The complex genetic basis of natural variation in cell growth

by

Naomi Ziv

A dissertation submitted in partial fulfillment

Of the requirements for the degree of

Doctor of Philosophy

Department of Biology

New York University

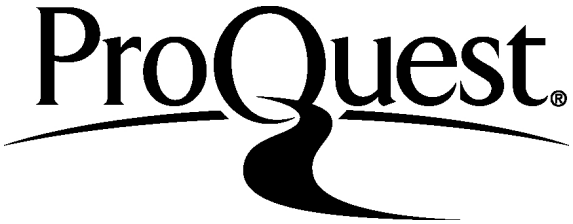
May, 2015

_____

David Gresham, PhD

_____

Mark Siegal, PhD



Why is a raven like a writing desk?

Lewis Carroll

DEDICATION

To Rony and Alma Cohen

ACKNOWLEDGEMENTS

I wish to thank my two advisors, David Gresham and Mark Siegal, for being wonderful mentors on every level. Gloria Coruzzi, Matthew Rockman and Frederick Cross served on my committee; each has become a unique role model for me. I acknowledge all past and present members of both my labs for scientific discussion and friendship, in particular, Sasha Levy, Nathan Brandt, Christina Lyman, Benjamin Neymotin, Christopher Bauer, Kerry Samerotte, Jungeui Hong, Rodoniki Athanasiadou, Shuang Li, Darach Miller, Joyce Kao, Russell Durrett and Yevgeniy Plavskin. I thank Alicia Mastrocco, Christopher Celentano, Bentley Shuster and Bukhtawar Waqas for allowing me to be their mentor. I thank the rest of the biology department at NYU, including faculty, graduate students and post-docs, in particular, Steve Small, Michael Purugganan, Christine Rushlow, Claude Desplan, Daniel Tranchina, Jane Carlton, Nathalie Neriec, Evan Baugh, Andres Mansisidor, Annalise Paaby, Martina Bradic Wei Yuan, Helene Fradin, Melody Foo and Hong Mu. I thank past and present members of the genomics core facility, including Paul Scheid, Tara Rock and Pui-leng Ip. For funding, I wish to thank Dr. Andrew J. Perlman, Curt Strand, Jacob Cohen and the Kazianis family. I thank my mentors from Tel-Aviv University, Marcelo Ehrlich, Nechama Smorodinsky, Galit Horn and Roy Yaniv. The TA I had for genetics was good and the TA I had for animal behavior was the first to teach me the difference between noise and bias. I thank my high school biology teacher Victor; he was cool. I acknowledge all good advice, vibes and voodoo. Most importantly, I thank my loving family, Arnon, Gila and Tamar Ziv and Rony and Alma Cohen.

ABSTRACT

Complex networks of interacting genetic and environmental factors regulate the initiation of cell growth and subsequent proliferation rates. In microbes, growth rate is a major component of fitness and likely a target of selection. Determining the genetic architecture of traits can lead to insights concerning underlying molecular mechanisms and evolutionary dynamics. In this dissertation, I have dissected the genetic basis of cell growth using natural isolates of the budding yeast (*Saccharomyces cerevisiae*). I have placed an emphasis on uncovering the genetic basis of phenotypic variability, the extent of variation between genetically identical individuals within the same environment. In chapter 1, I review the use of quantitative genetics to map the determinants of phenotypic variance and give an overview of the complexity of cell growth regulation. In chapter 2, I quantify differences in cell growth distributions for a set of natural isolates across a range of nutrient concentrations, using a high-throughput microcolony growth assay. In chapter 3, I use different mapping approaches to describe the genetic architecture of both the central tendency and dispersion of cell growth distributions, with particular attention to gene-environment and gene-gene interactions. In chapter 4, I explore the genetic basis of variance in growth rate distributions using an additional pair of isolates. Together, these studies underscore the complexity and context dependency of genotype to phenotype mapping. The implications of these findings are discussed in chapter 5.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1: INTRODUCTION

Signaling pathways connecting nutrient sensing, growth regulation and cell cycle progression are conserved from yeasts to mammals (Fontana et al. 2010; Cross et al. 2011) and are important for normal development and disease. Following decades of research, there is a wealth of knowledge about the molecular components needed for regulated cell growth and division. However, questions concerning the genetic and environmental determinants of variation in growth in natural populations remain largely unexplored. As microbial cell growth entails the integration of nutrient availability and intracellular cues it is ideally suited to dissection using quantitative genetics.

1.1: Regulation of cell growth

Jacques Monod defined the study of cell proliferation as a basic method of Microbiology (Monod 1949), emphasizing the capacity of cell growth and division to integrate all aspects of cell physiology and biochemistry. During the G1 phase of the cell cycle, cells increase in volume and protein content. When a critical threshold is reached, cells irreversibly commit to cell division. In budding yeast, this transition depends on the cell reaching an appropriate size (Johnston et al. 1977). Consequently, variation in growth rate is thought to mainly derive from variation in the length of G1 (Zaman et al. 2008), which in turn is regulated by the rate of cell growth (mass accumulation). Extensive research, particularly using the budding yeast

(*Saccharomyces cerevisiae*) has elucidated the molecular networks governing cell growth in response to nutrient availability.

Glucose is primary source of energy for many eukaryotic cells (Ferretti et al. 2012). In budding yeast, replenishment of glucose causes transcriptional changes for over 40% of genes (Zaman et al. 2008). The effect of glucose is mediated by a number of signaling networks. The Protein Kinase A (PKA) pathway is responsible for a majority of transcriptional changes leading to an increase in biosynthetic capacity (via induction of ribosomes) and repression of the stress response (Broach 2012). PKA is regulated by levels of cyclic AMP, which are in turn regulated by both internal (via Ras1/2 GTPases) and external (via the G protein coupled receptor system Gpr1/Gpa2) concentrations of glucose (Smets et al. 2010). In addition to PKA, the multiprotein TOR complex 1 (TORC1) constitutes a major signaling pathway regulating cell growth (Smets et al. 2010). In response to nutrient quantity and quality, TORC1 regulates protein and mRNA synthesis and degradation, ribosome biogenesis, nutrient transport, and autophagy which collectively determine the accumulation of mass (Loewith and Hall 2011). The mechanisms by which TORC1 senses the nutrient environment remain unknown. The repression of genes in the presence of glucose needed for metabolism of alternative carbon sources is mediated by the Snf1 kinase (Broach 2012). Snf1 is inactivated by the presence (and metabolism) of glucose, leading to de-repression of the transcriptional repressor Mig1 (Smets et al. 2010). Finally, various glucose transporters are regulated to match external glucose levels by the Rgt network (Broach 2012). Sensing of glucose by Snf3 and Rgt2 leads to the phosphorylation and inactivation of the repressor Rgt1 (Zaman et al. 2008). Rgt1 is phosphorylated in both a PKA-dependent and Snf1-

dependent manner leading to inactivation or activation respectively (Zaman et al. 2008). Furthermore, extensive crosstalk between pathways can lead to mutual antagonism between TOR and PKA (Ramachandran and Herman 2011). These interactions underscore the complexity of the sensing, signaling and metabolic networks determining the nutrient response.

Cell growth and division are coordinated processes. Cell cycle progression is regulated by sequential activation of specific cyclins and the activity of cyclin-dependent kinases (Bloom and Cross 2007). The first cyclin to be expressed during G1 is Cln3, leading to initial inactivation of the repressor Whi5 and a positive feedback loop which enforces commitment to cell division (Cross et al. 2011). Cln3 activation and accumulation is dependent on internal and external environmental signals such as nutrient concentration and signaling molecules, modifying the length of G1 and coupling cell division to growth and nutrient availability. Cln3 transcription is activated within minutes by nutrients such as glucose, seemingly by the same transcription factors that regulate ribosomal gene expression (Shi and Tu 2013). This activation is dependent on glucose metabolism (Newcomb et al. 2003) and was recently shown to be initiated by the presence of Acetyl-CoA (Shi and Tu 2013). However, both TOR and PKA regulate Cln3 at the translational level. Systematic analysis of growth in different conditions can shed light on how the complex interplay between nutrient sensing and metabolism and cellular signaling networks generates variation in growth rates.

1.2: Methods for measuring and manipulating cell growth

In microbes, many methods exist for high-throughput growth phenotyping, each with different advantages (Blomberg 2011). Growth rates for hundreds of thousands of microcolonies can be measured using a recently developed assay utilizing time-lapse microscopy (Levy et al. 2012). Microcolony growth rates are estimated from the change in area over time obtained using custom image analysis software. Manipulation of genetic background and growth conditions is straightforward as cells are grown in liquid on the bottom of glass 96 or 384 well plates. A main advantage of this method is the ability to measure growth rate variance in clonal populations. An additional advantage is the ability to grow cells in extremely low nutrient concentrations due to the small number of cells that can be studied simultaneously.

Continuous culturing using chemostats enables extrinsic control of cell growth rate by nutrient limitation. The method of continuous culturing was independently described by Monod (Monod 1950) and Novick & Szilard (Novick and Szilard 1950) over 65 years ago. Cells are grown in a fixed volume of media that is continually diluted by addition of new media and simultaneous removal of old media and cells. At steady state, the growth rate and density of the culture and the concentration of the growth-limiting nutrient remain constant. By varying the dilution rate it is possible to establish large steady-state populations of cells at different growth rates and under different conditions of nutrient limitation. Using chemostats in combination with other analytical methods allows investigation of how the rate of growth impacts fundamental processes in the cell and conversely how the cell regulates and coordinates cellular process with its rate of growth

1.3: The genetics of complex traits

Quantitative genetics is the study of the inheritance of continuously distributed traits, which are defined by small quantitative differences in phenotype between individuals in contrast to discrete phenotypic classes. The amount of variation in such traits across individuals is classically partitioned into genetic and environmental variance components (Falconer and Mackay 1996).

Genetic variance constitutes phenotypic differences attributable to genetic variation and can be further partitioned into additive and epistatic (non-additive) components. In practice, most quantitative genetic studies focus on searching for loci that contribute to additive genetic variance. This means the locus has a non-zero effect on trait values when averaged over variation at all other genomic loci, leading to differences in mean trait value for different genotype classes. However, contribution to additive variance does not necessitate that genes act independently of one another. Epistasis is defined as a difference in the magnitude or sign of a locus effect dependent on the genotype at another locus and is commonly found in model organisms using various experimental approaches (Lehner 2011). Epistatic loci contribute to both additive and epistatic variance components depending on the details of the interaction and allele frequencies (Cheverud and Routman 1995). Focusing on marginal additive effects obscures the distribution of locus effects, challenging phenotypic prediction and the replication of loci between studies. Searching for pairwise or higher order interactions in quantitative genetic studies will lead to a better understanding of the mechanistic basis and evolutionary relevance of epistasis, and provide a more complete picture of the genetic basis of complex traits (Mackay 2013; Taylor and Ehrenreich 2015).

Environmental variance constitutes phenotypic differences not attributable to genetic variation. While environmental variance describes the proportion of phenotypic variance that is not due to genetic differences, the extent of environmental variance ('phenotypic plasticity') can be under genetic control (Scheiner and Lyman 1989; Hill and Mulder 2010). When considering experimentally controlled environmental differences, this control manifests as gene-environment interactions. In this case, loci have different effects (including an absence of effect) depending in which environment the trait was measured. In addition to this aspect of environmental variance, genetically identical individuals can have different phenotypes even in the same environment (Raser and O'Shea 2005; Losick and Desplan 2008; Levy et al. 2012), termed 'phenotypic variability' (Geiler-Samerotte et al. 2013). Direct estimates of variability require measurements of individuals in clonal populations or repeated measurements. This aspect of environmental variance may also be controlled by distinct genomic loci, adding an additional layer of complexity to genotype to phenotyping mapping. These loci are best identified when mean-independent estimates of variability can be measured and treated as quantitative traits. As with identifying gene-gene interactions, identifying the genetic basis of environmental variance can lead to better prediction of phenotypes based on genotype information.

Large-scale genetic mapping of quantitative traits first became feasible following the utilization of molecular polymorphisms as genetic markers (Botstein et al. 1980) and development of analytical methods of analysis (Lander and Botstein 1989). Genetic mapping relies on identifying statistical associations between allele identity and trait values. Studies typically use genetic variation either segregating in families (linkage mapping) or present in natural

populations (association mapping). The number of individuals needed for quantitative trait loci (QTL) detection depends on the effect size and allele frequency of the locus while for QTL localization it depends on recombination frequency (Mackay et al. 2009). Linkage mapping has an advantage in detection due to the absence of rare alleles. Alternatively, association mapping has an advantage in localization as it uses ancestral recombination events (Mackay et al. 2009). Different mapping strategies can improve detection and localization of QTL. For example, pooled analysis of large numbers of individuals with extreme phenotypes improves detection of small effect QTL. Crossing strategies that incorporate several generations improve mapping resolution by introducing additional recombination events into the mapping population. Advancements in experimental and analytical methods of genetic mapping provide an opportunity to rigorously study the genetic basis of variation in complex traits.

1.4: Quantitative analysis of cell growth

The work presented in this dissertation makes use of natural variation to investigate the genetic and non-genetic determinants of cell growth. Specifically, I have used high-throughput individual-based phenotyping to quantify variation in cell growth distributions between natural isolates of *Saccharomyces cerevisiae*. I subsequently used two different crosses and a combination of mapping strategies to explore the genetic basis of cell growth variation and identify causative loci.

I have worked on various additional projects throughout my graduate studies, a number of which have resulted in publications. I have contributed to a review about phenotypic variability; the

extent of variation between genetically identical individuals within the same environment (Geiler-Samerotte et al. 2013) and participated in a scientific video protocol outlining various applications of continuous culturing using chemostats (Ziv, Brandt, et al. 2013). I assisted in the initial description of the microcolony growth rate assay which focused on growth rate variability in benign conditions (Levy et al. 2012) and helped characterize the genetic factors required for initiation of cell quiescence using a reverse genetics approach (Gresham et al. 2011). These studies are not contained within this thesis.

CHAPTER 2: GENETIC AND NON-GENETIC DETERMINANTS OF CELL-GROWTH VARIATION ASSESSED BY HIGH-THROUGHPUT MICROSCOPY

This chapter was published in the journal *Molecular Biology and Evolution* as:

Ziv, N., M. L. Siegal, and D. Gresham. "Genetic and Non-Genetic Determinants of Cell-Growth Variation Assessed by High-Throughput Microscopy." *Molecular Biology and Evolution*, August 11, 2013. doi:10.1093/molbev/mst138.

2.1: Abstract

In microbial populations, growth initiation and proliferation rates are major components of fitness and therefore likely targets of selection. We used a high-throughput microscopy assay, which enables simultaneous analysis of tens of thousands of microcolonies, to determine the sources and extent of growth-rate variation in the budding yeast (*Saccharomyces cerevisiae*) in different glucose environments. We find that cell growth rates are regulated by the extracellular concentration of glucose as proposed by Monod (Monod 1949), but that significant heterogeneity in growth rates is observed among genetically identical individuals within an environment. Yeast strains isolated from different geographic locations and habitats differ in their growth-rate responses to different glucose concentrations. Inheritance patterns suggest that the genetic determinants of growth rates in different glucose concentrations are distinct. In addition, we identified genotypes that differ in the extent of variation in growth rate within an environment despite nearly identical mean growth rates, providing evidence that alleles controlling phenotypic variability segregate in yeast populations. We find that the time to reinitiation of growth (lag) is

negatively correlated with growth rate yet this relationship is strain-dependent. Between environments, the respirative activity of individual cells negatively correlates with glucose abundance and growth rate, but within an environment respirative activity and growth rate show a positive correlation, which we propose reflects differences in protein-expression capacity. Our study quantifies the sources of genetic and non-genetic variation in cell growth rates in different glucose environments with unprecedented precision, facilitating their molecular-genetic dissection.

2.2: Introduction

The rate at which a population of cells proliferates (i.e. the population growth rate) depends on both the rate of cell growth (increase in mass and volume) and the rate of cell division (increase in number). Understanding the physiological principles and molecular determinants governing cell proliferation rates is of broad importance in biology. Despite many decades of research, major questions remain regarding how cells regulate their rate of growth and how cell division, cell growth and diverse cellular processes including metabolism and macromolecular synthesis are coordinated. At the same time, new questions are emerging including the identities of naturally occurring genetic variants that underlie heritable variation in proliferation rates (Cubillos et al. 2011), the extent to which environmental conditions impact this variation (Liti and Louis 2012), and the molecular basis of heterogeneous growth strategies among genetically identical cells in the same environment (Levy et al. 2012).

In all organisms, the rate of cell proliferation is sensitive to the status of environmental nutrients required for biomass accumulation and energy metabolism. In the single-celled microbe *Saccharomyces cerevisiae* (budding yeast), the molecular form and abundance of environmental carbon is a major determinant of proliferation rates. The addition of glucose to glucose-deprived cultures of *S. cerevisiae* results in dramatic changes in cell physiology and metabolism, as well as alterations in the expression of more than 40% of genes (Zaman et al. 2008). The major transcriptional changes include increased expression of genes involved in ribosome biogenesis and repression of genes required for respiration, and the metabolism of alternative carbon sources (Zaman et al. 2008), consistent with glucose lying upstream of a regulatory network that coordinates cell growth with metabolism. While the study of glucose regulation has typically entailed comparison of cells deprived of glucose with those provided with an abundance of glucose (2% w/v or 111 mM glucose in standard formulations), evidence suggests that cells modulate their responses to environmental glucose across a wide range of concentrations (Reifenberger et al. 1997; Yin et al. 2003; Kaniak et al. 2004).

In single-celled microbes, variation in cell growth rates has important implications for evolution (Blomberg 2011). A fast-growing lineage will rapidly outcompete even slightly slower growing lineages when nutritional resources are abundant. However, microorganisms often face nutritionally poor environments (Smets et al. 2010). How they respond to suboptimal nutrient availability and starvation and, conversely, the kinetics with which they respond to nutrient replenishment are also major components of fitness. Moreover, it is not just the average response that matters, but the variance matters as well. If two lineages have identical arithmetic-mean

growth rates, the lineage with the least individual-to-individual variation around that mean will outcompete the other during growth (Frank 2011). However, population heterogeneity might provide an advantage in fluctuating environments (Kussell and Leibler 2005; Frank 2011; Levy et al. 2012). To date, most studies of microbial fitness have focused on the population growth rate in nutrient-rich conditions (Giaever et al. 2002; Hillenmeyer et al. 2008). However, this laboratory condition is of unknown relationship to environments encountered by natural isolates of yeast and analyses restricted to nutrient-rich conditions are likely to miss important, and potentially adaptive, variation. At the same time, variation in proliferation rates among diverse natural isolates of yeast in suboptimal conditions may provide unique insight into the regulation of cell growth and how this variation has been shaped by ecological and geographic histories.

A more complete understanding of environmental and genetic determinants of cell proliferation rates requires surmounting two technical challenges: 1) accurate measurement of proliferation rates across a wide range of conditions including near-starvation conditions and 2) quantification of variation among genetically identical individuals. We recently developed a growth assay that measures individual cells growing into microcolonies comprised of up to ~100 cells that solves both of these problems (Levy et al. 2012). An important advantage of this approach over other high-throughput methods of growth-rate analysis is the capability of determining distributions of growth rates derived from thousands of individual microcolony growth-rate measurements.

In this study, we have used this approach to investigate cell growth in a range of glucose concentrations of natural isolates of *S. cerevisiae* with different ecological histories. We

extended our high-throughput microcolony assay to enable measurement of both growth rate and lag time in single cells. We find that cell growth rates vary with glucose concentrations in accordance with a deterministic model of substrate-limited growth (Monod 1949). We surveyed a panel of wild yeast isolates across these conditions and find prevalent genotype-by-environment interactions, suggesting that different genetic factors underlie growth-rate variation at different glucose concentrations. Isolates also differ in growth-rate variance independently of differences in mean growth rate. Using a fluorescent reporter of respirative metabolism, we find that although increased respiration is anti-correlated with growth rate between environments, within an environment increased respirative enzyme expression is correlated with increased proliferation rates, perhaps reflecting non-genetic variation in protein production capacity. By quantitatively analyzing variation in growth reinitiation, proliferation and metabolism in a spectrum of glucose-containing environments, we reveal a continuum of growth strategies among yeast populations that is amenable to genetic dissection.

2.3: Results

2.3.1: High-throughput analysis of environmental determinants of cell growth variation

The rate of proliferation of yeast cells is regulated in response to both the form and abundance of environmental nutrients. Using chemostat cultures it has been shown that populations of yeast cells can modulate their rates of growth across at least a tenfold range (Brauer et al. 2007). In batch cultures, growth in environments containing low nutrient concentrations that are equivalent to the steady-state concentrations in chemostats cannot be easily measured using conventional methods (i.e. optical density or particle counting).

We hypothesized that our recently developed high-throughput microcolony growth-rate assay would provide sufficient resolution to measure cell proliferation rates in low-nutrient environments. Our assay uses time-lapse microscopy to monitor individual cells undergoing a small number of divisions to form microcolonies in 96-well glass-bottom plates (Levy et al. 2012) (**Figure 2.1A**). Previously, we showed that the rate of change in microcolony area is highly correlated with the rate of change in cell number and thus provides an accurate estimate of microcolony growth rate (Levy et al. 2012). To study the effect of environmental glucose concentration on cell growth rates we used minimal, chemically defined media (Saldanha et al. 2004). Prior to each experiment, cultures were grown to stationary phase in carbon-limiting media to ensure cell-cycle arrest due to carbon starvation. Starting from growth-arrested cells, rather than exponentially growing cells, allowed us to observe the time to reinitiation of growth (i.e. lag) in each environment. Each microcolony growth profile is defined by two phases, a lag phase and a growth phase (**Figure 2.1B**). We used a sliding-window regression method to locate the maximal rate of proliferation (increase in log area) for each microcolony (**methods**). Lag duration was defined by the intersection of this maximal-proliferation line with a horizontal line defined by the initial cell size (**methods**). Using 96-well plates our assay enables us to measure the lag times and growth rates of as many as 80,000 individual microcolonies in a single 24-hour experiment.

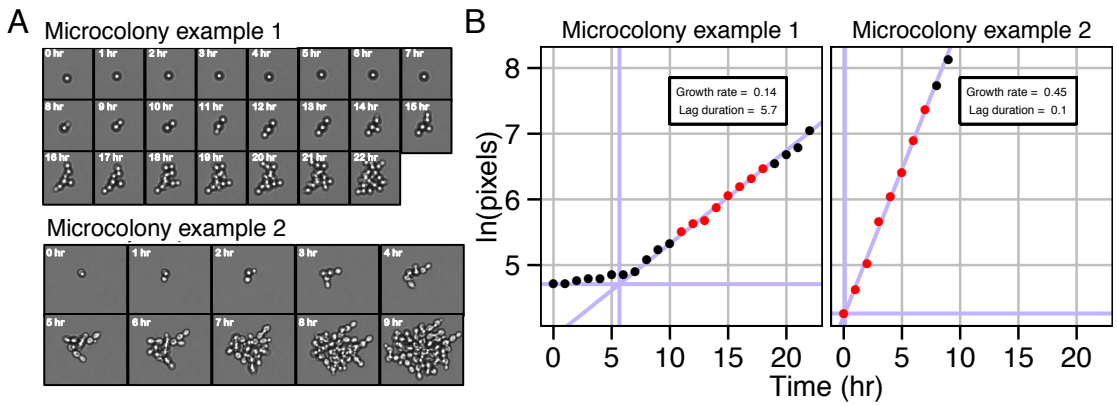


Figure 2.1: Calculation of growth parameters from microcolony growth profiles
Images (A) and growth profiles (B) for two representative microcolonies. Growth of microcolonies follows a simple two-phase log-linear model. The exponential growth phase for each microcolony is determined by a sliding-window regression. The window of eight consecutive time-points with the log-linear fit of greatest slope (and R^2>0.9) defines the maximal growth rate (red points). Lag duration is defined by the intersection of the line defining the growth phase with a horizontal line defined by the initial cell size

2.3.2: Microcolony growth rate is determined by glucose concentration in agreement with the Monod model for substrate-limited growth

We sought to determine: 1) the relationship between population growth rate and glucose concentration and 2) if the response to glucose concentration varies among natural isolates of *S. cerevisiae*. In preliminary studies, we found that mean growth rate was not affected in media containing ~25-fold less glucose than standard media. Therefore, we analyzed microcolony growth rates in seven different glucose concentrations, ranging from 0.05–4.44 mM glucose, for four different prototrophic diploid strains. The strains derive from the laboratory (FY4/5, isogenic to the reference yeast strain S288c; hereafter "lab"), a North American oak tree (BC248; hereafter "oak"), a Californian vineyard (BC241; hereafter "vineyard") and a cross between the

oak and vineyard strains (BC252, hereafter "oak/vineyard F1") (Gerke et al. 2006). The lab, oak and vineyard strains are homozygous throughout the genome. Each strain was grown in each glucose environment in three wells per plate and each plate was replicated four times, resulting in over 150,000 microcolony growth-rate measurements (**methods**). We confirmed that microcolony area is highly correlated with cell number for different strains growing in 0.22 mM glucose (**Figure 2.S1**).

To differentiate variation due to the factors of interest (i.e. genetic background and environment) from variation unique to individual wells and plates (which likely result from variation in illumination, focus and media preparation) we used mixed-effect linear modeling in which we included strain identity, glucose concentration and their interaction as fixed effects and the plate and well as random effects (see **supplementary note**). Estimates for each genotype-environment combination clearly showed growth rate to be a function of both genetic background and environment (**Figure 2.S2**). In order to combine growth rate measurements for a given genotype from different wells and plates, we normalized the data by subtracting plate and well conditional means estimated from the mixed model from each microcolony growth rate (see **supplementary note**). The normalized data were used for further analysis.

We aimed to model the growth-rate response to glucose concentration as a continuous function. Monod proposed that cell growth rate is related to the concentration of a limiting nutrient with saturating kinetics that resemble the Michaelis-Menten function (Monod 1949). Using non-linear least-squares regression, the normalized data for each strain were fit to the Monod model

(**Figure 2.2A, Figure 2.2B, Figure 2.S3, Table 2.S1**). This model requires two parameters: the maximum growth rate (μ_{max}) and the glucose concentration at which growth rate is half-maximal (K_s). Our estimates of μ_{max} (0.43-0.52 hr^{-1}) and K_s (0.1–0.2mM) are similar to values estimated for *S. cerevisiae* strains using bulk population growth rates in batch cultures and chemostats (Snoep et al. 2009) respectively. As our estimates are generated from a large number of measurements (28,000 – 42,000 growth rates per strain), these parameters are estimated extremely accurately with standard errors on the order of 10^{-4} (i.e. three orders of magnitude smaller than the parameters).

