RESEARCH STRATEGY

1. Background and Significance

Organisms are subject to random changes in their external environments and in their internal components. My long-term goal is to understand how this variation impacts living systems and their evolution.

Many cellular and developmental processes are robust. That is, they operate with high fidelity to produce stereotyped outcomes despite environmental and genetic perturbations [1]. A major thrust of my research has been to identify and characterize mechanisms that increase robustness against different types of perturbation, using high-throughput phenotyping of individual cells of the budding yeast, *Saccharomyces cerevisiae*. As described below in Recent Progress, we have made the most headway on mechanisms that buffer the effects of microenvironmental variation (stochastic fluctuations in the internal or external cellular environment) [2-4], and we have challenged conventional wisdom on mutational buffering [3,5].

Buffering mechanisms represent one end of a spectrum of adaptations to cope with uncertainty. At the other end are mechanisms that generate heterogeneity, allowing a population to maximize its long-term fitness by hedging its bets [6]. As described below in Recent Progress, a second major thrust of my research has been to study a form of adaptive heterogeneity that my lab discovered in *S. cerevisiae*, in which clonal populations contain fast-growing and slow-growing cells that are, respectively, sensitive or resistant to acute stress [7].

Both robustness and adaptive heterogeneity might have important consequences for human health. A compelling case has been made that genetic risk factors for increasingly common diseases (including heart disease, type II diabetes and some psychiatric conditions) are best viewed as alleles that used to be buffered but, due to modern changes in diet and lifestyle, now have consequences [8]. Studying how mutations are buffered and how buffering might fail therefore could contribute to combating multifactorial diseases. Adaptive heterogeneity is relevant to populations of both pathogenic microbes and tumor cells. In both cases, subpopulations of slow-growing cells can survive severe drug treatments and regenerate fast-growing cells when the treatments are withdrawn [9-13]. Our characterization of a similar phenomenon in a model eukaryotic microbe thus holds the potential to advance treatment of difficult fungal infections and of cancer.

Key Gaps in Understanding and Potential for Innovation and Advance

In my view, the most glaring gap in understanding robustness is that no specific gene product has been shown to confer robustness against the effects of naturally occurring mutations, above some baseline level of robustness that would exist in the absence of the gene product [3,5,14]. There is a long history of studies showing that various perturbations, including loss or impairment of gene products termed capacitors, reveal previously hidden (cryptic) genetic variation [1,14-23]. However, there is a logical flaw in any attempt to draw conclusions about robustness from such studies [14]. As originally noted by Hermisson & Wagner [14] and as I have discussed in our work [3,5], the flaw stems from the fact that cryptic genetic variation existing in nature has survived natural selection, which will tend to preserve buffered variation preferentially over nonbuffered variation. Any meaningful test of whether a gene product increases robustness to mutations must assay mutations that have not yet been subject to natural selection [14]. As noted below in Recent Progress, there has only been one report of such a test, which we conducted on a putative capacitor, histone variant H2A.Z in *S. cerevisiae* [3]. We found that although H2A.Z interacts extensively with spontaneous mutations, wild-type H2A.Z function does not confer greater robustness against mutational effects than absence of H2A.Z function [3]. I have argued that this finding does not diminish the importance of cryptic genetic variation, but instead suggests that the focus should shift away from attempting to find mechanisms of mutational robustness and toward gaining better understanding of genetic interactions (epistasis) in nature [3,5,24].

The study of epistasis is ripe for innovation and advance [25,26]. Epistasis has been studied at very large scale using laboratory-derived loss-of-function mutations. For example, ~5 million pairs of loss-of-function mutations have been analyzed as double mutants in *S. cerevisiae*, revealing many cases of non-additive effects [27]. A few excellent studies have shown that epistasis constrains adaptive evolutionary trajectories [e.g. 28-30]. Nevertheless, epistatic interactions between loci harboring natural polymorphisms, which might have much more subtle effects than loss-of-function mutations or mutations that fix under strong selection, are relatively poorly studied [31]. There is evidence from model organisms that epistasis is widespread, but the number of cases where such interactions have been mapped down to the relevant DNA sequence polymorphisms is small [31]. It has also been argued that progress in understanding the contributions of

epistasis to natural variation has been hindered, particularly in human genetics, because the existence of abundant additive contributions to quantitative-genetic variation has led some to assume epistasis is not important (even though the variation statistically partitioned into the additive component can contain a major contribution from epistatically interacting loci) [25,26]. An additional, important hurdle is that there is very low statistical power to detect such interactions in standard quantitative-trait mapping approaches [26,32,33].

