Chapter 1

EFFECTS OF EVOLUTIONARY SELECTION ON NOISE

Motivation

Many biological traits are quantitative: levels of gene expression, sizes of appendages, and abundances of cellular components can vary over a wide range. Mean values of such phenotypes are generally genetically encoded; therefore, they are subject to the forces of selection. Recently, it has become clear that such phenotypes are fundamentally noisy: that is, genes specify a distribution of possible values for the trait, rather than a precise value. More significantly, the variance of this distribution, like the mean, is under genetic control,[5] and the variance, at least in stress response genes, may be genetically independent of mean.[1,3] This data suggests that both the mean and variance of a quantitative phenotype are influenced independently by selective pressures that act on the phenotypes expressed in individual cells. However, an individual in a population is subject to selection only on its particular phenotypic value, not on the mean and variance that specify its phenotypic distribution. Here we ask how positive directional selection affects variance in a simple quantitative trait.

A notable feature of this question is that there are only three possible outcomes. The variance can increase, decrease, or stay the same. Conceptually, one might expect that as you continue to select for a specific phenotype, you select against other phenotypes, decreasing the variance in the population[6] (Fig.1A). Conversely, one can imagine a scenario in which selection on the tail end of a distribution can yield an extremely heterogeneous population. Here the initial population is very unlikely to pass the selection threshold unless there is a large variance (Fig. 1B). Finally, it is possible that as the selection increases the mean, the phenotypic variance remains the same (Fig. 1C). To distinguish between possibilities, we conducted an *in silico* experiment.

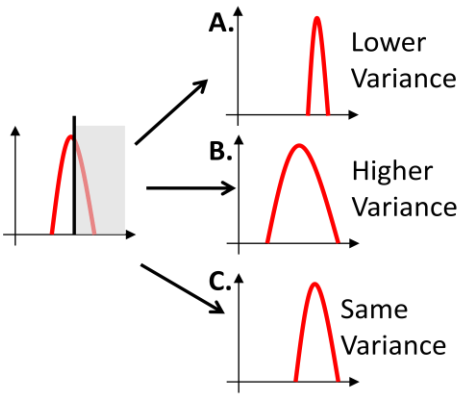

Fig. 1. **How does Selection Affect Variance?** There are three possibilities: Variance can decrease (A), increase (B), or stay the same (C), in response to directional selection.

In Silico **Result**

We considered a phenotype *x*, whose distribution is a Gaussian, parameterized by two genetically determined independent values (μ,σ) (other distributions, including log-normal and gamma, work as well):

$$P(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Here, μ and σ represent the mean phenotypic level, and variance, respectively. Either can vary under mutation. We asked how the population means of these parameters, <μ> and <σ>, change under rounds of mutation, selection, and growth (Fig. 2B). The number of cells, or potentially organisms, was kept large to minimize any effects of genetic drift. The initial population was left homogeneous, i.e. both genetic traits were identical in all cells, for simplicity. During mutation, the σ and μ for a given cell are assigned new values from a distribution centered on their previous values. During selection, we assume that only a certain percentage of cells with the highest phenotypes can survive. For example, under tight selection, the threshold may be chosen to allow only the highest phenotypes, e.g., the top 5%, to survive, while weak selection permits the top 55% to survive. After selection, surviving cells replicate to restore the original population size, completing one cycle of directional selection (Fig. 2B).

The result, shown in Fig. 2C, indicates that tight selection favors increasing variance while weak selection favors decreased variance. This result can be understood qualitatively by considering two individuals with different values of μ and σ, as shown in

