Evolutionary adaptation proceeds through a small number of phenotypic modules

Mohammad Hossein Donyavi[1,2], Reza Ghelich[1,2], Kara Schmidlin[1,2], Grant Kinsler[3], and Kerry Geiler-Samerotte[1,2,*]

[1]Biodesign Center for Mechanisms of Evolution, Arizona State University, Tempe AZ
[2]School of Life Sciences, Arizona State University, Tempe, AZ
[3]Department of Bioengineering, University of Pennsylvania, Philadelphia, PA
[*]Corresponding author: kerry.samerotte@asu.edu

December 17, 2025

Abstract

Understanding how the myriad molecular impacts of mutation percolate to influence higher-order traits and ultimately fitness requires compressing a many-to-many mapping into something tractable. Decades of theoretical work suggest this may be possible because biological systems are modular: effects of perturbation are often funneled through particular pathways or subsystems rather than propagating freely through the organism. Yet few empirical systems have been able to demonstrate such modularity at scale. Prior low-dimensional models hinted that mutation–fitness relationships can collapse onto a small number of latent axes, but these studies relied on simpler datasets, leaving open whether such structure generalizes to more genetically and environmentally complex systems. Here we show that, even across 774 diverse yeast lineages, fitness variation across 12 drug environments is organized by a strikingly low-dimensional structure defined by only a few inferred phenotypic axes that capture the main patterns of variation. Consistent with these lineages having evolved under strong selection pressure, they reveal a striking asymmetry in the genotype–phenotype map: their mutations exhibit broad pleiotropy, affecting nearly all inferred phenotypic axes, yet fitness in any given drug depends on a much sparser subset of the phenotypic modules these axes reflect. This architecture aligns with central expectations of evolutionary theory. Strong selection often favors mutations with broad physiological effects, whereas long evolutionary history shapes organisms into modular systems in which only certain trait combinations matter to fitness in particular environments. By compressing many-to-many relationships, this low-dimensional framework exposes the modular fitness space that constrains the pleiotropic effects of adaptive mutants. It also lays the groundwork for identifying the key phenotypic modules that matter for fitness across different environments.

1 Introduction

Predicting how genetic variation shapes phenotype and fitness across diverse contexts remains one of biology's most enduring challenges, with implications for understanding disease, engineering microbes, and anticipating evolution [1–5]. This difficulty arises from the complexity of the genotype–phenotype–fitness map, where mutations can be pleiotropic, altering many molecular and physiological phenotypes [6–9]. Further, the fitness consequences of these phenotypic changes can vary across environments [7, 10, 11]. These many-to-many mappings, linking each mutation to many phenotypes, and each phenotype to fitness in multiple environments, make it difficult to predict fitness effects [4, 12].

Surprisingly, despite this complexity, recent work shows that variation in mutant fitness across environments can often be captured by low-dimensional linear models [7, 13–16]. These models infer abstract phenotypic "axes" or "dimensions" from fitness data alone that summarize shared patterns of response among many mutations and environments. Their success implies an underlying structure in the genotype–phenotype–fitness map, often described in terms of modularity and sparsity (Figure 1).

Modularity (Figure 1A) refers to the extent to which mutational effects converge on a limited set of phenotypic axes that mediate fitness [6, 7, 17, 18]. In a modular system, the myriad consequences of mutation are funneled through a few physiological modules, such as stress-response activation or efflux regulation. The small number of components makes fitness easier to predict. Sparsity (Figure 1B), by contrast, describes whether each mutation influences many phenotypic modules, and whether each module affects fitness broadly or only in specific contexts [7, 16]. Sparsity improves interpretability because an inferred phenotype is easier to identify when only a specific class of mutations or pathways influences it.

The extent of modularity and sparsity in biological systems has been debated for decades [6, 8, 17, 19, 20]. These properties trace back to Fisher's early questions about how evolution works: how phenotypes are coupled, which "phenotypic dials" selection can turn independently, and how the structure of the genotype–phenotype map constrains or enables change [18, 21–23]. Understanding that structure is central not only for explaining evolution, but also for predicting, and ultimately steering, its outcomes [2, 24–26].

Drug resistance evolution provides an ideal testbed for probing how modularity and sparsity shape predictability. Mutations that confer resistance to one drug often produce trade-offs in others [15, 27, 28], and these tradeoffs differ across mutants [10, 28, 29]. Some mutants show non-monotonic fitness responses to changes in drug concentration, complicating predictions even across similar environments [10, 29–31]. This variability poses a critical challenge for the design of therapies that seek to trap pathogens in an evolutionary 'double bind' by switching to drugs that exploit the specific weaknesses of resistant lineages. If different mutants evolve via different mechanisms with distinct tradeoff profiles, then a single second-line therapy may fail to eliminate the diverse sub-populations that arise. We ask: Do all mutants evolving resistance to the same drug exhibit the same fitness landscape, or is the predictability of these tradeoffs obscured by the complexity of the underlying molecular mechanisms? But if this apparent complexity conceals modular and sparse structure, evolution may prove more predictable than it seems.

Here we test these ideas by analysing 774 yeast lineages previously evolved in one of twelve antifungal environments [10]. Despite extensive genetic diversity, including mutations involving efflux pumps, chaperones, signalling pathways, or membrane components, and despite differing fitness trade-offs and non-monotonic dose responses, variation in fitness across the twelve environments can be captured and predicted by a low-dimensional model. The success of this model indicates strong modularity: diverse mutations converge on a few shared physiological axes that drive fitness across environments. Yet the level of sparsity is asymmetric. The links from mutations to phenotypic modules are not sparse, each mutation perturbs multiple mod-

ules, indicating high pleiotropy. On the contrary, the links from phenotypic modules to each environment are sparse. This asymmetry is consistent with expectations from Fisher's geometric model: adaptation to strong or novel selection proceeds through large, pleiotropic effects, while longer-term or stabilizing selection favours modular architectures where phenotypes shape fitness in only a few environments [23, 32–36]. Our results demonstrate that, even in a system with high pleiotropy, the underlying map has a modular structure that makes evolutionary outcomes more predictable.

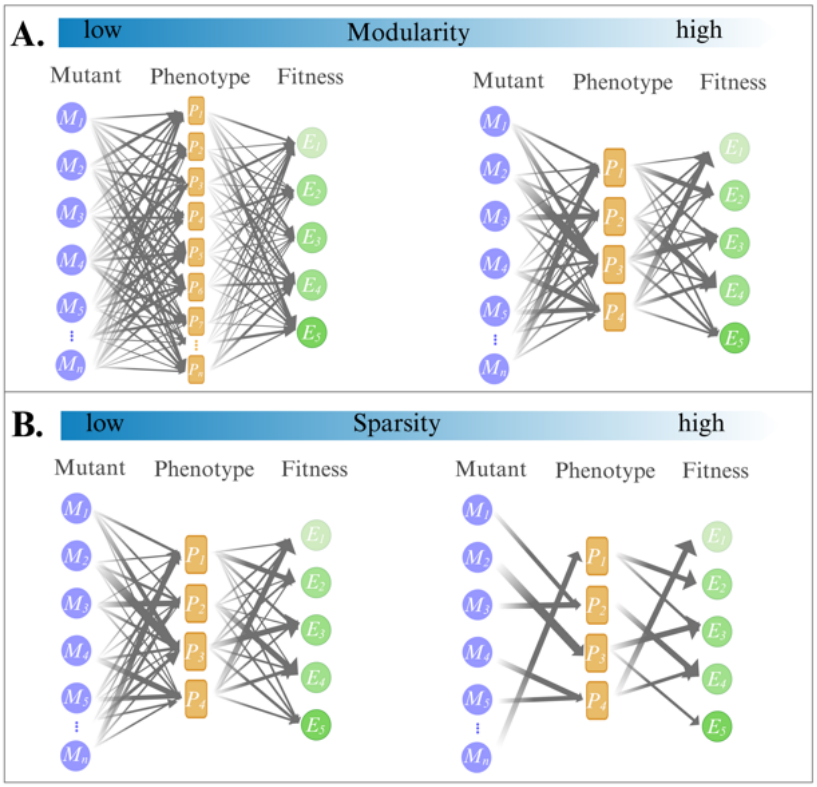

Figure 1: **A toy model of the genotype-phenotype-fitness (GPF) map that highlights two key features: modularity and sparsity.** These GPF map models are modified from previous work [7, 16] to highlight focal concepts in our study. Each map consists of different mutants (blue circles) that influence one or more orthogonal phenotypic modules (yellow squares) to differing degrees (thickness of grey arrows). The phenotypes each affect fitness in one or more environments (green circles) to differing degrees. **(A)** We use the term "modularity" to refer to the number of orthogonal phenotypic axes that mediate fitness (depicted as the number of yellow squares) [7]. In a map with low modularity (top left), the effects of mutations are diffuse, mapping onto a large and complex set of phenotypes that each contribute to fitness in many environments. This represents an architecture where predicting fitness would be more difficult. With high modularity (top right), the effects of different mutations converge on a few phenotypic axes that dictate fitness. This simplifies the map by reducing its effective dimensionality, which could make fitness in unseen environments more predictable. **(B)** We use the term "sparsity" to refer to the density of connections between mutants, phenotypes, and environments (depicted as the number of grey arrows) [16]. Connections are dense with a low sparsity structure (bottom left). In a high sparsity map (bottom right) a given mutant may only affect a few phenotypes, and a phenotype may only be relevant for fitness in a specific subset of environments. In this study, we first test for modularity, whether fitness variation can be captured by a few shared phenotypic axes, and later assess sparsity, which describes how densely mutations and environments connect to those axes.

2 Results

2.1 A low-dimensional model predicts the fitness effects of 774 diverse mutations

To explore the extent of modularity in the genotype-phenotype map, we asked whether low dimensional models could make accurate fitness predictions for diverse mutants. While earlier work focused on yeast mutants evolved under a single selection pressure [7], here we explore yeast evolved under 12 different concentrations or combinations of antifungal drugs [10]. Like previous work, each evolved mutant lineage we study differs from their shared ancestor by, on average, a single nucleotide mutation [7]. Unlike previous work, the targets of mutation among our 774 lineages are more diverse, including mutations to drug efflux pumps, protein folding chaperones, glucose sensing pathways and cell membrane components [10]. In addition to diversity at the genotype level, our lineages also have fascinating diversity in their gene-by-environment interactions. Some are broadly drug resistant, others are uniquely resistant or sensitive to narrow drug concentrations, and still others resist single drugs but not their combination [10, 29]. Despite this diversity, we found that low dimensional models can predict mutant fitness in drug combinations even when they are built using only single-drug data from a subset of 100 mutants.

Using Singular Value Decomposition (SVD) [7] (Figure 2A), we built low dimensional models summarizing fitness variation across a training set consisting of 100 randomly drawn mutants in 8 single-drug and no drug conditions. Then we predicted the fitness of the remaining 664 mutants in 4 double-drug combinations (Figure 2B). The fitness predictions we report in figure 2 represent the average across 100 iterations in each of which 100 mutants were randomly chosen for model training. In each iteration, we inferred latent phenotypic dimensions that minimized the mean squared error of the fitness predictions for the held out 664 lineages. Each latent phenotypic dimension represents an axis of variation in mutant fitness. These dimensions can be thought of as abstract phenotypes inferred from the data, somewhat like principal components inferred from high-dimensional phenotyping studies. We examine the biological interpretation of each inferred phenotypic dimension later. First, we focus on their utility for revealing the structure and modularity of the genotype–phenotype map. In essence, through this modeling, we are asking how well a small number of dimensions can capture how genotype translates to phenotype and fitness.

Even though our set of mutants and gene-by-environment interactions is diverse, we found that SVD models were effective at predicting fitness of held out mutants in held out environments (Figure 2C–D). Prediction accuracy across the four double-drug environments ranged from roughly 50% to 70%, comparable to previous studies that used similar approaches to show modularity in genotype–phenotype–fitness relationships [7, 14]. This predictive success implies that diverse mutational effects converge on a few shared phenotypic dimensions, the hallmark of modularity introduced in figure 1. The fact that the main axes of phenotypic variation learned from single-drug data could generalize to drug combinations suggests that fitness variation is structured by a small number of underlying processes that repeatedly shape outcomes across conditions. In other words, despite mutations affecting diverse genes and pathways, from efflux pumps to chaperones and signaling components, their phenotypic consequences appear to converge on a limited set of inferred phenotypic 'dimensions' or 'modules' that represent major contributors to fitness.

Previous work suggests that the lower prediction accuracy of SVD dimensional models in certain environments may arise because not all phenotypic modules are relevant in every condition. Some key phenotypic effects of mutation remain hidden when the training set lacks the environments where these phenotypes constrain fitness. In this view, only a small subset of phenotypes determines fitness in any given environment, while pleiotropic effects on other phenotypes remain latent until other contexts are considered [7, 14]. Thus, modularity both

enables and constrains predictability. The shared phenotypic modules make low-dimensional modeling possible, but context-specific ones define the upper bound on prediction accuracy. In the next section, we test these ideas by asking how prediction accuracy changes as we expand the number of environments, mutants, and model components used for training.