For these reasons I propose, in Overview of Future Research Program below, what I feel is sorely needed: a large-scale effort to map epistatic interactions underlying natural variation in complex traits. To do this, we will leverage the power of our high-throughput phenotyping platforms, as well as an experimental design that dramatically increases the ability to detect interaction effects. We also will study epistasis between spontaneous mutations, and between mutations that were adaptive during experimental evolution, to better understand how selection shapes the extent and nature of epistatic interactions that persist in populations.

There are two major gaps I aim to fill in the study of bet hedging. First, the molecular mechanisms underlying adaptive heterogeneity are poorly understood, particularly in eukaryotes. As described in Recent Progress below, we have built on our discovery of heterogeneity in growth rate and stress resistance in *S. cerevisiae* [7] by identifying a regulatory pathway that lies upstream of this heterogeneity. Work to characterize this pathway and understand the ultimate source of heterogeneity will constitute a major component of our research program. Because we are studying adaptive heterogeneity in a model eukaryote with powerful genetic tools, and because we have an innovative high-throughput experimental platform for simultaneously assaying growth, gene expression and stress resistance, we are in a unique position to advance the field. Second, different ecological pressures are expected to select for different extents and kinds of heterogeneity [6,34,35], but with few exceptions [e.g. 36] we lack information about differences in heterogeneity in natural populations. As described in Recent Progress below, we have shown that natural strains of S. *cerevisiae* differ in their growth-rate distributions [37]. We have built on that work to map quantitative-trait loci where the trait in question is variance in growth rate. As described in Recent Progress below, there is an unexpected, but potentially very interesting, connection between our work on epistasis and our work on bet hedging. Growth-rate variance in low glucose has an unusual genetic architecture with no additive-effect loci but with at least two pairs of epistatically interacting loci that contribute significantly to trait variation. It will therefore serve as an excellent test case for applying our proposed methods for mapping epistatic interactions.

2. Recent Progress

I elaborate here on recent progress most directly addressing the key gaps in understanding and motivating my future research directions. Some unpublished results are included here to indicate the direction our research is taking and to further support my proposed research program.

Testing candidates for buffering of genetic variation

We performed the first direct test of whether a gene product increases robustness against the effects of spontaneous mutations, and found that the histone variant H2A.Z does not [3]. To perform this test, we used *S.* cerevisiae mutation-accumulation strains, which had each experienced minimal natural selection for an estimated 2062 generations [38,39]. For each mutation-accumulation strain, we made a strain that was genetically identical except that *HTZ1*, the gene encoding H2A.Z, was deleted. We then analyzed cell morphology using high-throughput microscopy, and statistically partitioned the variance within and between mutation-accumulation strains with and without H2A.Z. WIthin-strain variance increases in the absence of H2A.Z, implying that H2A.Z buffers the effects of microenvironmental variation. Although H2A.Z does interact epistatically with accumulated mutations, its absence does not increase between-strain variance, implying that it does not increase robustness against the effects of new mutations [3] (Fig. 1).

H2A.Z had been an excellent candidate, because it does buffer the effects of microenvironmental variation [2,3] and because of a longstanding hypothesis in the field that posited a congruence between mechanisms that increase robustness against one form of perturbation and those that increase robustness against another form of perturbation [1,40]. By refuting the congruence hypothesis in this case, we also threw into question whether any gene product increases robustness against the effects of new mutations [3,5].