Alternatives to the Monod model have been proposed (Kovárová-Kovar and Egli 1998). Several of the alternatives are slight variations on the Monod model, containing an additional "maintenance" term representing the need for substrate even when cells are not growing (Kovárová-Kovar and Egli 1998). A conceptually different model is that of (Westerhoff et al. 1982), which is based on non-equilibrium thermodynamics and proposes a linear dependence of growth rate on the logarithm of the substrate concentration (Westerhoff et al. 1982). We fit the data on each of the four strains to various alternative models (**Figure 2.S3, Table 2.S1**). We compared model fits by Akaike information criterion (AIC), Bayesian information criterion (BIC) and, when appropriate, likelihood-ratio tests (**Table 2.S1**). With the exception of the lab strain, variants of the Monod model with an additional parameter slightly improved fit relative to the Monod model (**Figure 2.S3, Table 2.S1**). Moreover, for all four strains, the Monod model fit substantially better than the Westerhoff model (**Figure 2.S3, Table 2.S1**). The data therefore

support the Monod model, or slight variations of it, over the most prominent competing model of substrate-limited growth.

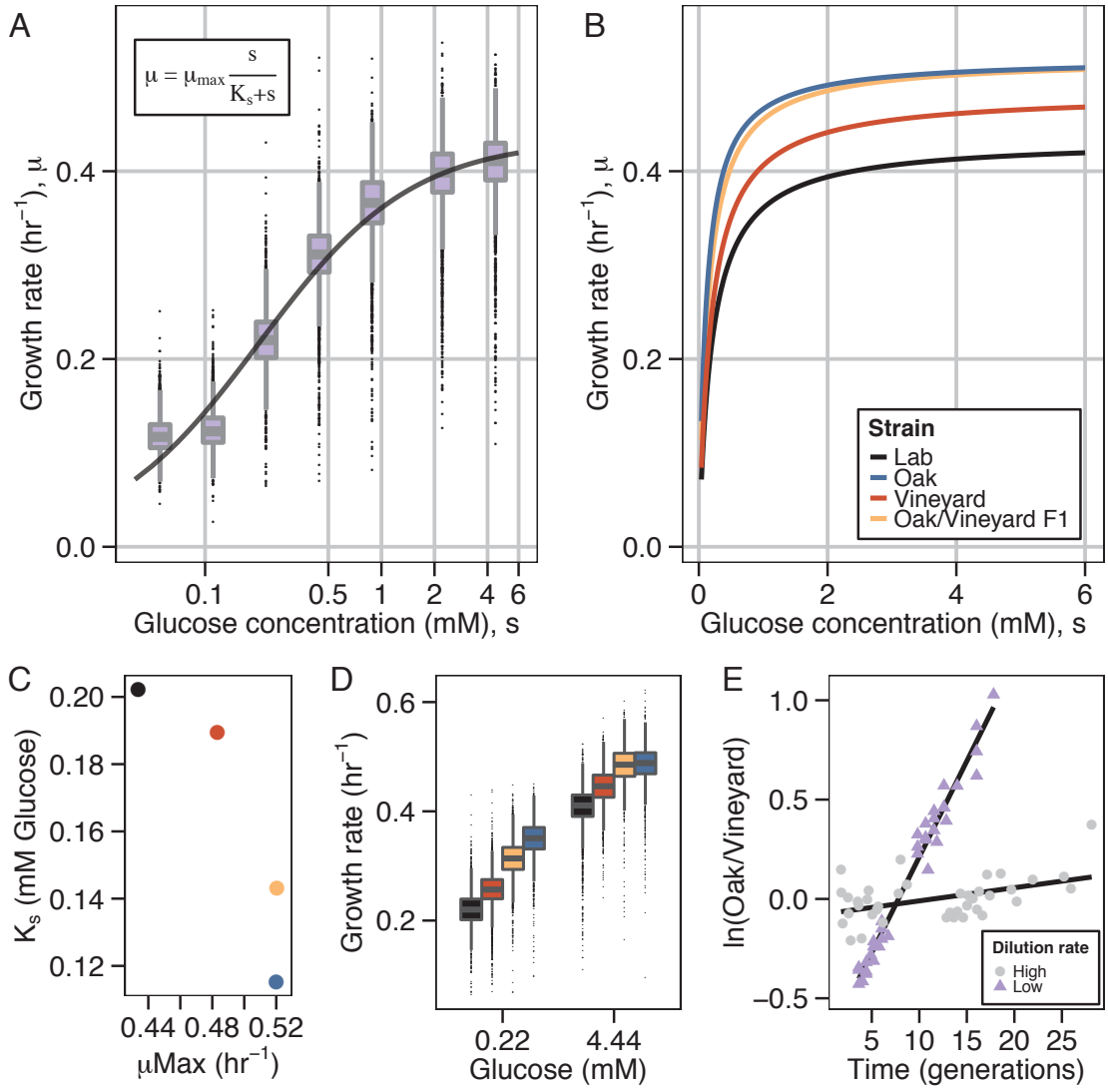


Figure 2.2: Growth rate is determined by glucose concentration in agreement with the Monod model for substrate-limited growth and varies among natural isolates of S. cerevisiae

(A) Growth-rate distributions for the lab strain in a range of glucose conditions. The line depicts the best fit of the Monod equation (inset) to the normalized data. Glucose concentration is shown

on a logarithmic scale for clarity. (B) Fits to the Monod equation, showing growth rate as a function of glucose concentration for four strains. (C) Ks and μmax estimates for four strains. (D) Growth rates at 0.22 mM and 4.44 mM glucose are proxies for Ks and μmax respectively. (E) Competitive growth rate assays between oak and vineyard strains at two dilution rates in chemostats (lower dilution rates correspond to lower glucose concentrations). Replicate experiments were centered by mean subtraction; lines depict linear regressions of log-transformed ratios against generations.

2.3.3: The growth-rate response to different glucose concentrations varies among yeast strains

Each of the four strains is defined by a unique combination of μ_{max} and K_s parameters (**Figure 2.2C**). The oak strain grows faster than both the vineyard and lab strains at all glucose concentrations; however the data also display genotype-by-environment interactions. The oak/vineyard F1 has an intermediate value for K_s compared to the parental strains, whereas its μ_{max} is identical to the oak parental strain (**Figure 2.2C**). This suggests distinct genetic effects underlying variation in these two parameters, which we estimated using the mid-parent heterosis (MPH) metric (Zorgo et al. 2012) (**methods**). In the case of K_s the net genetic effect is largely additive (MPH = 0.25) whereas in the case of μ_{max} the net genetic effect is primarily dominant (MPH = 1). Genetic variation in μ_{max} and K_s is reflected in growth rate in high and intermediate glucose concentrations respectively: the growth rate of the oak/vineyard F1 at 0.22 mM glucose yields a MPH = 0.22 whereas at 4.44 mM the MPH = 0.9 (**Figure 2.2D**).

We sought to independently confirm the effect of environmental glucose concentration on the growth rates of the oak and vineyard strains. Therefore, we measured the relative growth-rate

differences between the oak and vineyard strains using competitive growth-rate assays in chemostats (**methods**). Because the steady-state residual glucose concentration increases with increased dilution rate we performed competition assays in glucose-limiting media at a low (D= 0.18–0.2 hr^{-1}) and a high (D=0.35–0.39 hr^{-1}) dilution rate. Consistent with our microcolony growth rate results, the growth advantage of the oak strain at a low dilution rate is greater than at a high dilution rate (9.6%±0.28% vs 0.6%±0.2%) (**Figure 2.2E**). Thus, both competition assays in chemostats and our microcolony growth-rate assay reveal that growth-rate differences between these two strains are conditional upon environmental glucose concentration.

To more broadly survey genetic variation in the response of growth rate to glucose concentrations, we used the microcolony assay to analyze additional strains (**methods**) covering a range of genetic backgrounds and ecologies (Liti et al. 2009). Each strain was measured in 0.22 mM and 4.44 mM glucose, resulting in over 300,000 microcolony growth rates. These diverse strains exhibit a range of growth rates in both glucose concentrations (**Figure 2.3**), which is delimited by the fastest-growing oak strains and the slowest-growing lab strain. The majority of strains have similar, intermediate growth rates in 4.44 mM glucose, but show more pronounced differences in growth rates in 0.22 mM glucose. Similarity in growth rates does not appear to be determined solely by common ecologies or genetic relatedness (as defined by (Liti et al. 2009)).

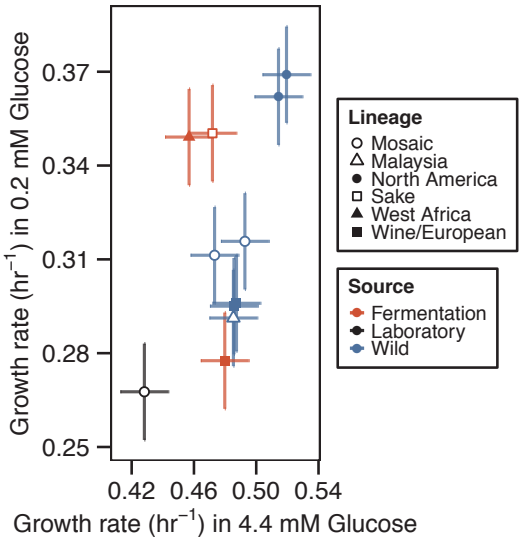


Figure 2.3: Natural variation in growth rates at two different glucose concentrations
Growth rates at 0.22 mM and 4.44 mM glucose for 11 diverse strains. Shapes and colors correspond to genetic lineage and source (as defined by (Liti et al. 2009)) respectively. Error bars represent 95% confidence intervals.

2.3.4: Natural variation exists in the distributions of growth rates within an environment

In addition to variation in the average growth-rate response to different glucose concentrations between genotypes, we observed substantial variation in growth rates within each environment for a given genotype. In contrast to our previous study in which we observed left-skewed distributions (Levy et al. 2012), growth-rate distributions among diverse strains and conditions are largely symmetric. Therefore, we studied the effect on the shape of the growth-rate distribution of ploidy, growth condition and the recent history of the cells. We find that diploid strains have fewer slow-growing cells than haploid strains, that growth in minimal medium yields fewer slow-growing cells than growth in rich media, and that these effects of ploidy and nutrient conditions are particularly strong in the lab strain genetic background (**Figure 2.S4**).

We aimed to determine if the variance of growth-rate distributions differs among the different strains and glucose concentrations. We used log-transformed absolute values of the residuals

from the mixed model as a measure of growth-rate deviation (see **supplementary note**). These deviations were then used as random variables in a new linear mixed model with the same structure as the original model. By using mixed modeling, we control for confounding technical effects of wells and plates on our estimates of average growth-rate deviations for each strain at each nutrient concentration.

We find significant differences in the variability of growth rates among strains (**Figure 2.4A**) that are independent of the mean growth rate. Consistent with a lack of correlation between growth-rate means and variances (**Figure 2.S5**), there is no clear relationship between the variances in growth rates in the two glucose concentrations (**Figure 2.4A**). Notably, two European soil strains, which have nearly identical mean growth rates, show significantly different deviations from the mean in both glucose concentrations (Wilcoxon test on normalized data, p-value < 2.2e-16). Comparison of their growth-rate distributions shows that the Dutch soil strain has a broader distribution than the Finnish soil strain, including both slower and faster growing cells (**Figure 2.4B**). These observations provide evidence that variability in growth rates within environments is genetically determined and may be affected by genetic factors that are independent of those factors that affect the mean growth-rate response.

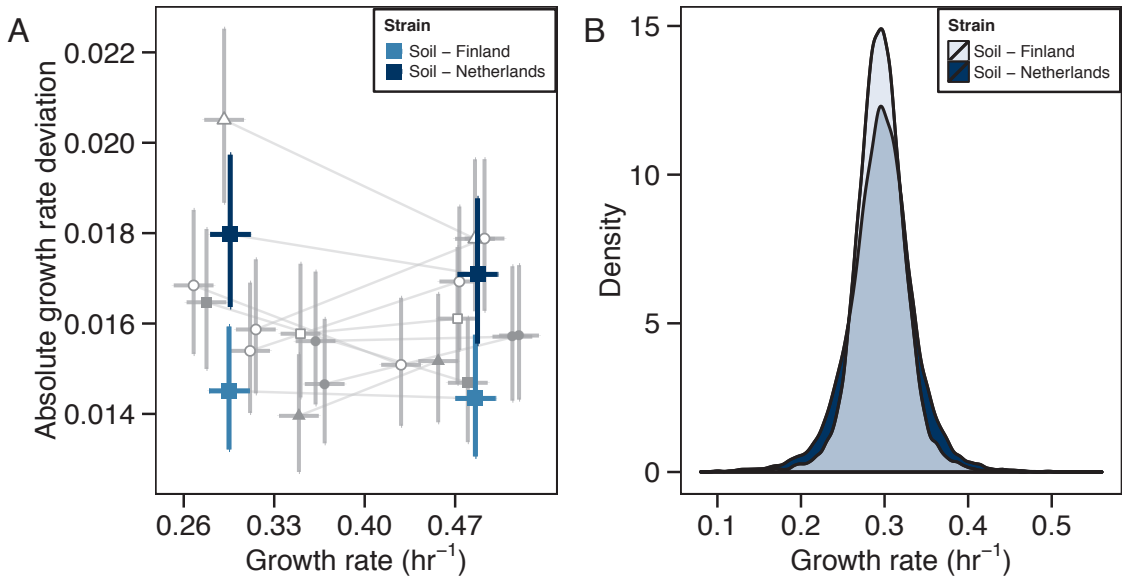


Figure 2.4: Variation in the extent of growth-rate variability

(A) Absolute growth-rate deviations are plotted against growth rate for 11 strains at 0.22 mM and 4.44 mM glucose. Lines connect estimates for the same strain at the two concentrations. Shapes indicate genetic lineage as defined in Figure 2.3. The two European soil strains that exhibit the same mean but different absolute growth-rate deviations are highlighted. Error bars represent 95% confidence intervals. (B) Growth-rate distributions for the two soil strains at 0.22 mM glucose.

2.3.5: Natural variation in time to reinitiation of growth

Our assay enables estimation of the time each cell takes to reinitiate growth in a defined environment (**Figure 2.1**). The fraction of cells that undergo a detectable lag decreases as glucose concentration increases (**Figure 2.5A**). In 4.44 mM glucose few cells lag whereas in lower glucose concentrations the majority of cells display a delay before initiating growth. The fraction of cells that lag also displays genetic variation, as a greater proportion of vineyard and lab cells lag than oak cells in almost all environments. The percentage of lagging cells (at 0.11

and 0.22 mM glucose) correlates with estimates of K_s across strains (Pearson correlation coefficient>0.969, p.value<0.03). Although 30% of oak strain cells do not have a detectable lag time (in 0.05 mM glucose), the unimodal distribution of lag times for all strains suggests that a nutrient concentration threshold exists at which all cells exhibit a lag regardless of genotype.

To quantify the difference in lag duration between strains we used mixed-effect modeling (**methods, supplementary note**). We find that the average duration of lag is inversely correlated with mean growth rate, yet this relationship is variable between strains (**Figure 2.5B**). Strikingly, the lab and vineyard strains have longer average lag durations than the oak strain even when the subsequent growth rate is similar. This observation suggests that reinitiation and proliferation rates are under distinct genetic control.

We wanted to determine if strains differ in the variability of lag time in addition to their differences in average lag duration. In contrast to growth rate, there is a strong relationship between average duration of lag and the associated variance within environments (**Figure 2.5C, Figure 2.S5A**). In low-glucose environments cells exhibit extremely heterogeneous behaviors with some cells initiating growth immediately whereas others lag for greater than ten hours before initiating growth. In order to control for this inherent relationship, we used smoothed local regression to estimate and control for the relationship between mean and variance. Specifically, median absolute deviations were regressed on medians and the residuals were compared between strains (**Figure 2.S6**). No significant difference was found between strains (F=1.618, p=0.185).

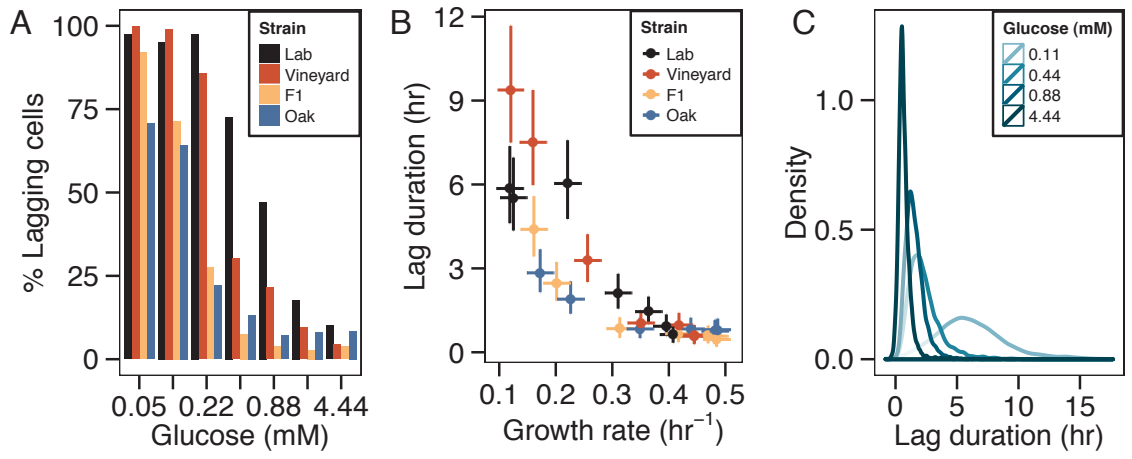


Figure 2.5: Variation in single-cell lag-duration distributions

(A) Percentage of cells with estimated lag durations greater than 1 hour for four strains grown in seven different glucose concentrations. (B) Estimates of lag duration versus growth rate. Error bars represent 95% confidence intervals. (C) Distributions of lag times for the lab strain at four glucose concentrations.

2.3.6: A marker of respiration correlates with growth rate negatively across conditions and positively within conditions

As environmental glucose concentrations are known to affect whether yeast cells ferment or respire, we sought to determine the metabolic states of microcolonies growing at different glucose concentrations. *CIT1* encodes a citrate synthase that catalyzes the first step in the TCA cycle. CIT1-GFP expression has been shown to correlate with the degree of respiration on different carbon sources (Fendt and Sauer 2010) and the relative abundance of CIT1 mRNA is negatively correlated with growth rate in glucose-limited chemostats (Brauer et al. 2007). Therefore, we used the average expression of a CIT1-GFP fusion protein (in the lab strain genetic background) as a marker of respiratory activity in growing microcolonies (**methods**).

To define the range of CIT1 expression we measured protein fluorescence and growth rates in the lab strain growing in: 1) a non-fermentable carbon source (acetate); 2) a high concentration of glucose (2%, 111 mM), in which respiration is minimal; and 3) galactose, in which cells simultaneously respire and ferment (Fendt and Sauer 2010). We then measured CIT1 expression and growth rate simultaneously in the range of glucose concentrations over which all strains exhibit glucose-dependent growth rate variation. Across all conditions CIT1 expression is negatively correlated with growth rate (**Figure 2.6A**). These data are consistent with a near-complete absence of respiratory activity in cells growing in concentrations as low as 4.44 mM glucose and a systematic increase in respiration as environmental glucose concentration decreases.

By contrast, within environments there is a positive correlation between CIT1 expression and growth rate in different carbon sources (**Figure 2.6B**) and different glucose concentrations (**Figure 2.6C**). The strength of this relationship increases as average growth rate decreases. That is, although the average growth rate in conditions that promote increased respiration (low glucose concentrations or alternative carbon sources) is lower, within these conditions cells that have higher expression of CIT1 tend to grow faster than cells with lower levels of CIT1 expression.

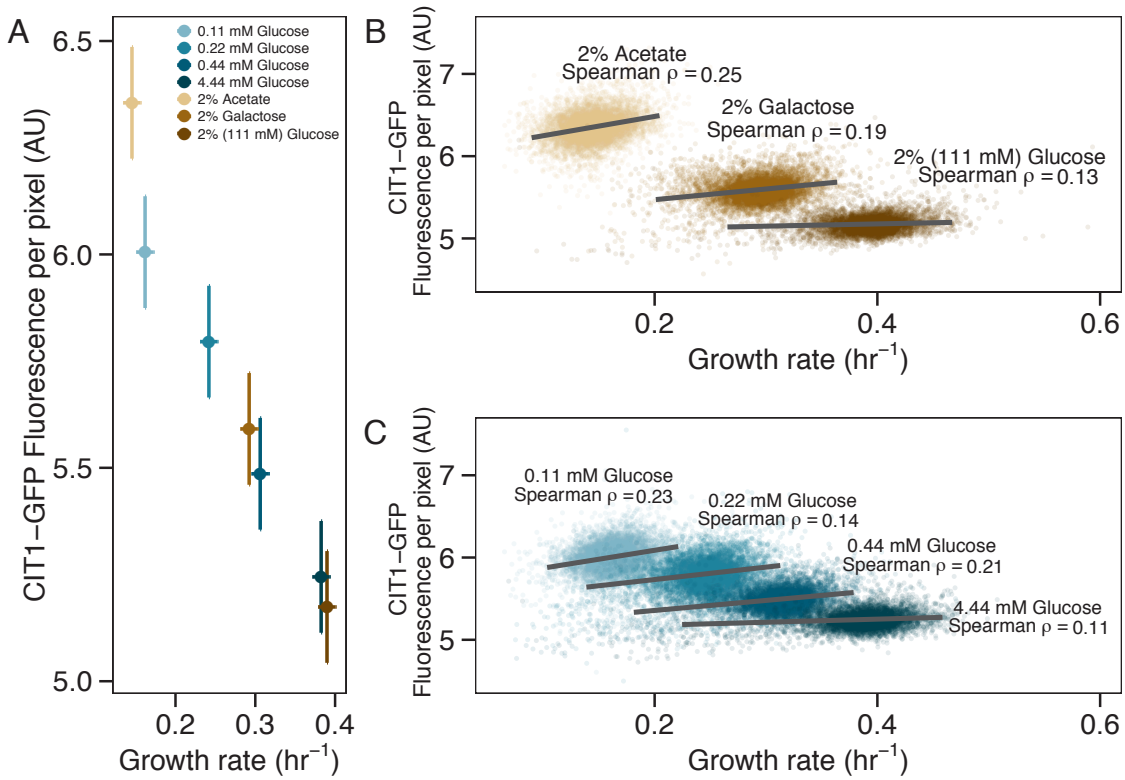


Figure 2.6: Respiration and growth rate are negatively correlated between environments but positively correlated within environments

(A) Estimates of CIT1 expression versus growth rate. Colors depict different environments. Error bars represent 95% confidence intervals. (B) Microcolony growth rates and CIT1 expression with glucose, galactose and acetate as carbon sources. Lines depict major axis type II regressions for the central 98% of growth-rate variation in each condition. Non-parametric correlation coefficients (Spearman ρ) were calculated for each condition. Bootstrapped standard errors range between 0.012–0.014. (C) Same as (B) for four different glucose concentrations.

2.4: Discussion

The quantitative analysis of microbial growth was initiated by Jacques Monod and colleagues in the middle of the twentieth century (Monod 1949). The advent of genomics and systems biology has stimulated renewed interest in understanding cell growth as recent advances make it clear

that the rate of cell growth is a major determinant of the transcriptional (Regenberg et al. 2006; Brauer et al. 2007; Gutteridge et al. 2010) and metabolic state (Boer et al. 2009; Gutteridge et al. 2010) of the cell. Moreover, modeling cell behavior requires incorporation of cell growth rate as a parameter (Scott and Hwa 2011). Here, we have extended and enhanced our recently reported microcolony growth assay (Levy et al. 2012), which combines the advantages of accurate measurement of variation in growth rates between individual cells with high-throughput capacity enabling investigation of growth rate distributions across genetic backgrounds and environments.

We have used this assay to study growth rate variation in response to changes in extracellular glucose concentrations. We have shown that the growth rate response of *Saccharomyces cerevisiae* to glucose concentration agrees with the Monod model of nutrient-regulated growth. The growth rate of yeast cells is continuously adjusted in response to the concentration of environmental glucose and maximal growth rates are achieved at low millimolar concentrations of glucose. Although it has been suggested that differences in genotype can influence the parameters of the Monod model (Ferenci 1999), these parameters had not been compared between different natural isolates in any organism. We found that the growth rate response to glucose shows natural variation among yeast strains. Comparisons of two homozygous parental strains and their F1 hybrid as well as a number of wild strains suggest that K_s and μ_{max} are under distinct genetic control. As fluctuating nutrient availability and nutrient limitation are probably important aspects of the natural habitats of microbes we expect that identifying the genetic factors underlying variation in growth rates at different glucose concentrations would prove

informative about both the genetic control of cell growth and the evolutionary histories of diverse yeast strains.

Although functional genomic studies have shown that deletion of many genes can affect cell growth rates in rich media conditions (Giaever et al. 2002; Hillenmeyer et al. 2008) there are few examples of natural alleles that underlie variation in cell growth rate. Natural variation in components of the RAS/cAMP pathway (IRA1 and IRA2) has been implicated in regulating quantitative growth at high temperature (Parts et al. 2011) and expression of growth-related transcripts in glucose and ethanol (Smith and Kruglyak 2008). The Monod constant (K_s) relates to the affinity of the cell for the nutrient, but its biological interpretation is a subject of debate (Liu 2007). K_s may be related to the K_m of the relevant transporter, but the relationship between K_s and K_m depends on assumptions about the kinetics of the transport system, and the extent of control that the transport step has on the growth rate can change between different substrate concentrations (Snoep et al. 2009). Increased expression of the high-affinity glucose transporters *HXT6* and *HXT7* has been shown to confer increased fitness in experimental evolution in glucose-limited conditions (Brown et al. 1998; Gresham et al. 2008) suggesting that growth rates at sub-maximal glucose concentrations may be largely determined by substrate-transport rates.

A unique aspect of our assay is the ability to monitor lag in individual cells. Lag duration in bulk populations is usually poorly defined and the accuracy of its estimation is limited by the power to detect growth at low cell density. In population growth curves the transition between lag and growth phases is typically smooth, due to variability between individual cells (Buchanan et al.

1997), making the determination of a single lag time somewhat arbitrary. When measuring single cells, our data are best described by a distinct transition between lag and the exponential growth phase. Our results support previous observations (Peleg and Corradini 2011) that these different phases of growth can vary independently. We observe significant heterogeneity in the duration of lag particularly as environments become increasingly poor in glucose. Understanding the molecular basis of heterogeneity in lag will provide insight into the processes that underlie exit from quiescence and reinitiation of cell growth and may have practical applications; for example, in the food industry (Swinnen et al. 2004), where outgrowth of a small number of individual cells is a major concern.