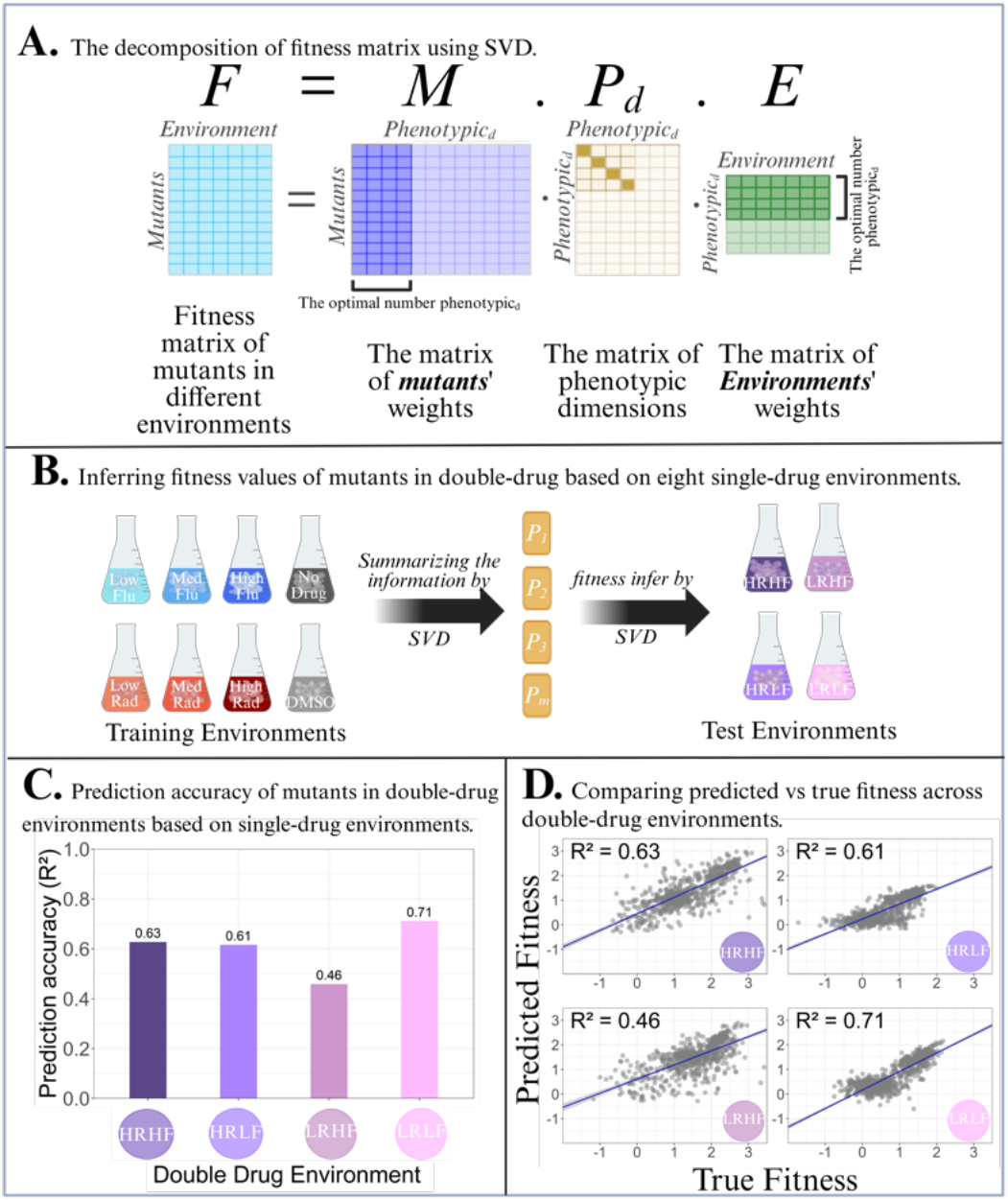

Figure 2: **A low-dimensional linear model predicts mutant fitness in double-drug environments from single-drug fitness.** **(A)** Schematic of Singular Value Decomposition (SVD) applied to a fitness matrix (F). The matrix is decomposed into mutant weights (M), phenotypic dimension weights (Pd), and environmental weights (E), reducing the data into a lower-dimensional space that captures key genotype–phenotype–fitness relationships. **(B)** Training and testing environments. The SVD model was trained on fitness measured in eight single-drug environments [10]: three Radicicol concentrations (red; 5 μM, 15 μM, 20 μM), three Fluconazole concentrations (blue; 4 μg/ml, 6 μg/ml, 8 μg/ml), and two controls (gray; DMSO, no drug). Model accuracy was tested in four double-drug environments (purple): HRHF (15 μM Rad + 8 μg/ml Flu), HRLF (15 μM Rad + 4 μg/ml Flu), LRHF (5 μM Rad + 8 μg/ml Flu), and LRLF (5 μM Rad + 4 μg/ml Flu). **(C)** Prediction accuracy of the SVD model in the four double-drug environments. **(D)** Scatter plots underlying the R^2 values in panel D, comparing predicted and measured fitness for each of 774 mutants. Predictions were generated by repeatedly training on random subsets of 100 mutants and testing on the remainder; each mutant's reported prediction is the average across all iterations where it was held out of training.

2.2 Rapid Saturation of Prediction Accuracy Reveals Modular Structure and Its Limits

To assess how modular structure constrains and enables fitness prediction across environments, we examined how prediction accuracy changes as we systematically increase the number of training environments. For each training set size, we repeated the analysis 100 times, randomly selecting a new subset of environments for training in each iteration. We performed this analysis in our dataset (12 environments; Figure 3A) and in a second, larger dataset (45 environments; Figure 3B) [7]. In both datasets, prediction accuracy increased sharply with the first few added environments before rapidly plateauing (Figure 3; right side). The number of phenotypic dimensions required for high predictive power showed a similar pattern, leveling off after only a few dimensions (Figure 3; left side). Even across model iterations where the 'optimal' number of dimensions (the number that minimized mean squared error) was greater, predictive power does not improve (Figure 3; 3D plots). This rapid saturation suggests that a small, representative subset of environments captures most of the predictable signal.

This behavior supports a modular view of the genotype–phenotype–fitness map [7], in which a limited number of causal phenotypes consistently govern fitness across diverse conditions. Once those shared phenotypic modules are learned, adding more environments yields little additional predictive power. At the same time, the plateau exposes a fundamental limit: beyond a small amount of training data, prediction accuracy no longer improves. This ceiling could arise for at least two reasons. First, it may reflect an inherent limitation of linear models such as SVD, which cannot capture nonlinear relationships between phenotypes and fitness, including threshold or saturation effects known to shape adaptive landscapes [13, 37]. Second, as explained above and emphasized by Ghosh et al. [14], not all phenotypic functions contribute to fitness in every environment. Thus, some remain hidden from SVD models unless the environments where they matter are represented in training. In this "limiting functions" view, expanding environmental diversity should reveal these latent modules and, in turn, improve prediction accuracy in the contexts where those modules constrain fitness. If that logic holds, then datasets spanning more diverse environments should require more training environments to reach saturation in predictive power.

Yet, when we compared datasets, the opposite occurred. The larger, 45-environment dataset actually reached saturation in prediction accuracy faster than the smaller, 12-environment one, contradicting the expectation that broader environmental diversity uncovers more phenotypic modules. This pattern is clear in figure 3: prediction accuracy in the 45-environment dataset (bottom) rises more steeply with the number of training environments and plateaus earlier than in the 12-environment dataset (top). Further, a direct comparison of how prediction accuracy improves with each additional training environment shows that, beyond the third added environment, gains in the 45-environment dataset were consistently smaller than in the 12-environment dataset (Figure S1). In other words, the 45-environment dataset required fewer training environments to saturate prediction accuracy than did the 12-environment dataset.

The most straightforward explanation, that the 45 environments are simply more similar to one another, is not supported by the data. The 45 environment set spans a wide range of conditions, including antifungal drugs, salts, oxygenation levels, carbon sources, and metabolic states [7, 38]. The difference, therefore, cannot be explained by environmental redundancy. Something else must make predictions easier in the 45 environment dataset. We posit that the key difference lies not in the environments but in the mutants themselves. Our 12-environment dataset includes 774 mutants evolved under many distinct selective pressures, while the 45-environment dataset includes 292 mutants that all evolved under a single one: glucose limitation [39, 40]. Because latent phenotypes can only be revealed when they are perturbed, a model trained on a genetically uniform set of mutants will fail to detect modules that those mutations do not affect. If the mutants in the 45-environment dataset collectively perturb fewer phenotypes, then SVD applied to that dataset should uncover fewer phenotypic modules, even though it spans more

environments.

Indeed, this is what we observe. When we ask how many phenotypic dimensions are required to capture 95% of fitness variation in the full datasets (with no data held out), that number is smaller (three) for the 45-environment dataset and larger (four) for the 12-environment dataset (Figure S2). Similarly, when we hold out data and systematically add it back, the number of inferred phenotypic dimensions rises faster in the dataset with more mutants, not the one with more environments (Figure S2). This indicates that mutant diversity, in addition to environmental diversity, can drive the discovery of new phenotypic dimensions. This observation reflects something deeper, and wonderfully circular, about evolution. Selection has shaped cellular systems to be modular, and thus the solutions that evolution discovers are themselves modular. Mutants isolated from a single selective pressure tend to perturb a smaller number of phenotypic modules than mutants drawn from twelve distinct selective pressures. Still, in both cases, a small number of shared phenotypic modules underlie fitness. We next sought insights about those phenotypic modules by asking which environments reveal them and which mutants define them.

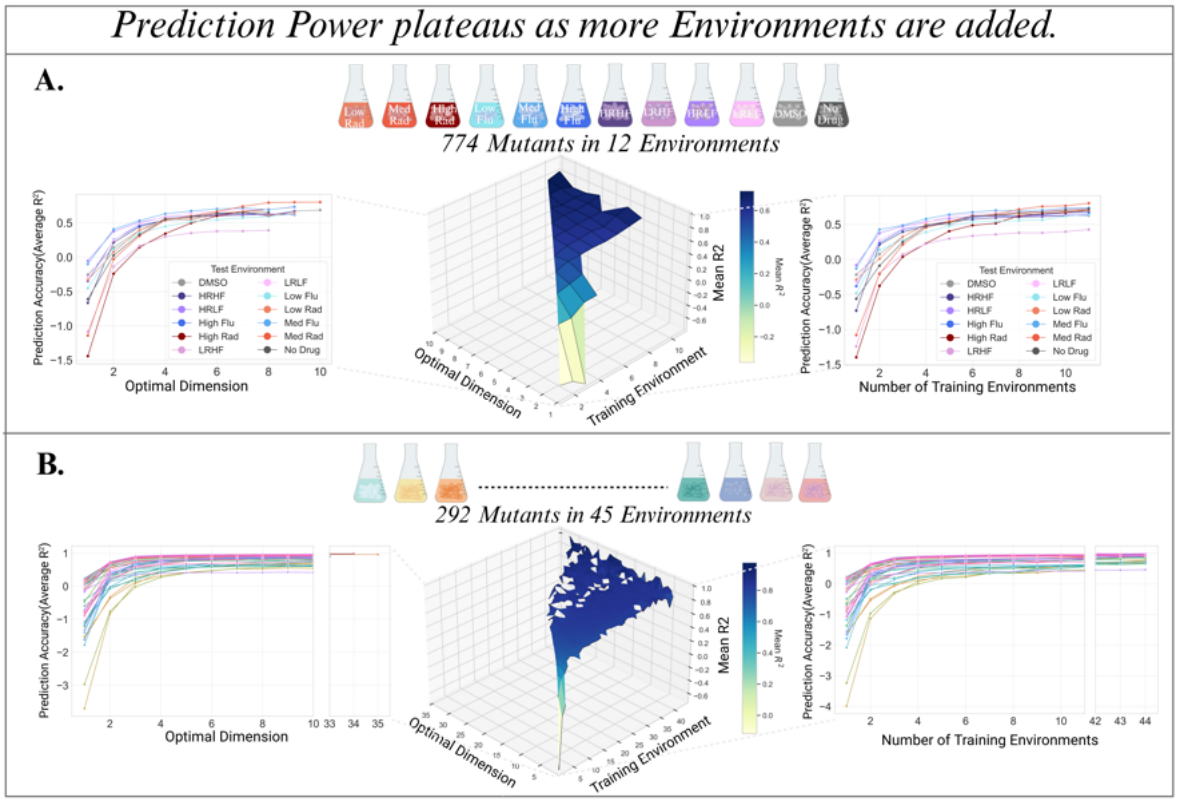

Figure 3: **Prediction accuracy saturates rapidly, revealing a low-dimensional and modular genotype-phenotype map.** To test how prediction accuracy depends on training data and model complexity, we varied the number of training environments (k) from 1 to $n-1$, where n is the total number of environments. For each k, 100 iterations were performed in which k environments were randomly selected for training, and predictive accuracy (R^2) was calculated on the held-out environments and averaged across iterations. Training mutants were chosen as in previous work [7], by selecting a small but diverse subset representing different G×E interactions. This differs from Figure 2 where training mutants were chosen at random, with 100 iterations. We applied this analysis to assess the impact of adding more training environments to two datasets: **(A)** fitness of 774 mutants across 12 environments [10], and **(B)** fitness of 292 mutants across 45 environments [7]. The right-most plots show that prediction accuracy rises rapidly and then saturates after just a few training environments; fewer in the bottom graph. This rapid saturation indicates that a small subset of environments captures the majority of the information needed to predict fitness. Similarly, the left-most plots show that prediction accuracy plateaus after just a few phenotypic dimensions, fewer in the bottom graph. This suggests the fitness effects of these mutations are explained by a small number of phenotypic dimensions. The central 3D surface integrates the left and right plots, showing the prediction accuracy (average R^2) as a function of both the number of training environments and optimal dimensions (i.e. the number of dimensions chosen after each iteration of bi-cross validation). The surface rises quickly and then forms a broad plateau, demonstrating that high predictive accuracy is achieved with a limited number of both environments and dimensions, with further additions making negligible improvement. This reinforces the conclusion that the system's dimensionality is low.

2.3 Interpreting phenotypic modules by examining informative environments and mutants

Having shown, like previous studies [7, 14, 41], that the map from genotype to phenotype to fitness appears to be modular, we next ask about the underlying biology that is captured by each phenotypic dimension. To better understand the inferred phenotypic dimensions that drive fitness in multidrug environments, we asked which single-drug environments contribute most to predictive accuracy. Our results demonstrate that individual environments can disproportionately reveal specific phenotypic modules. Specifically, low-dose fluconazole exposes a distinct axis of gene-by-environment interaction that shapes multidrug fitness.

Fluconazole-containing environments were the most informative in general. Removing all fluconazole environments from the training set caused a steep drop in predictive accuracy, while removing radicicol had more modest effects (Figure 4A). The most dramatic impact came from low-dose fluconazole (4 μg/ml): removing this single environment reduced prediction accuracy, while adding it back increased R^2 by ~ 0.2 for the LRLF condition (Figure 4B).

To understand why low fluconazole had such a strong effect, we examined the mutants whose predicted fitness changed the most when low fluconazole was excluded from the training data. Two strikingly different groups emerged (Figure 4C–E). The first group, mutants underpredicted without low fluconazole, had high fitness only at low fluconazole (4 μg/ml) but not in higher doses (Figure 4C, blue). The second, mutants overpredicted without low fluconazole, showed the opposite pattern, performing poorly only at low doses but well in all others (Figure 4E, red).