Because our H2A.Z results ran counter to conventional wisdom, we next aimed to test whether another very strong candidate, the chaperone Hsp90, increases robustness against mutations (Geiler-Samerotte et al. submitted). Hsp90 was the first gene product to be described as a capacitor and its impairment reveals

cryptic genetic variation in a wide range of species, including *S. cerevisiae* [18,19,41-44]. Using a similar experimental design as for H2A.Z, but with Hsp90 inhibition instead of H2A.Z mutation, we found a similar result: Hsp90 buffers microenvironmental variation (corroborating a prior study [45]), but does not increase robustness against new mutations. For Hsp90 we also measured variation among natural strains with or without Hsp90 inhibition and found that Hsp90 impairment does tend to reveal cryptic genetic variation. These results provide the first direct evidence that selection preferentially eliminates nonbuffered interactions.

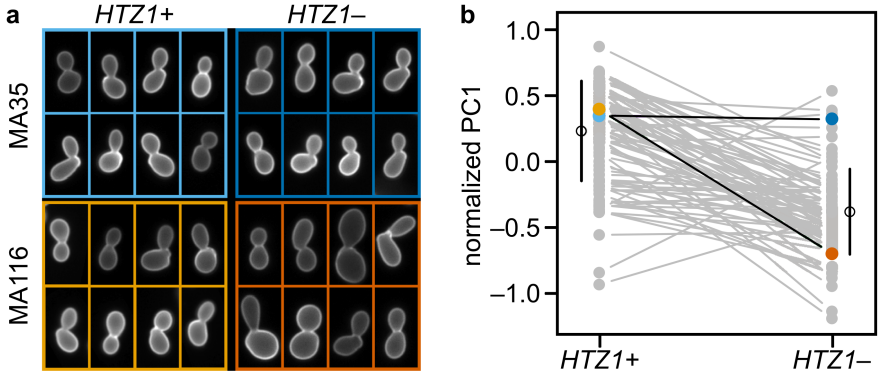

Figure 1. Epistasis between *HTZ1* and new mutations. Mutation-accumulation (MA) strains with or without *HTZ1*, the gene encoding H2A.Z, were analyzed by high-throughput microscopy. (a) Eight representative cells each from two strains (MA35 and MA116) with wild-type (*HTZ1+*) or absent (*HTZ1–*) H2A.Z are shown (concanavalin A-FITC stains the cell surface). H2A.Z absence has little effect in MA35 but a large effect on cell size in MA116. These strains are also highlighted in blue or orange, respectively in (b), which shows strain means for 79 *HTZ1+* MA strains and their *HTZ1–* derivatives for the normalized first principal component (PC1) of 99 morphological phenotypes. Cell-size phenotypes load heavily onto PC1. Black circles and bars show among-strain means and standard deviations for *HTZ1+* and *HTZ1–*.

Bet hedging and natural variation in growth strategies

We discovered that clonal populations of *S. cerevisiae* contain cells growing at vastly different rates, and that a cell's growth rate correlates negatively with its survival of acute heat stress and with its expression of a gene product (TSL1) involved in the production of trehalose, a stress protectant [7] (Fig. 2). Our discovery was enabled by a microscopy method we developed for automated, highly parallel measurement of micro-colonies founded by individual cells. The method tracks as many as 200,000 micro-colonies per experiment, and can simultaneously assay growth, fluorescent-protein expression and survival of stress [4,7,37].