We observed phenotypic variability in both lag and growth phases. The advantage of increased cellular variability in the face of novel and fluctuating environments is relevant to the evolutionary rates of cancer progression and drug resistance (Frank and Rosner 2012). Although the extent of nongenetic phenotypic variation has important implications for evolutionary dynamics, it has been understudied in part because of the difficulties of accurately assessing phenotypes of individual cells. Consequently, the mechanisms regulating nongenetic phenotypic variability remain poorly understood (Pelkmans 2012; Geiler-Samerotte et al. 2013). One possible source of phenotypic variability is differences in gene expression between individual cells. Possible mechanisms by which these differences can be enhanced or reduced include regulation of chromatin modification (Levy and Siegal 2008), promoter structure (Ferguson et al. 2012) and variability in the inheritance and functionality of mitochondria (Johnston et al. 2012). Furthermore, variation in protein translation capacity between cells (possibly as a result of

variation in ribosomal content), may contribute to variation in growth rates (Scott and Hwa 2011). For cell growth in particular, variability in the lengths of different cell-cycle stages had been observed (Di Talia et al. 2007; Son et al. 2012), but the impact of natural variation on mechanisms controlling variability of cell-cycle timing has not been explored. Our identification of strains that differ substantially in their growth-rate variances despite nearly identical means, presents an ideal scenario for identifying the genetic and molecular basis of natural variation in growth-rate variability.

We find that a decrease in cell growth rate corresponds with increased respirative activity, as measured by CIT1 expression, consistent with previous studies of the diauxic shift and yeast metabolic cycle (Brauer et al. 2005; Silverman et al. 2010). Because cells increase their respiratory activity at low glucose concentrations, differences in the efficiency of respiration between the oak and vineyard strains (Gerke et al. 2006) may underlie variation in K_s. Intriguingly, in contrast to the negative correlation between CIT1 expression and growth rate between glucose environments we find that within a glucose environment, increased expression of CIT1 is correlated with increased growth rate. We suggest that this correlation may represent global differences in rates of protein production between cells that are not necessarily specific to CIT1. The fastest growing cells within the same environmental conditions may have greater translational capacity. In this scenario, the environmental conditions specify the metabolic state of the cell, but inter-individual variation in protein production capacity underlies heterogeneity in growth rates. As a result, different combinations of metabolic and translational capacity can specify the same growth rate. Continued investigation of variation both within and between

environments will provide a deeper understanding of the genetic and non-genetic sources of this variation and how cells optimize their growth potential in a particular environment.

2.5: Materials and methods

2.5.1: Yeast strains and media

All strains used in this study are prototrophic diploids, with the exception of two prototropic haploids used to investigate the effect of ploidy on the shape of the growth-rate distribution (**Figure 2.S4**). All wild strains were obtained from the lab of Barak Cohen (Washington University). The strains used in this study are the oak (BC248), vineyard (BC241), the F1 hybrid (BC252), an additional North American oak strain (YPS126), two European strains isolated from soil samples (DBVPG1373 and DBVPG1788), three genetically diverse strains isolated from plants/fruit in Malaysia, Hawaii and the Bahamas (UWOPS03-461.4, UWOPS87-2421 and UWOPS83-787.3), an African palm wine isolate (Y12) and a West African Bili wine isolate (DBVPG0644). Strains shown in **Figure 2.S4** are the prototrophic diploid laboratory strain FY4/5; its haploid parent of mating-type a, FY4; the oak strain from which BC248 was derived YPS606; and its MAT**a** haploid spore (after HO was knocked out), YPS2056.

For competitive growth rate assays, the HO locus was replaced with the mCherry fluorescent protein and a nourseothricin resistance marker (natMX4) in both the oak and vineyard strains using high-efficiency transformation as in (Gerke et al. 2006). A haploid CIT1-GFP strain (from the yeast-GFP collection (Huh et al. 2003)) was purchased from Invitrogen and mated to FY5 (a prototropic alpha haploid), creating a functionally prototrophic diploid. All media used in this

study were chemically defined carbon limiting media (Saldanha et al. 2004; Brauer et al. 2005) without amino acid or nucleotide supplements.

2.5.2: Inoculation of microcolonies

For all growth-rate experiments, frozen cell stocks were streaked out on YPD plates and single colonies were used for inoculation. Cells were cultured in 4.44 mM glucose media for ~24 hr, diluted 1:300 into fresh 4.44 mM glucose media and cultured for ~48 hr. This procedure reduced variability in culture density at the initiation of growth-rate experiments and ensured that cells were starved for carbon. Cells were then diluted to a concentration of 1–2 x 10^4 cells/ml in fresh media containing defined glucose concentrations and plated in 96-well glass-bottom plates coated with concanavalin A. Each well was loaded with 400 µl of diluted cells (i.e. ~6000 cells). The glucose concentration is assumed to remain approximately constant throughout the experiment due to the low cell density and large media volume within each well.

2.5.3: Microscopy and automated image analysis

Experiments were conducted as described in (Levy et al. 2012), including all equipment and software for computing and tracking microcolony areas over time. The focusing routine was updated to a manual assignment for each well based on a single field (which took ~10 minutes per 96-well plate). Images were taken every hour for 2,880 fields (30 fields per well) for 20–24 hours. CIT1-GFP fluorescence was captured for 2 s at 10x gain; due to the long exposure time, these experiments contained only 32 wells per plate and 20 fields per well (640 fields total).

2.5.4: Growth profiles and CIT1 expression analysis

Microcolony growth profiles were analyzed according to a two-phase log-linear model. The use of a simple model is preferred (Buchanan et al. 1997; Peleg and Corradini 2011), particularly as microcolony growth profiles do not exhibit the smooth transition between phases caused by variability among cells in population growth curves (Buchanan et al. 1997). For each microcolony, a sliding-window approach was used to determine the phase of maximal growth. The natural log of microcolony area was regressed against time for each set of eight time points and growth rate was calculated as the greatest slope of a regression with $R^2>0.9$. Subsequently, lag duration was estimated as the intersection of this regression with a horizontal line determined by the area of the microcolony at the first time-point. Lag duration was not calculated for colonies that were not tracked in the first time-point. A small proportion of microcolonies with aberrant growth parameters (growth rate < 0.075 hr^{-1}, lag duration > 15 hr or initial size > 250 pixels) were omitted from further analysis. All calculations and analysis were conducted in R.

CIT1 fluorescence was averaged over area and time. First, total fluorescence intensity was measured for each microcolony and divided by the area of the microcolony, resulting in a measurement of fluorescence per pixel for each microcolony at each time point. Measurements were then averaged across four time points, during the period of maximal growth rate. Values were also log-transformed in order to reduce heteroscedasticity.

2.5.5: Mid-parent heterosis (MPH) metric

Mid-parent heterosis was calculated as:

$$MPH = \frac{X_{F1} - \frac{X_{Oak}+X_{Vineyard}}{2}}{\frac{X_{Oak}-X_{Vineyard}}{2}}$$

Where X is either the estimate of K_s and μ_{max} or the combined fixed-effect parameters (see **supplementary note**) for growth at 0.22 or 4.44 mM glucose.

2.5.6: Competitive growth rate assays

Competitive growth rate assays were performed in chemostats as described (Gresham et al. 2008). Each strain was competed against a mCherry-labeled strain in reciprocal experiments. Competitions were initiated with equal proportions of each strain and samples were obtained every 3-6 hours over 20 generations. The proportion of each strain at each time point was measured using flow cytometry and the relative growth-rate difference was determined by linear regression of ln(strain1/strain2) against time (measured in generations). The slope of the regression is the proportional difference in growth rate (i.e. the fitness advantage) of one strain relative to the other. Each competition was performed in replicate and data from replicate and reciprocal competitions were normalized by mean subtraction and pooled. Competitions were performed in chemostats at two different dilution rates, approximately 0.2 hr^{-1} (low) and 0.4 hr^{-1} (high).

2.5.7: Statistical analysis

All statistical analyses used R (Team 2012). Mixed-effect modeling was performed using the package *lme4* (Bates et al. 2011: 4) to analyze all measurements obtained from the microcolony growth assay, including distributions of growth rates, lag durations and fluorescent measurements. We use mixed effect modeling to estimate parameters and eliminate various aspects of technical variation on cell growth measurements. A discussion on the use of mixed-effect models including determination of terms and evaluation of parameters is contained in a supplementary note. Reproducibility of the analysis presented in this paper is also discussed in the supplementary note. Estimation of the non-linear regression parameters for the Monod growth model was performed using the *nls* function in R. These parameters' standard errors were estimated both by linearization and by bootstrapping, which yielded similar estimates. Correlation between growth rates and CIT1-GFP fluorescence was visualized by type II ranged major axis regression (Legendre 2011: 2), as both measurements are dependent variables, using the *lmodel2* package.

2.6: Acknowledgements and funding information

We thank members of the Siegal and Gresham labs for valuable discussions and two anonymous reviewers for helpful comments. We thank Barak Cohen for providing strains used in this study. This work was supported by the National Institute of Health (GM097415 to M.L.S) and the National Science Foundation (MCB-1244219 to D.G.) and a Dupont Young Professor award to D.G.

2.7: Supplementary material – Note about mixed effect modeling

2.7.1: Model construction

Mixed-effect models contain both fixed and random terms. The levels of fixed-effect terms (such as concentrations of glucose) are repeatable and of interest in and of themselves, rather than instantiations drawn randomly from a larger population. By contrast, the levels of random terms (such as plates) are sampled from a potentially infinite population. For all models the fixed terms were genotype, environment and the genotype-environment interaction and the random terms were wells (explicitly nested within plates), plates and the interactions of plate with genotype and environment.

The parameters of a mixed model are the fixed terms' regression coefficients, the random terms' variances and the error variance. All measurements (after appropriate transformation) were modeled with Gaussian distributions for error and random effects. Parameters were estimated using restricted maximum likelihood (REML), using the *lmer* function in the *lme4* R package. The complete dataset and an R script containing the analysis are provided as supplementary material.

We determined the significance of terms by performing a series of likelihood ratio tests, using the *anova* function, on nested models fit to the same dataset. For analyses of lag duration and absolute growth-rate deviation, log transformations were used to reduce heteroscedasticity and non-normality of the measurements, improving the accuracy of estimations and their associated errors. **Table 2.1** presents datasets and models analyzed in this paper.

Table 2.1: Datasets and analytical strategy

The complete study was performed and analyzed as four datasets.

Dataset	Structure	Models discussed
LWO – Lab, Vineyard, Oak and Oak-Vineyard F1 strains in a range of glucose concentrations *	4 genotypes X 7 environments X 3 wells per plate X 4 plates	Growth rate Lag duration
WILD – A range of different wild yeast isolates in two glucose concentrations **	12 genotypes X 2 environments X 4 wells per plate X 4 plates	Growth rate Growth-rate deviation Lag duration
CIT1 – GFP fusion in Lab strain background in a range of glucose concentrations ***	1 genotype X 8 environments X 4 wells per plate X 3 plates	Growth rate CIT1 florescence
SOIL – Two wild yeast strains in two glucose concentrations ****	2 genotypes X 2 environments X 24 wells per plate X 2 plates	Growth rate Growth-rate deviation

*Experiments also contained an additional environment (12 wells per plate) of base media (containing no added glucose) in which we variably observed cells growing very slowly or not at all. Data from these wells were excluded from the analysis.

**In addition to strains reported in the main text, we studied a strain that we understood to be DBVPG6765, but we later suspected to be an additional Pennsylvanian Oak strain (possibly YPS129) based on an oak-like phenotypic profile and adjacent well positions in a frozen stock plate. Data for this strain were included in the analysis (i.e., contributed to normalization) but not presented.

***These experiments only contain 24 wells per plate due to long exposure time in the fluorescent channel. Experiments also contained a base media environment with no added glucose. In contrast to the LWO dataset, the CIT1-GFP strain consistently grew slowly in the base media. Data for these wells were included in the analysis but not presented.
****This dataset was used to assess the reproducibility of growth rate deviation estimates (see below).

2.7.2: Fixed-effects estimation

Each model contains coefficients for each level of the fixed-effect terms. To estimate a specific genotype-environment combination, the relevant parameters were combined. For example, the estimate for strain A in environment 1 would be calculated by adding 4 parameters (the regression intercept and the coefficients for Strain A, Environment 1 and the Strain A-Environment 1 interaction). The error of this estimate is a combination of the errors of the individual parameters. Specifically, the standard error is the square root of the summed parameter variances and twice each of the co-variances. For the example of strain A in environment 1, this calculation would include four variances and six co-variances. 95% confidence intervals were calculated as plus or minus 1.96 multiplied by the standard error. These estimates and confidence intervals are shown in **Figures 2.3-2.6** of the main paper. If data were transformed for modeling, as in the case of log transformations for lag duration and absolute growth rate deviation, estimates and intervals were transformed back to the original scale for presentation.

Some genotype-environment combinations were reproduced in different datasets. Six combinations were shared between the LWO and WILD datasets and four were shared between

the WILD and SOIL datasets. Although experiments for different datasets were conducted 6–12 months apart, the combined parameter estimates are highly reproducible (**Figure 2.S7**).

2.7.3: Random-effects estimation

Unlike fixed effects, random effects are not modeled with a parameter for each level. Instead, the variances of the random effects are parameters in the model. However, an estimate for each level of each random term can be calculated conditional on the model parameters. These conditional means are the best linear unbiased predictors (BLUPs). The assumption of normality of the random effects was confirmed by assessing quantile-quantile plots (**Figure 2.S8**). Confidence intervals on the BLUPs are calculated from the conditional variance-covariance matrices (posterior variances).

2.7.4: Adjusted pooled distributions

The fixed-effect terms represent the factors of interest (genotype and environment) whereas the random-effect terms capture various aspects of technical variation (focus, illumination, media preparation and cell preparation). Although our estimates of the fixed-effect parameters were sufficient for some analyses (e.g., comparison of average growth-rate deviation between the Netherlands and Finland soil strains), other analyses required distributions of measurements for each microcolony (e.g., correlation between CIT1 expression and growth rate within conditions). In the latter cases, the relevant random-effect conditional means were subtracted from each original measurement, creating adjusted values that could be pooled by condition across wells and plates. For example if a specific well designated A1.110316 contained strain A and

environment 1, four values would be subtracted from each original microcolony measurement (the effects for well A1.110316, plate 110316, strain A-plate 110316 interaction and environment 1-plate 110316 interaction). After this normalization, we find that the means of all wells with the same condition are nearly identical (**Figure 2.S9**). This approach is analogous to using linear mixed modeling to normalize data from cDNA microarray experiments (Wolfinger et al. 2001).

2.8: Supplemental figures

Table 2.S1: Assessment of alternative models relating growth rate to nutrient concentration

Model	Strain	Parameters			Fit		
		μ_{max}	K_s		AIC	BIC	
$\mu = \frac{\mu_{max}\ s}{K_s + s}$	Lab	0.433	0.202		-161830.7	-161804.7	
	Oak	0.519	0.115		-103657.9	-103633.2	
	Vineyard	0.483	0.189		-169043.2	-169017.2	
	Oak/Vineyard F1	0.520	0.143		-150138.7	-150113	
		μ_{max}	K_s	a	AIC	BIC	P-value*
$\mu = \frac{\mu_{max}\ (s - a)}{K_s + s - a}$	Lab	0.433	0.202	-0.0001	-161828.8	-161794.2	0.729
	Oak	0.515	0.101	0.011	-104353.7	-104320.7	< 2.2e-16
	Vineyard	0.478	0.168	0.0149	-170171.2	-170136.5	< 2.2e-16
	Oak/Vineyard F1	0.518	0.135	0.005	-150355.7	-150321.4	< 2.2e-16
		μ_{max}	K_s	a	AIC	BIC	P-value*
$\mu = \frac{\mu_{max}\ s - K_s\ a}{K_s + s}$	Lab	0.433	0.202	-0.0003	-161828.8	-161794.2	0.729
	Oak	0.515	0.090	0.062	-104353.7	-104320.7	< 2.2e-16
	Vineyard	0.478	0.153	0.046	-170171.2	-170136.5	< 2.2e-16
	Oak/Vineyard F1	0.518	0.130	0.022	-150355.7	-150321.4	< 2.2e-16
		μ_{max}	K_s	a	AIC	BIC	P-value*
$\mu = \frac{\mu_{max}\ s - K_s\ a - s\ a}{K_s + s}$	Lab	0.433	0.202	-0.0003	-161828.8	-161794.2	0.729
	Oak	0.577	0.090	0.062	-104353.7	-104320.7	< 2.2e-16
	Vineyard	0.524	0.153	0.046	-170171.2	-170136.5	< 2.2e-16
	Oak/Vineyard F1	0.540	0.130	0.022	-150355.7	-150321.4	< 2.2e-16
		a	b		AIC	BIC	P-value*
$\mu = a + b\ \ln(s)$	Lab	0.333	0.073		-142529.7	-142503.7	NA (models not nested)
	Oak	0.434	0.076		-80730.29	-80705.55	
	Vineyard	0.376	0.080		-139222.8	-139196.8	
	Oak/Vineyard F1	0.424	0.081		-119618.4	-119592.7	

* Likelihood-ratio test comparing fit of alternative model to that of first (basic Monod) model.

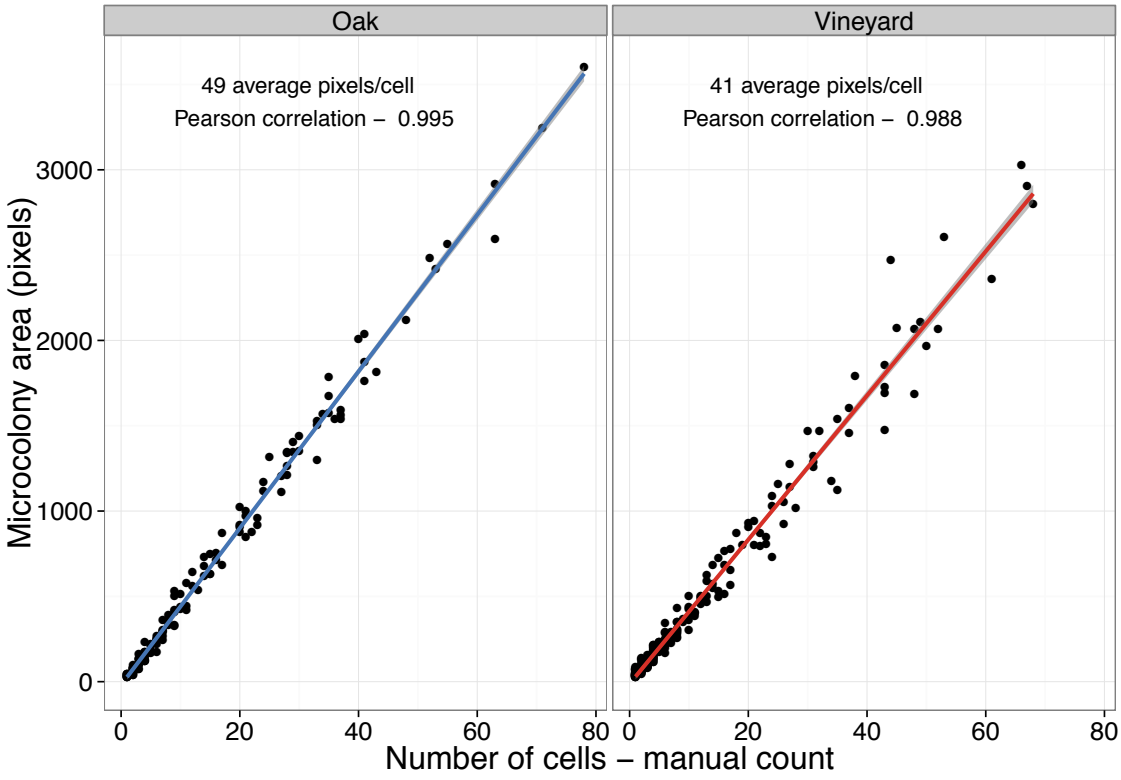


Figure 2.S1: Microcolony area is correlated with cell number for different strains growing in low glucose conditions

The number of cells within each microcolony was counted at each time point for ten microcolonies for both the oak (left panel) and vineyard (right panel) strains growing in 0.22 mM glucose. The average number of pixels per cell and Pearson correlation coefficient between the manual cell counts and automated area estimations are indicated in the upper left corner of each panel.

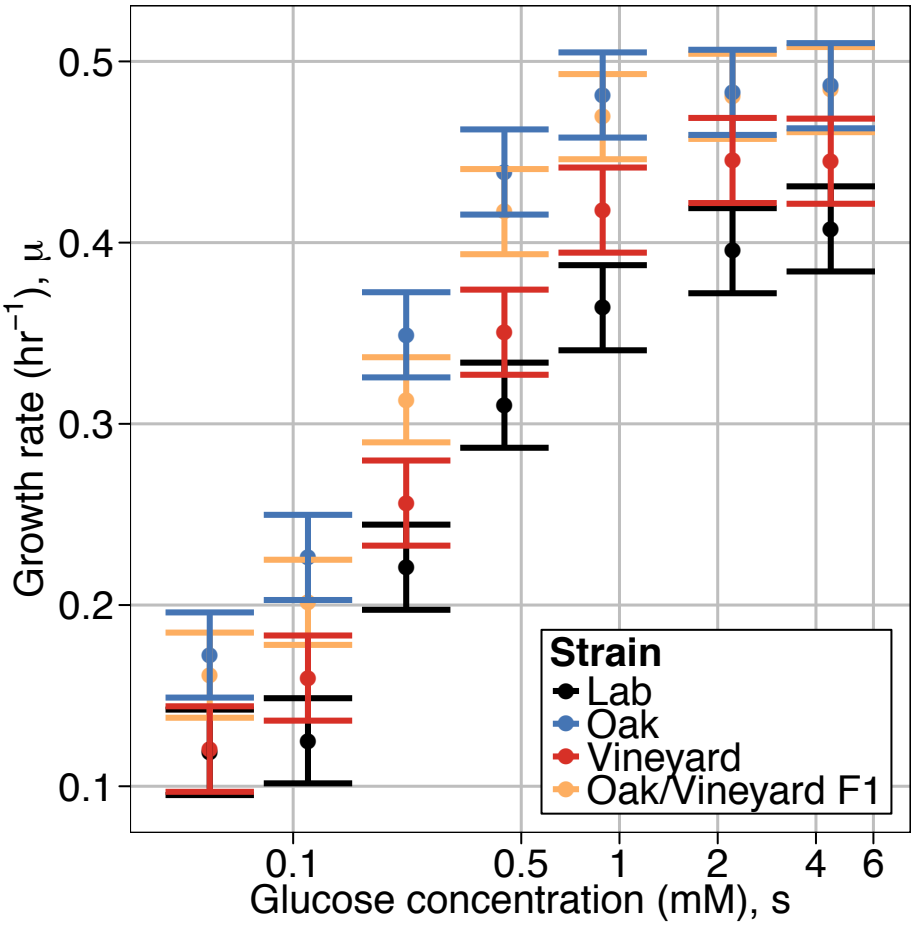


Figure 2.S2: Growth rate is determined by glucose concentration in a genotype specific manner

Growth rate estimates for each genotype and environment combination based on mixed effect modeling. Estimates are calculated as a combination of relevant fixed effect parameters, estimated using restricted maximum likelihood (REML) (see supplementary note). Glucose concentration is shown on a logarithmic scale for clarity. Error bars represent 95% confidence intervals.

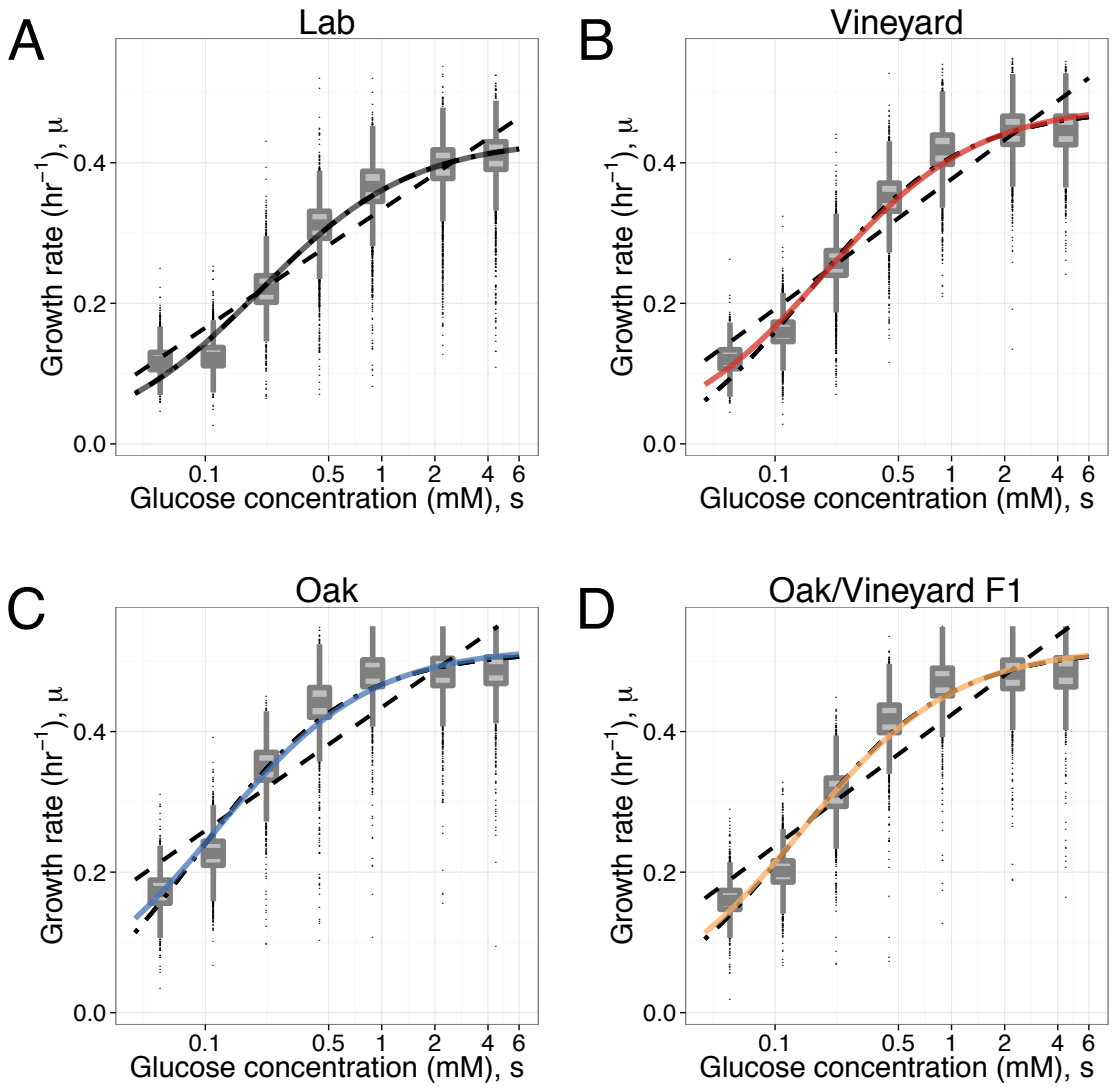


Figure 2.S3: Growth rate dependence on glucose concentration fits the Monod model for different strains

Growth-rate distributions for four strains (A-Lab, B-Vineyard, C-Oak and D-Oak/Vineyard F1) over a range of glucose concentrations. Solid curves depict the best fit of the Monod equation to the normalized data. Fits of the Westerhoff model are shown as dashed curves. Three variants of the Monod model with an additional parameter give virtually indistinguishable fits and are shown as a single dot-dashed curve in each panel. The glucose concentration is shown on a logarithmic scale for clarity.

