[Title TBD]

Grant Kinsler, Kerry Geiler-Samerotte, Dmitri Petrov - [chris?]

**Introduction**

[conceptual question – what’s a phenotype? How do we measure them? What *matters* to fitness

A central question in biology is identifying the genetic and phenotypic basis of organismal fitness in a given environment. … Identifying the phenotypes important for fitness and adaptation is a crucial step in understanding both “how organisms work” and the process of adaptation itself. However, identifying the direct links between phenotypes and fitness is a difficult problem. [correlations between traits, correlation between trait and fitness, pleiotropy]. Geiler-Samerotte et al. [CITE] demonstrate that this is very hard, and that the relationship between measurable traits is complicated by the choices of when, how, and what to measure. Thus, an approach of measuring phenotypes and associating those measurements to fitness effects is complicated by two primary issues: (1) the choice of which traits to measure and (2) spurious correlations between non-causal traits and fitness. How then can we begin to understand what phenotypic changes are important to fitness and how the relationships between such traits affects evolution (modularity, pleiotropy, etc.).

One possible approach is to instead approach the problem from the other direction – measure changes in fitness across environments to understand which phenotypes are important to fitness in a given environment. There are clear problems with this approach when measuring fitness across very different environments: the environmental contribution to phenotype is likely to have a larger influence and different phenotypes are likely to have very different contributions in these various environments. Instead, consider subtle environmental changes. In such environments, rather than a complete change in the phenotypes important to fitness, there is instead a subtle shift in the relative importance of the phenotypes that influence fitness.

Our approach develops the notion of *independent, fitness-relevant phenotypes*, that are affected by genetic variation and have significant and independent contributions to fitness. These phenotypes are not necessarily measurable traits in the traditional conceptualization of “phenotypes”, rather each fitness-relevant phenotype is likely influenced by complex combinations of measurable (and perhaps currently unmeasurable) features of that organism.

**Previous approaches to identify phenotype space**

*Antigenic Cartography?*

*Fisher’s Geometric Model*

One common framework used to model and understand phenotypic evolution is Fisher’s Geometric model(Fisher 1930) (see Fig 1A). In this model, traits are represented by orthogonal axes in a *d*-dimensional space, with the number of dimensions, *d,* indicating the number of traits that have independent contribution to fitness. Organisms are represented by points in this space, with their position determined by the organism’s particular combination of trait values. The fitness of an organism in a particular environment is determined by a function of the organism’s position and the position of the optimum combination of traits in that particular environment (hereafter “the optimum”). Any such function can be considered, but typically a simple Gaussian function is assumed for mathematical convenience, and has some empirically-supported properties, including diminishing-returns epistasis (de Visser et al. 1999; Chou et al. 2011) (but see (Wiser et al. 2013)). Thus, evolution can be seen as proceeding through “adaptive walks” from an ancestral organism that occupies some location in phenotype space. Genetic mutations can cause phenotypic changes, moving the mutant to a new location in this space. If this new location is favored by the fitness function, then it may increase in frequency and give rise to additional mutations.

Using this general framework, estimating the number of independent traits needed to explain the behavior of adaptive mutants is equivalent for estimating the number of dimensions in Fisher’s model. Previous work has aimed to measure the dimensionality of this space for mutations in general (and more specifically deleterious mutations), rather than the dimensionality of adaptation. These studies primarily fall into three categories. One class of methods derives the dimensionality of the space from the distribution of fitness effects(Martin & Lenormand 2006) (Fig 1B). Assuming the ancestor is at the optimum, a Gaussian mutation distribution in phenotype space, and a multivariate Gaussian distribution for the fitness function, Martin and Lenormand (Martin & Lenormand 2006) derive the expected distribution of fitness effects (DFE) taking the form of a gamma distribution. One can then derive the effective dimensionality of the system utilizing the moments of the DFE. Using directed mutation and mutation-accumulation techniques, these methods find that phenotypic dimensionality is relatively low (less than three dimensions) for the model organisms considered.

A second class of methods uses drift load (Poon & Otto 2000; Tenaillon et al. 2007). These also assume that populations are relatively close to the optimum, but instead explicitly use distance from an optimum as a metric to quantify the phenotypic dimensionality. Intuitively, this is accomplished by observing that large populations are able to effectively purge most deleterious mutations. However, small populations are unable to remove mildly deleterious mutations that drift to fixation, thus carrying a “drift load”. Moreover, the number of such deleterious mutations available to a population can be worked out as a function of the number of phenotypic dimensions, assuming some mutation distribution. Combining these relationships, this class of methods can then observe the fitness of populations across a gradient of effective population sizes to estimate the effective number of phenotypic dimensions. From this set of measurements, researchers found that there were relatively many dimensions (more than five) for the organisms observed.