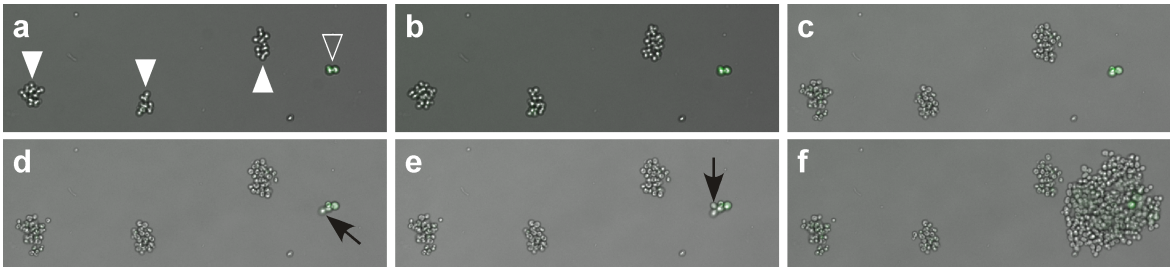

Figure 2. Identification of adaptive heterogeneity in yeast populations. Growth rates of individual micro-colonies were monitored before and after cells were heat shocked. The strain encodes a fusion of GFP with TSL1. Before heat shock (a), three large micro-colonies have low GFP fluorescence (solid arrowheads), whereas one small micro-colony has high GFP fluorescence (open arrowhead). The same colonies are shown 1 hr later (b); the large colonies have clearly kept growing, whereas no cell in the small colony has yet divided. At successive time points (0, 2, 7 and 13 hrs) after heat shock (c–f), the large colonies do not survive, as they do not increase in cell number over 13 hrs. In the small colony, by contrast, one cell divides to give rise to a daughter (arrow) with low fluorescence (d). This daughter then divides quickly (e) and eventually produces many descendents (f).

Heterogeneous expression of TSL1 correlates with heterogeneous expression of several other stress-responsive genes that show evidence of being regulated by the protein kinase A (PKA) pathway [46]. This pathway affects transcription of downstream genes in response to stress by modulating oscillations of the nuclear entry and exit of the paralogous and partially redundant transcription factors MSN2 and MSN4 [47-49]. We therefore hypothesized that MSN2 and MSN4 not only regulate the canonical stress response but also have variable nuclear dwell times under benign conditions and thereby regulate bet hedging against acute stresses. One prediction based on this hypothesis is that slow-growing cells will be lost upon deletion of the gene encoding MSN2 or MSN4. Our results indeed support this prediction for both MSN2 and MSN4 (Li & Siegal, unpublished results) (Fig. 3a). Another prediction is that variable TSL1 expression will be lost upon deletion of the genes encoding MSN2 or MSN4. Our results support this prediction for MSN2 (Li & Siegal, unpublished results) (Fig. 3b).

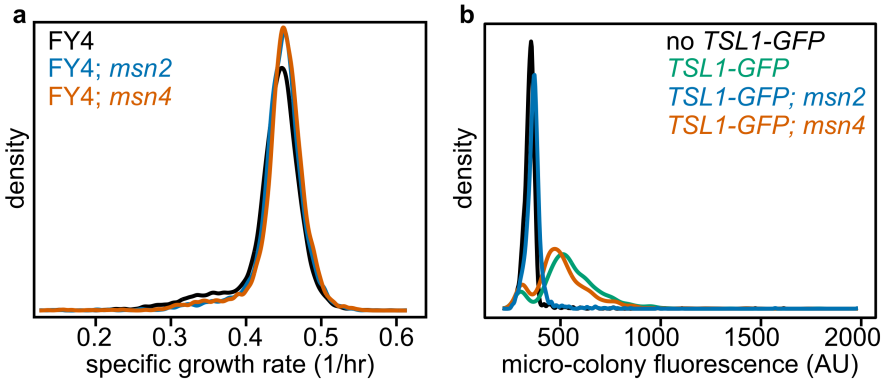

Figure 3. MSN2/4 regulate heterogeneity. (a) Micro-colony growth assays were performed on a wild-type strain (FY4) and its *msn2*- or *msn4*-deleted derivatives. Slow-growing cells are not as abundant in *msn2* and *msn4* mutants. (b) TSL1 abundance was measured with a *TSL1-GFP* fusion-protein construct. Fluorescence was measured at the level of micro-colonies. Due to autofluorescence a no-GFP control was used. *msn2* mutants had a GFP-fluorescence distribution indistinguishable from the no-GFP contol. *msn4* mutants had a distribution similar to that of the wild-type *TSL1-GFP* strain.

