**Title:**

*Ideas:*

1. Using subtle environmental shifts to probe the number of phenotypes that contribute to fitness

2. Adding subtle environmental shifts to Fisher’s Geometric Model reveals…

3. Subtle environmental perturbations reveal the number of fitness-relevant phenotypes

4. Subtle environmental perturbations reveal the number of fitness-relevant phenotypes in adaptation to glucose limitation

Something more relevant to phe

**Authors:**

**Affiliations:**

**Abstract:**

**Rough Outline of Introduction:**

# **Paragraph One: Adaptation is awesome, but it shouldn’t be.**

# Single punchy (less technical) sentence that captures main idea of paper:

# Organisms have an amazing ability to adapt to diverse challenges, yet recent observations suggest most mutations affect many traits and most of these effects are deleterious. How is adaptation possible if mutants that influence one trait in a beneficial way influence many other traits in a deleterious way?

* 1. We know it is indeed possible: examples of adaptation’s awesome power
     1. Ben Good showing that microbes never run out of adaptive mutations
     2. Which other examples?
  2. But how is adaptation possible? Conclude by defining the ‘cost of pleiotropy’ and explaining that the more pleiotropy there is, the greater the constraint on adaptation.

1. **Section Two: How much pleiotropy is there and why is it so hard to find out?**
   1. Long history of debate on this topic.
      1. Is pleiotropy omnigenic/universal or modular? If modular, by learning relationships among traits one might be able to predict which suites of traits are jointly influenced by genetic changes.
      2. Though modularity seems reasonable and many studies suggest cellular systems have a modular organization, GWAS studies detect ungodly amounts of pleiotropy suggesting every gene has the potential to influence every trait.
   2. This debate persists because of challenges defining and measuring phenotypes, and integrating information about phenotypes at different levels of biological organization
      1. There are too many phenotypes, potentially millions or an infinite number depending on how creative you are
      2. Phenotypes are often related, effects cascading through levels of biological organization, introduce genotype-phenotype-phenotype map idea
   3. Conclude by explaining the ‘impossible task’: Quantifying all of these traits and enumerating relationships between them to understand which suites of traits (or whether every trait) will be jointly influenced by a mutation is too much work.
2. **Section three: Clever solutions leveraging barcoding**
   1. We have a new approach to understanding the extent of pleiotropy among adaptive mutations that circumvents the problem. We do not measure any of these phenotypes, instead we repeatedly measure fitness.
   2. Briefly explain the idea in general terms, subtle perturbations, FGM
   3. Explain the large data set we will use to test the idea, and how this is only possible because (1) these strains are barcoded, (2) NGS gives us lots of power to measure fitness, and (3) we cleverly use information about fitness to learn something about phenotypes.
3. **Section four: What do we find? Punchlines**
   1. The number of ‘traits’ affected by an adaptive mutation depends on context
      1. Adaptive mutations influence a limited number of ‘traits’ in environments close to the one in which they are adaptive.
      2. As we perturb the environment farther from this one, we reveal additional pleiotropic effects of adaptive mutations
      3. This suggests resolution to the paradox, how pleiotropy can be pervasive and adaptation can still happen
   2. Our method is really cool.
      1. Talk about all the cool things you could learn using this epic approach.

**Results**

***Results Section 1: A method to identify fitness-relevant traits***

To identify fitness-relevant phenotypes of mutants

Figure 1: Conceptual figure, explaining method to identify fitness-relevant “traits”, also explain cross-validation/prediction scheme here? [do we want to use FGM figure or SVD figure? – FGM more intuitive, SVD more accurate to actual inference]

***Results Section 2: Our method makes accurate predictions on simulated data***

Figure 2: Simulations: our method works in principle on simulated data with a couple of complications – (1) measurement error sets what we can detect, (2) spread of mutants/conditions also affects what we can detect

***Results Section 3: We can accurately predict fitness with ~5-6 FRPs***

*[transition to talking about the evolution and identifying FRPs in evolutionary context]*

We focus on identifying the fitness-relevant traits of a collection of adaptive mutations that arose in a barcoded evolution experiment [Levy et al]. Despite knowing the genetic basis of adaptation for many of these mutants, with the majority of adaptive mutations being either auto-diploidization or loss-of-function mutations in nutrient-response pathways [Venkataram et al], there remains the question of what phenotypic routes these mutations are taking. Do all the mutations in the nutrient-response pathways represent the same adaptive strategy in this environment? Do particular mutations exhibit specific adaptive strategies?

