns as environments and rows as mutants) so that we could understand environmental variation. We projected each environment onto the first two PCs and used `scipy.spatial.ConvexHull` to generate the convex hulls around each base environment.

4. Calculation of z-score

To quantify the magnitude of our perturbations, we used batch replicates as a baseline for variation. Each batch of our fitness measurements included one "unperturbed" base. For each base B, we took the collection of batch replicates and calculated the mean fitness for each mutant i, $\mu_{i,B}$. We obtained an observed standard deviation ($\sigma_{i,o}$)from these replicates, and we used the average inferred standard deviation from the fitness inference for the replicates ($\hat{\sigma}_i$), together to represent variation across batch replicates. For each mutant i in each environment e belonging to base B, we calculated a z-score as

$$z_{i,e} = \frac{|X_{i,e} - \mu_{i,B}|}{\sqrt{(\sigma_{i,o}^2 + \hat{\sigma}_i)^2}} \tag{22}$$

To characterize the entire environment, we took the average $z_{i,e}$ across all mutants, and the standard deviation across all mutants. For the external bases, we used the same equation for $z_{i,e}$ as above, but inserted environments corresponding to different bases.

5. Generating the δX matrix

To generate the δX matrix, get an average fitness value for each mutant in each base, $\bar{X}_{i,B}$. Then, for each perturbation p, we define

$$\delta X_p^{(B)} = X_p^{(B)} - \bar{X}_{i,B} \tag{23}$$

We propagate the error in each measurement (both in generating the standard error of the base environment mean, and in the subtraction of fitness values to generate δX_p.

6. Calculating deviation from 1:1 line of perturbations on different base environments

For each mutant, we compared the change in fitness due to the same perturbation, on different base environments (as shown in Fig. 3b). We calculated a deviation from the 1:1 line by subtracting one base's δX_p from another. We took $\delta X_p^{1\text{ Day}} - \delta X_p^{2\text{ Day}}$, $\delta X_p^{1\text{ Day}} - \delta X_p^{\text{Salt}}$, $\delta X_p^{2\text{ Day}} - \delta X_p^{\text{Salt}}$, and $\delta X_p^{\text{Rep 1}} - \delta X_p^{\text{Rep 1}}$. This last comparison was to have a point of reference for the deviation to be expected under biological and measurement noise, and not due to ExE interactions.

7. Generating δX correlation matrices

To generate the correlation matrices in Fig. 3e, we used correlations generated from Fig. 3e, but for all pairwise combinations of environments. Brielfy, we asked whether the δX of all mutants in two different environements were correlated with each other, and generated a correlation coefficient, or Pearson's r, with `scipy.stats.pearsonr`. So we have a correlation matrix of environment to environment δX correlation. We first clustered rows of this matrix by perturbation type, and then by batch. We next clustered mutants by base, and then by batch. We included these partitions in Fig. 3e as dashed lines on top of the correlation matrix.

8. Quantifying block diagonality of correlation matrices with modularity score, Q

To quantify the block-diagonality of each matrix, we used a tool from network theory called the "modularity" score, Q. Typically used to quantify the modularity of a network, Q can intuitively be understood as a measure of how much more density falls within a partition than would be expected of a random matrix, where the row sums are conserved. First, we scaled the values of the correlation matrices, which initially lay between -1 and 1, to fall between 0 and 1. Then, for each partition of the data, we calculated the difference between the observed and expected weight under a model of randomly distributed weights. As the matrix approaches perfect block-diagonal form, Q approaches 1. We calculated Q for the scaled correlation matrix with perturbation partitions and base-environment partitions, and found that as expected, Q is significantly higher for the base partition. We got values of Q=0.35 for the base-clustered matrix, and Q=0.08 for the perturbation clustered matrix.

9. Relationship between model in section C and SVD

Given a collection of mutants, we can write the collection of $\vec{\phi}_{i,b}$ vectors from Eq. (3) as a matrix, Φ_b. Therefore, if we have measured δX_p for a collection of mutants, we use singular value decomposition (SVD) to factorize the δX_p matrix and infer Φ_b and $\beta_{b,p}$ for

each base b. Here, $\Phi_b = U\sqrt{\Sigma}$ and $B_b = \sqrt{\Sigma}V^T$, where the SVD of $\delta X = U\Sigma V^T$.