Using our micro-colony growth-rate assay, we found that natural strains of *S. cerevisiae* vary in growth-rate mean and variance [37]. We next aimed to map loci that contribute to these differences. Using segregants from a cross of a vineyard isolate and an oak-tree isolate of *S. cerevisiae* [50,51], we mapped quantitative trait loci affecting mean or variance of growth rate (Ziv, Gresham & Siegal, unpublished results). Notably, a major-effect locus for mean growth rate in low-glucose (0.22 mM) medium corresponds to a hexose-transporter-encoding gene that commonly increases in copy number in experimental evolution of yeast strains in low-glucose conditions [52,53]. This correspondence between natural variation and experimental evolution is an important validation of experimental evolution as a strategy for understanding adaptation, although interestingly the natural variation does not appear to be copy-number variation and instead is likely a more subtle alteration of function. Another important finding of our mapping study is that there are no significant additive-effect loci for growth-rate variance in low glucose, yet there are at least two pairs of interacting loci that contribute significantly to differences in growth-rate variance, accounting for ~10% (for the pair involving chromosomes I and X) and ~8% (for the pair involving chromosomes III and VI) of the variance in growth-rate variance among the segregants (Fig. 4). We do not yet know the identities of the interacting genes or sequence variants within those genes, but this unusual genetic architecture presents an opportunity to apply our proposed methods of finding them while advancing our goal of understanding natural variation in bet hedging.

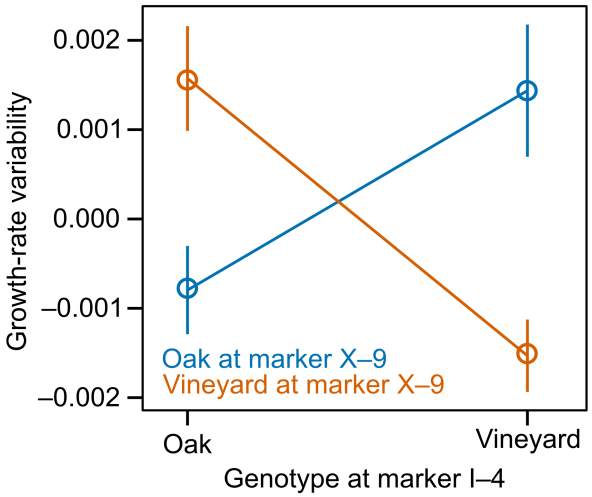

Figure 4. An epistatic interaction underlying variation in growth-rate variance. In 374 segregants from a cross between oak and vineyard isolates, genotyped at 225 markers [50,51], growth-rate variability (a mean-corrected measure of dispersion) at low (0.22 mM) glucose was measured using our micro-colony growth assay. A genome-wide pairwise scan revealed a significant interaction between loci on chromosomes I and X (near markers I–4 and X–9, respectively). Shown are means and standard errors of growth-rate variability for the four genotypes (genotype at marker I–4 on horizontal axis and genotype at marker X–9 indicated by blue or orange line for oak or vineyard, respectively). Neither locus shows a significant additive effect (consistent with symmetry of the two lines).

3. Overview of Future Research Program

Large-scale study of epistasis between naturally occurring mutations

Finding epistasis in nature

We aim to find epistatic interactions that contribute to natural variation in complex traits. The main obstacle is that power to detect epistasis, particularly between loci with no additive effects, is quite low, because the number of possible interactions to test is so high [26,31-33]. Overcoming this obstacle will require advances in experimental design, scale and analysis. We aim to make such advances using our high-throughput methods for measuring single-cell traits in *S. cerevisiae*. These methods allow us to assay many independent traits [2-4,13,37,54,55], and thereby will produce a rich picture of the genetic architectures of complex traits.

Even with such advanced phenotyping methods, however, progress will not be made on mapping interactions without addressing the typically low power of such efforts. Our key innovation in the immediate future will be

to create and use chromosome substitution strains (CSSs) to detect and map interactions [56]. Creating a CSS involves replacing one chromosome with an intact, homologous chromosome from another strain. Mouse CSSs have provided overwhelming evidence that epistatic interactions are abundant, although these interactions have in general not been mapped down to the actual genes involved [57-59]. CSSs have not been used in *S. cerevisiae*, although they are straightforward to produce using heterokaryons to transfer single chromosomes, without recombination, from one parental strain to another [60-63].