These mirror-image, nonmonotonic dose–response curves suggest that low-dose fluconazole is not simply a weaker version of drug stress but engages a distinct cellular regime. Similar dose-specific shifts have been observed in molecular studies, where low and high drug concentrations activate qualitatively different physiological responses [30, 31, 42]. Here, we uncovered this unique biology directly from an abstract genotype–fitness map where predictive performance pinpoints low-dose fluconazole as uniquely informative.

Consistent with the idea that low-dose fluconazole triggers a distinct cellular state, mutants with nonmonotonic fluconazole dose–response curves recur across our independent analyses (Figure S3, S6), and prior work has shown that different adaptive lineages emerge at low versus high fluconazole concentrations [30, 31, 43]. Including low fluconazole consistently improves multidrug prediction across model variants (Figure S4), and fitness in this environment is least correlated with any other condition (Figure S5), underscoring the unique role of low fluconazole in defining a key axis of the genotype–phenotype–fitness map. Together, these findings illustrate how abstract, low-dimensional models can be used not only to predict evolutionary outcomes, but to expose the environments and mutations that define key phenotypic modules.

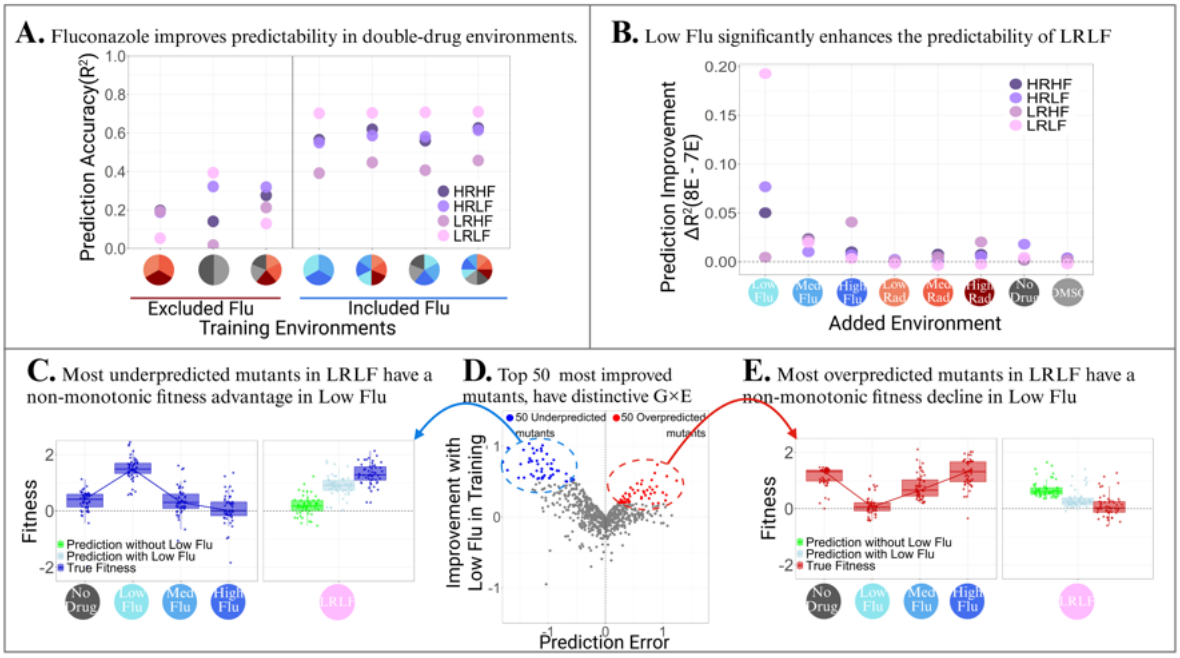

Figure 4: **Low-dose fluconazole (4 μg/ml) is a uniquely informative environment.** To identify which environments provide the most information for predicting fitness in double-drug combinations, we systematically eliminated environments from the training datasets. **(A)** Prediction accuracy improves when fluconazole (Flu) environments are included in the training set. Accuracy was quantified as R^2 between predicted and measured fitness, averaged over 100 training runs, where in each run a random set of 100 mutants was used to predict the fitness of the remaining 664 (as in Figure 2). **(B)** The specific contribution of each single-drug environment to prediction accuracy in each double-drug environment. The y-axis shows the improvement in R^2 when a given environment on the x-axis is added to a base training set of 7 single-drug environments. Adding Low Flu (4 μg/ml) provides the highest improvement in prediction accuracy, specifically for the LRLF test environment (light pink). **(C)** Fitness profiles of the 50 most underpredicted mutants (points) when Low Flu is removed from the training set display a non-monotonic response to fluconazole concentration, with a distinct fitness advantage in Low Flu. The model trained without Low Flu fails to capture this advantage and underpredicts fitness in LRLF (green). Including Low Flu in the training set allows the model to better approximate fitness in LRLF (light blue). **(D)** Relationship between prediction error when Low Flu is excluded from training, and prediction improvement when Low Flu is included (for the LRLF condition). Colored points display the 100 mutants for which predictions improve the most upon adding Low Flu to the training set (blue when excluding Low Flu leads to underprediction, red when it leads to overprediction). **(E)** Fitness profiles of the 50 most overpredicted mutants from panel D. These mutants exhibit an opposite non-monotonic pattern to those in panel C, with a fitness decline in the Low Flu environment compared to other Flu concentrations. The model trained without Low Flu misses this decline and overpredicts their fitness in LRLF (green).

2.4 Our map appears less sparse and more asymmetric than other datasets

So far, our analyses reveal that the genotype–phenotype–fitness map is modular in that fitness variation can be captured by a few shared phenotypic axes. We next ask whether these axes are broadly or sparsely connected (Figure 1). Whereas modularity explains why fitness variation can be captured by a few shared dimensions, sparsity determines how broadly those dimensions are used, whether each mutation or environment engages many modules or just a few. In a sparse

system, pleiotropy is limited: each mutation changes only a subset of the phenotypic modules. Further, each phenotypic module matters for fitness in only certain environments. This kind of sparse organization is theorized to arise to offset the "cost of complexity," the idea that when a mutation affects many traits simultaneously, it can make adaptation slower or less efficient [32]. At the same time, sparse structure may support evolvability, allowing different parts of an organism to change and adapt more independently [17, 18]. Thinking about sparsity in this way provides a framework for understanding whether adaptation to the drug environments we study here occurred through a handful of distinct routes or through broad, system-wide changes.

Here we study the degree of sparsity following previous work [16]. First we applied SVD to the complete matrix of 774 mutant fitness values across 12 environments, without holding out any data (Figure S7). Then we applied a sparsity-constrained variant of SVD, Sparse Structure Discovery (SSD) [16] (Figure 5, S7). SSD enforces sparsity by pruning links in the genotype–phenotype–fitness map (Figure 5B; grey arrows). This improves interpretability by highlighting which links are most meaningful.

In datasets with truly sparse structure, random rotations of the fitness matrix "F" (Figure 2A) will destroy the alignment between mutants (or environments) and the inferred phenotypic dimensions. As a result, more inferred dimensions are required to achieve similar reconstruction accuracy to the original, unrotated data. This is indeed what we observe for the mapping from phenotypic dimensions to environments, where there is a clear gap between the reconstruction errors on the original versus rotated data (Figure 5C). This suggests a sparse structure, where not every phenotype affects fitness in every environment. This corresponds to fewer grey arrows connecting phenotypes to environments in our conceptual model (Figure 5B).

By contrast, rotation tests had less effect on the alignment from mutants to inferred phenotypic dimensions (Figure 5A). This lack of a significant gap between the original and rotated data suggests that the mapping from mutants to phenotypes is less sparse. In other words, pleiotropy is higher, such that each mutation affects many phenotypic dimensions, as depicted by the many grey arrows connecting mutants to phenotypes in our conceptual model (Figure 5B).

This asymmetric pattern, high pleiotropy on the mutation-to-phenotype side, but sparsity on the phenotype-to-environment side, departs from prior maps in which both sides showed sparse structure [16]. To understand why, we step back to foundational theory. Under Fisher's Geometric Model, adaptation to novel environments is predicted to proceed via large-effect, often pleiotropic mutations, whereas stabilizing selection should favour smaller and more modular changes [23, 32, 33, 44]. Empirical evolution studies echo this pattern: early adaptation to novel or harsh environments often proceeds via large-effect, highly pleiotropic mutations in global regulators or metabolic hubs [45–47], whereas subsequent adaptive steps and adaptation under more benign environmental shifts tends toward smaller-effect, more modular changes [36, 48]. This framework suggests that the amount of pleiotropy (and thus sparsity) on the mutant-to-phenotype side of the map depends on the selection regime under which the mutations evolved.

Our results present a nuanced variant of that paradigm. We observe little sparsity on the mutant-to-phenotype side of the map. Because our mutants arose under a strong, drug-selection laboratory regime, this pattern is consistent with adaptation dominated by large, pleiotropic leaps. Yet it is striking that, despite such broad pleiotropy from mutation to phenotype, the downstream mapping from phenotypes to environments remains consistently sparse. One possible reason is that the phenotype-to-environment mapping has been shaped by long evolutionary history and natural selection: organisms evolve to produce a limited set of phenotypic modules, each relevant to fitness in only a subset of environments. This idea echoes theory and data showing that fluctuating environments favor modular, sparsely connected phenotypes [34, 35, 49] and that selection for evolvability tends to localize phenotypic effects rather than dis-

tribute them globally [18, 33]. In contrast, the mutation-to-phenotype mapping depends more directly on the immediate mutational input and selective regime. Thus, the consistent sparsity of phenotype-to-environment mapping may reflect deep-time selection for modular environmental responsiveness, while the mutation-to-phenotype side remains more variable and contingent on current evolutionary conditions. In this way, our findings not only support Fisher's geometric model–derived expectations about pleiotropy and effect size, but also suggest that the architecture of genotype–phenotype–environment maps is shaped by both past selective histories and present-day selective contexts.

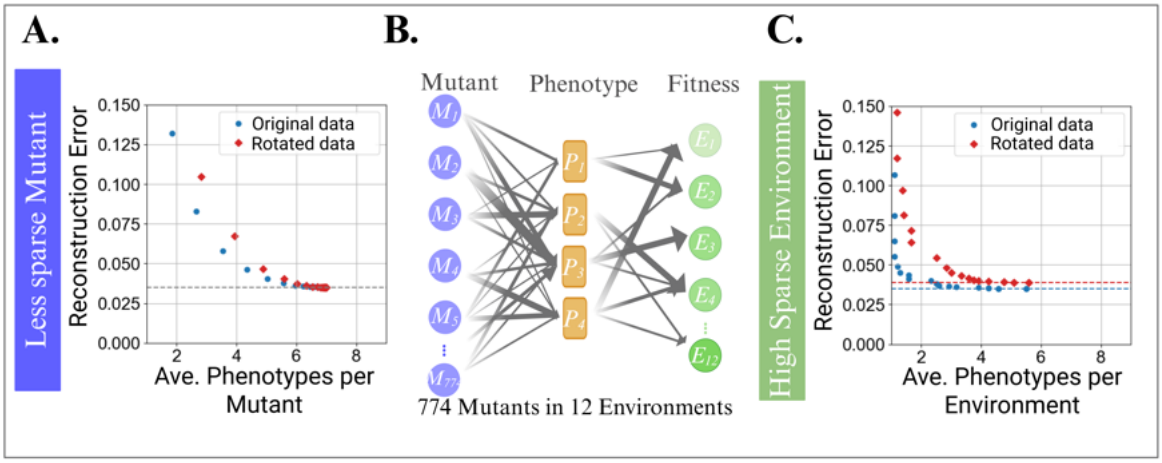

Figure 5: **A rotation test reveals a less sparse mutant-to-phenotype map (blue) but an inherently sparse environment-to-phenotype map (green).** To formally test for the presence of inherent sparsity in our experimental dataset, we applied a rotation test using a sparse discovery algorithm (SSD) [16]. The rotation test tries to find a sparse, low-error solution for the original fitness matrix, F, versus a rotated matrix ($F' = FO$), which preserves the correlational structure but removes any special sparse basis. A significant gap between the solution curves for the original and rotated data would indicate the presence of a sparse structure. **(A)** The rotation test versus original data for the mutant-to-phenotype mapping. The solution curves for the original and rotated data largely overlap. **(B)** A conceptual genotype-phenotype-fitness map that represents the amount of sparsity found within our dataset. Many grey arrows connect mutants to phenotypes, while fewer connect phenotypes to environments, depicting how the latter mapping is sparser. **(C)** The rotation test versus the original data for the phenotype-to-environment mapping. Each point represents a solution for a different regularization parameter. This plot shows the trade-off between reconstruction error and the average number of inferred phenotypes per environment. A distinct gap exists between the solutions for the original data (blue) and the rotated data (red). The original data can be explained by a much sparser model (fewer phenotypes per environment) without reducing accuracy, a clear hallmark of an inherently sparse structure in the environment-phenotype map.

2.5 Phenotypic axes correspond to gene-by-environment interaction patterns

Finally, to connect the modular and sparse structure of the genotype–phenotype–fitness map to underlying biology, we examined which mutants and environments define each phenotypic axis. The phenotypic axes recovered from SSD largely overlapped with those from standard SVD (Figure S7), regardless of which sparse solution we chose. This means the original vectors returned by SVD are not arbitrary but represent a stable, biologically relevant solution that the sparsity-constrained search of SSD independently confirms. This robustness is signif-

icant because it suggests the low-dimensional structure we identified is not a fragile statistical abstraction but a primary feature of our studied system.

Unlike other studies [16], the phenotypic dimensions we recover are difficult to interpret in terms of discrete cellular functions, perhaps due to low sparsity in the mapping from mutants to inferred phenotypes. For example, although 35 of our 774 mutants perturb regulators of multidrug efflux, there is no single phenotypic dimension that is affected by exclusively these mutants. They predominantly fall into group 2 (Figure 6A; green) and contribute strongly to phenotypic dimension 1 (Figure 6A-B). But consistent with lower sparsity, many other mutants have similar effects on this phenotypic dimension. These effects scale with average fitness advantage across drug environments. Thus, phenotypic dimension 1 seems to capture a common type of gene-by-environment interaction (G×E) whereby fitness is high in environments that contain drugs, and low in environments that do not (Figure 6B).

