**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 pleiotropy?

**Authors:**

Grant Kinsler, Dmitri Petrov, Kerry Geiler-Samerotte

**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***

To demonstrate that our approach to identify fitness-relevant phenotypes works, we first perform a simulation study to understand if, how, and when our method is able to correctly identify fitness-relevant phenotypes and, perhaps more importantly, understand when our method breaks down. To do this, we first simulate data according to a particular number of phenotypes . When all phenotypes contribute equally

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 measured the fitness of mutants across a range of X environmental conditions (see Methods and [venkataram et al] for fitness estimation details). After filtering for inclusion in the particular condition and sufficient coverage to acquire reasonable fitness estimates, we have a collection of 425 mutants for downstream analysis. Of these, Y have been sequenced, and D of them are diploids (the result of autodiploidization, see Venkataram et al). Of the remaining mutants, X of them are clearly adaptive (see methods), with various mutations

Our X environmental 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 adaptive 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.

A total of 425 mutants had enough coverage to be able



**Figure 3. Measuring fitness for a collection of adaptive mutants across many environments has the potential to reveal real biological signal. (A)** 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. **(B)** For each mutation type, we take the average fitness across the evolution condition batches and the standard deviation of this behavior –Shaded regions represent two standard deviations away from the mean amongst the batch conditions per mutation type. On the right are kernel density estimates for the distribution of each mutation type across all conditions (two standard deviations around batches are shaded).

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

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 that are more than two standard deviations away. First, we will construct a phenotype space from the subtle perturbation set, using cross-validation to identify the number of phenotypes to include in the space that gives us maximum explanatory power without overfitting measurement error. Next, using this phenotype space trained only on subtle perturbations, we make predictions about the fitness of mutants in the strong environmental perturbations.

For a given collection of mutants, we can identify a phenotype space using the subtle environmental perturbations.. the key assumption is that the behavior of particular mutants is *informative* for other mutants (the space created by some subset of the data informs the location of new mutants in that same space and helps predict the behavior of those new mutants).

Why is the space of recurrent + others so different than the recurrent one?

Are the spaces similar if you use the same dimensionality?

Are others mostly neutral + noise (and so we’re fitting noise) – then curate these in the training set?

We want to construct a space using the subtle perturbations and a set of training mutants. Then, using this constructed space, we want to predict the behavior of other mutants in other conditions.

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

To identify the phenotypes relevant to fitness in the evolution condition, we construct a phenotype space according to the method described above. In order to accurately identify how many phenotypes to include in our model, striking a balance between explanatory power and avoiding overfitting, we use cross-validation. We divide the mutants into a “train” and “test” set, taking care to For each of these partitions, we fit phenotypic models with every number of phenotypes for the training set of mutant and conditions, independently fit locations for the test mutants and new conditions, and evaluate our ability to capture the fitness values of the test mutants in the test conditions for each number of phenotypes included. Models where the number of phenotypes is too low will show worse performance than those with additional, real phenotypic components. Additionally, models that include an additional phenotype that represents overfitting to measurement noise should have worse performance due to assigning importance to random effects. Thus, we should select models that give the best predictive power for the test set (for details see the bi-cross validation portion of the Methods). This procedure was repeated for every possible partitioning of the pairs of subtle perturbations into the train and test sets, and the average fit to the data was evaluated, as a function of the number of phenotypes included. On average, the three phenotype model was best supported. Additionally, amongst all train and test pairs, six phenotypes was the most common number selected as the best model (Fig. 4B). For these reasons, we select three to be the best supported number of phenotypes to include in our phenotypic model.

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

We construct a three-phenotype space using the entire set of subtle environmental perturbations and the training mutants. Using the phenotypic components of the mutants and weightings of subtle conditions in this space, we explain up to X% of variation in fitness for this set of mutants and conditions, indicating we capture a lot of the subtle behavior. We can use this subtle perturbation space to make predictions of fitness in strong environmental perturbations by independently fitting the locations of the test mutants and the strong perturbation conditions and then evaluating our ability to predict the fitness of these test mutants in the strong environmental perturbations based on this initial space. For all the test mutants, we generally explain more variance in the fitness values than a model with only a single phenotypic component or the average of permutations that remove the correlational signal of mutants and conditions in the training set, indicating that we are learning about phenotype beyond average fitness across the subtle environmental perturbations, and that these subtle environmental perturbations do in fact contain information about the behavior of these mutants that allows us to make concrete predictions of fitness in other contexts. This ability to make predictions, however, is limited, and our predictive power declines as perturbations become stronger, suggesting that the phenotypic effects observed in the local neighborhood of conditions around the evolution condition may be unable to explain behavior in very different environments and that this predictive power may be limited to an intermediate scale.



**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).

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

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

Though it is clear that this phenotypic model can make predictions of aggregate mutant behavior across a variety of conditions an intermediate distance away from the evolution condition, it remains to be shown that we are learning real phenotypic differences and behavior for these mutants. In particular, our six phenotype model accurately estimates fitness for the cases where recurrently hit mutations exhibited clear fitness differences from the evolution condition (Fig. 5). This includes cases where only a subset of the mutations showed such a difference as in the case of GPB2 in 1 Day environment.

Despite mutation identity being hidden from the model as it assigns new mutants to locations in the phenotype space, mutations in similar genes tend to cluster near each other in phenotype space (distance figure, dots especially GPB2, PDE2, Diploids), indicating that phenotype space location gives information about the phenotypic identity of our mutants. However, there are cases (IRA1 missense and nonsense) in which the mutation types appear to be more distant from each other than other like types. This represents phenotypic heterogeneity in mutations in these genes, which is also observed from the fitness effects of these particular mutations in even just the evolution condition alone.

Furthermore, our six phenotype model identifies strong differences between mutation types that are not immediately clear from behavior in a one component model or from fitness alone. In particular, GPB2 and PDE2 mutants have similar fitness in the evolution condition, and from a one component model, the geometric median of the locations of the mutants of these types appear to be located in similar locations in the space. However, in the full six component model from the subtle perturbations, the mutants are as different from each other as other pairs of recurrent mutations, suggesting they do have distinct phenotypic effects. This phenotypic prediction for these mutations types, learned from subtle perturbations alone, is observed in the strong perturbations including differences between the genes in the 1 Day transfer condition, and the high salt concentration conditions.



**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.*

[paragraph: it’s not important that it’s exactly 6 dimensions but rather that it’s not 1 and not infinite, there are a finite number of phenotypes that matter to fitness in this environment]

Do we have enough mutants from IRA2 to look at them?

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?

One example of a phenotype that clearly matters in other conditions (but not in M3): stationary phase performance (as shown by Yuping’s work).

Another important aspect? We use a very curated training set of mutants – does this prevent us from predicting other mutants well?

*[…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)]*

*[worth noting that we don’t have a complete (nor unbiased) collection of adaptive mutants, so additional routes could exist – these just seem to be the most commonly taken – something about mu x s?]*

**References formatted for Cell**