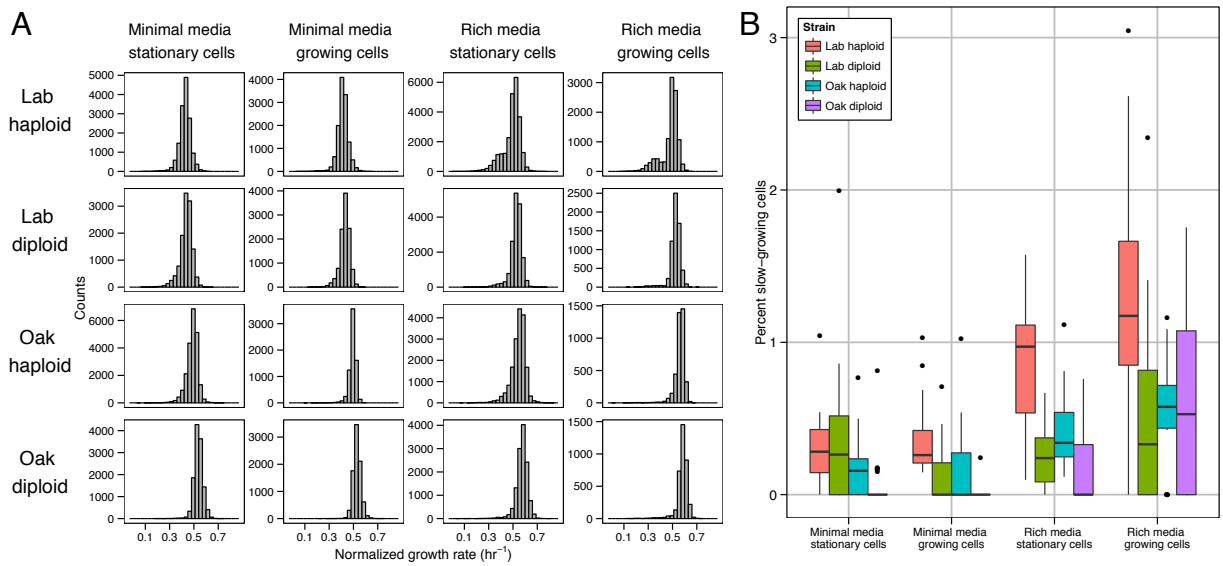


Figure 2.S4: Factors affecting the shape of the growth-rate distribution

(A) Growth-rate distributions for four strains (haploid or diploid cells in the lab or oak genetic background) in four media/growth history conditions. Minimal medium is chemically defined carbon-limiting media with 0.08% glucose and rich medium is synthetic complete (SC) with 2% glucose. Cells were plated for imaging either from stationary cultures (~48 hours after 1:300 dilution) or actively growing cultures (4 hours growth after 1:50 dilution of the stationary culture). Distributions are from data pooled across four replicate plates, each with six wells per strain/condition combination, after normalization using mixed modeling. (B) Box-plots of percent slow-growing cells, calculated as previously (Levy et al. 2012) as the percentage of cells growing at less than half the median for each well.

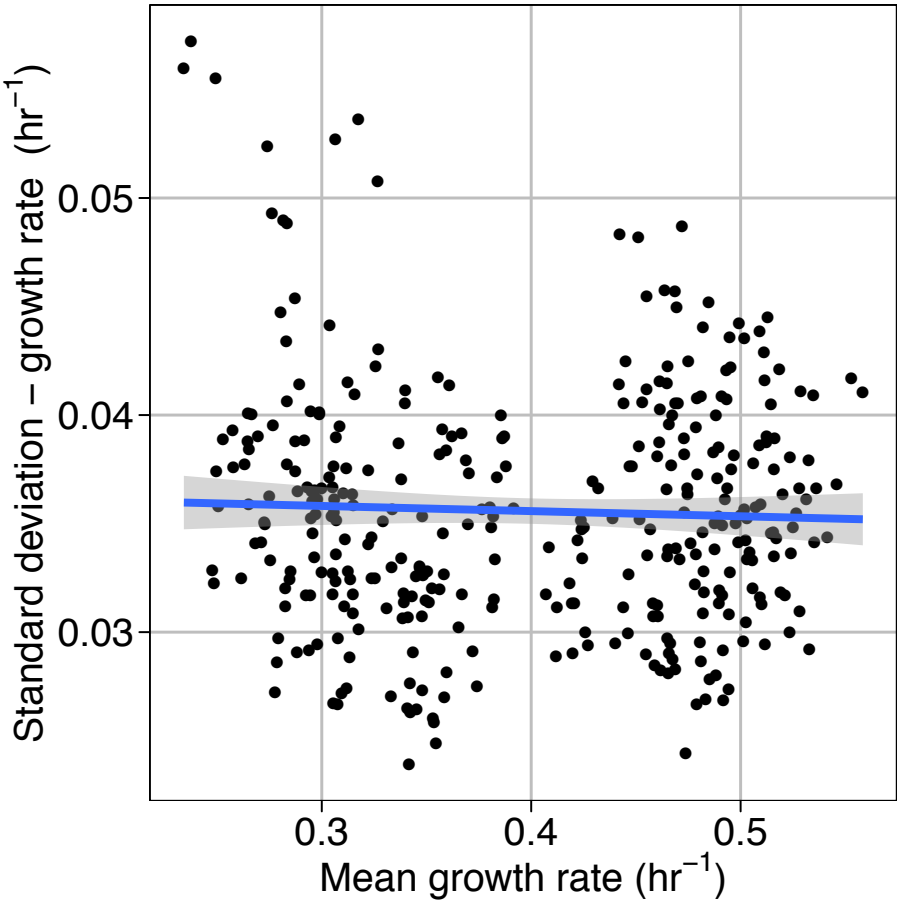


Figure 2.S5: Absence of mean-variance correlation for growth rate distributions
Standard deviations verses means for 352 growth rate distributions representing eleven genotypes and two environments. The line depicts linear least squares regression.

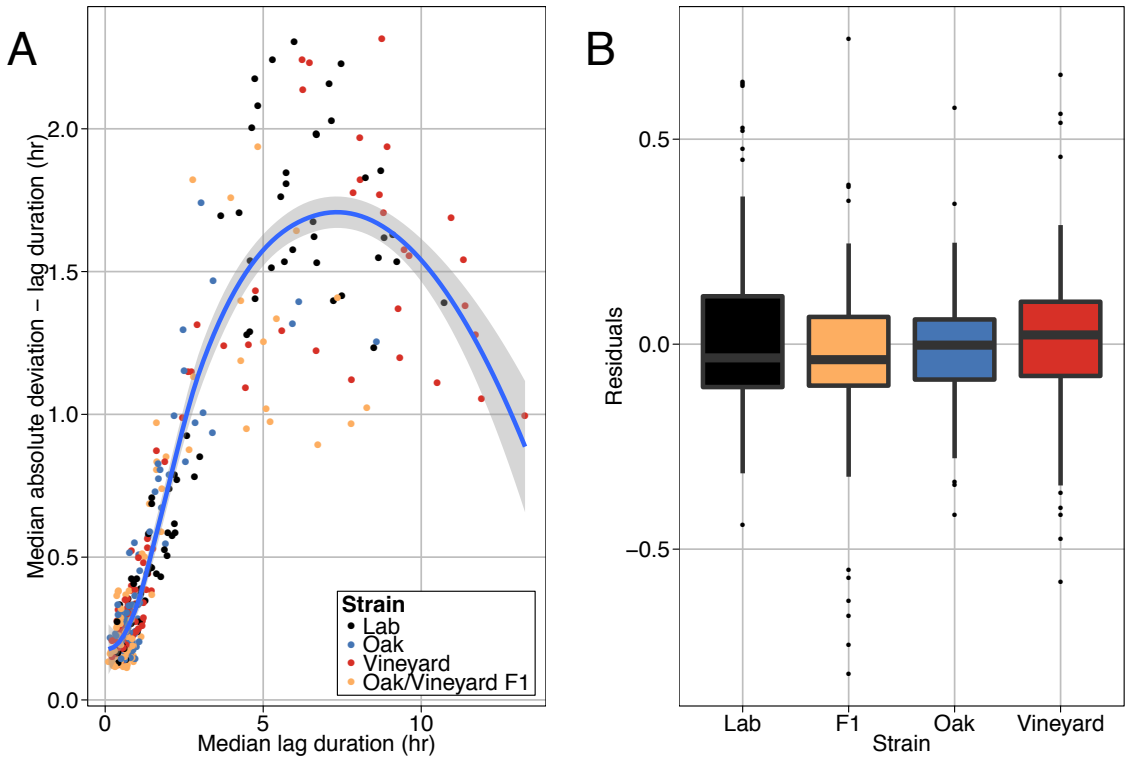


Figure 2.S6: No median independent genotype specific difference in lag duration variation
(A) Median absolute deviations (MAD) verses medians for 333 lag duration distributions representing four genotypes and seven environments. The line depicts a local loess regression. (B) Distributions of residuals from loess regression shown in A, grouped by genotype. There is no significant difference between genotypes.

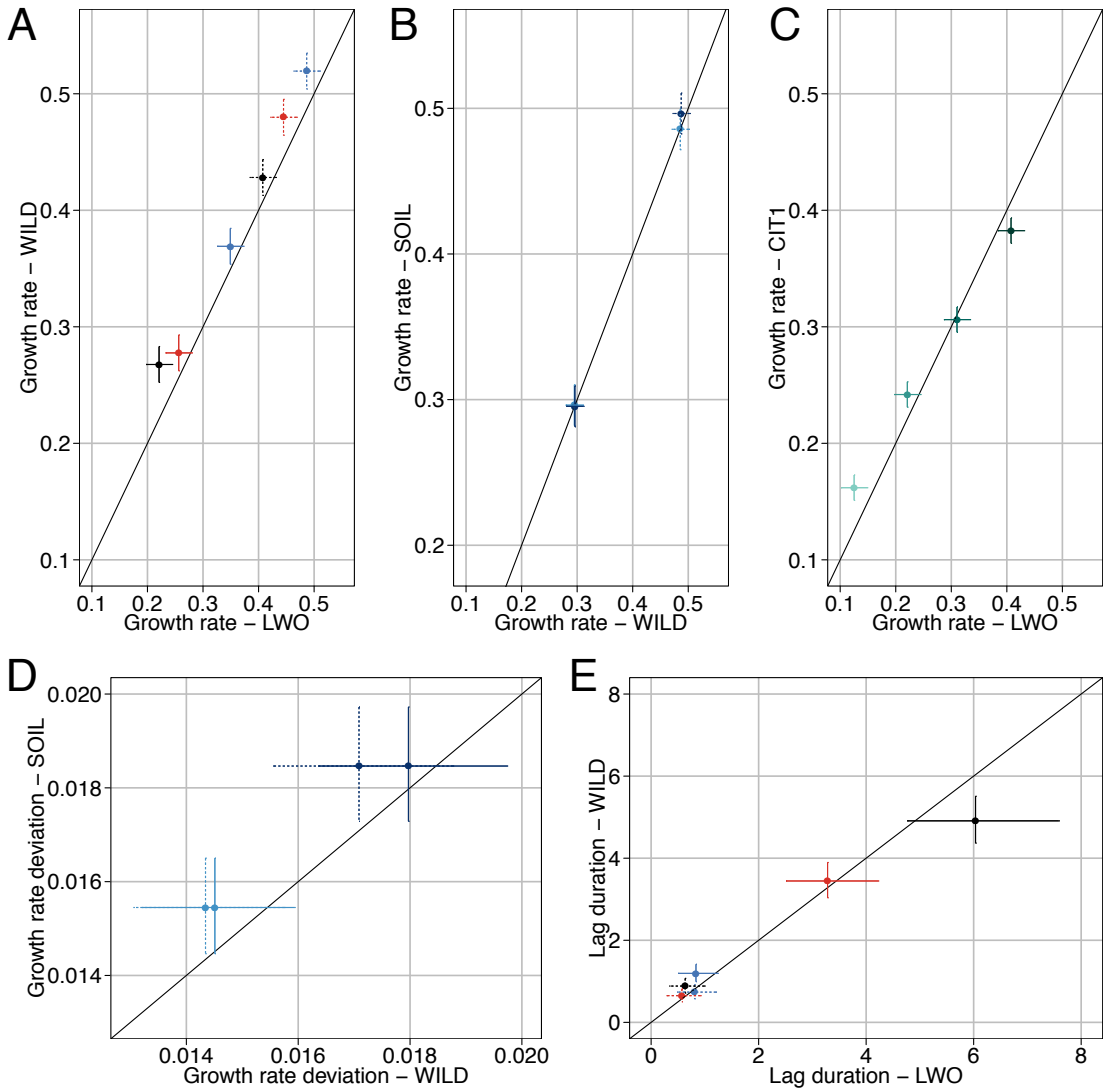


Figure 2.S7: Reproducibility of fixed effect parameters

For growth rate (A-C), growth rate deviation (D) and lag duration (E). In all panels, error bars represent 95% confidence intervals and the solid black line indicates x=y. (A and E) Comparison of growth rate (A) or lag duration (E) estimates for 6 genotype-environment combinations included in both the LWO and WILD datasets. The line type depicts glucose concentration (solid – 0.22mM, dashed – 4.44mM) and color depicts genotype (blue – Oak, red – Vineyard, black – Lab). (B and D) Comparison of growth rate (B) or growth rate deviation (D) estimates for 4 genotype-environment combinations included in both the WILD and SOIL datasets. The line

type depicts glucose concentration (solid – 0.22mM, dashed – 4.44mM) and color depicts genotype (light blue – Finland, dark blue – Netherlands). (C) Comparison of growth rate estimates for 4 similar genotype-environment combinations between the LWO and CIT1 datasets. While the environments compared are identical, the genotype in the two datasets differs slightly: the CIT1-GFP strain differs from the lab strain used in the LWO experiments, as it is heterozygous for the CIT1-GFP fusion and a number of auxotrophies. Color depicts glucose concentration (from light to dark green – 0.11, 0.22, 0.44 and 4.44 mM).

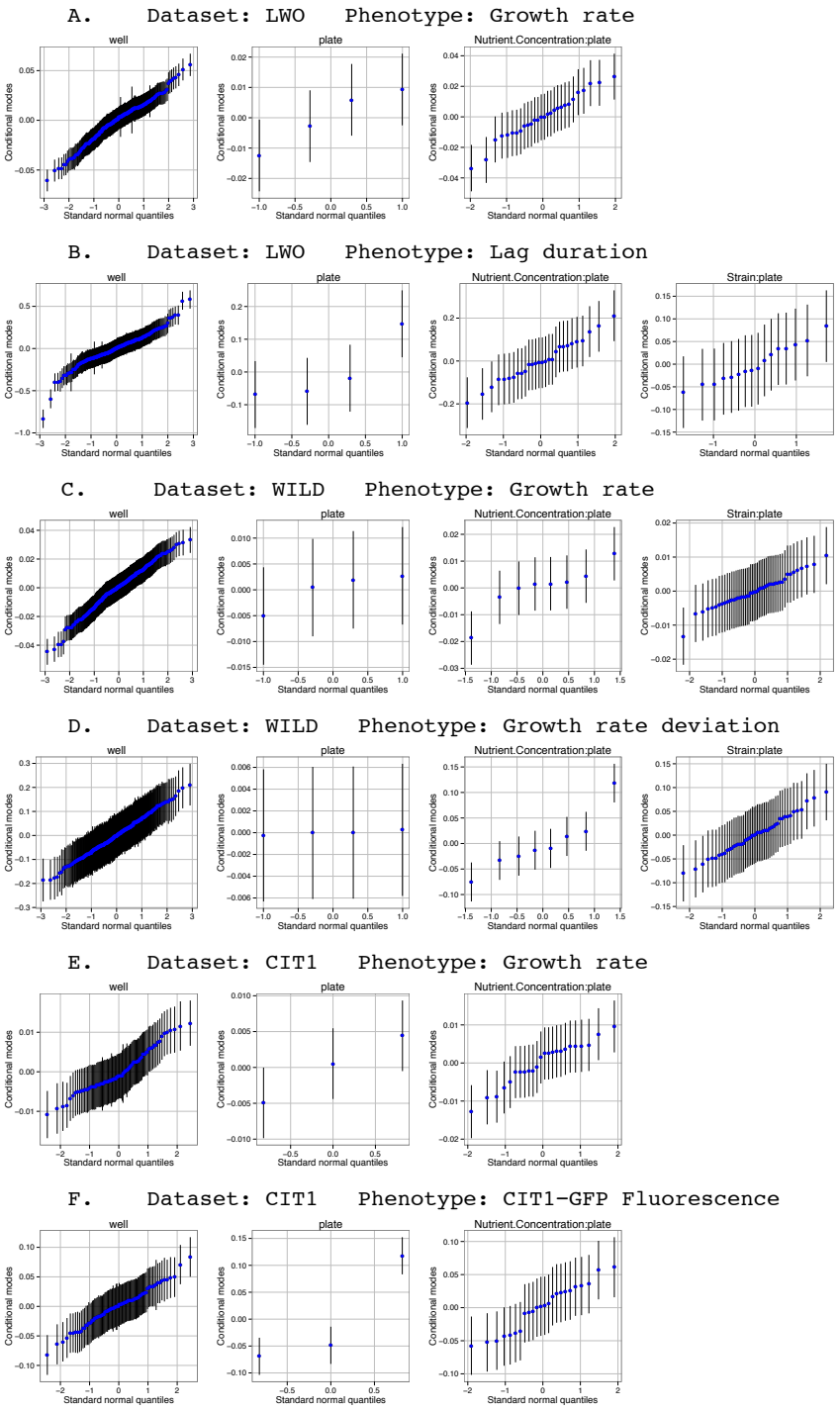


Figure 2.S8: Random effect conditional means

Quantile-quantile plots comparing random effect conditional means for each random effect term in 6 distinct mixed models to standard normal quantiles. Panels represent models of growth rate

(A, C and E), lag duration (B), growth rate deviation (D) and CIT1-GFP fluorescence (F). Panels also represent the three main datasets: LWO (A and B), WILD (C and D) and CIT1 (E and F). Vertical columns represent different random effect terms, from left to right: well, plate, environment-plate interaction and genotype-plate interaction (not present in all models). Blue dots represent estimates and black lines represent 95% confidence intervals.

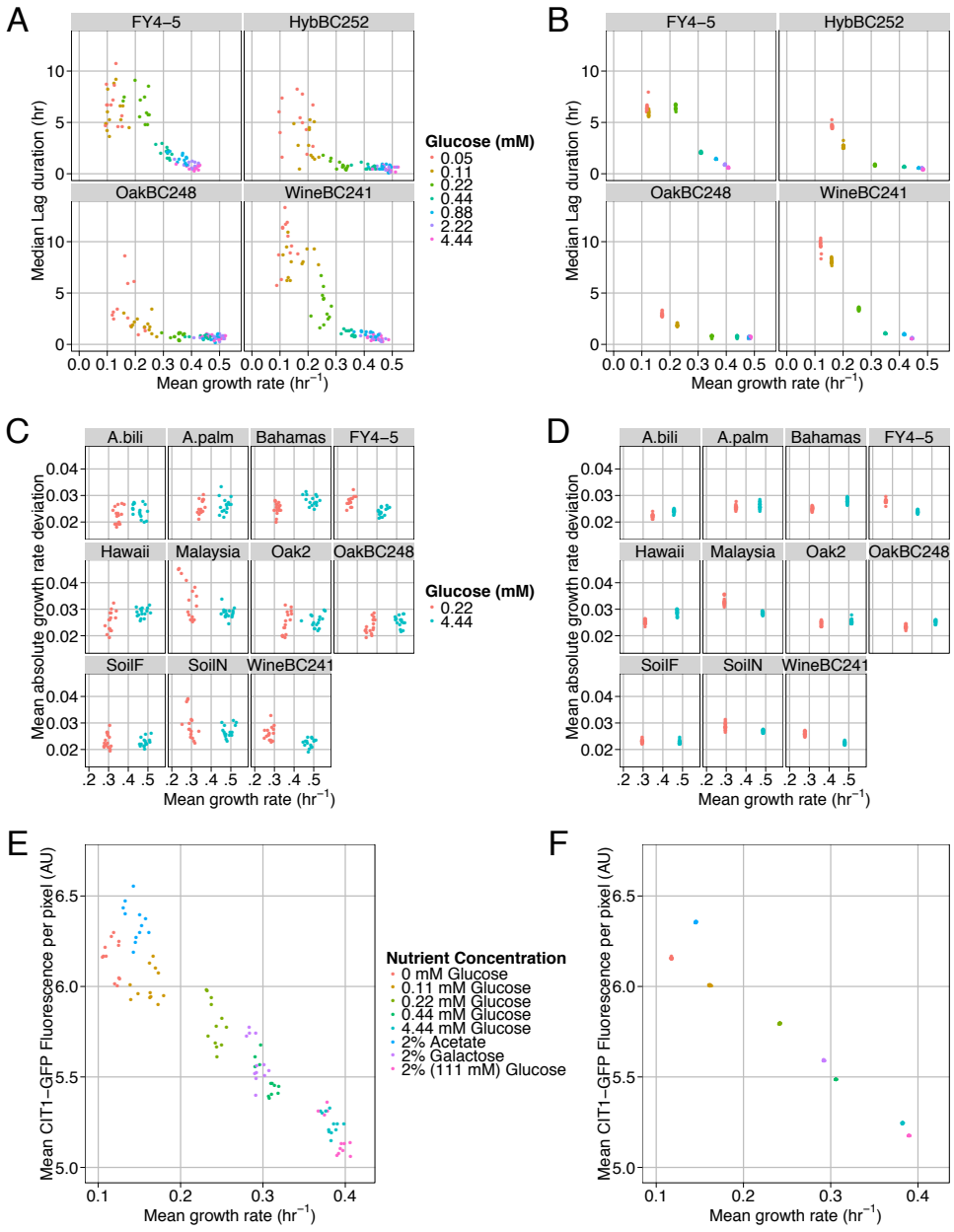


Figure 2.S9: Consequences of normalization by subtraction of random effect conditional means

In all panels, a single dot corresponds to mean or median estimates of distributions originating from a single experimental well. Left panels (A, C and E) depict original estimates before adjustment; right side panels (B, D and F) depict corresponding estimates after adjustment. The

panels represent estimates of growth rate verses lag duration in the LWO dataset (A and B), growth rate verses growth rate deviation in the WILD dataset (C and D) and growth rate verses CIT1-GFP fluorescence in the CIT1 dataset (E and F). In some panels (A-D), estimates from the same genotype are grouped and represented as different labeled facets. In all panels colors correspond to different environments, which are specified in a legend between each pair of panels. For lag duration and growth rate deviation, estimates are presented on the original scale despite modeling log-transformed distributions, resulting in greater spread between same-condition estimates.

2.9: Correction

We plan to publish a correction to the preceding chapter as it appeared in the journal *Molecular Biology and Evolution*. The correction will state that the strain assumed to be DGVPG1373 and regarded to as the "Dutch soil strain" or "Soil-Netherlands" strain in the text and figures is now believed to be a strain from the Malaysian lineage (Liti et al. 2009).

CHAPTER 3: GENE-GENE AND GENE-ENVIRIONMENT INTERACTIONS UNDERLIE VARIATION IN CELL GROWTH

3.1: Abstract

The growth of microbial populations is characterized by a lag phase (the time until the culture initiates growth) and a growth phase (characterized by an exponential growth rate). These parameters can vary both between environmental conditions and between single cells in a given environment. We used a high-throughput microscopy assay, which enables characterizing growth of hundreds of thousands of individuals, to map genetic loci determining variation in lag and growth rate distributions in distinct glucose concentrations, using natural isolates of the budding yeast (*Saccharomyces cerevisiae*). Some quantitative trait loci (QTL) are shared between traits or environments while some are unique, exhibiting gene by environment interactions. Furthermore, whereas variation in the central tendency (mean growth rate or median lag duration) is explained by many additive loci, variation in phenotypic variability can be explained by genetic interactions. We used an analogous mapping strategy to increase QTL resolution, consisting of bulk segregant analysis of complex mixtures of genotypes under selection, utilizing whole genome sequencing and growth in chemostats. We find that sequence variation in the high affinity glucose transporter HXT7 contributes to variation in growth rate and lag duration. Allele replacements of the entire locus as well as a single amino acid reveal the dependence of variation in HXT7 on the genetic background and inter-locus context. Our study highlights the complex nature of genotype to phenotype mapping across environments even in model organisms.

3.2: Introduction

Variation in continuously distributed traits is a consequence of variation in multiple genetic and environmental factors. Genetic mapping has been successful in locating loci important for specific traits and even dissected to single nucleotide resolution in budding yeast (Gerke et al. 2009). As more studies are performed, it is becoming clear that most associated loci are context dependent (Mackay et al. 2009; Liti and Louis 2012), their effect depending on variation at other loci or in the environment.

An additional factor effecting variation in complex traits is phenotypic variability (Geiler-Samerotte et al. 2013): the variation between individuals of identical genotypes in the same environment. Despite increased appreciation of the importance of phenotypic variability, many questions remain regarding its genetic and molecular basis. Specifically, it remains unclear the extent to which loci determining phenotypic variability and loci determining average trait values across different environments overlap.

Complex networks of interacting genetic and environmental factors regulate cell growth in microbes and multicellular organisms, making it an ideal system to dissect the genetic basis of complex traits. The rate at which a cell grows is the result of myriad cellular processes including nutrient sensing and transport, signal transduction, macromolecular synthesis and metabolism. In microbes, growth can be separated into distinct phases, specifically lag phase, a period of adaptation in which cells do not grow and exponential phase in which cells grow at a constant rate (Monod 1949). Although correlated, the phases vary independently in natural populations

between different conditions (Cubillos et al. 2011; Ziv, Siegal, et al. 2013). Our recently developed high-throughput microcolony growth rate assay (Levy et al. 2012; Ziv, Siegal, et al. 2013) permits accurate estimation of variability within each of these phases. Variation in cell growth is important from an evolutionary perspective as a major component of fitness in microbes (Blomberg 2011). Dissecting the genetic basis of cell growth variation in ecologically relevant environments may illuminate the prevalence of adaptation in natural populations.

Budding yeast is a tractable model for the analysis of complex traits (Liti and Louis 2012). Genetic mapping of complex traits involves detecting and localizing quantitative trait loci (QTL). Detecting QTL ultimately relies on associating variation in genotype and phenotype in a mapping population. The detection of QTL depends on the effect and frequency of alleles while localization depends on the frequency of recombination (Mackay et al. 2009). Individual segregant analysis involves genotyping and phenotyping individual segregants and searching for associations (Steinmetz et al. 2002; Gerke et al. 2009; Cubillos et al. 2011; Bloom et al. 2013). Alternatively, bulk segregant analysis involves selecting a portion of the population based on extreme trait values and looking for a deviance in allele frequency from the entire population (Michelmore et al. 1991; Ehrenreich et al. 2010; Swinnen et al. 2012). The advantage of bulk segregant mapping is the increased power to detect loci with small effect due to increased sample size (Ehrenreich et al. 2010). However, only analysis of individuals can identify genetic interactions (Wilkening et al. 2014) which may be important for explaining trait variance. Localization of QTL can be improved by using an advanced intercross population, as increased recombination breaks up linkage (Darvasi and Soller 1995; Illingworth et al. 2013). Advanced

intercross populations have been used successfully in yeast in combination with bulk segregant analysis (Parts et al. 2011; Cubillos et al. 2013).