Similarly, the other phenotypic dimensions seem to map to G×E interactions more so than underlying physiological mechanisms by which mutants affect fitness. For example, phenotypic dimension 2 separates mutants based on relative advantage in fluconazole versus radicicol (Figure 6C). Dimensions 3 and 4 capture patterns of G×E related to changes in fluconazole concentration (Figure 6D–E).

These results suggest that the extracted dimensions likely reflect shared fitness outcomes of mutations, rather than the unique cellular process that underlie those outcomes. Interpreting the underlying biological mechanisms that link genotype to fitness may therefore require integrating additional information, such as mutant genotypes or transcriptional profiles. Still, the fact that just four dimensions explain the vast majority of fitness variation, despite the genetic diversity present across this collection of 774 mutants, implies that genotype-by-environment interactions are canalized. In other words, every genetic perturbation does not have a distinct pattern of G×E. Instead, a small number of G×E patterns appear to emerge that describe all mutants [10]. This canalization might arise from constraints imposed by regulatory network architecture or trade-offs in resource allocation, which may in turn restrict the evolutionary routes available for drug resistance. Whatever the cause, this structure is encouraging because it shows that even though mutations are inherently pleiotropic, in the sense that they can influence countless molecular-level traits, their fitness effects often funnel through a small number of shared axes. This implies that mapping the underlying molecular mechanisms onto these axes, identifying the core processes that shape fitness, and ultimately predicting evolution, may be more tractable than one might expect given the apparent complexity of all that a mutation could, in principle, alter inside a cell.

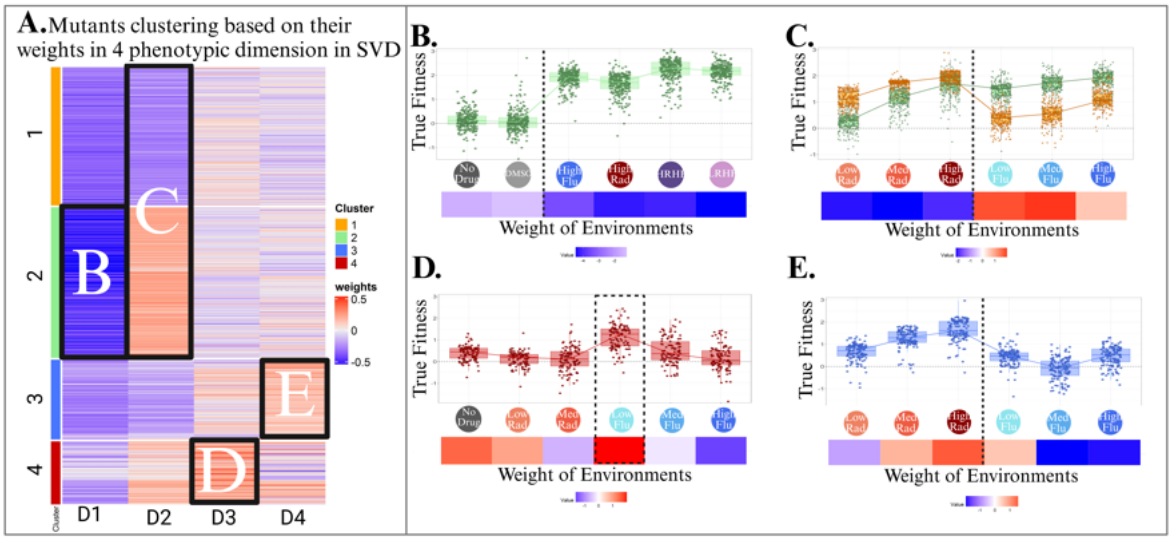

Figure 6: **A four-dimensional model decomposes the fitness matrix into phenotypic axes that capture gene-by-environment interactions.** Singular Value Decomposition (SVD) was applied to the complete fitness matrix of 774 mutants across all 12 environments to interpret the underlying structure of the genotype-phenotype map. **(A)** A heatmap of the contribution of all 774 mutants to each inferred phenotypic dimension (red vs. blue indicate strong contribution in opposite directions). Mutants are clustered based on their weight profiles (calculated based on $M \times \sqrt{P}$). Boxes highlight the mutant clusters detailed in panels B–E. **(B)** This phenotypic dimension is most strongly affected by mutants that have fitness advantages across environments containing drugs that are lost in environments without drugs. Mutants with amino acid changes in the PDR1 and PDR3 genes, which are regulators of drug efflux pumps, largely fall into this category. **(C)** This dimension is strongly influenced in opposite directions by two groups of mutants: one that resists flu more than rad (green), and another that resists rad more than flu (orange). **(D)** This dimension captures a specific response to low-dose fluconazole, which has the strongest positive environmental weight. Mutants with high weights in this dimension show a distinct fitness advantage in this specific environment. **(E)** This dimension reflects a distinct nonmonotonic response to fluconazole: fitness is lowest at intermediate concentrations. This pattern differs from the nonmonotonic responses in Figure 4, where mutants are instead most sensitive in low fluconazole.

3 Discussion

Our results show that, despite extensive genetic and environmental complexity, the fitness effects of hundreds of antifungal-resistance mutations can be captured by a strikingly low-dimensional structure. A small number of inferred phenotypic axes explain most variation in mutant fitness across drugs and doses, and axes learned from single-drug environments generalize to predict fitness in drug combinations. This supports a strongly modular genotype–phenotype–fitness map. Rather than each mutant having its own idiosyncratic genotype-by-environment (G×E) pattern, evolution in multidrug environments is governed by a few repeatable ways that mutants interact with drugs.

Our results echo ideas from the omnigenic perspective on complex traits, which holds that mutations in many genes can shape trait variation through dense regulatory networks [19, 50]. In our case, mutational effects did not scatter unpredictably but instead collapsed onto a small set of shared axes. A recent quantitative omnigenic model formalized this principle, showing that network-level complexity can be summarized in low-dimensional, interpretable structures

[51]. The key difference is that while the quantitative omnigenic framework is a bottom up approach deriving axes from regulatory network topology, our top-down SVD approach infers them from observed fitness patterns alone. Yet both converge on the idea that molecular complexity funnels through a limited set of modules that enable prediction.

In our dataset, the architecture of this modular map is asymmetric. Our sparsity analysis shows that the mapping from mutations to phenotypic modules is dense, consistent with high pleiotropy. However, the mapping from phenotypic modules to environments is sparse, with each module affecting fitness in a restricted subset of conditions. In this view, our experiment exposes a highly pleiotropic "top half" of the map sitting on top of a relatively stable, modular phenotype–environment mapping. This asymmetric architecture may be common under adaptation to novel environments or strong stressors. This asymmetry paints a picture of the genotype-phenotype-fitness map as a 'funnel'. The wide top of the funnel represents the immense diversity of potential genetic changes and their myriad molecular effects (dense mutation-to-phenotype mapping), which are then channeled through a narrow neck of physiological constraints (the low-dimensional phenotypic modules).

Our analysis also highlights how the inferred dimensionality of any given genotype-phenotype map depends as much on which mutants are assayed as on which environments are sampled. We showed that increasing genetic diversity can reveal new phenotypic modules even when environmental coverage is modest, whereas expanding environmental panels without diversifying mutants may uncover relatively fewer modules. This has practical implications for experimental design: to map genotype–phenotype–fitness structure efficiently, it may be important to study many kinds of mutants as well as many environments. Even with pervasive pleiotropy, mutations differ in how strongly they affect each phenotypic module, so sampling diverse mutants is essential for resolving the underlying structure.

Low dimensional models can do more than reveal the structure of the genotype-phenotype relationships, here we show that they can also be used as discovery tools. Low-dose fluconazole emerged as a particularly informative condition that likely perturbs a unique physiological response. Excluding it from the training set degraded predictive performance despite other fluconazole concentrations being included in the model. This mirrors molecular evidence that low and high drug concentrations can induce distinct physiological states, rather than a simple graded stress response. Crucially, we detected this potentially unique physiological response, and the specific mutants that define it, purely from patterns of predictive success and failure.

These findings have immediate implications for the management of antimicrobial resistance, particularly for collateral sensitivity cycling. Our data definitively show that tradeoffs are not uniform; mutants selected by the same drug can reside in opposing regions of the fitness landscape in a second environment. Consequently, a 'one-size-fits-all' second-line drug may effectively treat one cluster of resistant mutants while acting as a breeding ground for another. Effective multidrug therapies must therefore move beyond population averages and towards 'cluster-specific' strategies, where the specific phenotypic fingerprint of the resistant sub-population dictates the next move in the therapeutic sequence.

Looking forward, while our linear model successfully captures the bulk of the fitness variation, its specific failures are as illuminating as its successes. The identification of environments and mutant subsets that defy prediction points to potential non-linear interactions that shape fitness. Several questions remain. How general is the low-dimensional modular structure across timescales, selective pressures, or organisms [6, 7, 17, 32]? What molecular or physiological variables underlie the axes we infer, and how can we map them to interpretable traits [12, 16, 52]? To what extent do nonlinear or environment-specific modules emerge in more complex conditions [14, 53]? Addressing these questions will require integrating multi-omics data and nonlinear modeling frameworks. Such approaches may bridge the gap between predictive low-dimensional structure and the underlying mechanisms, illuminating how mutations ripple through networks to produce fitness effects.

4 Methods

4.1 Fitness data, yeast strains and experimental design

The yeast strains and experimental methods employed in this study were previously described in detail [10]. Briefly, all yeast strains originated from a single ancestral "landing pad strain" (SHA185), a $MAT\alpha$ haploid strain characterized by the genetic background: *ura3Δ0, ybr209w::Gal-Cre-KanMX-1/2URA3-loxP-Barcode-1/2URA3-HygMX-lox66/71.*

To generate the mutants, around 300,000 unique neutral DNA barcodes were introduced into this strain using a Cre-lox recombination system [40]. These barcoded yeast were evolved in 12 environments. These 12 evolution experiments were conducted in a glucose-limited M3 medium lacking uracil, supplemented with fluconazole, radicicol, or DMSO at the following concentrations: Low Flu (4 μg/ml), Med Flu (6 μg/ml), High Flu (8 μg/ml), Low Rad (5 μM), Med Rad (15 μM), High Rad (20 μM), LRLF (5 μM + 4 μg/ml), LRHF (5 μM + 8 μg/ml), HRLF (15 μM + 4 μg/ml), HRHF (15 μM + 8 μg/ml), 0.5% DMSO (0.5%) and No Drug (no supplement). Drug concentrations were chosen to apply mild selective pressures without dramatically reducing barcode diversity [10]. Herein, we refer to the 4 environments containing both fluconazole and radicicol as *"double-drug environments"* and all others as *"single-drug environments."* In previous work, adaptive yeast mutants, isolated at intermediate evolutionary time points to ensure maximal barcode diversity, were combined from all 12 evolution experiments into a single pooled population. Then they were competed against an ancestral strain in each of the 12 aforementioned experimental conditions. Fitness was inferred from log-linear changes in each mutant's barcode frequency, over time, relative to the ancestor. Barcode frequencies were measured by Illumina sequencing after sequential batch culture transfers. Only barcodes meeting strict sequencing coverage criteria ($\geq$500 counts across time points) were included, resulting in a dataset comprising fitness values for 774 adaptive yeast mutants across all 12 conditions.

Sixty of these 774 adaptive mutant lineages have been previously sequenced [10]. Most differ from the ancestor by only one or two amino acid changes. Thirty five have (different) amino acid changing mutations to either PDR1 and PDR3, known regulators of a drug efflux pump (PDR5). These 774 adaptive mutant lineages have also been previously clustered based on their G×E interaction profiles [10]. These clusters were used to pick a balanced set of mutants with diverse and representative G×E interactions for the analysis described in Figure 3.

For a complete description of yeast strain construction, barcode library generation, and detailed fitness assay protocols, please refer to Schmidlin et al. 2024 [10].

4.2 Dimension reduction of the fitness matrix via singular value decomposition (SVD)

We applied singular value decomposition (SVD) to reduce the dimensionality of a fitness matrix consisting of 774 mutants measured across 12 environments. In matrix notation,

$$F = U\Sigma V^T \tag{1}$$

where F is the $N \times E$ fitness matrix ($N = 774$, $E = 12$), U is an $N \times r$ matrix of left singular vectors (mutant loadings), V is an $E \times r$ matrix of right singular vectors (environment loadings), and Σ is an $r \times r$ diagonal matrix of singular values.

In this framework, the columns of U describe how individual mutants load onto each latent dimension, while the columns of V describe how environments load onto those same dimensions. The diagonal entries of Σ quantify the amount of variation explained by each dimension. Together, these matrices define abstract phenotypic dimensions that summarize systematic patterns of genotype–environment interactions.

Depending on the analysis, different combinations of mutants and environments were used to train the SVD model. These differences are described in the main text, the figure legends, and below. All analyses were conducted using custom Python scripts. The full code, including a Jupyter notebook demonstrating the SVD workflow, has been made publicly available at `https://github.com/Russellrezagh/SVD-SSD-genotype-phenotype-mapping`.

4.3 Bi-cross-validation of SVD models

In most figures, a bi-cross-validation approach was used to assess the predictive performance of the SVD-based model [7, 54]. This approach involves choosing a subset of mutants and environments as the training data, building a low dimensional model, and then using that model to predict the fitness of the held out mutants and environments, which are referred to as the test data.

In bi-cross-validation, the fitness data matrix (matrix F) is divided into four submatrices: (submatrix A) test mutants in test environments, (submatrix B) test mutants in training environments, (submatrix C) training mutants in test environments, and (submatrix D) training mutants in training environments. SVD is applied to the training submatrix (submatrix D), producing singular values and singular vectors that represent latent dimensions in the training data. Predictions for the held-out submatrix (submatrix A) are then generated using the relationship

$$\hat{A} = B(D^{(k)})^{+}C \tag{2}$$

where $D^{(k)}$ is the rank-k approximation of D, obtained by retaining the top k singular values and corresponding vectors, and $(D^{(k)})^{+}$ is its Moore–Penrose pseudo-inverse. This pseudo-inverse step is equivalent to solving a least-squares regression problem, in which the test mutants (matrix B) and test environments (matrix C) are projected into the training-dimensional space identified by SVD.