10. Singular value decomposition on δX matrices

To generate our model, we use singular value decomposition to factorize the δX matrix. For each base, we perform SVD using `numpy.linalg.svd`. We determine fraction of variance explained by component in Fig. 4 by taking each singular value squared, and dividing by the sum of squared singular values.

$$\text{Fraction variance explained by component } i = \frac{\sigma_i^2}{\sum_j \sigma_j^2} \tag{24}$$

11. Calculating a limit of detection for dimensionality

To calculate the number of dimensions that exceed noise in our data, take the same approach as (22). We use the inferred standard error in our fitness estimate, and generate 100 random matrices of the same shape as our data, centered at 0. Each entry is drawn from a normal distribution with mean 0 and standard deviation equal to the inferred standard error at the corresponding index in our data. From this, we can ask how much variation in the true data would have been explained by components derived from the noise-only matrix, by taking the singular values squared from the noise matrix and dividing by the sum of squared singular values in the data. The maximum amount of variation explained by noise matrices represents our limit of detection.

$$\text{Fraction of variance explainable by noise} = \frac{\bar{\sigma}_{1n}^2}{\sum_j \sigma_j^2} \tag{25}$$

where $\bar{\sigma}_{1n}$ is the average first singular value from each fold of the noise-only matrices, and the σ_j in the denominator come from the δX matrix.

12. Calculation of entropy as a proxy for dimensionality in scree plots

To calculate Shannon entropy, we use the standard formula for the entropy of a distribution,

$$H = -\sum_i p_i \log p_i \tag{26}$$

Here, we use the "distribution" of variance explained, so

$$p_i = \frac{\sigma_i^2}{\sum_j \sigma_j^2} \tag{27}$$

We use only components that fall above the noise limit for the calculation of entropy.

13. Linear regression on fitnotypes

On a methodological note, it might seem unnecessary to approach overlap between vectors via prediction, rather than directly looking at some quantity like the inner product, or cosine of the angle between the two $\vec{\phi}$ vectors. We take this approach because the method we use to infer our $\vec{\phi}$ vectors is matrix factorization, which is under-determined. This means that any rotation or linear transformation of our latent fitnotype space will be equally good at reconstructing our data, so we must focus on properties of our model that are invariant under linear transformation. Prediction error is one such quantity.

We use prediction error for a linear model trained in one base, and tested in held out environments, to asses model success and overlap in latent space. First, we do SVD on δX all environments in one base, but we leave out a single perturbation. Let B1 be our focal base, and B2 is a target base. For a perturbation p, we have the reduced matrix δX_{-p}^{B1}, where we have removed perturbation p. We then do SVD on this reduced matrix.

$$\delta X_{-p}^{B1} = U\Sigma V^T \tag{28}$$

For each value of k up to K=8, we use the k'th column of U as a feature in a linear regression, fitting coefficients for the held out perturbation p within base, but also for all environments in the target base, $B2$. We can sum the variance from each component, plus the residuals, to equal the total variance, because our regressors are orthogonal (they are singular vectors, after all). This allows us to neatly partition the variance by component. For each value of k, we hold out each perturbation, fit, and generate a prediction for the environments within the focal base, $B1$. The we use the normal δX^{B1} matrix as the design matrix (without reducing it) for a regression in the target base. We can also ask how much variation is explained within a base if we use SVD with an 8 dimensional model, which is the lowest-error 8 dimensional linear model (14). This represents an ideal, within-training-set error. We can then compare predictive power across components and across bases, as in Fig. 5b and c. Note, within base, we are actually training on slightly different models for each held out prediction, and aggregating predictive power across. By comparing the unexplained variation across bases, and the weights of each component, we can glean information about the underlying fitnotype spaces, even though they are latent.

DATA AND CODE AVAILABILITY

Source code for the data analysis and figure generation is available at Github (`https://github.com/omghosh/limiting-functions`).