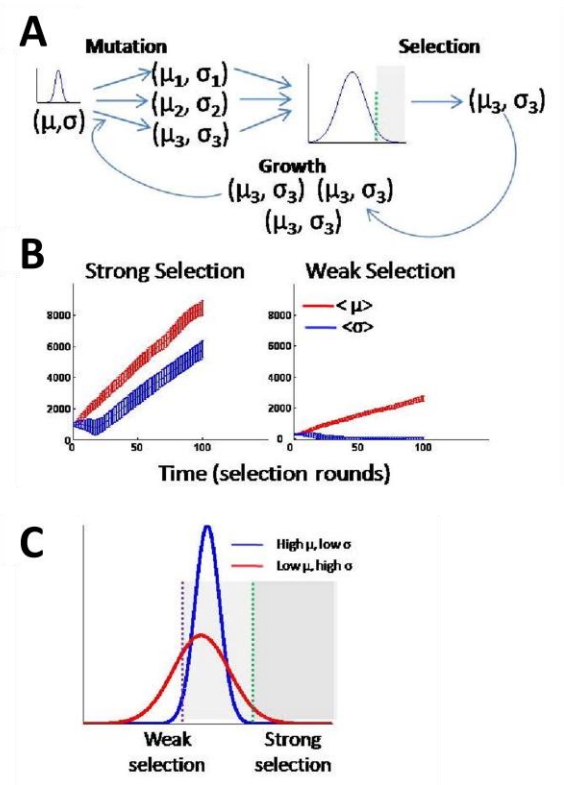

Fig. 2. Phenotypic Selection
A) A flow-chart of the *in silico* experiment. An isogenic population (μ,σ) is mutated so that now each cell has a different genotype (μ_n,σ_n), giving a broader distribution of expression levels than the original population. A threshold selection is imposed (green dashed line). Cells are grown to the original population size. The selected cells are then mutated again and the cycle continues. B) In a simulation, 20,000 isogenic cells with a mean of 1000 and a variance of 300 have been mutated and selected such that the top 5% (left) or 55% (right) of phenotypes are allowed to survive for 100 selection rounds. Strong selection favors increasing <σ> (red lines) while weak selection favors decreasing <σ> (blue lines). <μ> increases in both scenarios. Error bars indicate one standard deviation over the population. C) Two individual distributions, one with a high mean and low variance (blue) and one with a lower mean and higher variance (red) are shown. If strong selection, (cells with the top 5% of all phenotypes are selected, (green dashed line)) is imposed on both distributions, the red distribution will be favored. However, if weak selection (the top 55% of all cells will be selected, shown with the magenta dashed line) is imposed, the red distribution will be favored.

Fig. 2C. The tighter selection (dotted green line) favors the longer "tailed," or high variance individual, despite its lower mean level. Meanwhile, the weaker selection (dotted purple line) favors the individual with the higher mean and lower variance.

This finding has several insights. As expected, the mean phenotype, <μ>, always increases under directional selection, irrespective of the strength of selection, and the rate of increase is proportional to the mutability, the amount that the mean can increase in each round of mutation. More interestingly, <σ> behaves differently depending on the strength of selection. When more than half the cells survive selection, variance increases, but when less than half survive, phenotypic variance decreases to a basal level.

In summary, assuming that a phenotypic distribution is the product of two independent genes, computational analysis predicts that directional selection in which less than half of all cells survive, yields an increase in phenotypic variance.

Analytic Model

Although the intuitive diagram shown in Fig 2c explains the aforementioned *in silico* result, it does not explain the precise effect of selection strength. Why is it that selection pressures in which <50% of cells yield an increase in variance while those in which >50% of cells yield a decrease in variance? What is so special about 50%? We constructed an analytic model hoping to resolve this question.

We assumed the quantitative phenotype (x) controlled by two independent genetic traits, has a phenotypic response is given by (μ,σ). Hence, for a given value of genetic traits the phenotype distribution is given by:

$$(1)\quad P(x)=\frac{1}{\sqrt{2\pi}\sigma}e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

The genotype distribution function $\rho_t(\mu,\sigma)$ is defined as the probability to find the genotype (μ,σ) within the population at generation *t*. (In more formal terms, the probability is given by $\rho_t(\mu,\sigma)d\mu d\sigma$).

The goal here is to obtain an expression for $\rho_t(\mu,\sigma)$ after many generations of mutation, selection, and growth. More specifically, we are interested in what happens to $\langle\sigma\rangle$ and $\langle\mu\rangle$, defined as the mean and variance of the phenotype, respectively, after many generations under the influence of different types of selection.

In order to do that, we need to write the dependence of $\rho_{t+1}(\mu,\sigma)$ on $\rho_t(\mu,\sigma)$. The process which occurs in each generation is as follows:

(2) $\rho_t(\mu,\sigma) \to \rho_t^m(\mu,\sigma) \to \rho_{t+1}(\mu,\sigma)$,

where $\rho_t^m(\mu,\sigma)$ is the distribution after the mutation phase. The first arrow in (2) corresponds to the mutation phase and the second arrow corresponds to the selection phase.

Mutation phase

In the mutation phase, we allowed the values of μ and σ of each genotype to change a little. We can define a mutation function:

(3) $M(\mu,\sigma;\Delta\mu,\Delta\sigma) = \frac{1}{\sqrt{2\pi}\Delta\mu} e^{-\frac{\mu^2}{\Delta\mu^2}} \frac{1}{\sqrt{2\pi}\Delta\sigma} e^{-\frac{\sigma^2}{\Delta\sigma^2}}$,

where $\Delta\mu$ and $\Delta\sigma$ are fixed parameters that correspond to mutation ranges in μ and σ in each generation.