4.4 Predicting fitness in double-drug conditions from single-drug data (Figure 2)

Training data: In figure 2, we implemented bi-cross validation (described above) while using eight single-drug environments in the training data. These environments included three concentrations of fluconazole (Low, Medium, and High) and three concentrations of radicicol (Low, Medium, and High), together with two control environments (No Drug and DMSO). We also included 100 mutants in the training data, selected randomly from the full set of 774. The selection of training mutants was resampled 100 times, and all subsequent procedures were repeated each time, to ensure that results did not depend on the particular mutants chosen.

Test data: In each iteration, the remaining 674 mutants were held out as the test set, as were the four multidrug environmental conditions (LRLF, LRHF, HRLF, HRHF).

Choosing the optimal number of phenotypic dimensions: In each iteration, we chose the number of latent phenotypic dimensions (r) that minimized the mean squared error on the fitness predictions. We chose this approach because prediction accuracy initially improves as dimensions are added because biologically meaningful patterns are captured. However, the prediction accuracy eventually begins to decline with k when additional phenotypic dimensions incorporate noise rather than signal. Therefore, the optimal number of dimensions was identified by minimizing mean squared error (MSE), as in previous work [7]. The optimal number of dimensions varied across our 100 iterations, as did the fitness predictions for each held out mutant in each held out environment. We averaged the fitness predictions pertaining to each mutant across all iterations where it was held out. Comparing these averages to the measured fitness values provides the robust estimates of predictive performance that we report in Figures 2D–E.

4.5 Evaluating the effect of environmental complexity on predictive power (Figure 3)

To test whether genotype–fitness relationships are organized by a small set of phenotypic modules, we asked how quickly prediction accuracy improves, and then plateaus, as (i) we add more training environments and (ii) more latent dimensions are used in the model. Rapid saturation along either axis is expected if a limited set of phenotypic processes consistently shapes fitness across conditions.

We performed the analysis on two independent datasets including the one that is the focus of the rest of the study (774 drug-resistant mutants × 12 drug environments) [10] and our previous dataset which surveys fitness across a larger number of environments (292 low-glucose-evolved mutants × 45 diverse environments) [7].

Since we wanted to focus on the effects of altering the number of training environments, we held the number of training mutants constant. We also balanced the training mutants to contain an even number from different mutants classes. Using balanced mutants reduces the chance that missing a mutant class masquerades as a need for more environments. From the 774 drug-resistant mutants, we chose 100 mutants that evenly represent previously defined mutant clusters with characteristic patterns of G×E [10]. From the 292 low-glucose-evolved mutants, we chose the same balanced set as was used previously, which includes 60 mutants balanced across genotypes [7]. These training mutant sets were fixed (unlike in Figure 2, where training mutants were repeatedly re-sampled at random).

Here, the training environments were repeatedly re-sampled at random. For each dataset with n total environments, we varied the number of training environments (k) from 1 to $(n-1)$. For each k, we drew 100 random subsets of k environments for training; the remaining $n-k$ environments were held out for testing. For each iteration, we performed bi-cross validation (see above) choosing the number of latent phenotypic dimensions (r) that minimized the mean squared error on the fitness predictions. We then recorded the corresponding R^2 between predicted and measured fitness for all test mutants in the test environments. Averaging R^2 across 100 iterations for each level of training environments (for $k = 1$ through $k = n-1$) gives R^2 for each level of training environments, which is displayed in the rightmost plots in Figure 3. Averaging across all iterations for each number of optimal dimensions (for $r = 1$ through $r =$ maximum r) gives R^2 for each number of optimal dimensions, which is displayed in the leftmost plots in figure 3. To capture both effects together, we also plotted the average R^2 as a function of both k and r; these central surface plots show how predictive accuracy rises quickly with additional environments and dimensions before reaching a broad plateau.

4.6 Evaluating the effect of environmental complexity on number of dimensions (Figure S2)

To continue examining the effect of adding training environments on model complexity, we performed a similar analysis to that in Figure 3 in figure S2. Here, instead of evaluating how predictive power (R^2) changes as we add training environments, we evaluate how the number of dimensions required for a reconstruction of the training data changes as we add training environments (Figure S2). For each number of environments ($k = 2$ through $k = n-1$), we determined the minimum number of phenotypic dimensions necessary to capture 95% of the data variability. For robustness, each incremental increase in environment count was iterated 100 times, with a different random subset of environments selected for each iteration. Since this was a reconstruction, no mutant sampling was performed, all 774 or 292 mutants were included, depending on the dataset. The number of dimensions required to reconstruct the training data increased with the number of environments more quickly in the dataset pertaining to the 774 drug-resistant mutants × 12 drug environments [10]. The same result was true when we hold the number of environments constant (at n) and test the effect of increasing the number of

mutants on model complexity. The dataset pertaining to drug resistant mutants requires more dimensions to capture 95% of the variation in fitness data than does the dataset pertaining to mutants evolved under glucose limitation (Figure S2).

4.7 Evaluating each training environment's influence on prediction accuracy (Figure 4)

To assess how certain types of training environments impact predictive performance of SVD, we trained our model with smaller training sets including some that omitted radicicol, some that omitted fluconazole, some that omitted the control conditions, and some other combinations (Figure 4A). To assess how specific training environments impact the predictive performance of our SVD model, we systematically included and excluded each single-drug environment from the training set and calculated the prediction accuracy using the coefficient of determination (R^2). We assessed the improvement (ΔR^2) in prediction accuracy for each double-drug environment that resulted upon adding individual single-drug environments to the 7-environment training set. We do so 100 times, each time using 100 random mutants in our training set. For each mutant, we averaged the resulting fitness predictions. then calculated the ΔR^2 values between each mutant's average prediction and measured fitness. The values on the vertical axis of figure 4B represent these averages.

After identifying that including the low fluconazole environment in the training set notably improved predictions in the LRLF condition, we further analyzed which mutants contributed most significantly to this improvement. Specifically, we a identified two groups of mutants: the 50 mutants whose fitness was underpredicted when low fluconazole was excluded and the 50 mutants whose fitness was overpredicted when low fluconazole was excluded. We examined their gene-by-environment interaction profiles and observed characteristic behaviors that echo those detected in an earlier study using different methods (Figure S3) [10].

To robustly evaluate the effect of low fluconazole, we extended our drop-out analysis beyond single environments. In addition to testing the effect of adding one environment at a time to the baseline training set, we also systematically added all unique pairs and triplets of environments. For each case, the improvement in prediction accuracy (ΔR^2) was recalculated relative to the 5- or 6- environment baseline (Figure S4).

4.8 Using SVD as an interpretation tool rather than for making predictions (Figure 5–6)

We conducted a comprehensive dimensionality analysis using the entire dataset to evaluate the interpretative power of singular value decomposition (SVD) in capturing phenotypic drivers of fitness. Rather than using bi-cross-validation, here we applied SVD directly to the entire fitness dataset, without separating training and test subsets, to identify abstract phenotypic dimensions. We performed SVD on the complete data matrix (774 mutants by 12 environments) and identified the number of phenotypic dimensions needed to explain at least 95% of the observed variation in fitness. Our analysis revealed that just four dimensions were sufficient to capture this threshold of variation. To quantify the overall reconstruction accuracy of our SVD model, we calculated the coefficient of determination (R^2) by comparing the predicted (reconstructed) fitness values against the observed values across the entire dataset. This resulted in an accurate reconstruction (R^2 overall was 0.87) (Figure S7). Reconstruction accuracy is lower than 0.95 due to measurement noise. We also computed R^2 values for each individual environment independently to assess how effective our linear dimensionality reduction method works at capturing patterns of fitness variation within each environment (Figure S7). Having reconstructed a comprehensive genotype-phenotype-map that was modular (only four dimensions) we next applied SSD to test it for sparsity.

4.9 Analysing sparsity with SSD models (Figure 5)

To enhance our ability to interpret the biological meaning of the inferred phenotypic dimensions, and confirm the robustness of the identified dimensions, we sought to resolve the inherent ambiguity in standard SVD. While SVD excels at identifying the optimal low-dimensional subspace that captures the maximal variance in fitness, its solution is rotationally invariant. This means that any orthogonal rotation of its basis vectors (the phenotypic axes) within that subspace is equally valid from a reconstruction standpoint, leaving the specific axes returned by the algorithm biologically arbitrary. To arrive at a more biologically meaningful set of axes, one could introduce a guiding prior that selects for a particular kind of structure.

Here, we adopt the principle of sparsity as a biologically motivated prior, operationalized through the Sparse Structure Discovery (SSD) framework [16]. This approach assumes that the genotype-phenotype-fitness map is not a dense, all-to-all network. Instead, it posits that fitness in any given environment is shaped by only a small subset of core phenotypic axes (environment-sparsity), and conversely, that any given mutation perturbs only a few of these axes (mutant-sparsity). This framework translates the search for a sparse decomposition into a concrete optimization problem. We seek to find matrices for environmental weights (W) and mutant weights (M) that minimize the objective function:

$$C(W, M) = ||F - WM||_F^2 + \lambda_W ||W||_1 + \lambda_M ||M||_1 \tag{3}$$

where the first term is the squared Frobenius norm representing the reconstruction error, and the subsequent terms are L_1-norm penalties that enforce sparsity on W and M, respectively. The regularization parameters, λ_W and λ_M, allow us to tune the trade-off between reconstruction accuracy and the desired level of sparsity.

Crucially, the appropriateness of imposing such a sparsity prior is not an axiom but an empirically testable hypothesis. We validated this assumption for our dataset using the rotation test [16]. This procedure tests whether sparsity is an intrinsic, basis-dependent property of the data. By applying a random orthogonal rotation to the fitness matrix (e.g., F' = FO), we preserve its underlying correlational geometry and SVD reconstruction error while shattering any privileged sparse basis. If the original data matrix F can be approximated by a substantially sparser model than its rotated counterpart F' for a given error budget, it provides strong evidence for an inherent sparse structure.

Our analysis revealed a striking asymmetry in the structure of our system. The rotation test for the environment-phenotype map demonstrated a significant sparsity-error gap, confirming that fitness in each environment is indeed governed by a small number of phenotypic axes (Figure 5C). In contrast, the test for the mutant-phenotype map showed no clear gap; the original and rotated data were statistically less distinguishable, indicating a less sparse mapping (Figure 5A). This finding suggests that the drug-resistant mutations in our dataset have pervasive pleiotropy, each influencing nearly all of the underlying phenotypic axes. Informed by these results, we leveraged SSD to derive a model that respects this observed asymmetrical structure, yielding a more constrained but not necessarily more interpretable representation of the genotype-phenotype-fitness architecture (Figure S7).

4.10 Visualizing loading matrices to interpret the meaning of each phenotypic axis (Figure 6)

To facilitate biological interpretation, we computed the loading matrices, which represent the contribution (weights) of each mutant and environment to these four abstract phenotypic dimensions. Specifically, mutant loadings were calculated by multiplying the mutant singular vectors (M) by the square root of the diagonal singular-value matrix (Σ), and environmental loadings were calculated by multiplying the environmental singular vectors (E) by the square

root of Σ. These loading matrixes allowed us to interpret each phenotypic dimension regarding underlying biological mechanisms related to drug resistance (Figure 6).

Supplementary Material

.1 Figure S1:

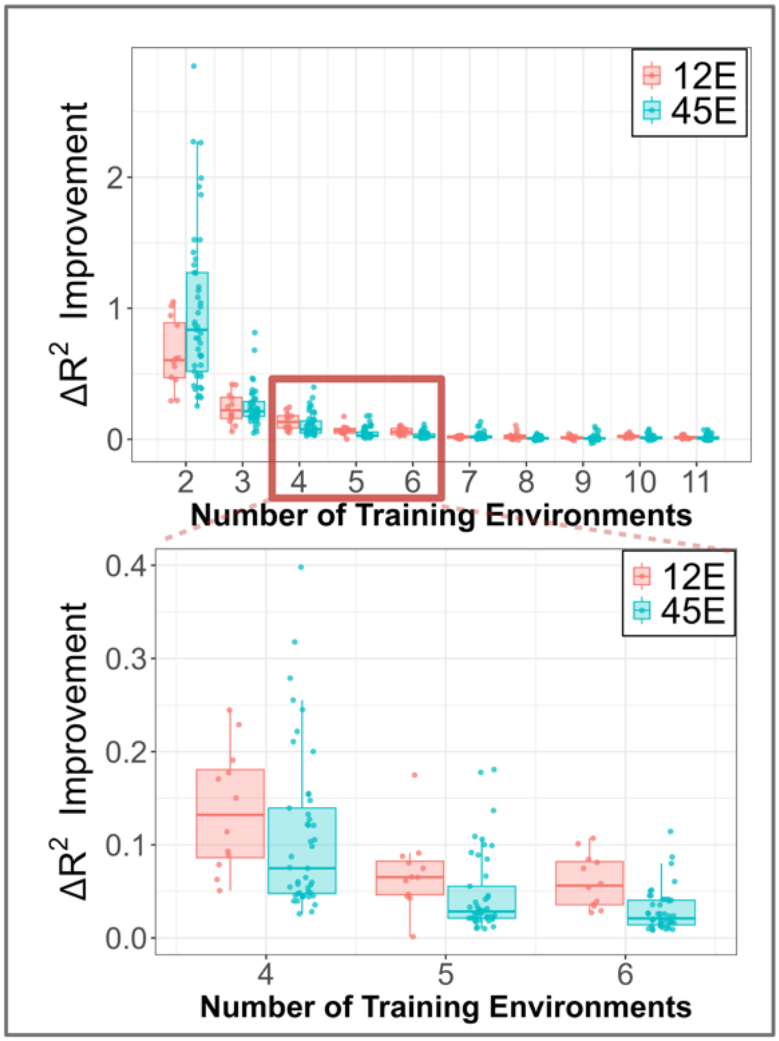

Figure S1: **A dataset with 45 environments (Kinsler et al 2020; blue) shows faster saturation of prediction accuracy than one with only 12 environments (Schmidlin et al. 2025; red).** This may suggest the phenotypes underlying fitness variation in the 45 environment dataset are less complex than those in the 12 environment dataset. The y-axis represents the improvement in prediction accuracy (ΔR^2), calculated as the increase in the average R^2 when the number of training environments is incrementally increased. The boxplots summarize the distribution of ΔR^2 values obtained from 100 randomized iterations across each incremental increase, where the number of environments specified on the horizontal axis are chosen randomly to comprise the training set. **Top:** The improvement in R^2 is on average highest when adding the first few training environments and diminishes rapidly as more environments are added. This decline is faster for the 45 environment dataset. **Bottom:** A zoomed-in view of the 4 to 6 training environment range shows that the improvement for the 45 environment dataset converges toward zero more rapidly than the 12 environment dataset.