Attempting to reconcile the different measures of phenotypic dimensionally observed by the two classes of methods, Lourenco(Lourenço et al. 2011) conducted analysis that suggests that DFE approaches are downwardly biased by universal pleiotropy assumptions, and that methods using drift load are relatively insensitive to the pleiotropy of mutations. Lourenco proposes a model of “partial pleiotropy” to explain the observed differences, suggesting that single mutations only affect a small subset of traits, but that mutations altogether affect many traits.

A third class of methods utilizes datasets generated to study epistasis (Fig 1C). Assuming a Gaussian mutation distribution around the ancestor and additive effects of mutations in phenotype space, epistasis is expected to be reflected in the nonlinearity of the phenotype to fitness map. Previous studies have used the distribution of angles between pairs of mutations (Weinreich & Knies 2013) and approximate Bayesian computation(Blanquart & Bataillon 2016) on epistasis datasets to understand phenotypic dimensionality.

These approaches require the use of strict mathematical assumptions to derive tractable mathematical expressions and predictions. Most (if not all) of these assumptions are likely to be violated. It is unlikely that the ancestor is optimally fit, since adaptation is constantly proceeding (Wiser et al. 2013). Moreover, assuming additivity of mutants in phenotype is likely to be wrong. This is particularly clear in the extreme example of pairs of loss of function mutations in the same pathway. Finally, a local Gaussian distribution for mutations in phenotypic space is also unlikely loss of function mutations can drastically change the expression of proteins and thus move a cell’s phenotype very far from the initial set of phenotypes. Additionally, there have been calls for the use of partial pleiotropy in these models to reconcile conflicting conclusions by some of these studies(Lourenço et al. 2011; Bataillon & Bailey 2014).

Recent technological advances allow us to measure the fitness of many mutants with high precision in a high-throughput manner (Levy et al. 2015) and take an approach that removes the necessity for some of these assumptions. In particular, this allows for fitness measurements of a large collection of mutants in a range of conditions (Venkataram et al. 2016; Li et al. 2018). Our approach builds on the phenotypic intuition supplied by Fisher’s geometric model and takes advantage of this new technology by measuring this dimensionality using subtle environmental perturbations (Fig 1D). That is, we slightly change the environment that mutant linages grow in, and use precise measurements of their fitness relative to the ancestor to triangulate both the mutants and the optima that represent the optimal combination of phenotypes in a given environment. We find the space that best fits our data for each number of dimensions, and then evaluate which number of dimensions gives us the best ability to predict withheld data using cross validation.

****

**Figure 1. Fisher’s geometric model and approaches to infer dimensionality. A.** In Fisher’s geometric model, traits have independent effects on fitness, and mutations are represented by vectors in phenotype space, where the new mutant genotype is represented by a new location. Fitness is determined by the distance from the optimum (denoted by gray circle). Thus, everything the same distance away from the origin as the ancestor is neutral (on dashed line), and everything closer is adaptive. **B.** Estimating the number of dimensions using the distribution of fitness effects assumes a fixed mutation distribution and that the ancestor is optimally fit. **C.** Estimating the number of dimensions using epistasis data assumes a mutation distribution and also that the effect of mutations on phenotype is additive, i.e. the vector representing a double mutation is the sum of the vectors of the subsequent single mutations. **D.** Our approach. We use subtle environmental perturbations to triangulate mutants and conditions. Relative fitness measurements give us information about the distance between a mutant and the condition’s optimum, relative to the distance between the ancestor and that optimum.

**Methods**

We consider an explicit model of phenotypic evolution analogous to Fisher’s Geometric Model (Fisher 1930). Phenotypes are depicted as orthogonal axes in a *d*-dimensional space, with the number of dimensions, *d,* representing the number of traits possibly relevant in this space. Organisms are depicted as points in this *d*-dimensional space, with their position determined by the combination of phenotypes represented by that particular organism. An organism’s absolute fitness in a particular environment is determined by a function of its distance from an optimal phenotype.