Li et al. (hidden complexity) began to answer these questions by probing the performance of these adaptive mutations in lag, fermentation, respiration growth stages, all of which were experienced in the evolutionary condition, as well as the extent to which these mutants tradeoff in stationary phase, which was not experienced in the evolutionary condition. The adaptive mutants generally gained an advantage during all three phases experienced in the evolutionary condition, though particular mutants differed in their extent as well as relative importance of the phases in their improvement. This poses a more general question: is performance in these growth phases the only way in which the mutants have gained a benefit? Are there other hidden phenotypes these mutants exhibit that promotes their fitness in this environment or other environments?

*[describe the environmental perturbations we use, pointing out real biological behavior that we have potential to use/learn from our method, referencing Figure 3]*

We measure the fitness of these mutants relative to their ancestor (see Methods and [venkataram et al] for fitness estimation details) in X environmental conditions. These conditions range from subtle perturbations such as measurements of the original evolution condition done in different batches to strong perturbations of this environment by lengthening the transfer to include stationary phase or high salt concentration. Ranking conditions by the average deviation from the variation across batch conditions for a balanced set of recurrent mutations, we see a range of effect from less than 1 standard deviation, representing a very subtle environmental perturbation to extremely strong perturbations up to 15 standard deviations away from the evolution condition. Of the stronger conditions, many exhibit clear fitness differences for some or all of the recurrent mutations (Fig. 3B). Interestingly, some conditions exhibit interesting behavior for only a subset of the recurrent mutation types. For instance, only GPB2 mutants have a clear fitness difference from the evolution condition in the 1 Day Transfer environment. This demonstrates that there are real phenotypic differences amongst these various adaptive mutants, including between GPB2 and PDE2, despite them both being negative regulators in the RAS/PKA pathway having similar fitness effects in the evolution condition, when we consider more distant environmental perturbations.



**Figure 3. Measuring fitness for a collection of adaptive mutants across many environments has the potential to reveal real biological signal. (A)** A collection of ~500 adaptive mutants in the 2 day M3 evolution condition exhibit complex behavior across a range of environmental conditions. Mutations in recurrent genes are highlighted in colors, with the median for each mutation type emphasized by the line. **(B)** For each mutation type, we take the average fitness across the evolution condition batches and the standard deviation of this behavior – conditions in which the median of the mutation type was more extreme than 2 standard deviations are in bolder colors and represent clear biological differences that differ from the mutation type’s behavior across evolution condition batches. Bars on the right represent the mean and 2 standard deviations per mutation type. **(C)** Conditions are ordered based on similarity to the average across all batches. Conditions where the balanced recurrent mutations are less than two standard deviations different from the evolved condition are denoted in black and make up the subtle perturbation set. Conditions where the aggregate behavior exceeds two standard deviations are shown in red and make up the strong perturbations.

Though there aren’t clear fitness differences between the evolution condition and subtle environmental perturbations for recurrent mutations according to this crude measure, there remains the possibility that these subtle perturbations supply enough variation and signal to detect phenotypic differences and make fitness predictions in the conditions with strong, clear fitness differences. To test this, and to understand the phenotypes relevant to the evolution condition, we partition our data into two sets of conditions: (1) the “subtle perturbation set,” consisting of all M3 conditions and those within 2 standard deviations of the evolution condition average according to balanced recurrent mutation behavior, and (2) the “strong perturbation set,” which contains the remaining conditions