.2 Figure S2:

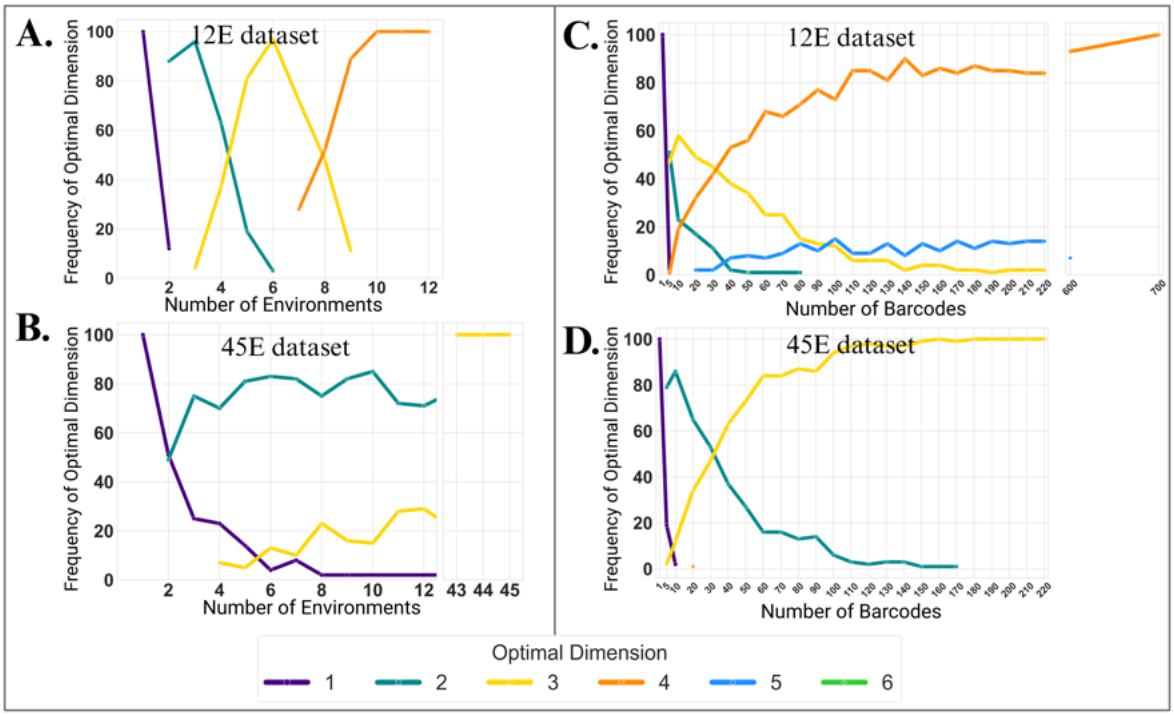

Figure S2: **A dataset with 45 environments (Kinsler et al 2020; bottom) is captured in less latent dimensions than a dataset with only 12 environments (Schmidlin et al. 2024; top).** This figure shows how the dimensionality of the fitness data changes when varying the number of environments (left) or the number of mutants (i.e. "barcode") (right). For this analysis, the "optimal dimension" was defined as the minimum number of dimensions required to explain 95% of the variance in the training data. On the left, the training data includes all mutants, and the number of environments on the horizontal axis. On the right, the training data includes all environments, and the number of barcoded mutants on the horizontal axis. For each value on the horizontal axis, the SVD analysis was performed 100 times by iteratively subsampling environments (left) or mutants (right). The y-axis shows the frequency of each level of optimal dimensions as a function of either the number of environments sampled (left) or the number of mutants sampled (right). In both cases, the 12E dataset requires up to four dimensions to capture variation in fitness, while the 45E dataset, despite spanning more environments, stabilizes at two or three dimensions.

.3 Figure S3:

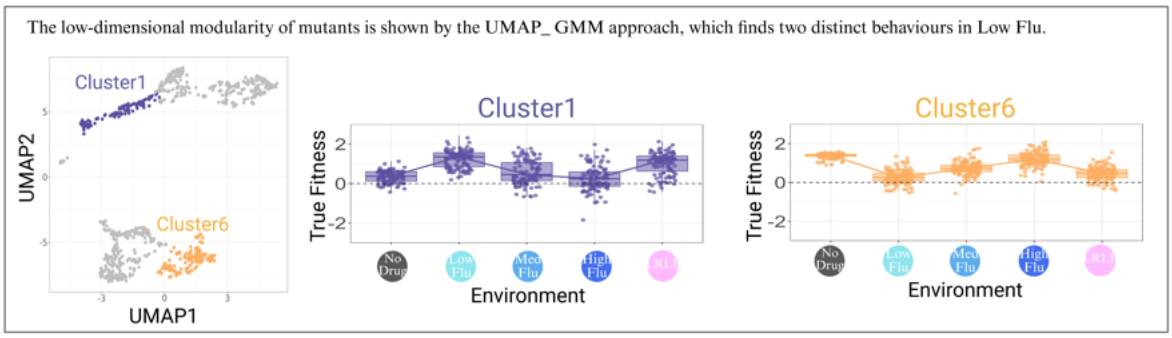

Figure S3: **Mutant groups identified by prediction error in figure 4 correspond to clusters independently discovered in a previous study.** To provide external validation for the novel mutant groups identified in Figure 4, we show that they correspond precisely to previously identified groupings [10]. These previous groupings were identified using an independent, unsupervised method (UMAP dimensionality reduction and Gaussian Mixture Model clustering) to categorize mutants based on their G×E interactions across the same 12 environments studied here. **Left:** The UMAP embedding of the groupings identified in previous work [10], where the 774 drug-resistant mutants are represented as points. Here, two of the previously reported clusters are highlighted, including Cluster 1 (purple) and Cluster 6 (orange). **Middle:** Cluster 1 mutants display a characteristic non-monotonic fitness advantage with a distinct peak in the Low Fluconazole environment. This behavior mirrors that of the "most underpredicted" mutants identified through our prediction-error analysis (Figure 4C). **Right:** The fitness profile of mutants in Cluster 6 show a non-monotonic fitness decline in Low Fluconazole. This pattern is identical to that of the "most overpredicted" mutants identified by our method (Figure 4E).

.4 Figure S4:

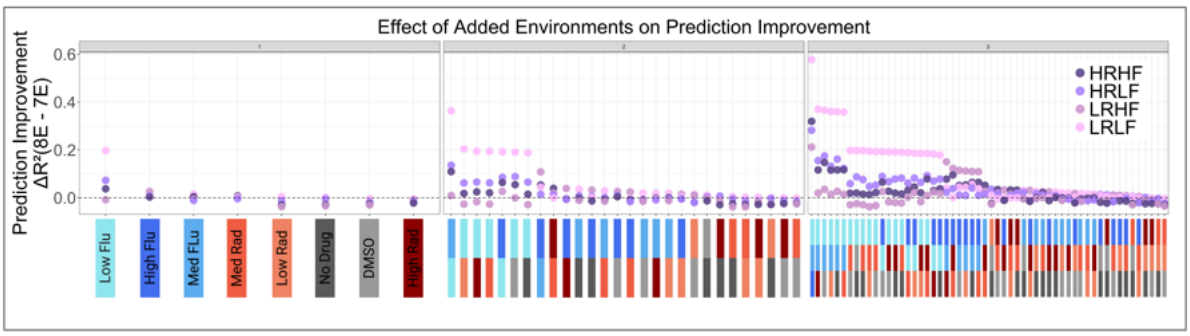

Figure S4: **Low Fluconazole consistently provides the greatest prediction improvement regardless of training set composition.** To robustly test the informational value of individual single-drug environments, this figure expands upon the analysis in Figure 4B. Here, we systematically hold out environments from the training set, and then report the improvement in prediction accuracy upon adding those training environments back. The y-axis shows the improvement in prediction accuracy (ΔR^2). The colored bars at the bottom of each panel indicate the specific environment or combination of environments that was held out and then added back to the training set. Within each panel, the combinations are sorted from left to right by their impact on prediction improvement. **Left:** Shows the effect of adding each of the 8 single-drug environments individually to a 7-environment training set. As shown in Figure 4B, adding Low Fluconazole (light blue bar) provides the largest single improvement, especially for the LRLF Environment (light pink dot). **Middle:** Shows the effect of adding all unique pairs of environments to a 6-environment training set. The combinations yielding the highest prediction improvement (ΔR^2) are consistently those that include Low Fluconazole. **Right:** Shows the effect of adding all unique triplets of single-drug environments to a 5-environment training set. Again, the most informative combinations are overwhelmingly those that contain Low Fluconazole.

.5 Figure S5:

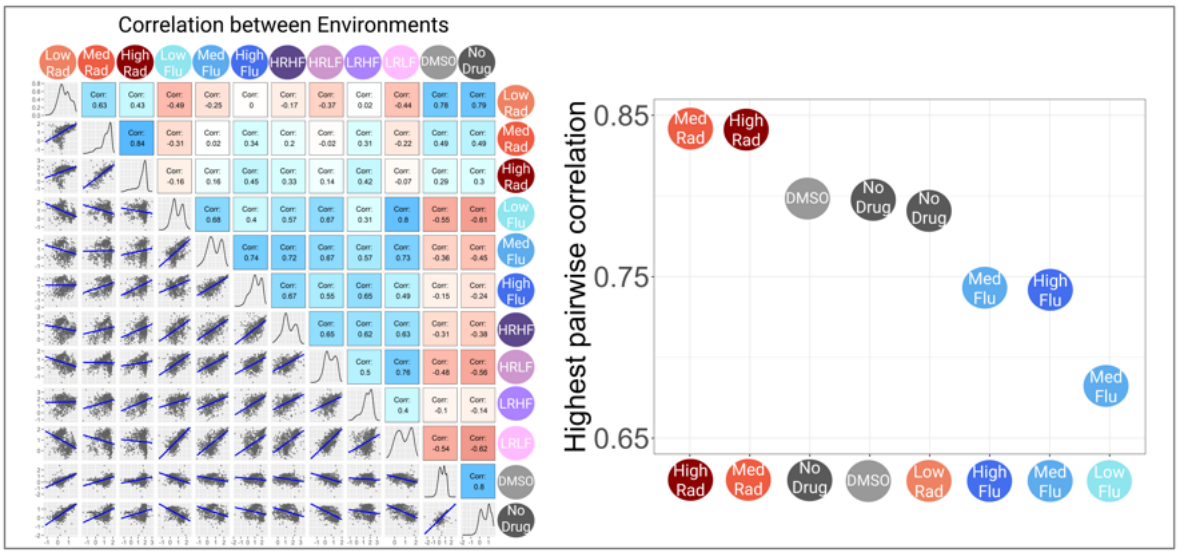

Figure S5: **Low Fluconazole exhibits the least correlation with other single-drug environments.** To further investigate why low-dose fluconazole is uniquely informative for predicting fitness, this figure analyzes the pairwise correlations of fitness between all 12 environments. **Left:** Shows a matrix of pairwise comparisons between all environments. The lower triangle contains scatter plots of fitness in each pair of environments, with a blue line showing the linear regression. The diagonal shows the density distribution of fitness values for each environment. The upper triangle displays the Pearson correlation coefficient for each pair, with the color indicating the strength and sign of the correlation, blue for positive, and red for negative. **Right:** Shows a summary plot displaying the highest pairwise correlation for each single-drug environment. For each environment on the x-axis, the y-value represents its maximum correlation with any other single-drug environment. This metric quantifies how redundant an environment is; a low value indicates high uniqueness. The analysis clearly shows that Low Fluconazole has the lowest maximum pairwise correlation (r = 0.68) compared to all other single-drug conditions. This indicates that the pattern of mutant fitness in Low Fluconazole is the most distinct and least redundant in the dataset.

.6 Figure S6:

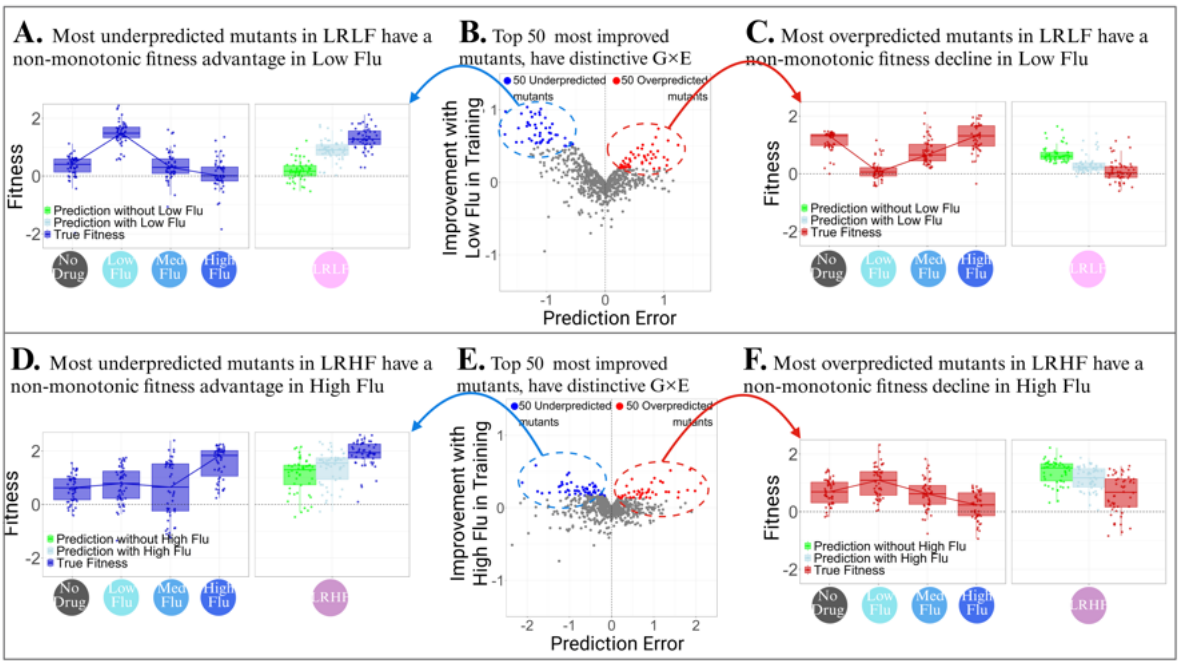

Figure S6: **The mutants with fitness predictions that rely on data from the high flu environment have nonmonotonic changes in fitness across fluconazole concentrations.** This figure extends the prediction-error analysis from figure 4 to investigate the LRHF environment, which is consistently difficult to predict. The top panel recapitulates figure 4 (LRLF) while the bottom panel pertains to the new LRHF analysis. The results suggest that some prediction difficulty in the LRHF environment is driven by nonmonotonic gene-by-environment (G×E) interactions across fluconazole, similar to those identified for the LRLF environment in figure 4. **(A–C)** Reproduction of figure 4. **(D)** Fitness profiles of the 50 most underpredicted mutants in LRHF as identified in panel E. These mutants display a non-monotonic fitness advantage that peaks in the High Fluconazole environment. The model trained without High Flu (green) is blind to this fitness peak, resulting in underprediction in LRHF. Adding High Flu in the training set improves the model's approximation of fitness in LRHF (light Blue). **(E)** The relationship between prediction error when High Flu is excluded from training, and prediction improvement when High Flu is included. The 100 most improved underpredicted (blue) and overpredicted (red) mutants are highlighted. **(F)** Fitness profiles of the 50 most overpredicted mutants in LRHF as identified in panel E. These mutants exhibit a non-monotonic fitness relationship with fluconazole concentration, peaking in low fluconazole, and plummeting in high fluconazole. These mutants overlap strongly with those in panel A. The model misses the fitness decline in higher flu when it is excluded from the training and overpredicts fitness in LRHF (green). When High Flu is added to the training environment, the prediction improves (light Blue).