In this study, we dissected the genetic architecture of cell growth using a combination of classical interval mapping and sequencing under selection. We focused on comparing the genetic architecture between two related and ecologically relevant environments, specifically, growth rate limiting (0.22mM) and non-limiting (4.44mM) glucose conditions. We decomposed cell growth by quantifying growth rate and lag duration distributions, mapping loci determining both the central tendency and variability of the traits. In addition to individual F2 segregant analysis that allowed us to discover genetic interactions, we used an advanced intercross population and bulk segregant analysis to increase the mapping resolution for additive QTL. Allele replacements confirm the effect of the candidate gene HXT7 but also reveal additional complexity.

3.3: Results

3.3.1: Distinct genetic architectures determine growth rate and lag duration distributions across environments

Genetic mapping of complex traits is seldom done in separate but closely related environments. We have previously shown that wild yeast isolates, specifically an oak strain (BC248; hereafter "oak") and a vineyard strain (BC241; hereafter "vineyard") differ in their growth rate and lag duration distributions across glucose concentrations (Ziv, Siegal, et al. 2013). Oak cells grow faster and lag for shorter amounts of time than vineyard cells. The difference in the response to increasing nutrient concentration can be characterized by growth in two conditions, 0.22mM

("growth limiting") and 4.44mM ("non-limiting") glucose. The higher glucose concentration results in maximum growth rates, despite being over an order of magnitude lower than standard lab media (Ziv, Siegal, et al. 2013). In order to identify quantitative trait loci (QTL), we used a panel of 374 recombinant segregants (Gerke et al. 2006) genotyped at 225 loci throughout the genome (Gerke et al. 2009). Each segregant was phenotyped in both conditions using a high-throughput microscopy based microcolony assay (Levy et al. 2012; Ziv, Siegal, et al. 2013).

The microcolony assay allows us to estimate for each strain, the central tendency and dispersion of growth rate and lag duration distributions. Growth rate distributions for wild isolates in both glucose concentrations are approximately normal and were characterized by their mean and standard deviation. On the other hand lag duration distributions tend to be right tailed and were characterized by their median and median absolute deviation (MAD). We disregarded lag duration distributions in the higher glucose concentration as most cells commence growth in less than one hour and thus do not have detectable lag times (Ziv, Siegal, et al. 2013). This analysis resulted in six traits amenable to genetic mapping. The segregant data within conditions recapitulated correlations observed between conditions in our previous study (Ziv, Siegal, et al. 2013). Specifically, there exists a strong positive correlation between lag duration median and lag MAD, weak correlations between growth rate mean and growth rate standard deviation and a negative correlation between mean growth rate and median lag duration (**Figure 3.S1**). There is also a positive correlation between mean growth rate in the limiting and non-limiting glucose concentrations. In order to obtain variability estimates that are independent of average trait

values, we performed loess regressions and used the residuals as trait values for mapping (**methods**).

Quantitative trait loci (QTL) are identified as statistical associations between genotype identity and trait variation. We used the R package RQTL (Broman et al. 2003) for mapping of QTL (**methods**). We first searched for additive QTL (**methods**) and found multiple QTL for most traits (**Figure 3.1, Figure 3.S2**). For mean growth rate and median lag duration, we found 5-9 QTL explaining 32-58 percent of trait variance using an additive model (**Figure 3.1, Figure 3.4**). As expected based on trait correlations, some QTL were shared, however each trait also had unique QTL (**Figure 3.1, Figure 3.S2, Table 3.S1**). The effect of each QTL was estimated and ranged between 4-23 percent of the difference in parental phenotypes (**Figure 3.S3**). We identified 1-2 transgressive QTL (where the vineyard allele is predicted to increase growth rate or decrease lag duration) for each trait (**Figure 3.1, Figure 3.S3**). In contrast, only 0-2 additive QTL, explaining 0-14 percent of trait variance were found for variability traits (**Figure 3.1, Figure 3.4**). Two out of the three QTL found for variability were also found for central tendency (**Figure 3.1, Table 3.S1**). In particular, a single locus on chromosome IV had an additive effect on four traits, growth rate central tendency in both glucose concentrations and lag duration central tendency and variability.

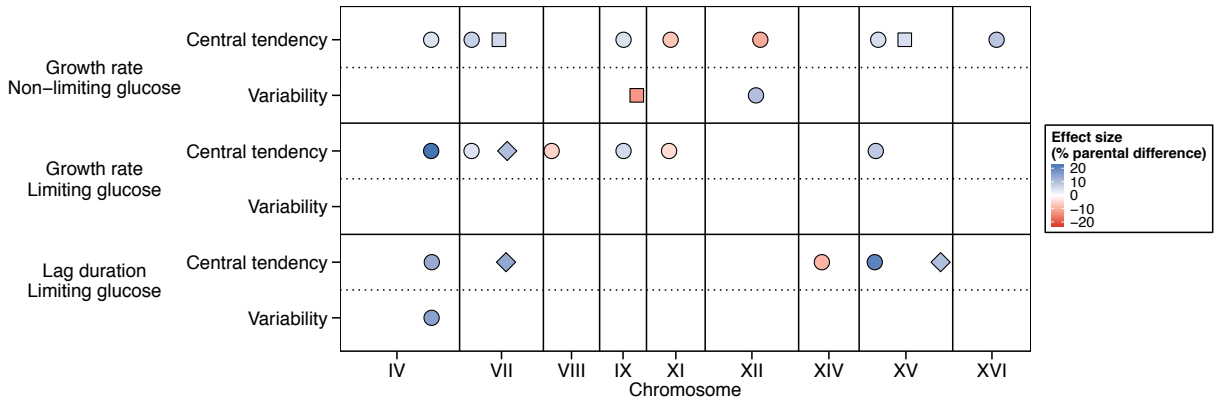


Figure 3.1: Multiple QTL underlie variation in cell growth

Additive QTL found for all traits. Chromosome size and QTL position correspond to genetic distance (measured in cM). Colors depict estimated effect sizes given the full additive QTL model. Different shapes are used to depict distinct QTL on the same chromosome. Positive effects correspond to oak alleles increasing growth rate traits or decreasing lag duration traits while negative effect sizes correspond to vineyard alleles increasing growth rate traits or decreasing lag duration traits.

We tested for genetic interactions using two-dimensional genomic scans (**methods**). We identified a total of 8 significant interactions across all traits (**Figure 3.2**). There was some overlap between loci found for different traits. A strong interaction, which was found between positions on chromosome I and chromosome X, was shared between two traits, growth rate variability and lag duration variability in the growth-limiting glucose concentration (**Figure 3.2, Figure 3.3**). Additionally, three loci found for lag duration variability, on chromosomes IX, XII and XV had effects on other traits. The chromosome XII position was found as part of an interaction effecting growth rate central tendency in the limiting glucose concentration, while the other two loci had additive effects on growth rate central tendency or all central tendency traits respectively (**Figure 3.1, Figure 3.2**). The effect size of interacting loci (given the genotype of

the interaction partner) was comparable to the effect size of additive loci (**Figure 3.S3**). It is interesting to note that most of the interactions were found for variability traits and were responsible for a large proportion of explained trait variance in the limiting glucose condition (**Figure 3.4**).

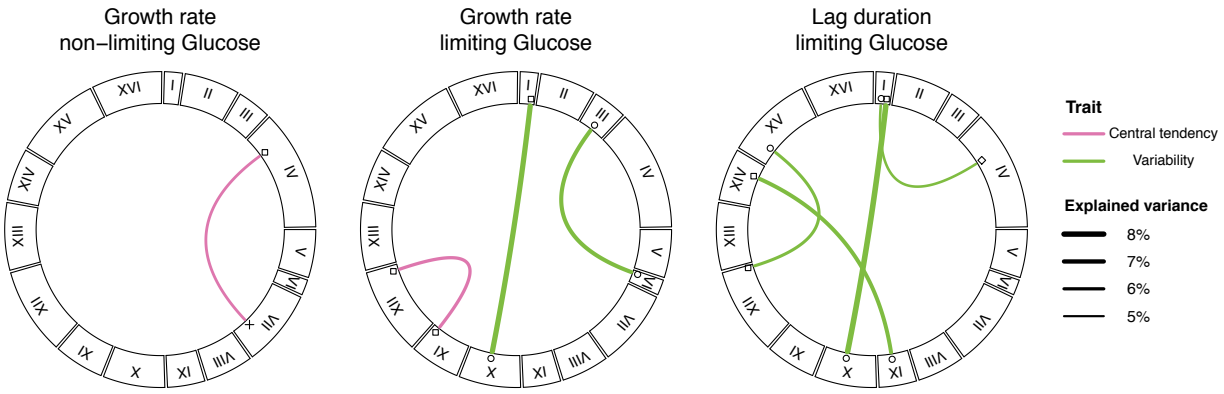


Figure 3.2: Genetic interactions underlie variation in cell growth variability

Genetic interactions found for all traits. Chromosome size and QTL position correspond to genetic distance (measured in cM). Line width corresponds to the percent trait variance explained when only the two interacting loci are modeled. Colors depict trait type (pink – central tendency, green – variability). Different shapes are used to depict distinct loci on the same chromosome.

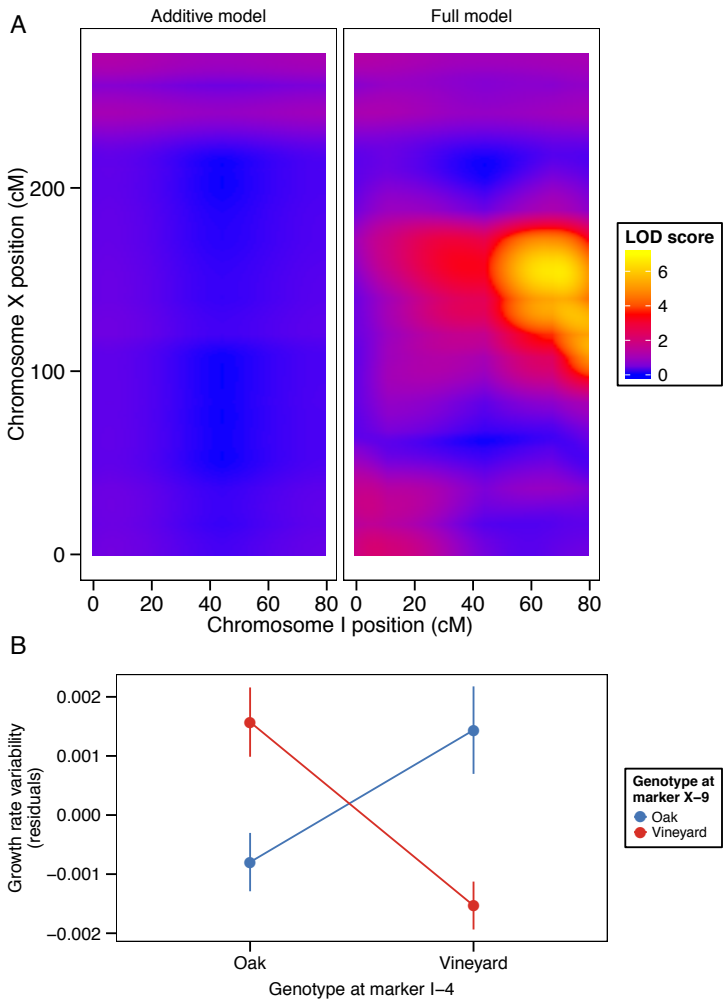


Figure 3.3: Example of a strong genetic interaction determining variation in growth rate variability in limiting glucose

(A) LOD scores corresponding to additive (two QTL) or full (two QTL and interaction) models for each combination of chromosome I and X positions, for growth rate variability in the limiting glucose condition. (B) Mean growth rate variability in the limiting glucose condition for the four genotype combinations corresponding to two closest markers to the maximum LOD score difference shown in (A), error bars represent standard errors.

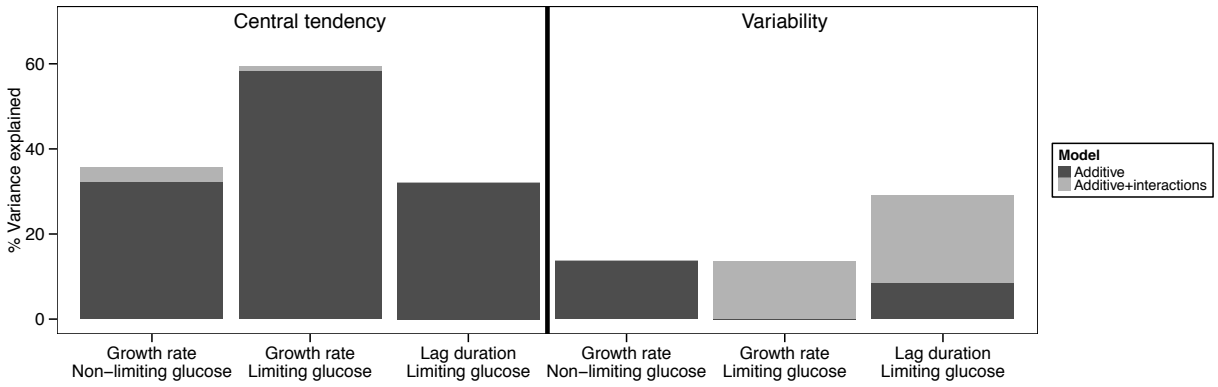


Figure 3.4: Variance in central tendency and variability traits are mainly explained by additive QTL or genetic interactions respectively.

Percent trait variance explained per trait by a model of all identified additive QTL (black) or all identified additive QTL and genetic interactions (grey).

3.3.2: An advanced intercross population and sequencing under selection increases mapping resolution

One of the challenges of QTL mapping is identifying the relevant gene and causative variation. We sought to improve the resolution of QTL mapping by using a complementary mapping approach. We used a variant of bulk segregant mapping in which an advanced intercross population is subjected to selection for a desired trait and sequenced (Parts et al. 2011).

We created an advanced intercross population for the oak/vineyard cross by repeated rounds of sporulation and mating (**methods**). Allele frequencies and linkage were characterized by sequencing the final population and three isolated clones. Allele frequencies in the final population deviated from the expected 0.5 in a number of genomic loci (**Figure 3.S4**). A relevant source of the deviation is selection pressure during the creation of the population due to variation

in sporulation efficiency and indeed we identified two of the three major sporulation QTL known to segregate in this cross (Gerke et al. 2009). Despite inadvertent selection, over 85% of single nucleotide polymorphisms still segregated with minor allele frequencies above 10%. Linkage in the intercrossed population was decreased compared with the F2 segregants as an average of 84.3 crossover events (60, 99 and 94) were identified in the three clones (**methods**), compared to an average of 31.9 (range: 13-55) in 374 F2 segregants. The increase in recombination frequency is consistent with the genetic map expansion observed in previous studies (Parts et al. 2011).

In order to enrich for QTL conferring increased growth rate, we grew the advanced intercross population in replicate glucose-limited chemostats with a low (D=0.18 hr^{-1}) or high (D=0.35 hr^{-1}) dilution rate. In order to directly assess the contribution of the advanced intercross population to QTL mapping, we also pooled the panel of F2 segregants and grew them in chemostats. We collected and sequenced multiple samples from each chemostat over 20-40 generations. We identified QTL by comparing the allele frequencies between early and late time-points using the MULTIPOOL software (Edwards and Gifford 2012) (**methods**) (**Figure 3.S5**).

We have previously shown that the effect of glucose concentration on the growth rate of the oak and vineyard strains can be recapitulated using glucose-limited chemostats (Ziv, Siegal, et al. 2013). However, this does not ensure that the same genomic loci are relevant when comparing growth in chemostats to the microcolony growth assay. In order to compare between the different experiments, we computed maximal LOD scores for each genomic interval flanked by markers used in the interval mapping (**Table 3.S3**). We define 209 such intervals, with a size

range of 14kb-139kb and a mean and median of 50kb. We find that both mapping approaches result in high LOD scores in the same regions of chromosomes IV (both glucose environments), VIII (only in low glucose) and XVI (only in high glucose) (**Table 3.S3**). In addition to these QTL, bulk segregant analysis of the panel of F2s also converged with the interval mapping, resulting in high LOD scores in some regions of chromosomes VII, XII and XV (**Table 3.S3**). The absence of a chromosome VII QTL when using the advanced intercross can be explained by strong selection in the same region during the creation of the advanced intercross (**Figure 3.S4**). One caveat is that the bulk segregant analysis of the F2 panel is potentially confounded by the small number of segregants in the population. In this case non-random associations between true QTL and unlinked loci may result in spurious signals. This is supported by the observation of regions with high LOD scores for the F2 pool not shared by the advanced intercross or the interval mapping of the F2 segregants (**Table 3.S3**).

The decreased linkage in the advanced intercross population had two consequences on QTL resolution. On the one hand, LOD scores decreased rapidly at individual QTL (**Figure 3.5A**). The size of 2-LOD drop intervals for the chromosome IV QTL decreased from more than 50kb and 235kb (for growth-limiting and non-limiting glucose respectively) using interval mapping of F2s to 9.3kb and 30.5kb using the advanced intercross. Most of the increased resolution was due to the use of the advanced intercross and not the bulk segregant approach, as intervals for the pooled F2s ranged between 22kb-78kb in low glucose and 66kb-203kb in high glucose based on choice of MULTIPOOL parameters (**methods**). *HXT6* and *HXT7*, which encode high affinity glucose transporters, lie within the region of increased resolution (**Figure 3.5A**). On the other

hand, the environment specific QTL on chromosome 8 appears to be composed of multiple linked QTL, resulting in a broad QTL region (**Figure 3.5B**). In this case, the single peak found in the F2 pool may be an example of a ghost QTL, where two linked QTL whose effects have the same sign give a maximum LOD score at a location in between the two loci (**Figure 3.5B**).

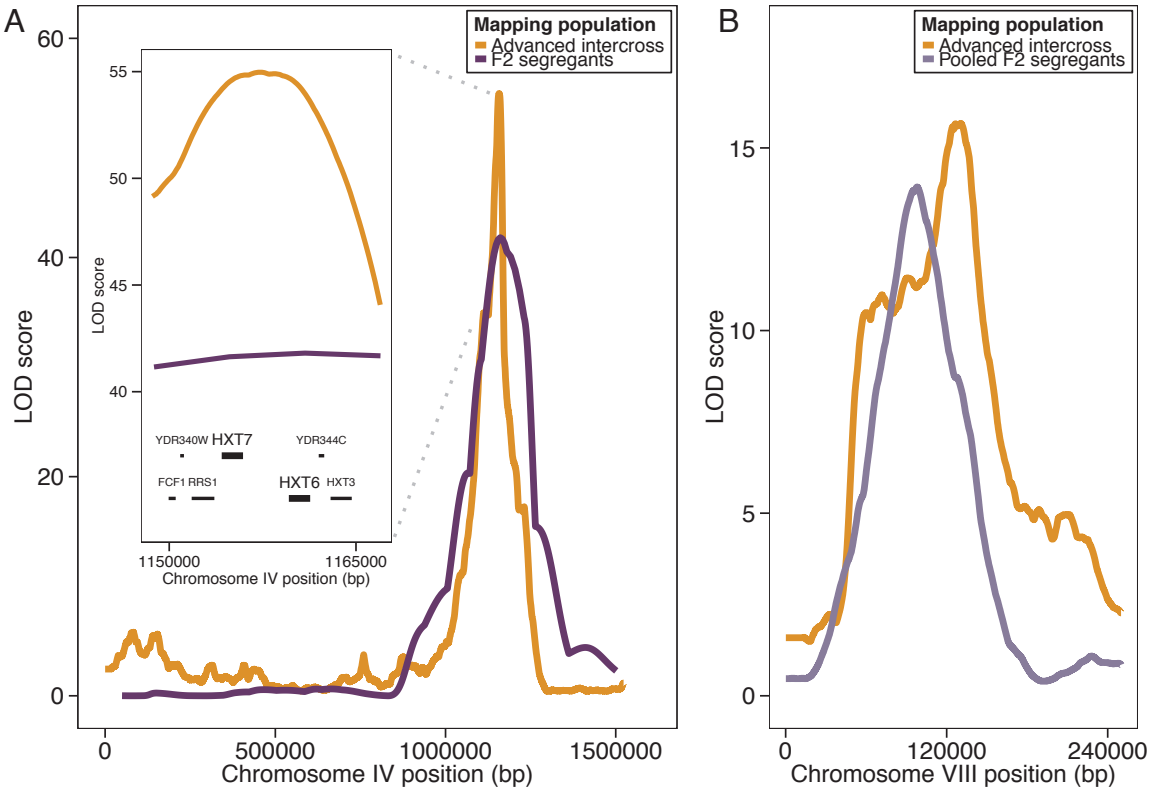


Figure 3.5: Increased QTL resolution due to decreased linkage in an advanced intercross population

(A) LOD score profiles for the entire chromosome IV obtained by interval mapping using F2 segregants (dark purple) or sequencing under selection of an advanced intercross population (orange). For the interval mapping profile, genetic distances were converted to physical distances based on marker positions. Inset shows profiles for 15kb directly under peaks, with gene positions. (B) LOD score profiles for a segment of chromosome VIII obtained by sequencing

under selection of a F2 pool (light purple) or an advanced intercross population (orange). MULTIPOOL parameters for both datasets are n=1000 and r=1000.

3.3.3: Sequence variation in HXT7 contributes to variation in growth

HXT6 and *HXT7* encode nearly identical high affinity glucose transporters, making them appealing candidate genes for the growth QTL on chromosome IV. Amplification of these two genes are frequently selected during carbon limiting experimental evolution (Brown et al. 1998; Gresham et al. 2008), however sequencing of the oak and vineyard strains did not reveal copy number variation. To assess the contribution of variation in *HXT6* and *HXT7* to variation in growth rate and lag duration, we created reciprocal hemizygote strains, in which one copy of *HXT6* or *HXT7* was knocked out in the F1 hybrid background. Phenotyping implicated *HXT7* and not *HXT6* in contributing to variation in growth (**Figure 3.S6**).

Analysis of the oak and vineyard *HXT7* alleles (**methods**) revealed 79 single nucleotide polymorphisms and 26 amino acid differences in the 1713 bp open reading frame (ORF). We created HXT7 allele replacement strains that contained no additional genetic modifications except the replaced allele (**methods**). Whole locus allele replacements replaced the entire ORF and 530 bp of upstream sequence (which contained 3 SNPs and 1 2bp indel). We also created single amino acid replacements by replacing a single amino acid (oak to vineyard T469Q) in each genetic background. This residue was chosen based on conservation and manual inspection of a structural homology model. The amino acid is positioned within the transporter channel and is predicted to interact with the sugar molecule. We also created strains homozygous for the oak or vineyard HXT7 allele in the F1 hybrid genetic background.

Phenotypic analysis revealed significant differences between the original and allele-replaced strains (**Figure 3.6**), consistent with the effects determined by linkage mapping. Of particular interest is the difference in the magnitude of the effect in the different genetic backgrounds. In the growth limiting glucose concentration, the strain containing the vineyard allele in the otherwise oak background grows slower and lags for a longer time than the original vineyard parent. In contrast, the oak allele in the vineyard background caused a small but significant increase in growth rate and decrease in lag duration. The single amino acid modification is only significant in the oak background and has a smaller effect than the whole locus replacement (**Figure 3.6**), suggesting additional loci within *HXT7* effect growth. Consistent with a smaller effect size in the non-limiting glucose concentration, only the whole locus replacement in the oak background had a significant effect on growth rate, causing a small decrease (**Figure 3.S7A**). No significant effects were found for lag duration variability (**Figure 3.S7B**). However, the small number of strains used in the regression to estimate median-independent lag variability reduces our power to detect a change (Levy and Siegal 2008).

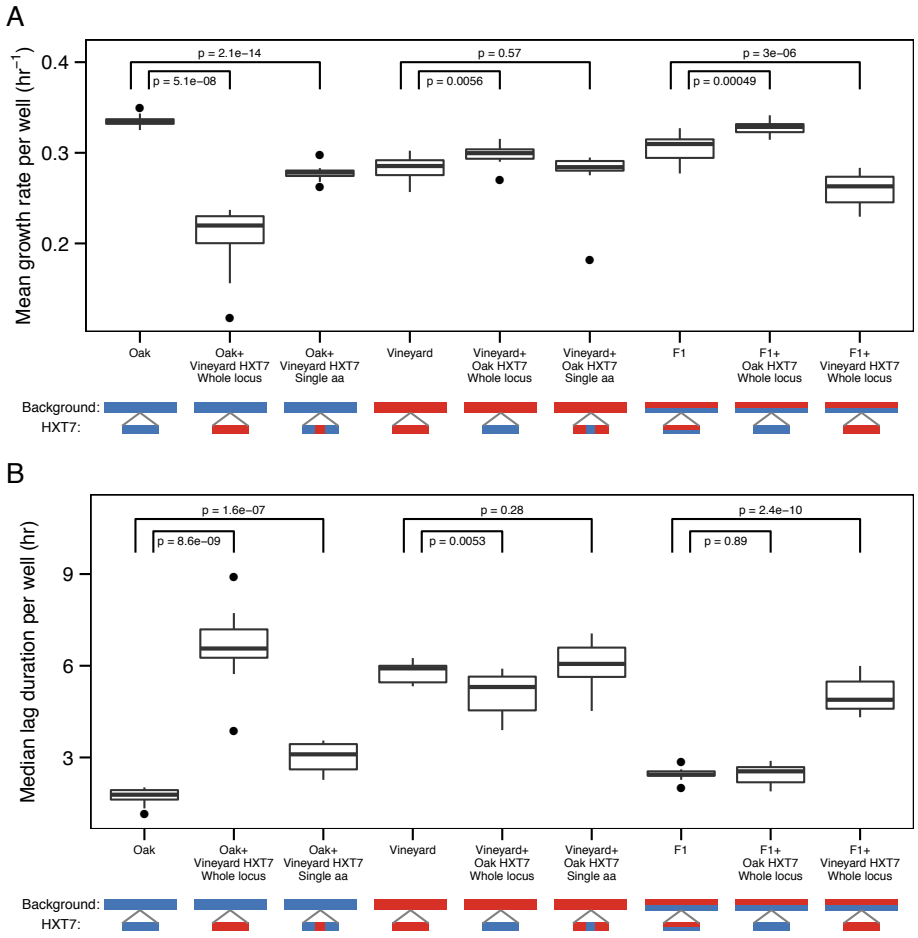


Figure 3.6: Effect of variation in HXT7

Distributions of mean growth rate (A) or median lag duration (B) for allele replacement strains grown in limiting glucose. P-values are for two sample t test, n=12 (4 wells X 3 plates) for each strain. In strain illustration, blue-oak, red-vineyard.