ACKNOWLEDGMENTS

The authors would like to thank members of the Petrov lab and Good lab for helpful discussions and feedback. We would like to thank Mihkail Tikhonov, Gabriel Mel de Fontenay, Anthony Thomas, Matthew Aguirre, Jonathan Pritchard, Seppe Kuehn, Douglas Stanford, Milo Johnson, Daniel Wong, and Sasha Khristich for helpful discussions. We would like to thank Mikhail Tikhonov, Clare Abreu, Shaili Mathur, Sasha Khristich, Anastasia Lyulina, Vivian Chen, José Aguilar-Rodriguez, Sophie Walton, Manuel Razo, and Zhiru Liu for their help with the experiments (Big Batch Bootcamp team). We thank the Stanford Research Computing Center for use of computational resources on the Sherlock cluster. B.H.G. acknowledges support from NIH NIGMS Grant No. R35GM146949. D.A.P. acknowledges support from NIH NIGMS Grant No. 5R35GM118165-07. O.M.G. acknowledges support from the NSF-GRFP. B.H.G. and D.A.P. are Chan Zuckerberg Biohub - San Francisco Investigators.

REFERENCES

[1] Aggeli, D., Y. Li, and G. Sherlock (2021), Nature Communications **12** (1), 5193.
[2] Aguirre, M., J. P. Spence, G. Sella, and J. K. Pritchard (2024), "Gene regulatory network structure informs the distribution of perturbation effects,".
[3] Altenberg, L. (2005), "Modularity in Evolution: Some Low-Level Questions," in *Modularity* (The MIT Press) pp. 99–128.
[4] Ardell, S., A. Martsul, M. S. Johnson, and S. Kryazhimskiy (2024), Science (New York, N.Y.) **386** (6717), 87.
[5] Ascensao, J. A., K. M. Wetmore, B. H. Good, A. P. Arkin, and O. Hallatschek (2023), Nature Communications **14** (1), 248.
[6] Bagheri-Chaichian, H., J. Hermisson, J. R. Vaisnys, and G. P. Wagner (2003), Mathematical Biosciences **184** (1), 27.
[7] Bersaglieri, T., P. C. Sabeti, N. Patterson, T. Vanderploeg, S. F. Schaffner, J. A. Drake, M. Rhodes, D. E. Reich, and J. N. Hirschhorn (2004), American Journal of Human Genetics **74** (6), 1111.
[8] Boyle, E. A., Y. I. Li, and J. K. Pritchard (2017), Cell **169** (7), 1177.
[9] Chan, Y. F., M. E. Marks, F. C. Jones, G. Villarreal, M. D. Shapiro, S. D. Brady, A. M. Southwick, D. M. Absher, J. Grimwood, J. Schmutz, R. M. Myers, D. Petrov, B. Jónsson, D. Schluter, M. A. Bell, and D. M. Kingsley (2010), Science (New York, N.Y.) **327** (5963), 302.
[10] Chesmore, K., J. Bartlett, and S. M. Williams (2018), Human Genetics **137** (1), 39.
[11] Cook, L. M., and I. J. Saccheri (2013), Heredity **110** (3), 207.
[12] Cooper, T. F., E. A. Ostrowski, and M. Travisano (2007), Evolution **61** (6), 1495.
[13] Diaz-Colunga, J., A. Sanchez, and C. B. Ogbunugafor (2023), Nature Communications **14** (1), 8055.
[14] Eckart, C., and G. Young (1936), Psychometrika **1** (3), 211.
[15] Eguchi, Y., G. Bilolikar, and K. Geiler-Samerotte (2019), Current Opinion in Genetics & Development Evolutionary Genetics, **58–59**, 95.
[16] Fisher, R. A. (1930), *The Genetical Theory of Natural Selection*, The Genetical Theory of Natural Selection (Clarendon Press, Oxford, England).