*[describe partitioning data into subtle and far set]*

Using environmental perturbations around M3, including measuring this set of mutants in M3 on many different days, as well as a range of other environmental perturbations. We can divide our conditions into 2 distinct sets: the “subtle perturbation set” corresponding of all M3 conditions, as well as those within 2 standard deviations of the average across the M3 conditions, and the “intermediate to far perturbations set”, consisting of the remaining conditions. To learn the number of phenotypes that matter to M3, we can estimate the number of phenotypes that give us predictive power and avoid over fitting in the subtle perturbation set. Furthermore, we divide the mutants into two sets: a training set, picked to have an even representation of all mutation types observed, and a testing set (see Methods for detailed description of the bi-cross validation scheme). To avoid cases where it is substantially difficult to detect …

*[cross validation to identify 6 FRPs from subtle perturbations]*

We divide our mutants into a minimal training and minimal testing set, divided such that there’s as even a distribution of mutation type as possible given our information. To learn the number of phenotypes that give us substantial predictive power and to avoid overfitting, we

Cross-validation amongst the subtle perturbation set conditions identifies six phenotypic components as the number of components that best represent (Fig. 4B)



**Figure 4. Cross-validation of subtle perturbations reveals 6 fitness-relevant phenotypes that can predict fitness of mutant behavior in more distant environments. (A)** A schematic of cross-validation scheme used for identifying the number of detectable fitness-relevant phenotypes. We divide take all of the subtle conditions divide them into pairs of conditions based on their deviation from the mean batch condition. Then, we take all possible combinations of one of each pair, fit a model, and use it to predict behavior in the other, held-out half of the conditions. **(B)** On average, a six-component model does best at predicting the held-out data. Blue dots represent the best model for each of the possible combinations of the subtle conditions. **(C)** We then fit a six-phenotype model to all of the subtle perturbation data and use this to predict the behavior of mutants in more different conditions. **(D)** A six-component model trained on subtle environmental perturbations can predict the fitness of held-out mutants in more distant environmental conditions. Predictions from the six-component model are better than the one-component mode (black dashed line) and the average of 1000 per permutations (solid black, each thin gray).

*[we can predict far conditions from the 6 FRPs trained only on subtle environmental perturbations]*

***Results Section 4: Uncovering real biology from our approach***

*[we can predict the interesting behavior that we pointed out earlier]*



**Figure 5. Our method can predict clear biological signals from subtle environmental perturbations.**Prediction of biological behavior (solid dots from Figure 3) from our approach. Black shows measured fitness, blue the one component model prediction, and red the six component model prediction. Small dots show measured and predicted fitness for each mutant in the mutation type, larger circle represents the median.



**Part of Figure 5/6? Mutants with similar evolution 1D distance can have surprisingly different phenotypic effects when using all 6 phenotypes.** Horizontal axis shows 1 dimensional distance (highly correlated with similarity of M3 fitness). Pairs of mutation types are represented by distance between the geometric medians of the mutation types (denoted by diamond shape). Circles denote the average pairwise distance for the representatives of the mutation type. *Note GPB2 and PDE2 which are similar in 1D distance and similar in distance to others pairs of mutation types in 6D space.*

^ also serves as a consistency check on our model. Mutants from the same gene are in fact close to each other (especially see average pairwise distance for GPB2, PDE2, Diploids) meaning we’re not extremely sensitive to measurement noise. [this needs to go somewhere]

Are IRA1 missense and GPB2 actually similar or is the median IRA1 missense kind of weird because IRA1 missense display a range of phenotypic behavior?

Are the particular pairs of points that appear to be close but actually have substantially different fitness effects across other conditions?? (for below part)

Are we learning anything more than what correlation patterns alone would show us?

*[…but there’s also some behavior that we fail to capture]*

Figure 6: Inability to detect particular behavior represents unpredictable pleiotropic effects not present/relevant to evolution in M3.

**Discussion**

*[possibility of using this approach to identify causal lower-level molecular phenotypes (via RNA-seq data or other phenotypic data)]*

**References formatted for Cell**