3.4: Discussion

The field of quantitative genetics was established nearly a hundred years ago, reconciling the inheritance of continuously distributed traits with Mendelian genetics (Nelson, Pettersson, and Carlborg 2013). With increased sample sizes and resources, genetic mapping studies are detecting more QTL but also uncovering surprising complexity, including condition specific loci,

gene by environment interactions, epistasis and linkage between causative loci (Mackay et al. 2009).

Gene by environment interaction is a result of genetic loci that have different effects on phenotypic variance in different conditions. Gene by environment interactions are frequently found and were shown to effect the expression of a third of yeast genes (Smith and Kruglyak 2008). Similar to previous studies, we also find that genetic interactions can be environment specific (Gerke et al. 2010). Considering additive loci, we find shared loci with different effects as well as environment specific loci when environmental glucose is varied 20-fold. Measuring changes in QTL effect sizes over a finer gradient of glucose concentration may be informative, analogous to dose dependent effects observed for chemical resistance (Wang and Kruglyak 2014). For example, the contribution of HXT7 may indicate the extent of control that the nutrient transport step has on growth rate, potentially relating the transporter K_m to the Monod constant (K_s) of the oak and vineyard strains (Ziv, Siegal, et al. 2013).

Recently, there has been increased interest in searching for 'variance QTL', found by comparing the difference in variance between genotypic classes, instead of difference in means (Ronnegard and Valdar 2011; Shen et al. 2012). While this type of analysis may uncover strong additive loci determining phenotypic variability, it is likely to uncover genetic interactions affecting trait means (Nelson, Pettersson, Li, et al. 2013). When clonal data and repeated measurements of individuals are available, variability can be directly estimated and mapped as a quantitative trait. Alleles determining variability segregate in natural populations and have been mapped in yeast

(Ansel et al. 2008), flies (Mackay and Lyman 2005), plants (Hall et al. 2007; Jimenez-Gomez et al. 2011) and mice (Fraser and Schadt 2010). To our knowledge, this study is the first to find genetic interactions determining phenotypic variability. Moreover, we find that genetic interactions make up a larger proportion of explained trait variance for variability traits compared to traits of central tendency. The prevalence of genetic interactions may indicate that changes in a single component of a biological system are less likely to have evolutionary stable effects on variability. It will be very interesting to see if this observation will generalize to growth rate variability under different conditions or variability in different phenotypes and systems.

Our study highlights the advantage of using different mapping approaches to dissect the genetic basis of complex traits. While individual segregant analysis has the advantage of detecting genetic interactions, the increased resolution of our bulk segregant approach was quite remarkable. The increased resolution is due to both an advanced intercross population and the increased sampling due to bulk segregant analysis. We note that the population was created using homothallic diploid strains with no auxotrophic or drug resistance markers.

We confirmed the effect of sequence variation in HXT7 on variation in growth rate and lag duration. However, we identified important distinctions between the segregant analysis and the allele replacements. Specifically, the differential effect of the allele depending on the genetic background indicates additional genetic interactions. These may be higher order interactions involving more than two loci (Taylor and Ehrenreich 2014). Alternatively, it may be the result of

an accumulation of undetected pair wise interactions. Although we did not find significant interactions between the chromosome IV position for mean growth rate and median lag duration, the estimated effect of the locus was consistently smaller in the vineyard background when considering the genotype at other additive loci, particularly in the higher glucose concentration. This observation emphasizes the difficulties of predicting phenotype from genotype. Particularly, the estimated marginal additive effect of a locus may not represent the actual effect in any given genetic background.

3.5: Materials and methods

3.5.1: Yeast strains and growth analysis

Parental oak (BC248) and vineyard (BC241) strains and the panel of segregants (Gerke et al. 2006) were obtained from the lab of Barak Cohen (Washington University). NCYC3606 and NCYC3591 (Cubillos et al. 2009) were used during allele replacements. All media was minimal chemically defined carbon limiting media (Saldanha et al. 2004; Brauer et al. 2005) without amino acid or nucleotide supplements. All growth conditions, microscopy and analysis of growth profiles are as described (Ziv, Siegal, et al. 2013).

3.5.2: Data normalization

Estimates for mean growth rate and median lag duration were corrected for plate effects by subtracting the mean parent phenotype and dividing by the distance between parent phenotypes for each plate. This has the effect of scaling segregants across plates, where 0.5 and -0.5 are the parent estimates. Parent phenotypes were calculated as the mean of all well estimates for that

parent on that plate. For variability traits, residuals of a loess regression (standard deviation on mean growth rate or median absolute deviation on median lag duration) were used. Data used for the regressions was original values per well, contained only the segregants (including segregants that were not genotyped) and was conducted within environment.

3.5.3: QTL mapping using R/QTL

We used the R package R/QTL for interval mapping (Broman et al. 2003). We performed genome scans with a single QTL model ('scanone' function), using a normal phenotype model, a 1cM step-size and the HK algorithm for all traits. QTL were identified using significance thresholds, based on an alpha of 0.1 and 10000 permutations. We also performed genome scans with a two QTL model ('scantwo' function) using the same parameters. 1000 two-dimensional permutations were performed using a 5cM step-size. Multiple additive QTL on the same chromosome were identified from the two dimensional scan by setting thresholds for additive and conditional additive (difference between additive and single QTL) LOD scores, based on the permutations and an alpha of 0.1. Additive multiple QTL models were fit ('fitqtl' function) and effect estimates and total percent variance explained were extracted. We identified genetic interactions for each trait using an alpha of 0.1 using the two-dimensional permutations for the interaction LOD scores (the difference between an additive and full model per pair of loci). Final models included all identified additive loci and genetic interactions. Both within and between traits, QTL were considered the same locus if they were within 30cM from one another.

3.5.4: Creation of an advanced intercross population

The advanced intercross population was created by 11 rounds of sporulation and mating starting with a hybrid of the oak and vineyard strains. Generally $2.5X10^8$ cells were sporulated for an average of 9 days at each iteration. Cells were sporulated by spinning (with the exception of 5 days following hurricane sandy) at room temperature in 1% Potassium Acetate at a density of $5X10^7$ cell/ml. For mating, the sporulated culture was resuspended in equal amounts of water and ether and vortexed for 10 min to kill unsporulated cells. Spores were separated using centrifugation, washed with water and incubated in 1 mg/ml Zymolase for 10 min at 30°C. Spores were resuspended in a large volume of 0.01% Triton and vortexed to increase spore dispersion. Spores were subsequently resuspended in a small volume of 0.01% Triton and plated at high density on multiple YPD plates (generally $1.5X10^8$ spores per plate). As both parental strains are homothallic, spores that do not mate with other spores will become diploids by mating type switching. After 19 hours of growth, cells were scraped off the plates and a portion were resuspended in 1% Potassium Acetate to begin a new round of sporulation. At the last iteration, spores were resuspended in liquid YPD and incubated overnight to reduce heterozygosity.

3.5.5: Whole genome sequencing and analysis

Libraries for sequencing were prepared and multiplexed using standard protocols and sequenced using an Illumina HiSeq. Reads were aligned to the reference genome (Ref.SGD020311.fasta) and single nucleotide polymorphisms (SNPs) were identified using BWA (Li and Durbin 2009) and SAMtools (Li et al. 2009). SNP alleles, position, quality and the number of high-quality reads mapping to reference or alternate alleles (DP4) were extracted from VCF files and

analyzed using R (Team 2012). Read depth was calculated as the sum of reads mapping to reference and alternate alleles. Oak and vineyard specific alleles were identified in each sample by comparing to SNPs found by sequencing the oak and vineyard strains. Data was formatted to reflect oak allele frequencies. Numbers of crossover events were identified in advanced intercross clones by identifying transitions between oak and vineyard SNPs. Data was filtered for read depth and SNP quality and two adjacent SNPs from the same parent were required to define a crossover. Numbers of crossovers in F2 segregants were based on transitions in marker genotypes.

3.5.6: QTL mapping using MULTIPOOL

For MULTIPOOL analysis data was filtered for read depth and SNP quality; SNPs with minor allele frequency less than 10% in at least one sample used for a comparison were also excluded. Samples used for comparison were separated by 12-14 generations (low dilution rate) or 20-26 generations (high dilution rate). Replicate chemostats were analyzed separately as well as combined by combining reads at each SNP. Analysis was run in 'contrast' mode, with the exception of the advanced intercross results shown in **Figure 3.S4B**. Each comparison was run with parameters n=1000 or n=200 (number of individuals) and r=1000 or r=2500 (length of cM). We find that different parameter combinations do not change the overall shape of the LOD profile, however the n parameter has a large effect on the magnitude of LOD scores. To assess significance, we performed null comparisons between early time-points of replicate chemostats. The null comparisons had LOD scores ranges of -0.3 to 0.69 for n=200 and 0.4 to 2.9 for n=1000.

3.5.7: Variation in HXT genes and allele replacements

The HXT6 or HXT7 genes were amplified individually using different primer pairs by PCR from the oak and vineyard strains and cloned in plasmids. The plasmids were Sanger sequenced to catalog genetic variation between the oak and vineyard genes and used during allele replacements. Reciprocal hemizygote strains were created by first replacing the HXT6 or HXT7 locus with a construct containing the G418 resistance marker (kanMX) in the oak and vineyard strains (BC248 and BC241). These strains were then mated to the opposite parental strain to create heterozygote knockout strains in the hybrid genetic background. Allele replacements were first created in haploid strains of the oak and vineyard genetic backgrounds (NCYC3606 and NCYC3591 (Cubillos et al. 2009)). Each locus was first replaced by the URA3 gene and subsequently replaced by the modified allele; overlapping PCR was used to create single amino acid modifications. These strains were crossed to the original oak and vineyard strain (BC248 and BC241) of the same background. The mated strains were sporulated and tetrads were screened to identify diploid prototrophs (inheritance of the functional HO and URA3 genes), sensitivity to G418 and resistance to hygromycin (inheritance of the sporulation marker found in the original parental strains). The HXT7 locus was sanger sequenced in this subset to identify final allele replacement strains (identical to the original strains with the exception of the replaced allele). Final allele replacement strains were also crossed to the opposite background oak or vineyard strain to create allele replacements in the heterozygote background.

3.6: Supplemental figures

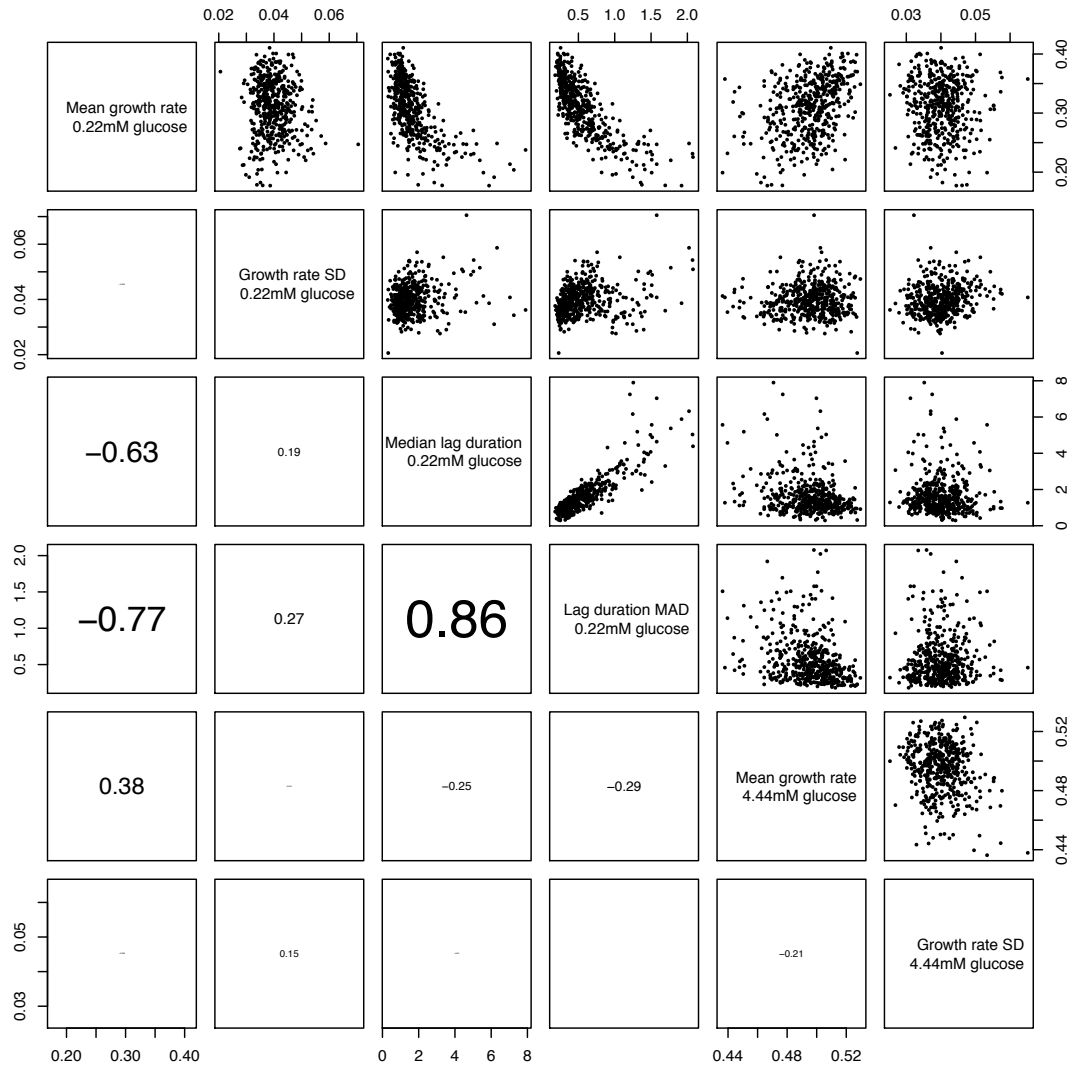


Figure 3.S1: Correlation between cell growth traits

Trait estimates are per well for all F2 segregants phenotyped. Data are in the upper triangle. Pair wise Pearson correlations are in the lower triangle, size of text corresponds to correlation estimate.

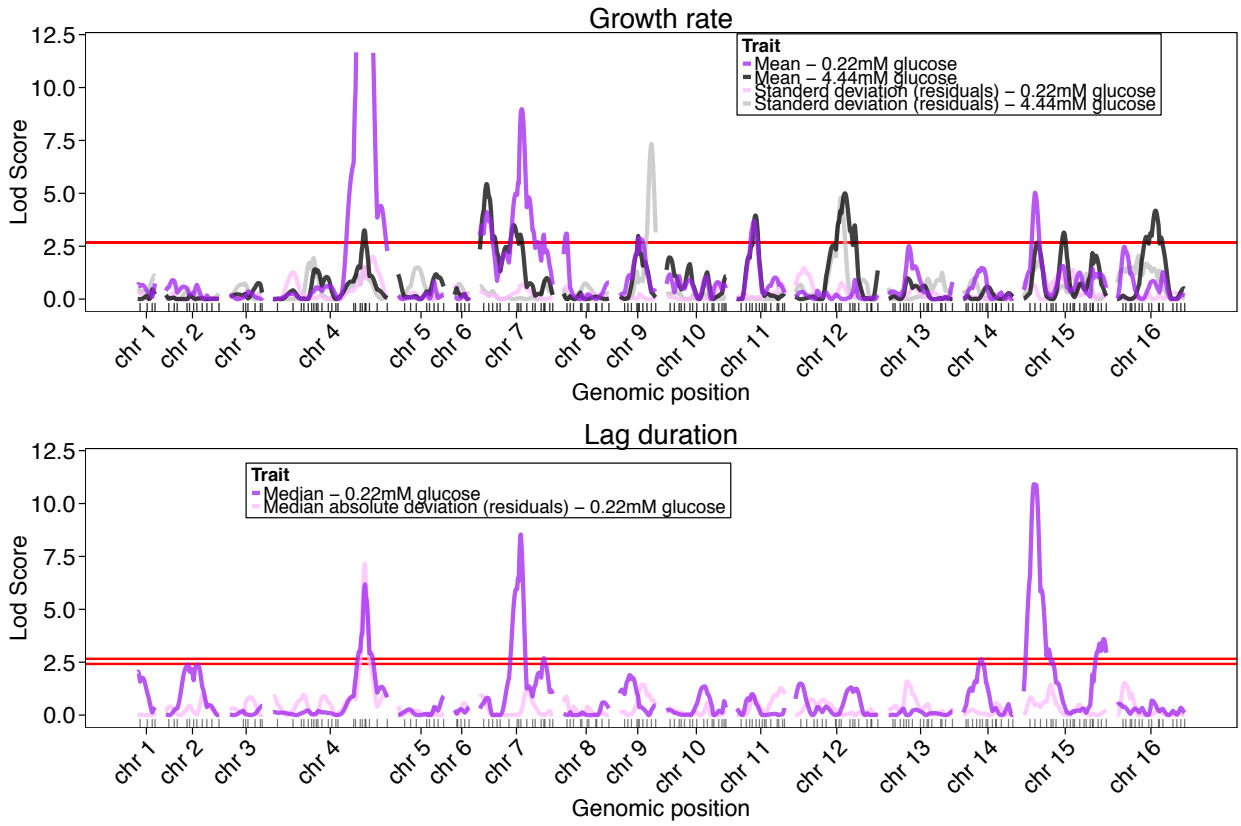


Figure 3.S2: Single QTL LOD profiles

LOD profiles using single-QTL model, based on interval mapping using 374 F2 segregants. Growth rate traits in upper panel, lag duration traits in lower panel. Colors represent combination of trait and environment. Y-axis is limited to a LOD score of 12 for clarity; 0.22mM glucose mean growth rate QTL on chromosome IV reaches a maximum of 41.8.

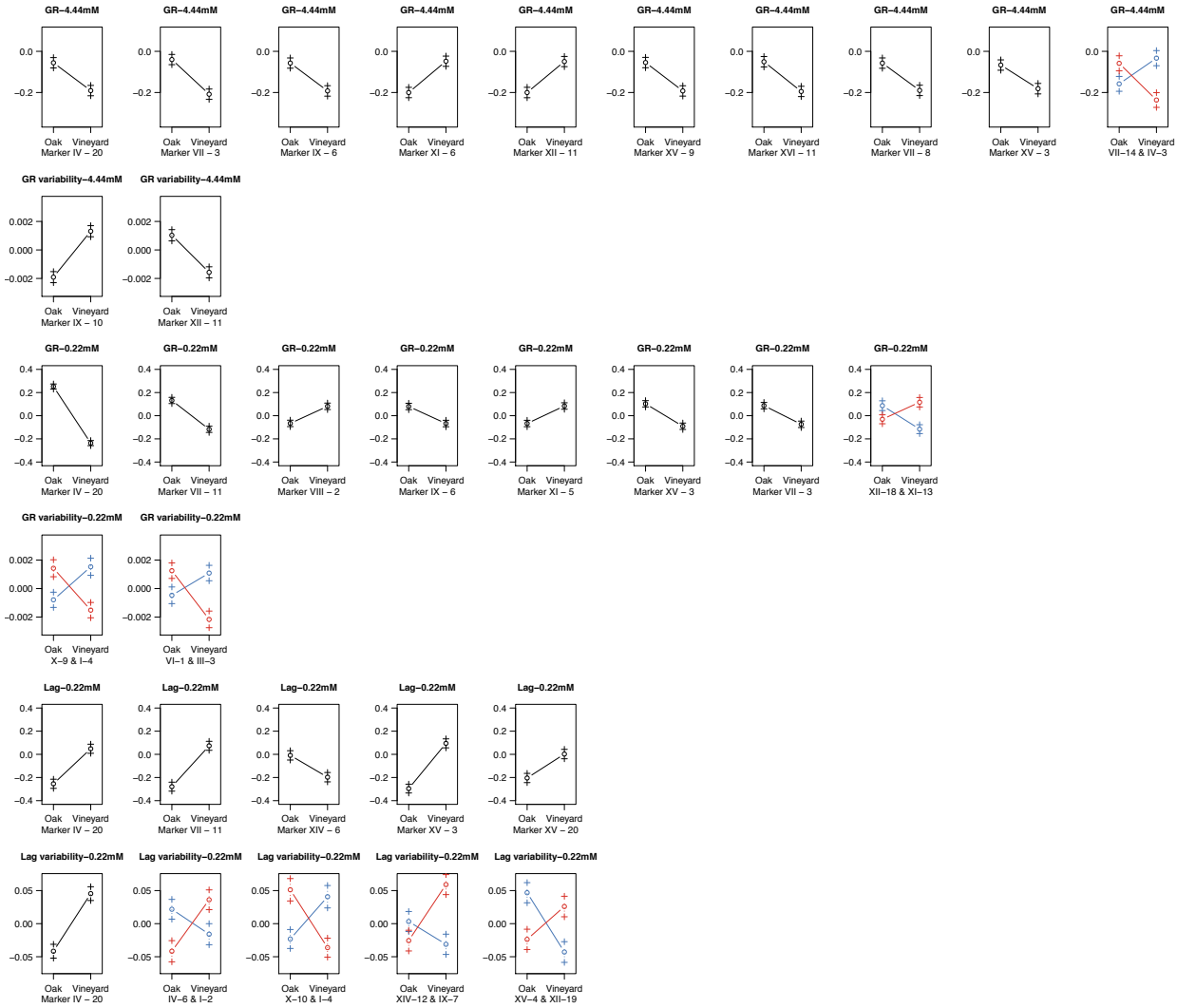


Figure 3.S3: Effects of additive loci and genetic interactions

Each row represents a single trait. Plots depicting additive loci correspond to trait means and standard errors for the oak and vineyard genotype class (at closest marker). Plots depicting genetic interactions correspond to trait means and standard errors for the four genotype combinations (at closest markers); X axis corresponds to first marker and colors represent second marker (blue-oak, red-vineyard). To interpret magnitude of effects, consider that for central tendency traits, normalization results in a difference of 1 unit (-0.5:0.5) between parental phenotypes and for variability, trait values represent residuals from a loess regression (see **methods**). GR-growth rate.

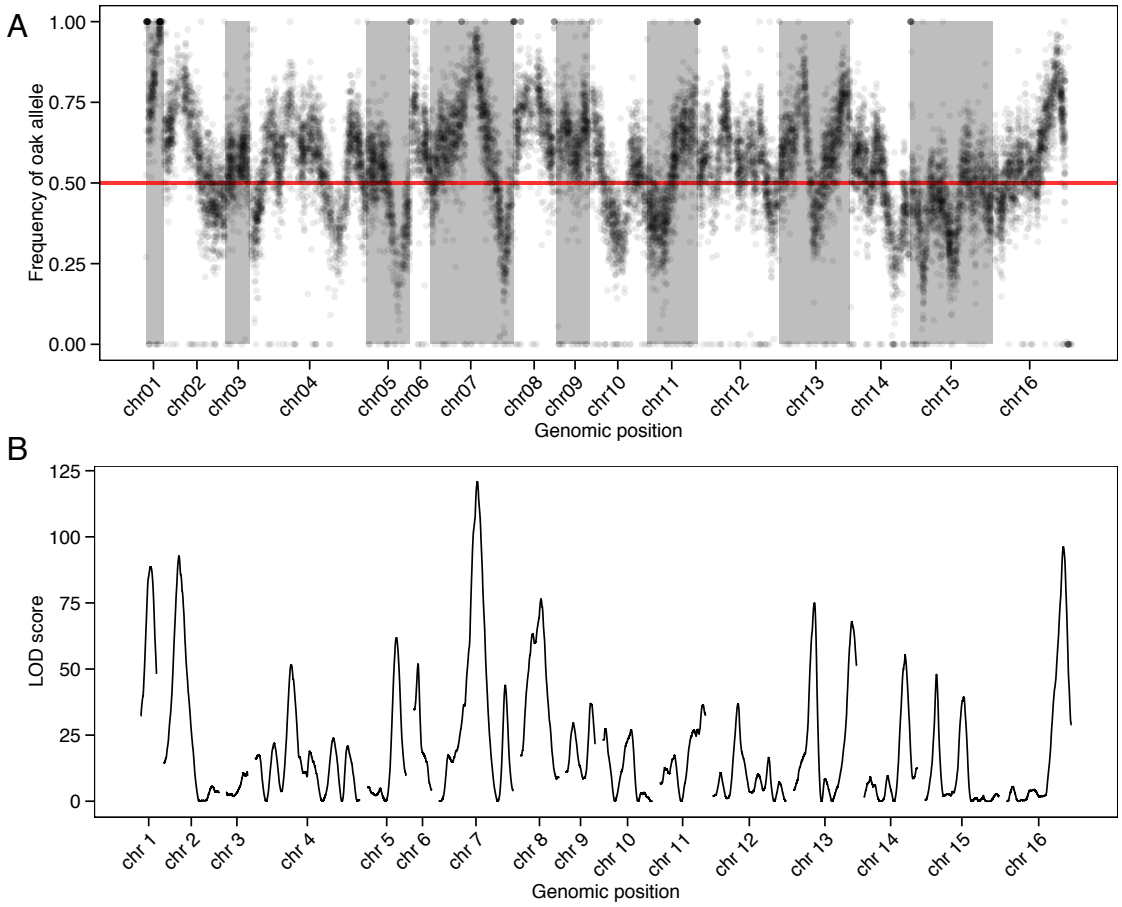


Figure 3.S4: Advanced intercross population

(A) Allele frequencies in the final advanced intercross population. SNPs with minor allele frequency <10% in the F1 heterozygote were excluded. Frequencies reflect the oak allele. (B) MULTIPOOL LOD score results for the advanced intercross population, run with parameters n=1000 (number of individuals), r=2500 (length of cM) and mode 'replicates'.