Our implementation of the model makes some inherent assumptions about the fitness function that determines fitness in each environment. First, it assumes that each trait contributes independently to fitness in any given environment – this can be done by transforming the space if considering a single optimum (Martin & Lenormand 2006), but is not generally true if the interactions between traits differ between conditions. Second, each trait is rescaled such that they have equal effect on fitness in a given condition. We assume this scaling holds for all conditions. Furthermore, we assume that absolute fitness in a given environment is a Gaussian function of the distance between an organism’s location in phenotype space and the location of the optimum for that particular environment:

where h\_k is the height, \sigma\_k^2 is the variance, x\_ij represents the ith coordianate of the jth mutant, o\_ik represents the ith coordinate of the kth optimum. This function is typically assumed for many investigations of Fisher’s model [cite a bunch] due to analytical tractability. Because we are interested in detecting the relevant phenotypic differences that lead to differences in fitness, a relevant fitness function must have the property that traits with no differences between organisms cannot factor into the relative fitness differences between these organisms. An exponential function of squared distance (of which the Gaussian is) is only class of functions with this property (see SI).

Our model makes several assumptions about phenotype space and fitness in phenotype space. The main key assumption is that of subtle environmental perturbations – more specifically, this assumption assumes that mutants have a fixed location in phenotype space (we ignore any environmental contribution to phenotype) and \sigma\_k is constant across environmental condition and dimension (ignoring any specific, correlated effect of phenotypes on fitness in a given environment).

*Simulation Methods*

To test our method for estimating the phenotypic space from fitness data, we perform a simulation study. First, we simulate data that fits our phenotypic model and then feed the corresponding data into our method to infer the phenotype space and number of fitness-relevant phenotypes. For simplicity and consistency with previous literature, we start with a Gaussian function of distance.

We use simulated data to test our method for estimating phenotype space from fitness data. First, we simulate data that fits our phenotypic model and feed this data into our method to infer the phenotype space. To do this, we use two independent simulation techniques. The first simulation technique is a classical Fisher’s Geometric Model simulation –

The second technique simulates an n−dimensional phenotype space by randomly placing the ancestor on the surface of the n−ball of unit distance. Then, we randomly generate “mutants” by sampling uniformly within the n−ball, to simulate only adaptive mutations. Finally, we randomly sample locations for the optima for new conditions in a ball of smaller radius, to simulate subtle environmental perturbations. We now have a set of mutants and conditions, for which we can generate relative fitness data by calculating a fitness function on this phenotype space for each mutant in each condition

To add realism to this simulated data, we also simulate Gaussian measurement noise to ensure our estimation can properly handle realistic levels of noise from fitness measurement studies (Venkataram et al. 2016).

*Estimation Methods*

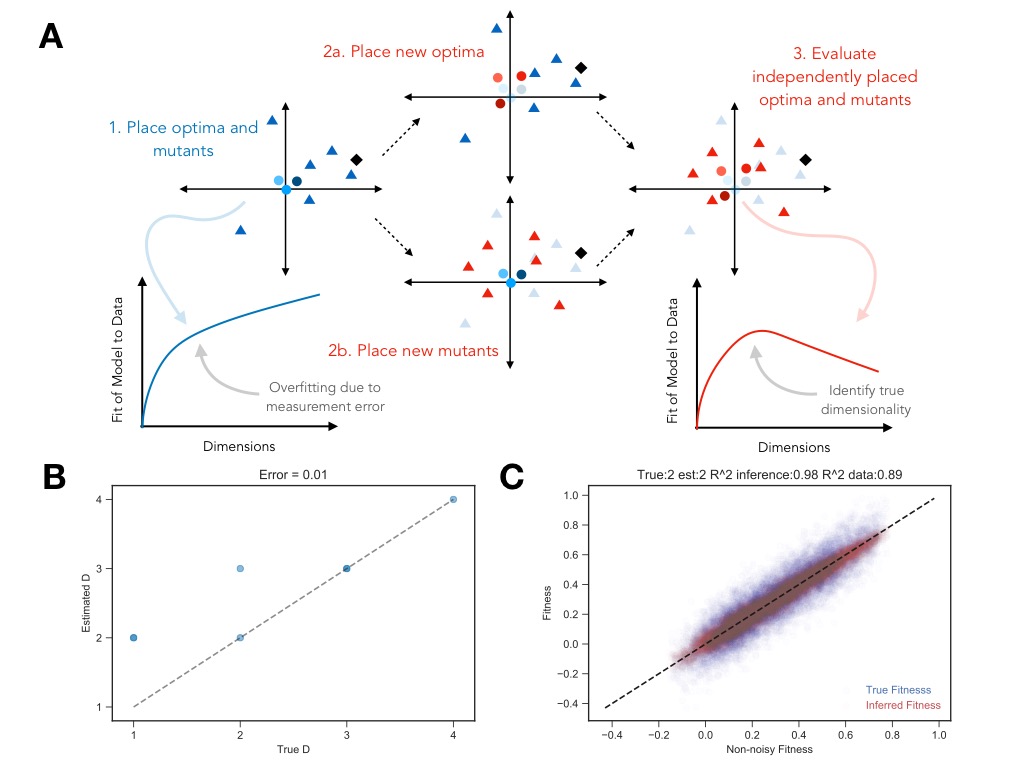
To infer the phenotypic space of our model from a set of mutants, we use an optimization technique on the relative fitness values of the set of mutants across subtle environmental perturbations. For a given number of traits *D*, we aim to find the positions of mutants, optima, and the ancestor that best fits our relative fitness data. We do this by finding the parameters that minimize the function:

where ,, and represent the locations of the mutants, optima, and ancestor, respectively. is the measured relative fitness of mutant in condition with a measure of the measurement uncertainty (see SI).

We use a global optimization technique to find the parameter values minimize this score for each number of dimensions. To estimate the number of dimensions and avoid overfitting, we use a 5-fold bi-cross validation scheme (Fig. 2) analogous to that used for Singular Value Decomposition(Owen & Perry 2009). We divide our data into 5 distinct sets (250 mutants into 5 groups of 50 and 50 conditions into 5 groups of 10). For each “fold” of the data, we exclude a set of mutants and conditions to use as a test set and first use the computational method to estimate the best space for the remaining set of data (See step 1 of Fig 2A.) for each value of D (the number of dimensions) we are interested in. If we were to only do this, we would expect the fit to continually increase as we increase the number of dimensions, because we will begin to fit measurement noise. Next, we fix this space and use only the locations of the original set of mutants as information to find the best location for the conditions in the test set (step 2a of Fig 2A), and separately, use the locations of the original set of conditions to find the best location for the mutants in the test set (step 2b). Finally, we evaluate the relative fitness values predicted from the estimated locations of the mutants and conditions in the test set that were placed independently (step 3). The model that has the best fit of the test set’s relative fitness values to the measured values is the correct number of dimensions that was not over-fit to the original data. This is repeated for each of the 5 test sets, and we pick the model with the best average predictive ability across all 5 folds.

[transition paragraph?]

**Results**



**Figure 2. Cross validation scheme and simulation results. A.** A representation of the cross validation scheme. First, we place a “training” set of optima and mutants in each d-dimensional space. Here, the fit of the model to the data will continuously increase as we add more dimensions. Second, we hold this space constant and place new optima using the information of the mutants in the training set and separately place new mutants using the optima in the training set. Finally, we evaluate the predicted fitness of the “test” set, which consists of the new, independently-placed optima and mutants. The fit of the model with the test set should decrease at some critical number of dimensions, representing where overfitting of the training set occurred, and serving as an estimate for the true underlying dimensionality. **B.** Comparison of the estimated dimensionality and true dimensionality across various simulations with simulated measurement error of (. **C.** Visualization of the fit of the model to true underlying simulated data compared to the “true” fitness with measurement error. Shown in red is the fit of the inferred fitness to the simulated data without noise (). Shown in blue is the fit of the true measured fitness with error to the simulated data without noise ().

*Result 1: In simplest cases, our method works*

*R1a: we capture the correct number of dimensions*

*R1b: we capture a reasonable space*

*R1c: we capture “true” fitness better than the simulated error*

Our computational method accurately estimates the true number of dimensions using this method (Fig 2B), particularly for a low number of dimensions in preliminary simulations. Moreover, it seems like our method, if anything, is prone to overestimating the number of dimensions, indicating that we should be confident that a result is an upper bound on the true underlying dimensionality.

In addition to our models estimating the true dimensionality of simulated data well, it is also able to accurately estimate the relative fitness of test data (Fig. 2C), fitting the underlying simulated data better than the noisy measurements passed into the model. This means that our cross validation scheme properly accounts for measurement error and finds the true underlying fitness values.