CSSs have three main advantages: 1) increasing the power to detect additive-effect loci by enabling crosses in which only a single chromosome segregates at a time; 2) allowing detection of interacting loci by enabling crosses in which all but a single chromosome segregate; and 3) converting the low-power task of testing all possible pairwise interactions into the task of testing single additive effects, which is much more powerful because there are far fewer tests and especially because one can use bulk segregant analysis (BSA), which offers high power and mapping resolution but ordinarily at the expense of obscuring interactions [64].

An additional advantage of CSSs is that the extent of epistasis can be measured without any mapping, by comparing the sums of effects of the CSSs to the difference between the parents [57-59]. We will apply our high-throughput morphometric and growth assays to the full set of CSSs. This will be a rather modest experimental effort, but will yield an expansive view of the contribution of epistasis to a large number of traits.

Ultimately our goal is to map epistatic loci underlying trait variation. For mapping, the main drawback of the CSS approach is that, instead of a single cross of the two parents, as many as 65 crosses might be desired (16 chromosomes X 2 reciprocal CSSs = 32 CSSs, each crossed to each parent, plus one simple parental cross). In this view, the 16 yeast chromosomes are a curse, but they are also a blessing, as epistatic loci are harder to detect when on the same chromosome. Moreover, a much-reduced set of crosses can be used when the simple parental cross yields evidence of epistatic interactions between particular chromosomes, as in our growth-rate variance case. For example, we can make just two crosses, between one of the parents and the respective CSSs for chromosomes I and X, measure growth-rate distributions for ~1000 segregants each, and perform BSA on growth-rate variance by pooling the 15% most extreme segregants for sequencing at 150X genome coverage, which is expected to yield small candidate intervals [65].

We will start with this kind of restricted approach, but it is not unreasonable to contemplate scaling up to 65 crosses over the longer term. Indeed, it is important to anticipate scaling up in this way, because the main technology currently limiting scale, DNA sequencing for genotyping, is clearly on a trajectory of increasing capacity and decreasing cost, especially when reduced-representation methods can be used [66-69]. When protocols for highly multiplexed DNA-sequencing libraries advance further and DNA sequencing costs decrease sufficiently, it will be possible to generate >1000 segregant strains per cross (e.g., to fill three 384-well plates) and genotype each one by sequencing in advance of any phenotyping. The up-front investments in labor and sequencing would yield huge benefits, because then any high-throughput phenotyping method (such as those we have developed and will continue to develop) could be applied to the plates and BSA could be done *in silico* by, in essence, virtual pooling of genotypes. The resulting high-resolution view of additive and epistatic contributions to variation in a large number of complex traits could be transformative.

Epistasis of new mutations and adaptation

To understand how evolution shapes standing genetic variation, it is first necessary to know the spectrum of genetic variation entering a population through spontaneous mutation. For this reason, we teamed up with Dmitri Petrov to sequence the mutation-accumulation lines we have been using and thereby produce the most comprehensive picture to date of the spectrum of spontaneous mutation in any organism [70]. We are currently following up with growth-rate assays to estimate (additive) fitness effects of these mutations.

This collection of ~1000 mutations also provides an opportunity to measure the extent and nature of epistatic interactions between spontaneous mutations. We will analyze these mutations by constructing strains carrying single mutations on the ancestral genetic background, as well as strains carrying pairs of these mutations. Hundreds of combinations can be tested to answer questions such as: What fraction of random pairs of mutations show interactions in their effects on growth rate? Are interactions typically synergistic or antagonistic? Do particular types of mutations show disproportionate tendency to interact?