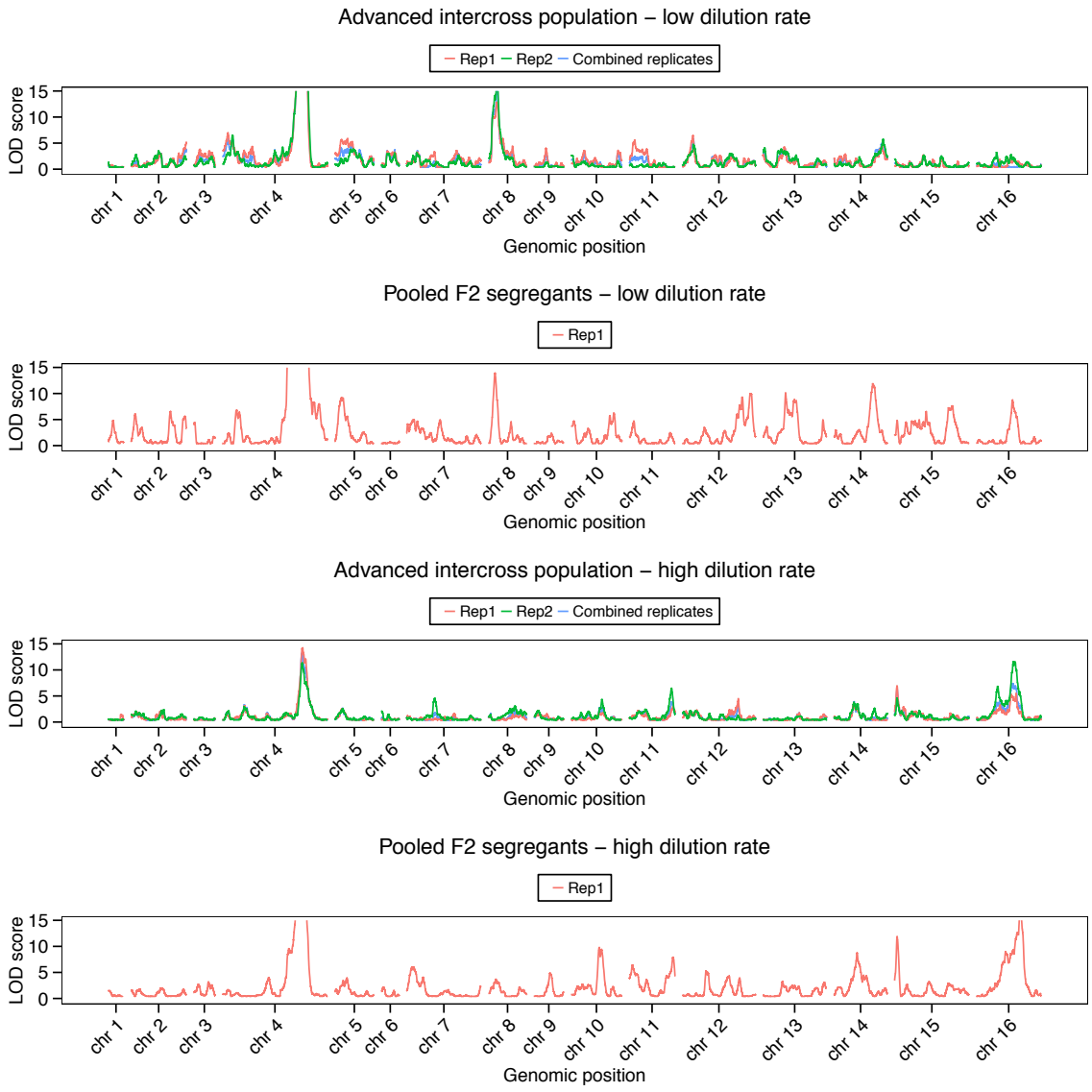


Figure 3.S5: MULTIPOOL LOD profiles

LOD profiles based on allele frequency comparisons over time during growth in chemostats. Panels represent different starting populations and chemostat dilution rates. For the advanced intercross panels, different colors represent replicates or a combined score. Replicates were combined by adding together reads per SNP. Y-axis is limited to a LOD score of 15 for clarity; advanced intercross low dilution rate QTL on chromosome IV reaches a maximum of 54.97. MULTIPOOL results are shown with parameters n=1000 (number of individuals) and r=1000 (length of cM).

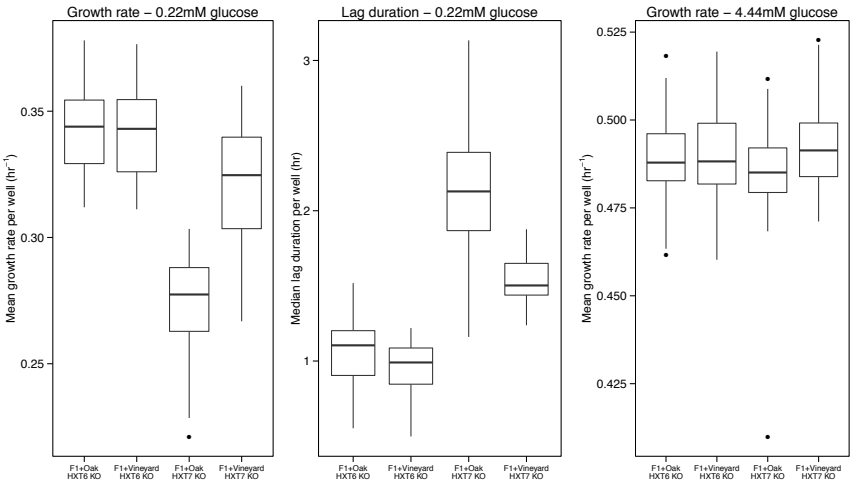


Figure 3.S6: Reciprocal hemizygote HXT6 and HXT7 strains

Distributions of mean growth rate or median lag duration per well for HXT6 and HXT7 reciprocal hemizygote strains. Estimates are for 47-48 wells (across 4 384-well plates) per strain and media combination.

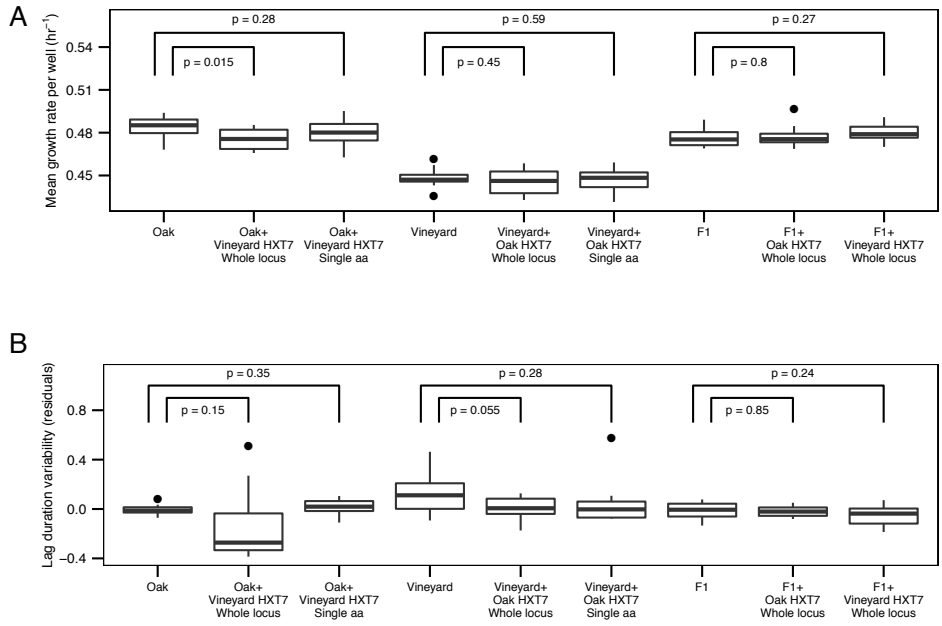


Figure 3.S7: Effect of variation in HXT7

Distributions of mean growth rate in 4.44mM glucose (A) or lag variability in 0.22mM glucose (B) for HXT7 allele replacement strains. P-values are for two sample t test, n=12 (4 wells X 3 plates) for each strain.

Table 3.S1: Single QTL model LOD scores found by interval mapping for all loci and traits

LOD scores for all loci identified in this study including additive QTL and genetic interactions. Positions are shown in cM as well as approximate physical location (in bp) based on position of markers. LOD scores are shown for all traits and are maximum value within 30cM of position. GR-growth rate, Chr.-chromosome, var.-variability.

Trait	Chr.	Position (cM)	Approximate position (bp)	GR 0.22mM	GR variability 0.22mM	Lag 0.22mM	Lag variability 0.22mM	GR 4.44mM	GR variability 4.44mM
GR 0.22mM	4	421	1160904	41.809	1.84	6.177	7.155	3.251	1.504
GR 0.22mM	7	194	604096	8.965	0.708	8.535	1.277	3.282	0.048
GR 0.22mM	8	13	62967	3.098	0.615	0.142	0.864	0.26	0.497
GR 0.22mM	9	85.209	224632	2.892	0.243	1.764	1.451	2.985	1.927
GR 0.22mM	11	79	268728	3.691	0.09	0.989	0.891	3.955	0.1
GR 0.22mM	15	51.783	150756	5.029	0.682	10.916	0.833	2.674	1.781
GR 0.22mM	7	30	137644	4.101	0.329	0.818	0.93	5.439	0.615
GR 0.22mM	11	200	588873	0.705	0.752	0.337	1.288	0.176	0.288
GR 0.22mM	12	360	1002860	0.007	0.008	0.09	0.284	1.354	0.704
GR var. 0.22mM	1	69	162928	0.737	0.142	1.23	0.106	0.468	1.113
GR var. 0.22mM	10	154	412254	1.054	0.086	1.361	0.464	1.171	0.281
GR var. 0.22mM	3	76	141885	0.357	0.273	0.203	0.85	0.367	0.739
GR var. 0.22mM	6	5	119284	0.097	0.606	0.274	0.43	0.516	0.712
Lag 0.22mM	4	424.142	1179958	41.809	1.94	6.177	7.155	3.251	1.504
Lag 0.22mM	7	188.801	594654	8.977	0.708	8.535	1.277	3.475	0.048
Lag 0.22mM	14	82	361686	1.473	0.161	2.644	0.199	0.594	0.944
Lag 0.22mM	15	47	137707	5.029	0.712	10.916	0.833	2.674	1.781
Lag 0.22mM	15	350	974737	1.214	0.224	3.381	0.212	2.049	0.26
Lag var. 0.22mM	4	423	1173031	41.809	1.894	6.177	7.155	3.251	1.504
Lag var. 0.22mM	1	27	88087	0.705	0.297	2.054	0.004	0.17	0.621
Lag var. 0.22mM	4	157	327970	0.349	0.275	0.214	0.537	1.228	1.924
Lag var. 0.22mM	1	67	156762	0.737	0.172	1.319	0.004	0.468	1.113
Lag var. 0.22mM	10	157	417725	1.054	0.129	1.361	0.002	1.242	0.281
Lag var. 0.22mM	9	92	253032	2.892	0.243	1.735	1.176	2.985	2.693
Lag var. 0.22mM	14	160	587945	0.315	0.249	0.329	0.525	0.341	0.105
Lag var. 0.22mM	12	380	1033949	0.194	0.004	0.014	0.003	1.354	0.012
Lag var. 0.22mM	15	77	214637	5.029	0.668	10.909	0.384	2.674	1.781

GR 4.44mM	4	421	1160904	41.809	1.84	6.177	7.155	3.251	1.504
GR 4.44mM	7	30	137644	4.112	0.329	0.818	0.93	5.429	0.615
GR 4.44mM	9	85.209	224632	2.892	0.243	1.764	1.451	2.985	1.927
GR 4.44mM	11	86	286752	3.691	0.09	0.989	0.891	3.955	0.1
GR 4.44mM	12	229	609824	0.183	0.76	1.309	0.1	5.003	4.806
GR 4.44mM	15	185	528176	1.564	1.203	1.525	0.461	3.145	0.621
GR 4.44mM	16	177	512955	1.018	0.229	0.683	0.247	4.181	1.543
GR 4.44mM	7	156	466129	8.006	0.596	8.229	0.093	3.496	0.065
GR 4.44mM	15	63	179172	5.029	0.682	10.916	0.826	2.669	1.781
GR 4.44mM	4	105	182578	0.251	1.273	0.103	0.844	0.269	0.124
GR 4.44mM	7	255	763417	4.802	0.434	2	1.194	0.162	0.251
GR var. 4.44mM	9	146.039	361957	2.062	0.08	0.463	1.331	1.753	7.335
GR var. 4.44mM	12	209	550479	0.21	0.76	1.116	0.27	5.003	4.806

Table 3.S2: Two QTL model LOD scores found by interval mapping for all genetic interaction loci and traits

Genetic interactions found for all traits. Each column specifies one interaction, with position for both loci in cM as well as approximate physical location (in bp) based on position of markers. Additive model and full model LOD scores are shown for the specific pair of positions for all traits. GR-growth rate, Chr.-chromosome, var.-variability, Add.-additive

Trait	GR 0.22mM	GR variability 0.22mM	GR variability 0.22mM	Lag variability 0.22mM	Lag variability 0.22mM	Lag variability 0.22mM	Lag variability 0.22mM	GR 4.44mM
Chr. #1	11	1	3	1	1	9	12	4
Position (cM) #1	200	69	76	27	67	92	380	105
Approximate position (bp) #1	588873	162928	141885	88087	156762	253032	1033949	182578
Chr. #2	12	10	6	4	10	14	15	7
Position (cM) #2	360	154	5	157	157	160	77	255
Approximate position (bp) #2	1002860	412254	119284	327970	417725	587945	214637	763417
GR 0.22 Add.	0.706	0.279	0.244	0.671	0.308	2.543	2.154	2.577
GR 0.22 Full	5.037	1.115	0.326	0.702	1.036	2.931	3.634	3.258
GR var. 0.22 Add.	0.744	0.198	0.998	0.197	0.182	0.496	0.016	1.29
GR var. 0.22 Full	0.746	6.627	5.707	2	6.512	0.523	0.037	1.347
Lag 0.22 Add.	0.029	1.39	0.217	1.97	1.452	0.536	5.936	1.392
Lag 0.22 Full.	1.722	1.45	0.395	1.98	1.478	0.643	5.987	1.875
Lag var. 0.22 Add.	1.315	0.002	0.708	0.544	0.005	1.665	0.386	0.803
Lag var. 0.22 Full.	2.141	6.538	1.104	4.492	6.879	5.599	4.888	0.913
GR 4.44 Add.	0.14	0.447	0.607	0.05	0.5	2.224	3.548	0.424
GR 4.44 Full	0.318	1.152	0.668	0.067	1.149	2.31	3.728	4.62
GR var. 4.44 Add.	0.267	1.123	1.362	1.661	1.098	1.002	1.535	0.053
GR var. 4.44 Full	0.669	1.265	2.847	1.802	1.312	1.033	1.54	0.348

Table 3.S3: Comparison of mapping approaches

List of intervals defined by each pair of adjacent markers used in the interval mapping. Each row specifies one interval which had a maximal LOD score >3 in at least one condition. Conditions are interval mapping in 0.22mM glucose or 4.44mM glucose, bulk segregant analysis of the advanced intercross grown in low or high dilution rate chemostats (scores are for combined replicate data) and bulk segregant analysis of the pool of F2 segregants grown in low or high dilution rate chemostats. MULTIPOOL results are shown with parameters n=1000 (number of individuals) and r=1000 (length of cM). GR-growth rate, Chr.-chromosome, var.-variability, Add.-additive, AIC-advanced intercross population.

Chr.	Marker #1	Marker #2	Position (bp) #1	Position (bp) #2	GR 0.22mM glucose	GR 4.44mM glucose	F2 low dilution	F2 high dilution	AIC low dilution	AIC high dilution
1	1	2	32304	65359	0.7	0.01	4.67	1.25	0.53	0.54
1	2	3	65359	112085	0.67	0.17	4.79	0.66	0.5	0.49
2	1	2	120437	157331	0.55	0.2	3.4	1.82	0.48	1.32
2	2	3	157331	232715	0.9	0.07	3.55	1	1.43	1.06
2	6	7	401143	470054	0.26	0.18	0.91	1.97	3.01	1.88
2	8	9	502224	575934	0.67	0.04	6.56	0.71	1.62	0.85
2	9	10	575934	656829	0.17	0.04	6.44	0.5	1.89	0.61
2	11	12	702460	771102	0.01	0.02	5.46	1.75	3.29	1.24
3	4	5	188786	225558	0.36	0.54	1.15	3.18	2.6	0.49
4	1	2	49679	114170	0.01	0.02	2.4	1.68	5.8	1.79
4	2	3	114170	151239	0.25	0.42	1.56	1.33	5.49	0.7
4	3	4	151239	246448	0.25	0.42	6.81	1.09	5.63	0.64
4	4	5	246448	280697	0.01	0.03	6.42	0.65	1.61	1.21
4	5	6	280697	346360	0.05	0.03	2.97	0.52	2.79	3.31
4	12	13	635526	696637	0.63	0.99	1	4	1.65	1.78
4	13	14	696637	767335	0.54	0.36	1.46	2.53	3.75	0.7
4	15	16	826861	938147	6.44	0.84	15.13	7.43	3.55	1.59
4	16	17	938147	1005614	9.79	1.04	24.21	9.57	4.95	1.56
4	17	18	1005614	1071559	20.35	1.55	31.55	16.85	16.17	1.73
4	18	19	1071559	1105164	30.71	2.49	44.3	21.64	32.03	4.81
4	19	20	1105164	1179958	41.81	3.25	60.1	21.83	54.97	13.31
4	20	21	1179958	1230646	40.53	3.19	59.81	18.84	27.15	10.91
4	21	22	1230646	1264123	34.55	2.61	42.53	13.22	17.25	6.69
4	22	23	1264123	1362236	15.45	1.11	11.52	4.28	4.55	4
4	23	25	1362236	1501556	4.41	0.45	8	0.89	0.71	0.82
5	1	2	49043	100056	0.17	1.17	9.21	2.49	4.14	2.21
5	2	3	100056	158171	0.1	0.1	9.19	3.43	3.81	2.65
5	3	4	158171	217279	0.1	0.14	5.62	3.91	3.95	1.42

Chr.	Marker #1	Marker #2	Position (bp) #1	Position (bp) #2	GR 0.22mM glucose	GR 4.44mM glucose	F2 low dilution	F2 high dilution	AIC low dilution	AIC high dilution
5	4	5	217279	338773	0.3	0.36	3.11	1.76	4.6	0.72
5	5	6	338773	399605	0.3	0.36	2.34	0.62	3.69	0.45
6	3	4	170876	201882	0.04	0.32	0.5	1.53	3.71	0.63
6	4	5	201882	223558	0.25	0.24	0.48	0.77	3.26	0.47
7	1	2	44796	96932	3.48	4.11	4.94	6.03	2.17	0.88
7	2	3	96932	171763	4.11	5.44	4.99	6	3.54	0.66
7	3	4	171763	230207	3.79	4.83	3.65	3.42	2.81	0.68
7	4	5	230207	284675	2.63	2.97	3.24	4.07	0.53	0.58
7	7	8	432466	490117	5	3.5	4.9	0.88	0.63	1.35
7	8	9	490117	545582	5.51	2.96	4.73	1.34	0.53	0.72
7	9	11	545582	594654	8.84	3.05	1.35	1.5	0.41	0.46
7	11	12	594654	644097	8.98	2.97	1.45	1.69	1.49	0.46
7	12	13	644097	726509	4.8	0.75	1.29	0.93	2.85	1.24
7	13	14	726509	763036	3.26	0.64	0.67	0.55	2.96	0.58
7	15	16	809363	863960	3.05	0.72	1.68	0.72	1.7	0.51
7	16	17	863960	892563	3.05	0.89	1.78	0.62	1.35	0.51
8	1	2	44529	66740	3.1	0.26	8.21	2.25	10.71	0.75
8	2	3	66740	117292	3.07	0.14	13.93	3.69	14.02	0.5
8	3	4	117292	153282	0.53	0.14	10.44	3.32	15.67	0.5
8	4	5	153282	198394	0.15	0.31	3.3	3.02	8.8	0.81
8	5	6	198394	218718	0.22	0.31	0.86	1.36	4.96	0.95
8	6	7	218718	256537	0.22	0.08	1.09	1.38	4.35	0.86
8	7	8	256537	297020	0.21	0.01	1.99	0.41	3.14	0.96
8	8	9	297020	396688	0.2	0.12	4.6	0.53	3.81	2.14
9	5	6	201167	224632	2.89	2.99	1.88	3.99	1.91	0.72
9	6	7	224632	250317	2.89	2.99	2.19	4.86	0.53	0.65
9	7	8	250317	276759	2.83	2.2	2.36	4.59	0.77	0.66
10	1	2	37500	66255	1.16	1.97	4.68	1.27	1.6	0.64
10	2	3	66255	101653	1.16	1.93	3.13	2.38	1.77	0.61
10	5	6	219800	270368	0.58	0.58	3.28	2.65	1.78	1.54
10	8	9	336551	385727	0.63	0.71	1.31	5.44	1.35	1.26
10	9	10	385727	444535	0.83	0.64	1.07	9.78	1.29	2.78
10	10	11	444535	465360	1.05	1.27	0.76	9.26	0.66	2.82
10	11	12	465360	527586	1	1.27	1.99	6.78	0.44	2.08
10	12	13	527586	565916	0.51	0.06	5.73	0.7	0.59	0.57
10	13	14	565916	619631	0.8	0.15	6.17	2.27	0.63	1.11
10	14	15	619631	660770	0.77	0.26	6.23	1.77	0.91	1.06
10	15	16	660770	690593	0.68	0.86	1.95	0.42	3.08	0.49
11	1	2	42555	98234	0.01	0.02	4.71	6.4	2.43	1.08
11	2	3	98234	153726	1.21	1.44	2.06	4.76	2.29	2.11
11	4	5	206087	249055	3.32	2.56	1.37	3.59	2	1.99
11	5	6	249055	295134	3.69	3.95	0.52	3.6	2.56	2.23

Chr.	Marker #1	Marker #2	Position (bp) #1	Position (bp) #2	GR 0.22mM glucose	GR 4.44mM glucose	F2 low dilution	F2 high dilution	AIC low dilution	AIC high dilution
11	6	7	295134	350464	3.03	3.9	0.56	1.69	0.84	0.48
11	11	12	504700	550500	1.02	0.16	0.87	4.39	0.62	0.99
11	12	13	550500	575225	0.94	0.09	1.49	4.19	0.43	1.94
11	13	14	575225	611629	0.71	0.18	2.48	5.61	0.45	3.87
11	14	15	611629	637800	0.52	0.18	1.88	7.92	0.44	3.88
12	2	3	89702	154678	0.34	0.05	0.57	0.83	4.96	2.04
12	3	4	154678	213608	0.17	0.11	0.74	1.02	5.85	2.2
12	6	7	311265	367241	0.33	1.07	3.48	5.29	0.81	0.66
12	7	8	367241	451216	0.33	2.54	2.12	5.01	2.36	1.03
12	9	10	492640	518152	0.21	3.64	1.28	0.71	2.06	0.41
12	10	11	518152	561785	0.1	4.17	0.84	0.69	2.39	0.67
12	11	12	561785	672637	0.07	5	2.86	4.19	2.22	1.45
12	12	13	672637	717712	0.32	3.29	1.31	4.29	2.88	1.68
12	13	14	717712	755968	0.91	3.49	2.87	2.26	2.84	1.59
12	14	15	755968	820804	0.91	3.49	6.47	3.23	2.08	3.24
12	15	16	820804	881789	0.06	0.87	9.29	3.92	0.8	2.54
12	16	17	881789	969370	0.06	0.05	7.95	0.51	1.74	0.8
12	17	18	969370	990749	0.07	0.02	9.98	0.46	1.37	0.81
12	18	19	990749	1034324	0.19	1.35	9.88	0.72	1.63	1.1
13	1	2	56276	102221	0.36	0.13	3.62	0.52	1.4	0.73
13	5	6	244520	308353	1.53	0.88	4.51	3.26	3.65	0.48
13	6	7	308353	341111	2.51	0.98	10.12	2.35	3.88	0.51
13	7	8	341111	396424	2.51	0.92	8.76	2.12	2.87	0.6
13	8	9	396424	502787	1.55	0.63	8.81	2.51	2.09	1.52
13	9	10	502787	555633	0.82	0.62	6.35	2.3	1.98	1.76
13	16	17	858414	914425	0.81	0.09	4.92	3.1	1.42	1.38
14	4	5	231538	288198	1.06	0.02	1.38	4.27	1.43	3.18
14	5	6	288198	340447	1.36	0.07	2.54	8.75	1.76	3.9
14	6	7	340447	387811	1.47	0.27	3.04	8.71	1.65	3.06
14	7	8	387811	437370	1.44	0.55	2.5	5.58	1.18	1.17
14	8	9	437370	466584	1.06	0.59	1.25	4.24	0.64	1.31
14	9	10	466584	502292	0.35	0.57	3.86	4.32	0.41	0.45
14	10	11	502292	558811	0.05	0.28	11.25	2.7	1.47	0.45
14	11	13	558811	635465	0.31	0.04	11.87	1.08	3.72	0.91
14	13	14	635465	694158	0.31	0.29	5.25	1.38	4.83	0.61
14	14	15	694158	745494	0.25	0.29	2.94	2.01	5.12	0.46
15	1	2	51154	88251	1.34	0.27	3.05	9.71	0.94	4.06
15	2	3	88251	150756	5.03	2.22	3.16	1.48	0.54	2.32
15	3	4	150756	215310	5.03	2.67	4.51	2.67	1.49	1.98
15	4	5	215310	280789	1.92	2.16	4.69	1.42	1.54	1.72
15	5	6	280789	346309	0.29	0.13	4.74	0.62	0.64	1.97
15	6	7	346309	407363	0.21	0	4.92	0.52	2.29	1.91

Chr.	Marker #1	Marker #2	Position (bp) #1	Position (bp) #2	GR 0.22mM glucose	GR 4.44mM glucose	F2 low dilution	F2 high dilution	AIC low dilution	AIC high dilution
15	7	8	407363	454325	0.55	0.27	6.4	1.13	2.67	1.47
15	8	9	454325	523989	1.56	3.13	6.54	3.22	1.38	0.93
15	9	10	523989	572941	1.55	3.15	2.57	2.66	1.26	0.93
15	13	14	693081	747648	1.25	0.72	3.31	1.47	1.88	0.65
15	14	15	747648	804445	1.25	0.73	7.25	2.91	0.46	1.19
15	15	16	804445	826784	0.48	0.48	7.5	2.61	0.45	1.33
15	16	17	826784	870951	0.29	1.05	7.62	2.46	0.48	1.35
15	17	18	870951	898552	1.44	2.18	4.7	1.83	0.53	0.95
15	18	19	898552	912655	1.44	2.18	3.51	1.79	1.33	0.91
16	6	7	266276	324298	0.43	0.7	0.45	3.36	1.16	3.73
16	7	8	324298	372267	0.05	1.95	1.05	6.21	0.85	4.06
16	8	9	372267	429004	0.54	2.95	0.75	8.96	0.52	2.91
16	9	10	429004	482737	0.82	3.12	3.72	8.8	0.59	3.22
16	10	11	482737	536003	0.85	4.18	8.75	9.35	0.41	7.4
16	11	12	536003	593104	0.41	3.48	7.96	12.65	0.5	7.26
16	12	13	593104	648617	1.28	2.91	5.39	16.5	0.68	4.21
16	13	14	648617	745578	1.28	2.5	0.83	16.23	1.18	2.87

CHAPTER 4: THE GENETIC BASIS OF GROWTH RATE VARIABILITY

4.1: Abstract

Phenotypic variability is present even when genetic and environmental differences between cells are absent. In this study we explore the genetic basis of natural variation in growth rate variability. We have identified wild isolates of *Saccharomyces cerevisiae* that differ in the variance of growth rate distributions measured for populations of clonal cells. The difference between the strains is apparent when cells are grown in two distinct glucose concentrations. We find that in both environments, the net genetic effect is primarily dominant, as the F1 hybrid has low variability. However, we also find that the genetic determinants of variability in each environment can be separated, indicating gene by environment interactions. Classic genetic mapping is complicated by reproductive isolation caused by low spore viability of the F1 hybrid, thought to be a result of extensive chromosome rearrangements in one of the parental strains. Growth rate variability in the higher glucose concentration can be separated from the presence of the aberrant chromosomal configuration and shows complex segregation. In contrast, there is an association between chromosomal configuration and growth rate variability at the lower concentration. Tracking the inheritance of spore viability in a set of backcrosses supports chromosome structure as the main determinant of reproductive isolation and reveals the creation of new chromosomal configurations by recombination and assortment.