[17] Geiler-Samerotte, K. A., S. Li, C. Lazaris, A. Taylor, N. Ziv, C. Ramjeawan, A. B. Paaby, and M. L. Siegal (2020), PLOS Biology **18** (8), e3000836.
[18] Good, B. H., M. J. McDonald, J. E. Barrick, R. E. Lenski, and M. M. Desai (2017), Nature **551** (7678), 45.
[19] Greenbury, S. F., A. A. Louis, and S. E. Ahnert (2022), Nature Ecology & Evolution **6** (11), 1742.
[20] Gu, X. (2007), Genetics **175** (4), 1813.
[21] Kacser, H., and J. A. Burns (1981), Genetics **97** (3-4), 639.
[22] Kinsler, G., K. Geiler-Samerotte, and D. A. Petrov (2020), eLife **9**, e61271.
[23] Kinsler, G., Y. Li, G. Sherlock, and D. A. Petrov (2024), PLOS Biology **22** (12), e3002848.
[24] Knecht, A. K., K. E. Hosemann, and D. M. Kingsley (2007), Evolution & Development **9** (2), 141.
[25] Kryazhimskiy, S. (2021), eLife **10**, e60200.
[26] Lande, R., and S. J. Arnold (1983), Evolution **37** (6), 1210.
[27] Levy, S. F., J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, and G. Sherlock (2015), Nature **519** (7542), 181.
[28] Li, Y., S. Venkataram, A. Agarwala, B. Dunn, D. A. Petrov, G. Sherlock, and D. S. Fisher (2018), Current Biology **28** (4), 515.
[29] Liebig, J. v. .-. (1841), *Die Organische Chemie in Ihrer Anwendung Auf Agricultur Und Physiologie*, 2nd ed. (Vieweg, Braunschweig).
[30] Lind, P. A., A. D. Farr, and P. B. Rainey (2015), eLife **4**, e07074.
[31] Lindgren, J., M. W. Caldwell, T. Konishi, and L. M. Chiappe (2010), PloS One **5** (8), e11998.
[32] Loegler, V., A. Friedrich, and J. Schacherer (2024), G3 (Bethesda, Md.) **14** (12), jkae245.
[33] Louis, A. A. (2016), Studies in History and Philosophy of Biological and Biomedical Sciences **58**, 107.
[34] Nguyen Ba, A. N., K. R. Lawrence, A. Rego-Costa, S. Gopalakrishnan, D. Temko, F. Michor, and M. M. Desai (2022), eLife **11**, e73983.
[35] Olivieri, M., T. Cho, A. Álvarez-Quilón, K. Li, M. J. Schellenberg, M. Zimmermann, N. Hustedt, S. E. Rossi, S. Adam, H. Melo, A. M. Heijink, G. Sastre-Moreno, N. Moatti, R. K. Szilard, A. McEwan, A. K. Ling, A. Serrano-Benitez, T. Ubhi, S. Feng, J. Pawling, I. Delgado-Sainz, M. W. Ferguson, J. W. Dennis, G. W. Brown, F. Cortés-Ledesma, R. S. Williams, A. Martin, D. Xu, and D. Durocher (2020), Cell **182** (2), 481.
[36] Orr, H. A. (1998), Evolution **52** (4), 935.
[37] Orr, H. A. (2000), Evolution **54** (1), 13.
[38] Paaby, A. B., and M. V. Rockman (2013), Trends in Genetics **29** (2), 66.
[39] Pan, J., J. J. Kwon, J. A. Talamas, A. A. Borah, F. Vazquez, J. S. Boehm, A. Tsherniak, M. Zitnik, J. M. McFarland, and W. C. Hahn (2022), Cell Systems **13** (4), 286.
[40] Pereira, A. G., and T. Kohlsdorf (2023), Genetics and Molecular Biology **46**, e20220384.
[41] Peter, J., M. De Chiara, A. Friedrich, J.-X. Yue, D. Pflieger, A. Bergström, A. Sigwalt, B. Barre, K. Freel, A. Llored, C. Cruaud, K. Labadie, J.-M. Aury, B. Istace, K. Lebrigand, P. Barbry, S. Engelen, A. Lemainque, P. Wincker, G. Liti, and J. Schacherer (2018), Nature **556** (7701), 339.
[42] Petti, S., G. Reddy, and M. M. Desai (2023), Genetics **225** (1), iyad127.