.7 Figure S7:

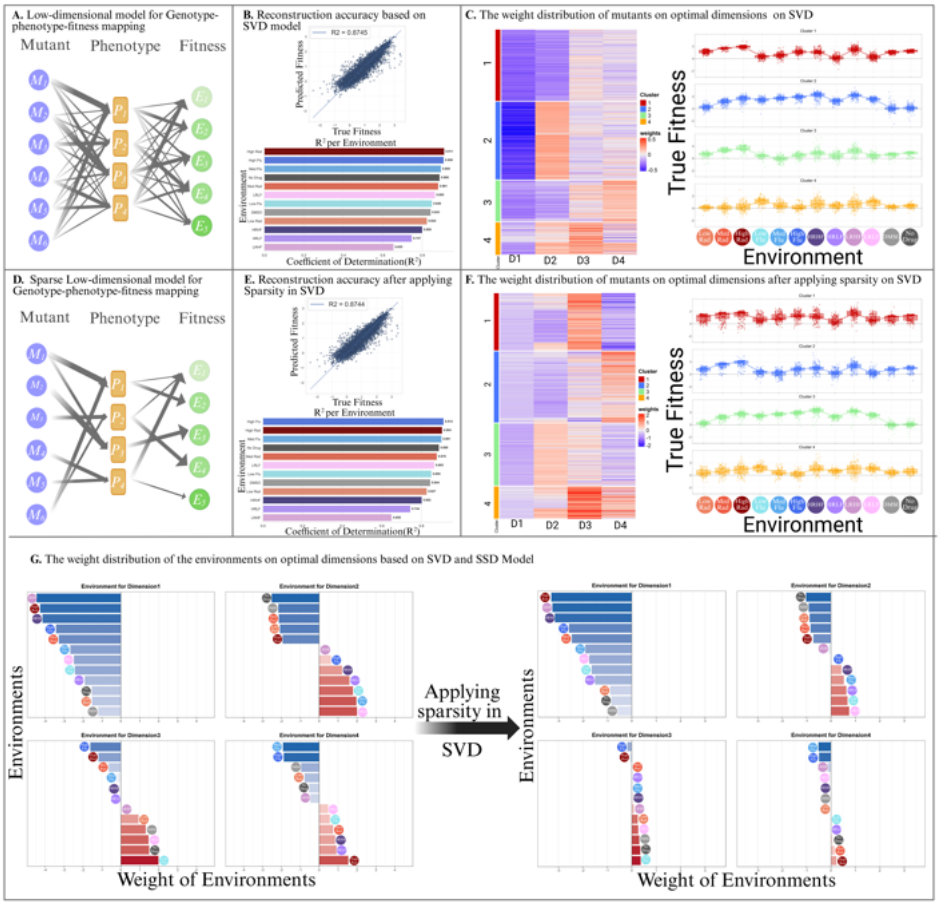

Figure S7: **A sparse decomposition model (SSD) indicates a similar structure of the phenotypic dimensions identified by standard SVD.** To assess the phenotypic dimensions identified in figure 6, this figure compares the standard Singular Value Decomposition (SVD) model with Sparse Structure Discovery (SSD). Both models were applied to the 12E fitness dataset (774 mutants $\times$ 12 environments) to interpret its underlying structure. Panels A to C indicate Standard SVD Model Analysis.**(A)** A conceptual schematic of a low-dimensional genotype-phenotype-fitness map inferred from SVD. **(B)** The model accurately reconstructs the original fitness values with an overall $R^2 = 0.8745$; also, the bottom plot represents the reconstruction accuracy based on the environment. **(C)** A heatmap of mutant weights across the four phenotypic dimensions, with the corresponding fitness profiles for the resulting mutant clusters. Panels D to F show the parallel results obtained using the SSD algorithm. **(D)** A conceptual schematic of a low-dimensional genotype-phenotype-fitness map inferred from SSD. **(E)** The reconstruction accuracy of the SSD model with overall $R^2 = 0.8744$ indicates that applying a sparsity constraint does not reduce the model's ability to explain the variance in the data. **(F)** The corresponding heatmap of mutant weights and cluster fitness profiles from the SSD model. The fundamental structure and biological interpretation of the clusters are similar to the SVD results. **(G)** Comparison of Environmental Weights based on SVD and SSD. This panel directly contrasts the environmental weights for each of the four dimensions as calculated by standard SVD on the left side and the SSD model on the right side. The analysis reveals that the core structure of each dimension, which includes the weights' order and sign of the environments, is similar between the two methods.

References

[1] Yuichi Eguchi, Gaurav Bilolikar, and Kerry Geiler-Samerotte. Why and how to study genetic changes with context-dependent effects. *Current opinion in genetics & development*, 58-59:95–102, 10 2019. ISSN 1879-0380. doi: 10.1016/J.GDE.2019.08.003. URL https://pubmed.ncbi.nlm.nih.gov/31593884/.

[2] Meike T. Wortel, Deepa Agashe, Susan F. Bailey, Claudia Bank, Karen Bisschop, Thomas Blankers, Johannes Cairns, Enrico Sandro Colizzi, Davide Cusseddu, Michael M. Desai, Bram van Dijk, Martijn Egas, Jacintha Ellers, Astrid T. Groot, David G. Heckel, Marcelle L. Johnson, Ken Kraaijeveld, Joachim Krug, Liedewij Laan, Michael Lässig, Peter A. Lind, Jeroen Meijer, Luke M. Noble, Samir Okasha, Paul B. Rainey, Daniel E. Rozen, Shraddha Shitut, Sander J. Tans, Olivier Tenaillon, Henrique Teotónio, J. Arjan G.M. de Visser, Marcel E. Visser, Renske M.A. Vroomans, Gijsbert D.A. Werner, Bregje Wertheim, and Pleuni S. Pennings. Towards evolutionary predictions: Current promises and challenges. *Evolutionary applications*, 16:3–21, 1 2022. ISSN 1752-4571. doi: 10.1111/EVA.13513. URL https://pubmed.ncbi.nlm.nih.gov/36699126/.

[3] Haoyang Zhang, Muhammad Kabir, Saeed Ahmed, and Mauno Vihinen. There will always be variants of uncertain significance. analysis of vuss. *NAR Genomics and Bioinformatics*, 6, 9 2024. ISSN 26319268. doi: 10.1093/NARGAB/LQAE154. URL https://dx.doi.org/10.1093/nargab/lqae154.

[4] Farhan Khodaee, Rohola Zandie, and Elazer R. Edelman. Multimodal learning for mapping genotype-phenotype dynamics. *Nature computational science*, 5:333–344, 4 2025. ISSN 2662-8457. doi: 10.1038/S43588-024-00765-7. URL https://pubmed.ncbi.nlm.nih.gov/39875699/.

[5] Michael Lässig, Ville Mustonen, and Aleksandra M. Walczak. Predicting evolution. *Nature Ecology & Evolution 2017 1:3*, 1:0077–, 2 2017. ISSN 2397-334X. doi: 10.1038/s41559-017-0077. URL https://www.nature.com/articles/s41559-017-0077.

[6] Kerry A. Geiler-Samerotte, Shuang Li, Charalampos Lazaris, Austin Taylor, Naomi Ziv, Chelsea Ramjeawan, Annalise B. Paaby, and Mark L. Siegal. Extent and context dependence of pleiotropy revealed by high-throughput single-cell phenotyping. *PLoS biology*, 18, 8 2020. ISSN 1545-7885. doi: 10.1371/JOURNAL.PBIO.3000836. URL https://pubmed.ncbi.nlm.nih.gov/32804946/.

[7] Grant Kinsler, Kerry Geiler-Samerotte, and Dmitri Petrov. Fitness variation across subtle environmental perturbations reveals local modularity and global pleiotropy of adaptation. *eLife*, 9:1–52, 12 2020. ISSN 2050084X. doi: 10.7554/ELIFE.61271.

[8] Annalise B. Paaby and Matthew V. Rockman. The many faces of pleiotropy. *Trends in genetics : TIG*, 29:66–73, 2 2013. ISSN 0168-9525. doi: 10.1016/J.TIG.2012.10.010. URL https://pubmed.ncbi.nlm.nih.gov/23140989/.

[9] David Houle, Diddahally R. Govindaraju, and Stig Omholt. Phenomics: the next challenge. *Nature reviews. Genetics*, 11:855–866, 12 2010. ISSN 1471-0064. doi: 10.1038/NRG2897. URL https://pubmed.ncbi.nlm.nih.gov/21085204/.

[10] Kara Schmidlin, Sam Apodaca, Daphne Newell, Alexander Sastokas, Grant Kinsler, and Kerry Geiler-Samerotte. Distinguishing mutants that resist drugs via different mechanisms by examining fitness tradeoffs. *eLife*, 13, 9 2024. doi: 10.7554/ELIFE.94144.

[11] Shi An A. Chen, Alexander F. Kern, Roy Moh Lik Ang, Yihua Xie, and Hunter B. Fraser. Gene-by-environment interactions are pervasive among natural genetic variants. *Cell genomics*, 3, 4 2023. ISSN 2666-979X. doi: 10.1016/J.XGEN.2023.100273. URL https://pubmed.ncbi.nlm.nih.gov/37082145/.

[12] Susanna Manrubia, José A. Cuesta, Jacobo Aguirre, Sebastian E. Ahnert, Lee Altenberg, Alejandro V. Cano, Pablo Catalán, Ramon Diaz-Uriarte, Santiago F. Elena, Juan Antonio García-Martín, Paulien Hogeweg, Bhavin S. Khatri, Joachim Krug, Ard A. Louis, Nora S. Martin, Joshua L. Payne, Matthew J. Tarnowski, and Marcel Weiß. From genotypes to organisms: State-of-the-art and perspectives of a cornerstone in evolutionary dynamics. *Physics of life reviews*, 38:55–106, 9 2021. ISSN 1873-1457. doi: 10.1016/J.PLREV.2021.03.004. URL https://pubmed.ncbi.nlm.nih.gov/34088608/.

[13] Jakub Otwinowski, David M. McCandlish, and Joshua B. Plotkin. Inferring the shape of global epistasis. *Proceedings of the National Academy of Sciences of the United States of America*, 115:E7550–E7558, 8 2018. ISSN 10916490. doi: 10.1073/PNAS.1804015115/SUPPL_FILE/PNAS.1804015115.SAPP.PDF. URL https://www.pnas.org/doi/abs/10.1073/pnas.1804015115.

[14] Olivia M. Ghosh, Grant Kinsler, Benjamin H. Good, and Dmitri A. Petrov. Low-dimensional genotype-fitness mapping across divergent environments suggests a limiting functions model of fitness. *bioRxiv : the preprint server for biology*, 4 2025. ISSN 2692-8205. doi: 10.1101/2025.04.05.647371. URL https://pubmed.ncbi.nlm.nih.gov/40291729/.

[15] Sarah M. Ardell and Sergey Kryazhimskiy. The population genetics of collateral resistance and sensitivity. *eLife*, 10, 12 2021. ISSN 2050084X. doi: 10.7554/ELIFE.73250.

[16] Samantha Petti, Gautam Reddy, and Michael M. Desai. Inferring sparse structure in genotype–phenotype maps. *Genetics*, 225, 8 2023. ISSN 19432631. doi: 10.1093/GENETICS/IYAD127. URL https://dx.doi.org/10.1093/genetics/iyad127.

[17] Günter P. Wagner, Mihaela Pavlicev, and James M. Cheverud. The road to modularity. *Nature reviews. Genetics*, 8:921–931, 12 2007. ISSN 1471-0064. doi: 10.1038/NRG2267. URL https://pubmed.ncbi.nlm.nih.gov/18007649/.

[18] Günter P. Wagner and Lee Altenberg. Perspective: Complex adaptations and the evolution of evolvability. *Evolution; international journal of organic evolution*, 50:967–976, 1996. ISSN 1558-5646. doi: 10.1111/J.1558-5646.1996.TB02339.X. URL https://pubmed.ncbi.nlm.nih.gov/28565291/.

[19] Evan A. Boyle, Yang I. Li, and Jonathan K. Pritchard. An expanded view of complex traits: From polygenic to omnigenic. *Cell*, 169:1177–1186, 6 2017. ISSN 1097-4172. doi: 10.1016/J.CELL.2017.05.038. URL https://pubmed.ncbi.nlm.nih.gov/28622505/.