We will also study the role of epistasis in adaptation using experimental evolution. In collaboration with Dmitri Petrov, who has worked with Sasha Levy and Gavin Sherlock to develop a system for tracking thousands of evolving lineages [71], we will investigate whether and how wild-type or impaired function of the capacitor

Hsp90 leads to different adaptive outcomes. Based on our findings on Hsp90 and mutational buffering, we hypothesize that Hsp90 impairment will alter the identities of adaptive mutations but will not alter the rate of adaptation. We will also investigate more generally the patterns of epistasis for mutations occurring along an adaptive lineage, similarly to how we will study the patterns of epistasis for mutations that have experienced minimal selection, to ask questions about the contingency of the evolutionary process [25,28-30].

Mechanisms of bet hedging and natural variation in growth strategies

Mechanisms of bet hedging

We aim to uncover the full molecular mechanism underlying heterogeneity in growth rate and acute stress resistance in *S. cerevisiae*. Our preliminary genetic analysis has implicated the PKA pathway's control of transcription factors MSN2 and MSN4 in generating the heterogeneity. Our short-term goal is to link MSN2/4 intracellular dynamics with growth rate and stress resistance. To do this, we are expanding our micro-colony growth assay to include much shorter time scales (oscillations of MSN2/4 in and out of the nucleus have period on the order of minutes, whereas micro-colony growth can be accurately measured using one-hour intervals) and to include measurement of subcellular localization of MSN2/4 with an mCherry-tagged MSN2 [49]. We are in the process of developing the imaging protocol (to avoid photobleaching and phototoxicity while maximizing sensitivity and precision) and the image-analysis routines (to capture nuclear fluorescence levels). We hypothesize that cells growing in benign conditions differ in their MSN2/4 nuclear dwell times and that these differences in turn activate stress-responsive genes to varying extents. Identifying the upstream factor that ultimately controls the heterogeneity will require further analysis of mutations in genes encoding proteins that act at different levels in the PKA pathway. We believe that being able to study a more proximal phenotype than growth or stress resistance, namely MSN2/4 localization, will help greatly in this regard.

Natural variation in growth distributions

Evolutionary theory predicts that bet-hedging systems will fine-tune their heterogeneity to match parameters of their selective environment [6,34,35]. We have proposed that transitions between bet-hedging systems, and even between robust sensing of environmental fluctuations and bet hedging against such fluctuations, might be possible through simple changes in regulatory networks [6]. We aim to study such transitions by identifying cases in which the extent of heterogeneity has changed (which we have done [37]) and then by mapping the genetic basis of such changes. Because we know that epistatic interactions underlie variation in growth-rate variance in low glucose, this trait will be the first we analyze using CSSs, with the goal of fine mapping down to the nucleotides that matter. We will then test the effects of those sequence differences on the growth-rate distribution and acute stress resistance under various conditions, and will use molecular-genetic analysis to elucidate how the sequence changes impact heterogeneity (and PKA signaling).

Possible additional directions

Because of the major gaps in understanding that surround them, epistasis and bet hedging are likely to generate interesting questions well into the future. I have a track record of adapting my research program and making methodological advances when called for, and I intend to continually take stock of whether we are addressing these questions most efficiently and rigorously. Choice of study organism is always key. Although I anticipate that *S. cerevisiae* will be advantageous for many years, there are other possibilities as well.

Recently we have begun to explore human genetic data as a way of identifying natural variants that affect trait variance [72] and to investigate related issues in the study of complex human traits [73,74]. Currently we are using family-based GWAS studies to detect such variants while controlling for population structure and mean/variance correlations, which other approaches fail to do adequately (Conley et al. submitted). This approach is underpowered on existing data sets, but as larger family-based studies are conducted, we will continue to test for variance-controlling loci, and should we find any compelling cases, we would likely try to get at mechanism by studying cell lines or orthologous genes in model organisms. (We do not intend to generate human GWAS data. Access to any controlled-access human genomic data will be requested through appropriate channels and the NIH Genomic Data User Code of Conduct will be strictly followed.)

In addition to my NIGMS grants, I currently have a small, NSF-funded project investigating cryptic genetic variation in *Drosophila melanogaster* (focusing on candidate fly genes that I chose based on our yeast work). I will continue to evaluate points of possible connection between the yeast and fly projects, and down the road, if appropriate, I may attempt to fold the fly work into MIRA.