[43] Piatigorsky, J. (2008), Evolution: Education and Outreach **1** (4), 403.
[44] Sackton, T. B., P. Grayson, A. Cloutier, Z. Hu, J. S. Liu, N. E. Wheeler, P. P. Gardner, J. A. Clarke, A. J. Baker, M. Clamp, and S. V. Edwards (2019), Science **364** (6435), 74.
[45] Sage, R. F. (2004), New Phytologist **161** (2), 341.
[46] Schmidlin, K., C. B. Ogbunugafor, and K. Geiler-Samerotte (2024), "Environment by environment interactions (ExE) differ across genetic backgrounds (ExExG),".
[47] Sella, G., and N. H. Barton (2019), Annual Review of Genomics and Human Genetics **20**, 461.
[48] Sivakumaran, S., F. Agakov, E. Theodoratou, J. G. Prendergast, L. Zgaga, T. Manolio, I. Rudan, P. McKeigue, J. F. Wilson, and H. Campbell (2011), American Journal of Human Genetics **89** (5), 607.
[49] Skwara, A., K. Gowda, M. Yousef, J. Diaz-Colunga, A. S. Raman, A. Sanchez, M. Tikhonov, and S. Kuehn (2023), Nature Ecology & Evolution **7** (11), 1823.
[50] Stern, D. L. (1998), Nature **396** (6710), 463.
[51] Stern, D. L., and V. Orgogozo (2008), Evolution; International Journal of Organic Evolution **62** (9), 2155.
[52] Szathmáry, E. (1993), Genetics **133** (1), 127.
[53] Tenaillon, O., A. Rodríguez-Verdugo, R. L. Gaut, P. McDonald, A. F. Bennett, A. D. Long, and B. S. Gaut (2012), Science **335** (6067), 457.
[54] Tikhonov, M., S. Kachru, and D. S. Fisher (2020), Proceedings of the National Academy of Sciences **117** (16), 8934.
[55] Turner, C. B., C. W. Marshall, and V. S. Cooper (2018), Evolution Letters **2** (4), 355.
[56] Venkataram, S., B. Dunn, Y. Li, A. Agarwala, J. Chang, E. R. Ebel, K. Geiler-Samerotte, L. Hérissant, J. R. Blundell, S. F. Levy, D. S. Fisher, G. Sherlock, and D. A. Petrov (2016), Cell **166** (6), 1585.
[57] Visscher, P. M., and J. Yang (2016), Nature Genetics **48** (7), 707.
[58] Wagner, G. P., and L. Altenberg (1996), Evolution **50** (3), 967, 2410639.
[59] Wolfe, J. M., J. Luque, and H. D. Bracken-Grissom (2021), BioEssays **43** (5), 2100020.
[60] Wong, A., and R. Kassen (2011), Microbiology **157** (4), 937.
[61] Wong, D. P. G. H., and B. H. Good (2024), Nature Communications **15** (1), 1605.

SUPPLEMENTAL INFORMATION (SI)