To obtain $\rho_t^m(\mu,\sigma)$, one needs to convolve $\rho_t(\mu,\sigma)$ with the mutation function:

(4) $\rho_t^m(\mu,\sigma) = \int \rho_t(\mu-\mu',\sigma-\sigma')M(\mu',\sigma',\Delta\mu,\Delta\sigma)d\mu' d\sigma'$.

(Please note that one can also write the continuous version of that with a free diffusion equation.) The new genotype distribution is broadened by the convolution with mutation.

Selection Phase

In the selection stage, we essentially look at the total phenotypic distribution and apply a phenotypic selection function S(x) to select for the surviving population. So:

(5) $\rho_{t+1}(\mu,\sigma) = \rho_t^m(\mu,\sigma)\int_x P(x;\mu,\sigma)S(x;\mu,\sigma)dx$,

Let's assume, for example, that we select by introducing a threshold T which is defined such that only a fixed fraction of the population ε_T remains after the selection. Equation (5) then becomes

(6) $\rho_{t+1}(\mu,\sigma) = \rho_t^m(\mu,\sigma)\frac{1}{\varepsilon_T}\int_{x>T} P(x;\mu,\sigma)dx = \rho_t^m(\mu,\sigma)\frac{1}{2\varepsilon_T} erfc\left(\frac{T-\mu}{\sqrt{2}\sigma}\right).$

Note that the term on the right hand side was normalized by ε_T. This corresponds to letting the population grow back to its original size after selection. The threshold T is defined by requirement that $\rho_{t+1}(\mu,\sigma)$ should be normalized, i.e,

(7) $\int \rho_t^m(\mu,\sigma)\frac{1}{2\varepsilon_T} erfc\left(\frac{T-\mu}{\sqrt{2}\sigma}\right)d\mu d\sigma = 1.$

The relation between selection rounds

Substituting Eq. (4) into Eq. (6) provides the required relation between two consecutive generations:

(8) $\rho_{t+1}(\mu,\sigma) = \frac{1}{2\varepsilon_T} erfc\left(\frac{T-\mu}{\sqrt{2}\sigma}\right)\left[\int \rho_t(\mu-\mu',\sigma-\sigma')M(\mu',\sigma',\Delta\mu,\Delta\sigma)d\mu' d\sigma'\right].$

The right hand side is just a mutated distribution (in square brackets) multiplied by a term we'll describe as the selection function. The function erfc(z) is the complementary error function:

(9) $erfc(z) = \frac{2}{\sqrt{\pi}}\int_z^\infty e^{-t^2}dt.$

The selection function becomes

(10) $erfc\left(\frac{T-\mu}{\sqrt{2}\sigma}\right).$

From the analysis of this expression, we have learned several important lessons. First, if the variance, or σ, is small and the mean phenotype, μ, is less than the selection threshold, T, there is no chance that individual will be selected. However, if the mean phenotype stays the same and the variance increases, that individual has some chance of being selected. The

50% selection threshold, at least in a Gaussian distribution is the point where the mean and the threshold are equal. Here the selection expression automatically becomes *erfc(0)*= 1. There is no longer any dependence on the variance in the expression.Next, in the cases when selection is stringent (ε_T <<1) and only a few cells pass selection, it is typical for most of the mean phenotypes to be less than the threshold value. Here it appears that the optimal strategy to pass selection is to have a large variance. In fact, the high variance might even act as an insurance policy, enabling cells to withstand mutations that lower their means and still pass selection. Finally, when selection is lenient, (ε_T >0.5) individuals with higher means and lower variances outcompete those with higher variances. High variance populations also have some very low phenotypic values far below the threshold, making them less fit than their low variance counterparts.

Discussion

This result is general; it applies to any quantitative phenotype under selection. For many selection experiments, the effects of variance generally have not been measured. However, for those experiments, in which distributions are plotted, it seems as if the variance does increase. For example, in 1957 Clayton and Robertson selected for increased and decreased abdominal bristle number in *Drosophila melanogaster*.[7] The resulting phenotypic distributions from their selection appear to be broader than their initial population (Figure 3).