[20] William G. Hill and Xu Sheng Zhang. On the pleiotropic structure of the genotype–phenotype map and the evolvability of complex organisms. *Genetics*, 190:1131, 3 2012. ISSN 00166731. doi: 10.1534/GENETICS.111.135681. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC3296247/.

[21] J. Draghi and G. P. Wagner. The evolutionary dynamics of evolvability in a gene network model. *Journal of evolutionary biology*, 22:599–611, 3 2009. ISSN 1420-9101. doi: 10.1111/J.1420-9101.2008.01663.X. URL https://pubmed.ncbi.nlm.nih.gov/19170816/.

[22] Massimo Pigliucci. Is evolvability evolvable? *Nature reviews. Genetics*, 9:75–82, 1 2008. ISSN 1471-0064. doi: 10.1038/NRG2278. URL https://pubmed.ncbi.nlm.nih.gov/18059367/.

[23] R. A Fisher. *The Genetical Theory Of Natural Selection*. 1930. URL https://archive.org/details/geneticaltheoryo031631mbp.

[24] J. Arjan G.M. De Visser and Joachim Krug. Empirical fitness landscapes and the predictability of evolution. *Nature reviews. Genetics*, 15:480–490, 2014. ISSN 1471-0064. doi: 10.1038/NRG3744. URL https://pubmed.ncbi.nlm.nih.gov/24913663/.

[25] Daniel Nichol, Peter Jeavons, Alexander G. Fletcher, Robert A. Bonomo, Philip K. Maini, Jerome L. Paul, Robert A. Gatenby, Alexander R.A. Anderson, and Jacob G. Scott. Steering evolution with sequential therapy to prevent the emergence of bacterial antibiotic resistance. *PLOS Computational Biology*, 11:e1004493, 2015. ISSN 1553-7358. doi: 10.1371/JOURNAL.PCBI.1004493. URL https://journals.plos.org/ploscompbiol/article?id=10.1371/journal.pcbi.1004493.

[26] C. Brandon Ogbunugafor. Mutations that enhance evolvability may open doors to faster adaptation. *Nature Communications 2023 14:1*, 14:6310–, 10 2023. ISSN 2041-1723. doi: 10.1038/s41467-023-41914-2. URL https://www.nature.com/articles/s41467-023-41914-2.

[27] Csaba Pál, Balázs Papp, and Viktória Lázár. Collateral sensitivity of antibiotic-resistant microbes. *Trends in microbiology*, 23:401–407, 7 2015. ISSN 1878-4380. doi: 10.1016/J.TIM.2015.02.009. URL https://pubmed.ncbi.nlm.nih.gov/25818802/.

[28] Daniel Nichol, Joseph Rutter, Christopher Bryant, Andrea M. Hujer, Sai Lek, Mark D. Adams, Peter Jeavons, Alexander R.A. Anderson, Robert A. Bonomo, and Jacob G. Scott. Antibiotic collateral sensitivity is contingent on the repeatability of evolution. *Nature communications*, 10, 12 2019. ISSN 2041-1723. doi: 10.1038/S41467-018-08098-6. URL https://pubmed.ncbi.nlm.nih.gov/30659188/.

[29] Kara Schmidlin, Brandon C. Ogbunugafor, Alexander Sastokas, and Kerry Geiler-Samerotte. Environment-by-environment interactions differ across genetic backgrounds. *Molecular Biology and Evolution*, 42:msaf287, 11 2025. ISSN 15371719. doi: 10.1093/MOLBEV/MSAF287. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC12631786/.

[30] Jessica A. Hill, Teresa R. O'Meara, and Leah E. Cowen. Fitness trade-offs associated with the evolution of resistance to antifungal drug combinations. *Cell reports*, 10:809–819, 2 2015. ISSN 2211-1247. doi: 10.1016/J.CELREP.2015.01.009. URL https://pubmed.ncbi.nlm.nih.gov/25660029/.

[31] Alexander Rosenberg, Iuliana V. Ene, Maayan Bibi, Shiri Zakin, Ella Shtifman Segal, Naomi Ziv, Alon M. Dahan, Arnaldo Lopes Colombo, Richard J. Bennett, and Judith Berman. Antifungal tolerance is a subpopulation effect distinct from resistance and is associated with persistent candidemia. *Nature Communications 2018 9:1*, 9:2470–, 6 2018. ISSN 2041-1723. doi: 10.1038/s41467-018-04926-x. URL https://www.nature.com/articles/s41467-018-04926-x.

[32] H. Allen Orr. Adaptation and the cost of complexity. *Evolution; international journal of organic evolution*, 54:13–20, 2000. ISSN 0014-3820. doi: 10.1111/J.0014-3820.2000.TB00002.X. URL https://pubmed.ncbi.nlm.nih.gov/10937178/.

[33] Günter P. Wagner and Jianzhi Zhang. The pleiotropic structure of the genotype–phenotype map: the evolvability of complex organisms. *Nature Reviews Genetics 2011 12:3*, 12:204–213, 2 2011. ISSN 1471-0064. doi: 10.1038/nrg2949. URL https://www.nature.com/articles/nrg2949.

[34] Nadav Kashtan and Uri Alon. Spontaneous evolution of modularity and network motifs. *Proceedings of the National Academy of Sciences of the United States of America*, 102: 13773–13778, 9 2005. ISSN 0027-8424. doi: 10.1073/PNAS.0503610102. URL https://pubmed.ncbi.nlm.nih.gov/16174729/.

[35] Merav Parter, Nadav Kashtan, and Uri Alon. Environmental variability and modularity of bacterial metabolic networks. *BMC Evolutionary Biology*, 7:169, 2007. ISSN 14712148. doi: 10.1186/1471-2148-7-169. URL https://pmc.ncbi.nlm.nih.gov/articles/PMC2151768/.

[36] Grant Kinsler, Yuping Li, Gavin Sherlock, and Dmitri A. Petrov. A high-resolution two-step evolution experiment in yeast reveals a shift from pleiotropic to modular adaptation. *PLOS Biology*, 22:e3002848, 12 2024. ISSN 1545-7885. doi: 10.1371/JOURNAL.PBIO.3002848. URL https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.3002848.

[37] Lilia Perfeito, Stéphane Ghozzi, Johannes Berg, Karin Schnetz, and Michael Lässig. Nonlinear fitness landscape of a molecular pathway. *PLOS Genetics*, 7:e1002160, 7 2011. ISSN 1553-7404. doi: 10.1371/JOURNAL.PGEN.1002160. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1002160.

[38] Yuping Li, Sandeep Venkataram, Atish Agarwala, Barbara Dunn, Dmitri A. Petrov, Gavin Sherlock, and Daniel S. Fisher. Hidden complexity of yeast adaptation under simple evolutionary conditions. *Current biology : CB*, 28:515–525.e6, 2 2018. ISSN 1879-0445. doi: 10.1016/J.CUB.2018.01.009. URL https://pubmed.ncbi.nlm.nih.gov/29429618/.

[39] Sandeep Venkataram, Barbara Dunn, Yuping Li, Atish Agarwala, Jessica Chang, Emily R. Ebel, Kerry Geiler-Samerotte, Lucas Hérissant, Jamie R. Blundell, Sasha F. Levy, Daniel S. Fisher, Gavin Sherlock, and Dmitri A. Petrov. Development of a comprehensive genotype-to-fitness map of adaptation-driving mutations in yeast. *Cell*, 166:1585–1596.e22, 9 2016. ISSN 1097-4172. doi: 10.1016/J.CELL.2016.08.002. URL https://pubmed.ncbi.nlm.nih.gov/27594428/.

[40] Sasha F. Levy, Jamie R. Blundell, Sandeep Venkataram, Dmitri A. Petrov, Daniel S. Fisher, and Gavin Sherlock. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, 519:181–186, 3 2015. ISSN 1476-4687. doi: 10.1038/NATURE14279. URL https://pubmed.ncbi.nlm.nih.gov/25731169/.

[41] Olivier Tenaillon, Alejandra Rodríguez-Verdugo, Rebecca L. Gaut, Pamela McDonald, Albert F. Bennett, Anthony D. Long, and Brandon S. Gaut. The molecular diversity of adaptive convergence. *Science (New York, N.Y.)*, 335:457–461, 1 2012. ISSN 1095-9203. doi: 10.1126/SCIENCE.1212986. URL https://pubmed.ncbi.nlm.nih.gov/22282810/.

[42] David B. Berry and Audrey P. Gasch. Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. *Molecular biology of the cell*, 19:4580–4587, 11 2008. ISSN 1939-4586. doi: 10.1091/MBC.E07-07-0680. URL https://pubmed.ncbi.nlm.nih.gov/18753408/.

[43] Benjamin D. Harrison, Jordan Hashemi, Maayan Bibi, Rebecca Pulver, Danny Bavli, Yaakov Nahmias, Melanie Wellington, Guillermo Sapiro, and Judith Berman. A tetraploid intermediate precedes aneuploid formation in yeasts exposed to fluconazole. *PLOS Biology*, 12:e1001815, 2014. ISSN 1545-7885. doi: 10.1371/JOURNAL.PBIO.1001815. URL https://journals.plos.org/plosbiology/article?id=10.1371/journal.pbio.1001815.

[44] John J. Welch and David Waxman. Modularity and the cost of complexity. *Evolution; international journal of organic evolution*, 57:1723–1734, 8 2003. ISSN 0014-3820. doi: 10.1111/J.0014-3820.2003.TB00581.X. URL https://pubmed.ncbi.nlm.nih.gov/14503615/.

[45] Gerda Saxer, Michael D. Krepps, Eric D. Merkley, Charles Ansong, Brooke L. Deatherage Kaiser, Marie Thérèse Valovska, Nikola Ristic, Ping T. Yeh, Vittal P. Prakash, Owen P. Leiser, Luay Nakhleh, Henry S. Gibbons, Helen W. Kreuzer, and Yousif Shamoo. Mutations in global regulators lead to metabolic selection during adaptation to complex environments. *PLOS Genetics*, 10:e1004872, 12 2014. ISSN 1553-7404. doi: 10.1371/JOURNAL.PGEN.1004872. URL https://journals.plos.org/plosgenetics/article?id=10.1371/journal.pgen.1004872.

[46] Tom M. Conrad, Nathan E. Lewis, and Bernhard O. Palsson. Microbial laboratory evolution in the era of genome-scale science. *Molecular systems biology*, 7, 2011. ISSN 1744-4292. doi: 10.1038/MSB.2011.42. URL https://pubmed.ncbi.nlm.nih.gov/21734648/.

[47] Douglas McCloskey, Sibei Xu, Troy E. Sandberg, Elizabeth Brunk, Ying Hefner, Richard Szubin, Adam M. Feist, and Bernhard O. Palsson. Evolution of gene knockout strains of e. coli reveal regulatory architectures governed by metabolism. *Nature communications*, 9, 12 2018. ISSN 2041-1723. doi: 10.1038/S41467-018-06219-9. URL https://pubmed.ncbi.nlm.nih.gov/30228271/.

[48] Kirsten Bomblies and Catherine L. Peichel. Genetics of adaptation. *Proceedings of the National Academy of Sciences of the United States of America*, 119:e2122152119, 7 2022. ISSN 10916490. doi: 10.1073/PNAS.2122152119/ASSET/D0D6AE5E-0F49-4B0E-8DA2-B701540C0C5A/ASSETS/IMAGES/LARGE/PNAS.2122152119FIG01.JPG. URL https://www.pnas.org/doi/abs/10.1073/pnas.2122152119.

[49] Nadav Kashtan, Elad Noor, and Uri Alon. Varying environments can speed up evolution. *Proceedings of the National Academy of Sciences of the United States of America*, 104: 13711–13716, 8 2007. ISSN 00278424. doi: 10.1073/PNAS.0611630104/SUPPL_FILE/11630APPENDIX.PDF. URL https://www.pnas.org/doi/abs/10.1073/pnas.0611630104.

[50] Xuanyao Liu, Yang I. Li, and Jonathan K. Pritchard. Trans effects on gene expression can drive omnigenic inheritance. *Cell*, 177:1022–1034.e6, 5 2019. ISSN 1097-4172. doi: 10.1016/J.CELL.2019.04.014. URL https://pubmed.ncbi.nlm.nih.gov/31051098/.

[51] Natália Ružičková, Michal Hledík, and Gašper Tkačik. Quantitative omnigenic model discovers interpretable genome-wide associations. *Proceedings of the National Academy of Sciences of the United States of America*, 121:e2402340121, 10 2024. ISSN 10916490. doi: 10.1073/PNAS.2402340121/SUPPL_FILE/PNAS.2402340121.SAPP.PDF. URL https://www.pnas.org/doi/abs/10.1073/pnas.2402340121.

[52] Carlos Martí-Gómez, Juannan Zhou, Wei-Chia Chen, Arlin Stoltzfus, Justin B. Kinney, and David M. McCandlish. Inference and visualization of complex genotype-phenotype maps with gpmap-tools. *bioRxiv*, page 2025.03.09.642267, 8 2025. ISSN 2692-8205. doi: 10.1101/2025.03.09.642267. URL https://www.biorxiv.org/content/10.1101/2025.03.09.642267v3https://www.biorxiv.org/content/10.1101/2025.03.09.642267v3.abstract.

[53] Prachee Avasthi, David G. Mets, and Ryan York. Harnessing genotype-phenotype nonlinearity to accelerate biological prediction. *The Stacks*, 9 2023. doi: 10.57844/ARCADIA-5953-995F. URL https://thestacks.org/publications/result-nonlinear-phenotypes.

[54] Art B. Owen and Patrick O. Perry. Bi-cross-validation of the svd and the nonnegative matrix factorization. *https://doi.org/10.1214/08-AOAS227*, 3:564–594, 6 2009. ISSN 1932-6157. doi: 10.1214/08-AOAS227. URL `https://projecteuclid.org/journals/annals-of-applied-statistics/volume-3/issue-2/Bi-cross-validation-of-the-SVD-and-the-nonnegative-matrix/10.1214/08-AOAS227.fullhttps://projecteuclid.org/journals/annals-of-applied-statistics/volume-3/issue-2/Bi-cross-validation-of-the-SVD-and-the-nonnegative-matrix/10.1214/08-AOAS227.short`.