*Result 2: Sensitivity/detectability (“dispersion” vs. measurement error drives this).*

The ability to detect the influence depends on the dispersion of mutants and optima in trait space, precision of measurement, and the combination of the two. If mutants and optima are less dispersed (mutants have similar phenotypes and conditions have subtle differences), then more precise measurements are needed (and vice versa). Because our model assumes that each trait contributes equally to fitness, traits that have lower contribution (in un-rescaled space) will also have the effect of lowering the dispersion of mutants for that trait in rescaled space.

*[TBD] but 2a could be selection and 2b could be mutation. Maybe show some figure to explain rescaling in terms of fitness. Probably include something about measurement error in here (and how measurement error x anisotropy effect sensitivity)*

*Somewhere, need a section on the results of non-subtle perturbations?*

*Result 3: New things we can learn from such an approach*

*R3a. Identification of “peculiar” behavior of particular combinations of mutants + conditions (weird uncles)*

Our method for inferring the space inherently relies on how the collection of mutants behave over a collection of conditions. How well can we detect the behavior of particular mutants and conditions? That is, is our method only good for inferring the commonalities but not particular behavior, and if so, how can we know that (and which) particular behaviors are not picked up?

*Result 4: Application to Hillenmeyer data?*

**Discussion**

[awesome discussion of this paper + upcoming amazing paper on 1bigbatch]

**References**

Bataillon, T. & Bailey, S.F., 2014. Effects of new mutations on fitness: Insights from models and data. *Annals of the New York Academy of Sciences*, 1320(1), pp.76–92.

Blanquart, F. & Bataillon, T., 2016. Epistasis and the structure of fitness landscapes: Are experimental fitness landscapes compatible with fisher’s geometric model? *Genetics*, 203(2), pp.847–862.

Chou, H.-H. et al., 2011. Diminishing Returns Epistasis Among Beneficial Mutations Decelerates Adaptation. *Science*, 332(6034), pp.1190–1192. Available at: http://www.ncbi.nlm.nih.gov/pubmed/21636771%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC3244271%0Ahttp://www.sciencemag.org/cgi/doi/10.1126/science.1203799.

Fisher, R.A., 1930. *The Genetical Theory of Natural Selection*, Oxford: Oxford University Press.

Levy, S.F. et al., 2015. Quantitative evolutionary dynamics using high-resolution lineage tracking. *Nature*, advance on, pp.1–78. Available at: http://dx.doi.org/10.1038/nature14279.

Li, Y. et al., 2018. Hidden Complexity of Yeast Adaptation under Simple Evolutionary Conditions. *Current Biology*, 28(4), p.515–525.e6. Available at: https://doi.org/10.1016/j.cub.2018.01.009.

Lourenço, J.M., Galtier, N. & Glémin, S., 2011. Complexity, Pleiotropy, and the Fitness Effect of Mutations. *Evolution*, 65(6), pp.1559–1571.

Martin, G. & Lenormand, T., 2006. A general multivariate extension of Fisher’s geometrical model and the distribution of mutation fitness effects across species. *Evolution; international journal of organic evolution*, 60(5), pp.893–907. Available at: http://onlinelibrary.wiley.com/doi/10.1111/j.0014-3820.2006.tb01169.x/abstract.

Owen, A.B. & Perry, P.O., 2009. Bi-cross-validation of the SVD and the nonnegative matrix factorization. *Annals of Applied Statistics*, 3(2), pp.564–594.

Poon, A. & Otto, S.P., 2000. Compensating for our load of mutations: freezing the meltdown of small populations. *Evolution*, 54(5), pp.1467–1479.

Tenaillon, O. et al., 2007. Quantifying organismal complexity using a population genetic approach. *PLoS ONE*, 2(2).

Venkataram, S. et al., 2016. Development of a Comprehensive Genotype-to-Fitness Map of Adaptation-Driving Mutations in Yeast. *Cell*, 166(6), p.1585–1596.e22.

de Visser, J.A.G.M. et al., 1999. Diminishing returns from mutation rate supply in asexual populations. *Science*, 283(January), pp.404–407.

Weinreich, D.M. & Knies, J.L., 2013. Fisher’s Geometric Model Of Adaptation Meets The Functional Synthesis: Data On Pairwise Epistasis For Fitness Yields Insights Into The Shape And Size Of Phenotype Space. *Evolution*, 67(10), pp.2957–2972.

Wiser, M.J., Ribeck, N. & Lenski, R.E., 2013. Long-Term Dynamics of Adaptation in Asexual Populations. *Science*, 342(6164), pp.1364–1367.