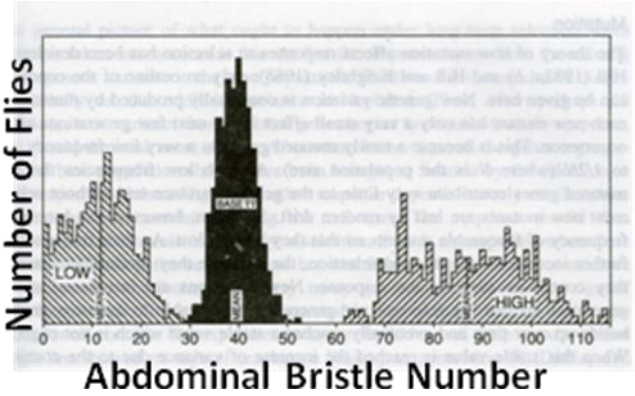

Fig. 3. **A Selection Example.** Clayton and Robertson selected for high and low abdominal bristle number in flies. They showed that flies under directional selection, exhibited higher variance than the base population, consistent with the proposed computational model.

Additionally, Peter and Rosemary Grant have been measuring several quantitative phenotypes in Darwin's finches on the Galapagos Islands for almost 40 years. They have

found that after one drought, during a single monsoon season, the beak depth of the finches increases. After just one generation, they found that the distribution of beak depths broadened.[8]

In a long-term evolution experiment with *E. coli*, Richard Lenski and colleagues found that while only a few mutations evolved at first, 6 of 12 replicate cultures independently mutated a DNA replicase creating hypermutable strains.[9] Due to selection pressures, this hypermutable mutation was only allowed to survive because it had the same fitness, or mean phenotype, as non-hypermutable strains. However, in half of the cell lines evolution selected for this increased ability to mutate, or higher variance in fitness.

Although these long-term evolution experiments are incredibly complicated because of the variety of selection pressures, the fact that we still see variance increasing in response to directional selection reiterates the potentially generality of our result. Furthermore, it shows that in the examination of how selective pressures affect phenotypes, it is important to consider how both the mean and variance of a given phenotype are affected.

Future Directions

As discussed in the introduction, our interest lies in gene expression noise and its potential causes and consequences. Using gene expression as a quantitative phenotype, we can apply our computational and analytic results towards noise. Specifically, we predict directional selection may provide a mechanism that can explain high levels of biological fluctuation. This seems even more plausible in lieu of recent computational[10] and experimental work.[11] Most notably, Kaneko and colleagues used stringent selection (top 0.1%) of a specific GFP to increase expression noise.[12]

Furthermore, we have designed a forward experiment to verify the strong evolutionary prediction: strong and weak directional selection for high expression level will select for high or low noise, respectively. One can take two clonal populations of cells with overlapping gene expression distributions and artificially impose different types of selection using flow cytometry. After several rounds of selection and re-growth, the library should be enriched for the clone with low or high noise in the weak or strong threshold cases, respectively (Fig 4). The only potential drawback of such an experiment is the

inability to account for other types of selection that could potentially interfere with the resulting data. For example, before and after every flow cytometry-imposed selection, there would be ample time for the populations to grow. During this interim period, when both clonal populations of cells would have to be cultured together, any small growth differences between the populations could have a tremendous impact on the experiment. To account for such errors, one could conduct a similar experiment without the flow cytometry-based selection. Ideally, in this experiment, both clonal populations would be tagged with a specific fluorophore, one with a CFP and one with a cherry. This way, one could use a plate reader or flow cytometer to measure the fraction of cells that are labeled with each color and determine the impact, if any, growth had on the ratios of the two populations. If there was an impact, its measured value could be used to deconvolve its effects from the flow cytometry imposed selection. Combined, both experiments could be used to measure the effect of directional selection on phenotypic distributions and, specifically, on phenotypic variance. We anticipate that such selection experiments can yield insight into the role different selective pressures play in causing phenotypic variance in gene expression and other quantitative phenotypes.

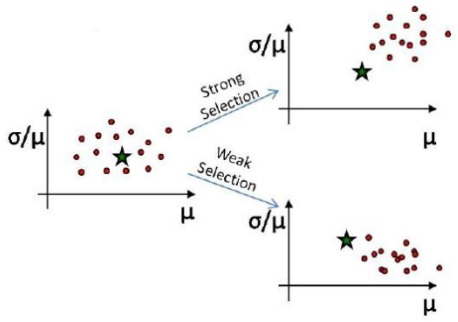

Fig. 4. How selection affects phenotypes.
We propose to impose artificial selection on gene expression via flow cytometry on pooled promoter mutant libraries. If we impose strong selection, we expect both the mean and cv of the pooled population to increase. If we impose weak selection, we expect the mean to increase, but the cv to decrease.

Acknowledgements

The analytic model presented in this Chapter was done with amazing assistance from Dr. David Sprinzak, a postdoctoral fellow in the Elowitz lab.