1. TABLE OF ALL PERTURBATIONS

Environment Name	Base media	0.5M NaCl?	Glucose concentratic	Extra component?	Flask shape	Transfer time	Temperature	Shaking	Batch
2Day base	M3	No	1.50%	No	Normal	48 hours	30C	Yes	2,3,4
2Day_1.4	M3	No	1.40%	No	Normal	48 hours	30C	Yes	2
2Day_1.4Baffle	M3	No	1.40%	No	Baffle	48 hours	30C	Yes	2
2Day_1.8	M3	No	1.80%	No	Normal	48 hours	30C	Yes	2
2Day_1.8Baffle	M3	No	1.80%	No	Baffle	48 hours	30C	Yes	2
2Day_Suc	M3	No	1.50%	1% Sucrose	Normal	48 hours	30C	Yes	3
2Day_SucBaffle	M3	No	1.50%	1% Sucrose	Baffle	48 hours	30C	Yes	3
2Day_Raffinose	M3	No	1.50%	0.5% Raffinose	Normal	48 hours	30C	Yes	3
2Day_RaffinoseBaffle	M3	No	1.50%	0.5% Raffinose	Baffle	48 hours	30C	Yes	3
2Day_No Shake	M3	No	1.50%	No	Normal	48 hours	30C	No	3
2Day_0.5% EtOH	M3	No	1.50%	0.5% Ethanol	Normal	48 hours	30C	Yes	4
2Day_4uMH89	M3	No	1.50%	4 uM H89	Normal	48 hours	30C	Yes	4
2Day_10uMH89	M3	No	1.50%	10 uM H89	Normal	48 hours	30C	Yes	4
2Day_10uMParomomyci	M3	No	1.50%	10 uM Paromomycin	Normal	48 hours	30C	Yes	4
2Day_50uMParomomyci	M3	No	1.50%	50 uM Paromomycin	Normal	48 hours	30C	Yes	4
1Day Base	M3	No	1.50%	No	Normal	24 hours	30C	Yes	1,2,3,4
1Day 30Baffle	M3	No	1.50%	No	Baffle	24 hours	30C	Yes	1
1Day 32	M3	No	1.50%	No	Normal	24 hours	32C	Yes	1
1Day 32Baffle	M3	No	1.50%	No	Baffle	24 hours	32C	Yes	1
1Day 28	M3	No	1.50%	No	Normal	24 hours	28C	Yes	1
1Day_1.4	M3	No	1.40%	No	Normal	24 hours	30C	Yes	2
1Day_1.4Baffle	M3	No	1.40%	No	Baffle	24 hours	30C	Yes	2
1Day_1.8	M3	No	1.80%	No	Normal	24 hours	30C	Yes	2
1Day_1.8Baffle	M3	No	1.80%	No	Baffle	24 hours	30C	Yes	2
1Day_Suc	M3	No	1.50%	1% Sucrose	Normal	24 hours	30C	Yes	3
1Day_SucBaffle	M3	No	1.50%	1% Sucrose	Baffle	24 hours	30C	Yes	3
1Day_Raffinose	M3	No	1.50%	0.5% Raffinose	Normal	24 hours	30C	Yes	3
1Day_RaffinoseBaffle	M3	No	1.50%	0.5% Raffinose	Baffle	24 hours	30C	Yes	3
1Day_No Shake	M3	No	1.50%	No	Normal	24 hours	30C	No	3
1Day_0.5% EtOH	M3	No	1.50%	0.5% Ethanol	Normal	24 hours	30C	Yes	4
1Day_4uMH89	M3	No	1.50%	4 uM H89	Normal	24 hours	30C	Yes	4
1Day_10uMH89	M3	No	1.50%	10 uM H89	Normal	24 hours	30C	Yes	4
1Day_10uMParomomyci	M3	No	1.50%	10 uM Paromomycin	Normal	24 hours	30C	Yes	4
1Day_50uMParomomyci	M3	No	1.50%	50 uM Paromomycin	Normal	24 hours	30C	Yes	4
Salt Base	M3	Yes	1.50%	No	Normal	24 hours	30C	Yes	1,2,3
Salt 30Baffle	M3	Yes	1.50%	No	Baffle	24 hours	30C	Yes	1
Salt 32	M3	Yes	1.50%	No	Normal	24 hours	32C	Yes	1
Salt 32Baffle	M3	Yes	1.50%	No	Baffle	24 hours	32C	Yes	1
Salt 28	M3	Yes	1.50%	No	Normal	24 hours	28C	Yes	1
Salt_1.4	M3	Yes	1.40%	No	Normal	24 hours	30C	Yes	2
Salt_1.4Baffle	M3	Yes	1.40%	No	Baffle	24 hours	30C	Yes	2
Salt_1.8	M3	Yes	1.80%	No	Normal	24 hours	30C	Yes	2
Salt_1.8Baffle	M3	Yes	1.80%	No	Baffle	24 hours	30C	Yes	2
Salt_Suc	M3	Yes	1.50%	1% Sucrose	Normal	24 hours	30C	Yes	3
Salt_SucBaffle	M3	Yes	1.50%	1% Sucrose	Baffle	24 hours	30C	Yes	3
Salt_Raffinose	M3	Yes	1.50%	0.5% Raffinose	Normal	24 hours	30C	Yes	3
Salt_RaffinoseBaffle	M3	Yes	1.50%	0.5% Raffinose	Baffle	24 hours	30C	Yes	3
Salt_No Shake	M3	Yes	1.50%	No	Normal	24 hours	30C	No	3
Salt_0.5% EtOH	M3	Yes	1.50%	0.5% Ethanol	Normal	24 hours	30C	Yes	4
Salt_4uMH89	M3	Yes	1.50%	4 uM H89	Normal	24 hours	30C	Yes	4
Salt_10uMH89	M3	Yes	1.50%	10 uM H89	Normal	24 hours	30C	Yes	4
Salt_10uMParomomycin	M3	Yes	1.50%	10 uM Paromomycin	Normal	24 hours	30C	Yes	4
Salt_50uMParomomycin	M3	Yes	1.50%	50 uM Paromomycin	Normal	24 hours	30C	Yes	4