4.2: Introduction

Genetic and environmental factors are the major determinants of natural variation in quantitative traits. However, even genetically identical individuals raised in nominally identical environments can display heterogeneity. This residual variation, or "phenotypic variability" (Geiler-Samerotte et al. 2013) can be an advantageous and even necessary feature of biological systems (Losick and Desplan 2008; Eldar and Elowitz 2010). The extent of variability can be genetically controlled (Hill and Mulder 2010) and hence subject to evolutionary selection. However, the prevalence of natural variation that modifies phenotypic variability is still unknown as only a few studies have searched for loci that alter the variance of traits (Geiler-Samerotte et al. 2013). Detection of such loci has the potential to shed light on both the mechanistic basis and the evolutionary significance of phenotypic variability.

Reproductive isolation restricts gene flow between two populations and is thought to be important for the onset of speciation. Although all members of the *Saccharomyces* genus can form viable F1 hybrids, hybrid spore viability is generally less than 1% (Hittinger 2013). Interspecies post-zyogotic isolation is mainly due to decreased recombination due to sequence divergence (Hittinger 2013), but translocations (Fischer et al. 2000) and nuclear/mitochondrial incompatibility (Lee et al. 2008) also contribute. In contrast, a recent study surveying reproductive isolation within *S. cerevisiae* found that reciprocal translocations explained cases in which hybrid spore viability is approximately 50% or 75% (Hou et al. 2014). Within *S. cerevisiae*, the Malaysian lineage (Liti et al. 2009), known to be reproductively isolated with 2.5% – 10% spore viability when crossed to other lineages of budding yeast (Naumov et al.

2006; Cubillos et al. 2011), has been shown recently to harbor many large-scale translocations (Marie-Nelly et al. 2014).

We previously identified two wild yeast isolates that differ in growth rate variability but not mean growth rate (Ziv, Siegal, et al. 2013). Here we show that these strains are reproductively isolated. By isolating, intercrossing and backcrossing viable segregants, we investigate the segregation of spore viability and growth rate variability in different environments. We explore connections between growth rate variability and aberrant chromosomal configuration and discuss possible mechanistic implications.

4.3: Results

4.3.1: Reproductive isolation of strains with different phenotypic variability

We have previously measured the growth rate distributions of 12 isolates of *S. cerevisiae*, using a high-throughput microscopy based assay (Levy et al. 2012) in two different glucose concentrations (Ziv, Siegal, et al. 2013). The two conditions, 0.22mM ("low") and 4.44mM ("high") glucose, were chosen as they characterize the growth response to variation in nutrient concentration. The high glucose concentration supports maximal growth rates despite being over an order of magnitude lower than standard lab media. We found that natural variation in phenotypic variability segregates in yeast populations. We were particularly interested in two strains that differed in variability in both glucose concentrations. DBVPG1788, which was originally isolated from soil in Finland (hereafter "Finland") had lower than average variability. The other strain was originally presumed to be DBVPG1373, a strain from soil in the

Netherlands, but was later discovered to be a strain from the Malaysian lineage (Liti et al. 2009) (hereafter "Malaysia"). The Malaysia strain has higher than average growth rate variability (Ziv, Siegal, et al. 2013). These strains also have nearly identical mean growth rate, making them ideal for identifying the genetic basis of natural variation in growth rate variability.

We crossed the two strains (**methods**) and sporulated the resulting hybrid. One hundred and thirty-two tetrads were dissected but only 44 spores formed colonies (8.3% spore viability). Many tetrads had 0 or 1 viable spores and the distribution of viable cells per tetrad fit well a Poisson distribution with a mean of 0.33 (**Figure 4.1A**). Different molecular scenarios could explain the observed distribution. However, the extent of spore inviability is not consistent with a small number of interacting alleles or structural differences (Hou et al. 2014). The Malaysian strain (UWOPS03-461.4) was recently shown to have extensive chromosomal rearrangements, consisting of eight large chromosomal translocations, four smaller subtelomeric translocations and a few small intrachromosomal inversions (Marie-Nelly et al. 2014). In contrast to the massive genomic rearrangement, sequence divergence is typical for an *S. cerevisiae* lineage (Liti et al. 2009).

It is likely that the low spore viability of the Finland/Malaysia hybrid is a consequence of the radically different chromosomal structure in the Malaysia strain. The potential effect of these rearrangements on growth rate variability is less clear. It is of interest to note that the hybrid, which contains both sets of chromosome configurations, has low growth rate variability when

grown in both glucose concentrations, suggesting that the increased variability is recessive (**Figure 4.1B**).

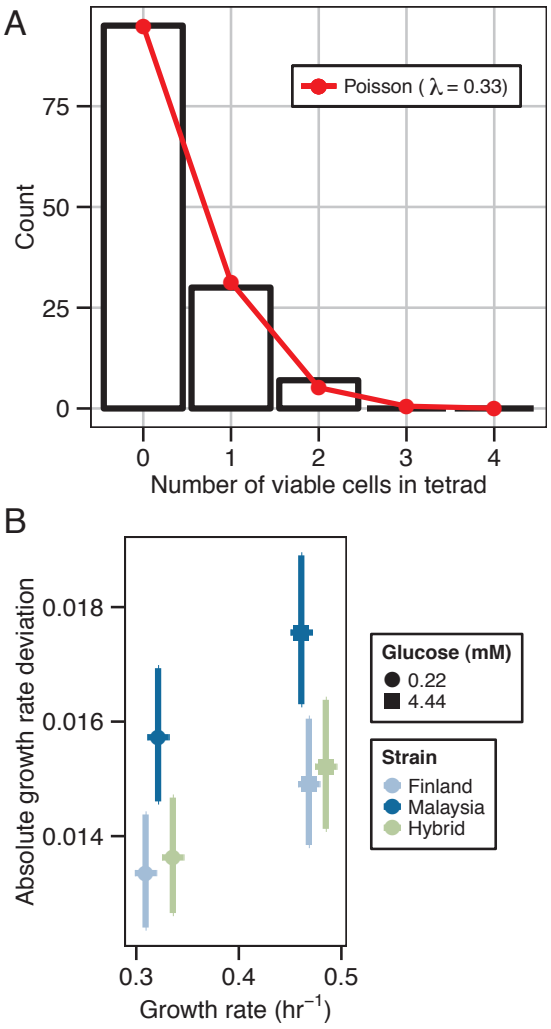


Figure 4.1: Finland/Malaysia hybrid has low spore viability and growth rate variability A) Distribution of number of viable cells per tetrad for the F1 hybrid. Mean and variance of distribution is 0.33. Red line depicts estimates based on a Poisson distribution. B) Absolute growth-rate deviations are plotted against growth rate; estimates for each strain/media combination are based on 25 wells, measured across 5 plates, analyzed using mixed effect modeling. Error bars represent 95% confidence intervals. As indicated, shapes represent media and colors represent strains.

4.3.2: Phenotypic variability of viable segregants suggest different genetic basis for variability between environments

Depending on the genetic basis of growth rate variability and mechanism of reproductive isolation, the rare viable segregants may or may not have variability differences. We first isolated a large number of segregants from the Finland/Malaysia cross by random spore analysis (**methods**). We define F2 segregants as the meiotic products of the Finland/Malaysia hybrid. Six F2 segregants were chosen that had low or high variability based on preliminary experiments and their growth rate distributions were measured in the two glucose concentrations. Some segregants had low or high variability in both conditions, resembling the parental strains (**Figure 4.2A**). However, some segregants had low variability in low glucose and high variability in high glucose (**Figure 4.2A**). This suggests that the genetic basis of variability is distinct between the two environments. The segregants did not have large differences in mean growth rate compared to the parental strains (**Figure 4.S1A**).

We performed intercrosses and backcrosses (**methods**) to determine if the spore inviability could be separated from the segregation of growth rate variability (**Figure 4.2B**). We used a subset of the segregants that had similar variability phenotypes to one or both parents. In this set of crosses, hybrids had either low spore viability (3% – 16%) or high spore viability (87% – 94%) (**Figure 4.2B**). Spore viability was associated with the difference in growth rate variability of the hybrid progenitors when grown in low glucose but not high glucose (**Figure 4.2C, Figure 4.2D**).

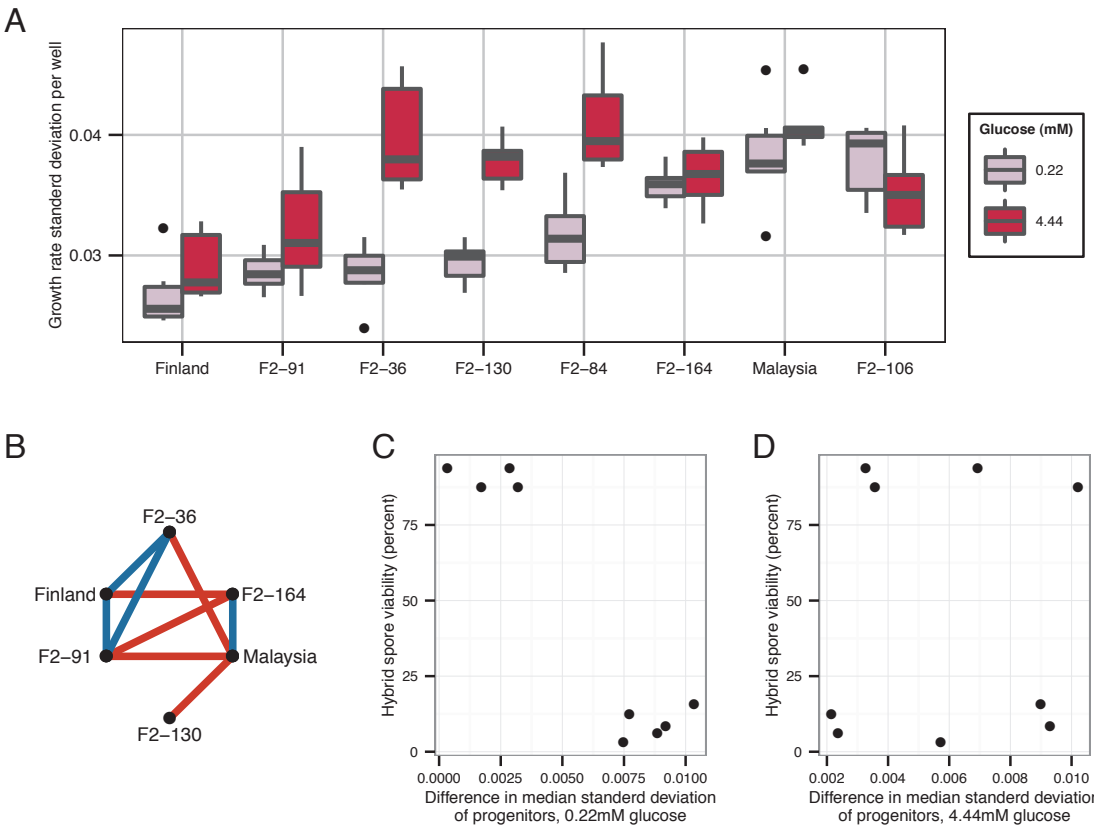


Figure 4.2: Phenotypic variability and spore viability for viable Finalnd/Malaysia segregants

A) Growth rate standard deviations (estimated for 6 wells per strain/media combination) for the Finland and Malaysia strains and six viable segregants. Strains are ranked by median standard deviation in 0.22 mM glucose. Colors represent media. B) Representation of crosses performed. Line colors represent spore viability of resulting hybrids (red – low, blue – high). C) Percent spore viability of hybrids is plotted against difference in the median standard deviation of the hybrid progenitors (shown in (A)). Standard deviation estimates are for cells grown in 0.22 mM glucose. D) Same as (C) for standard deviation estimates for cells grown in 4.44mM glucose.

4.3.3: Prospects for mapping growth rate variability in high glucose

Reproductive isolation complicates genetic mapping as it hinders the creation of a mapping population. Our results imply that the reproductive isolation phenomenon can be separated from

the segregation of alleles affecting growth rate variability in high glucose. Specifically, we had crossed strains (F2-36 and F2-91) that differed in growth rate variability but produced a hybrid with high spore viability (**Figure 4.2**). Using this cross to create a mapping population has some advantages. As it is an intercross, both the amount of segregating variation and the extent of linkage may be reduced.

Fifty-four segregants, consisting of 12 4-spore tetrads and 2 3-spore tetrads were collected from the F2-36/F2-91 cross. Growth rate analysis (**methods**) suggests complex segregation of phenotypic variability (**Figure 4.3A, Figure 4.3B**). Variability for a subset of the segregants was higher than the estimate for the F2-36 strain but comparable to that of the Malaysia strain analyzed in the same dataset (**Figure 4.3A**). This observation may indicate the importance of epistasis in determining growth rate variability. There was no correlation between growth rate mean and standard deviation in the segregants (**Figure 4.S1B**).

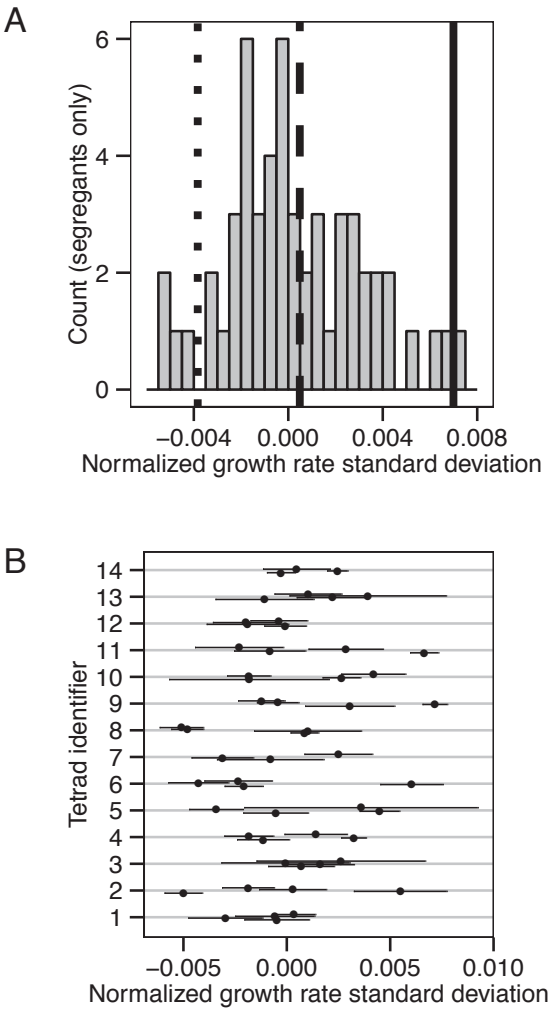


Figure 4.3: Segregation of phenotypic variability in the F2-36/F2-91 cross

A) Histogram of estimates of growth rate variability for 54 segregants. Estimates are mean standard deviation based on 3-6 wells, normalized between plates by mean subtraction of common strains. Vertical lines represent parental strains (dotted – F2-91, dashed – F2-36, solid - Malaysia). B) Same estimates as (A), grouped by tetrad, error bars depict standard errors.

4.3.4: Segregation of spore inviability likely depends on inheritance of chromosome structure

The association between reproductive isolation and growth rate variability at low glucose (**Figure 4.2C**) may indicate that shared or linked loci are responsible for both phenomena. This observation motivated us to follow the inheritance of spore viability. We utilized a backcrossing strategy starting with F2 segregants that had similar phenotypic variability as the Finland and Malaysia strains (F2-91 and F2-164) (**Figure 4.4**). Each F2 was backcrossed to the parent with

the opposite variability phenotype and the few viable segregants resulting from the backcross were again backcrossed in the same direction (**Figure 4.4**). The resulting strains had similar proportions of low and high spore viability. A cross that yielded low spore viability was chosen and the few surviving spores were again backcrossed and the procedure was repeated (**Figure 4.4**). This backcrossing strategy resulted in strains that were enriched for either the Finland or Malaysia genome but maintained low spore viability when crossed to Finland or Malaysia respectively.

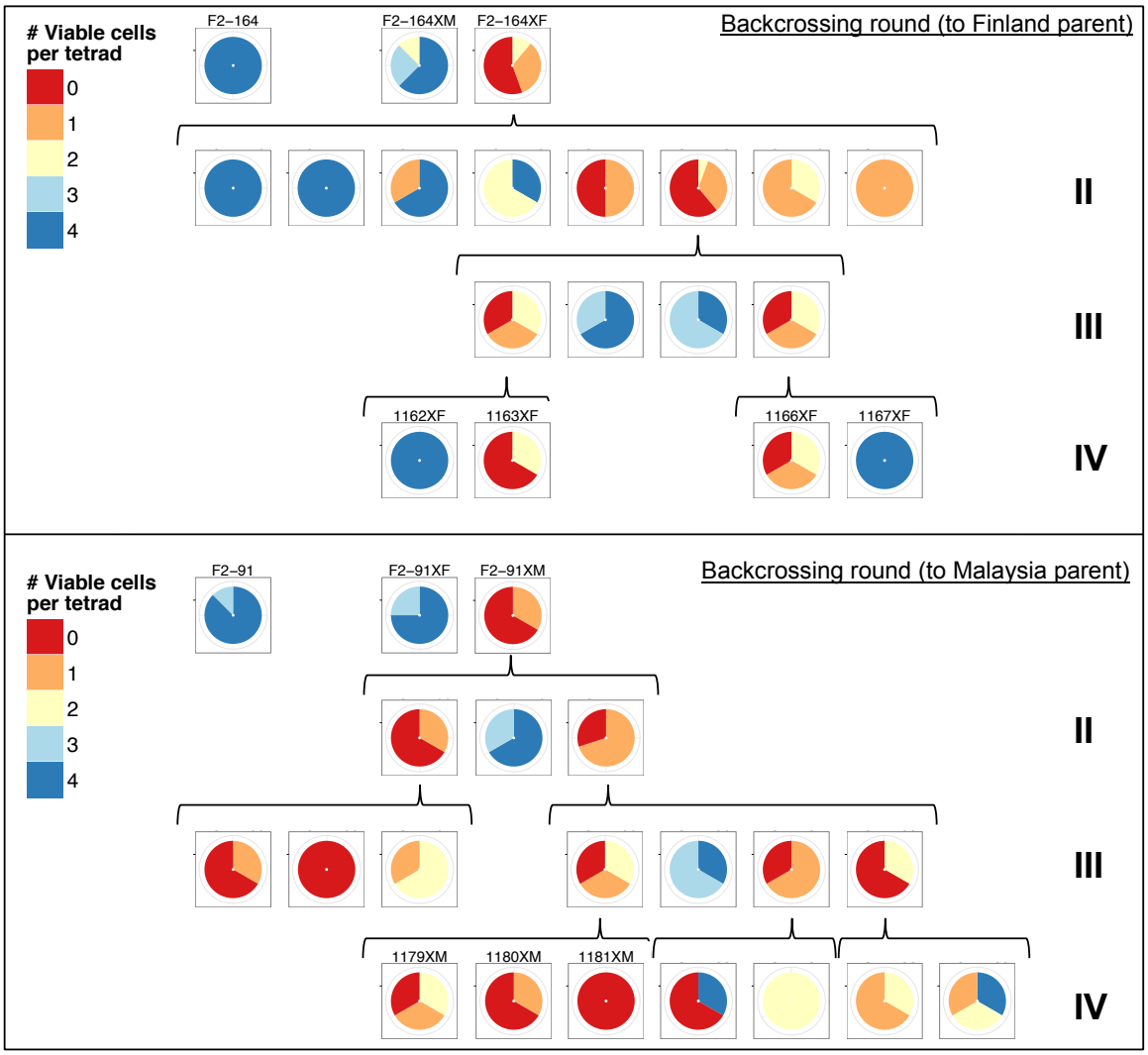


Figure 4.4: Backcrosses to follow the inheritance of spore viability

Pie charts represent the distributions of viable cells per tetrad for each backcrossed strain that was measured. Each row represents a generation; curly brackets represent ancestry. Strains with low spore viability are predominantly red and orange while charts for strains with high viability are mostly blue. Distributions are based on 2–18 tetrads per strain (mode=3). Strains were backcrossed at each generation in the Finland direction, starting with F2-164 (top) or in the Malaysia direction, starting with F2-91 (bottom).

The genomes of seven haploid segregants (DGY1162, DGY1163, DGY1166, DGY1167, DGY1179, DGY1180 and DGY1181), coming from the third generation of both backcrossing directions (**Figure 4.4**), were sequenced (**methods**). For comparison, the genomes of the Finland, Malaysia, F2-91 and F2-164 strains, as well as a pool of 170 viable F2s from the Finland/Malaysia cross (obtained from random spore analysis), were also sequenced. Sequencing revealed that the backcrossed strains were enriched for the expected genetic background (**Figure 4.6A**).

As the Malaysian strain (UWOPS03-461.4) harbors a number of large-scale chromosome rearrangements (Marie-Nelly et al. 2014), it is possible to deduce which chromosome configurations exist in the backcrossed strains. Inheritance of the Malaysian chromosome structure results in inheritance of Malaysian-specific SNPs around the translocation breakpoints. Segregant chromosomal configurations were inferred based on sequence analysis (**Figure 4.5A, Figure 4.5B**) and pulse field gel analysis (**Figure 4.S2**).

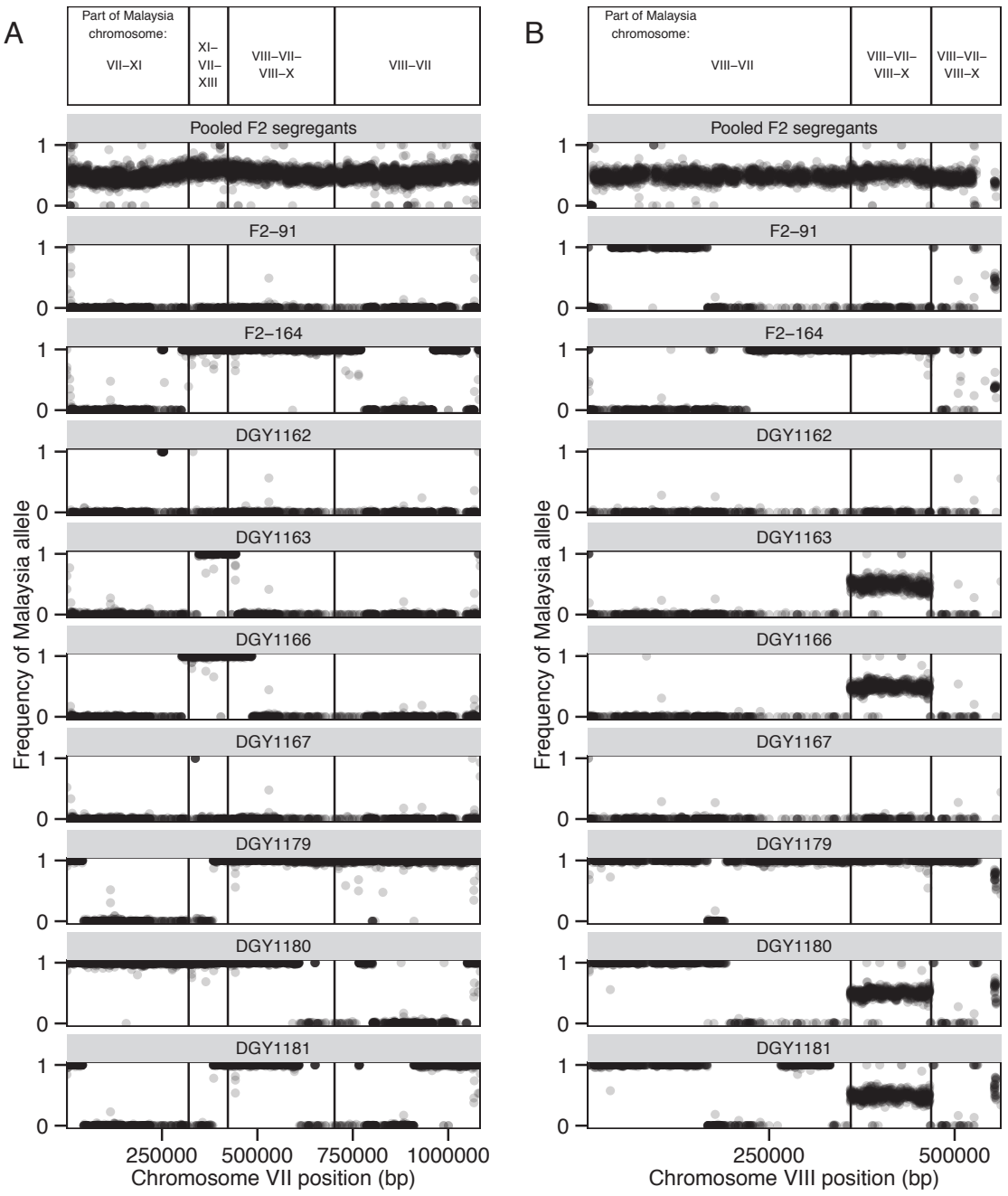


Figure 4.5: Sequence analysis of backcrossed strains

SNP allele frequencies in sequenced strains relative to the Malaysian strain for reference chromosomes VII (A) and VIII (B). Vertical lines depict translocation breakpoints. Top diagram relates chromosome fragments to Malaysia specific chromosomes (see **Figure 4.6B**).

Of the backcrossed strains that were sequenced, two had the Finland parental chromosome configuration and none had the Malaysian parental configuration (**Figure 4.6B**). In the remaining five strains, recombination and segregation have created new chromosomal configurations. Specifically, recombination events on chromosome VII created a VII-XIII chromosome that could segregate with a full-length chromosome XI (replacing Malaysian XI-VII-XIII and VII-XI). Similar events created a VII-VIII-X chromosome that could segregate with a full-length chromosome VIII (replacing Malaysian VIII-VII-VIII-X and VIII-VII) (**Figure 4.6B**). In this second case, the resulting strains have two copies of a 120kb region of chromosome VIII, one originating from each of the original parental strains. This was detected in the sequencing data as a doubling in read depth and the presence of both Finland- and Malaysia-specific SNPs at ~0.5 frequency (**Figure 4.5B**).

Intercrosses between the backcrossed strains (**Figure 4.6C**) support the idea that chromosome structure is the main determinant of reproductive isolation in budding yeast (Hou et al. 2014). Specifically, the cross between DGY1163 and DGY1181, which are enriched for Finland and Malaysia SNPs respectively, but have identical chromosome configurations, yields a hybrid with high spore viability. Due to mating type constraints, additional crosses (such as between DGY1166 and DGY1181) were not performed.

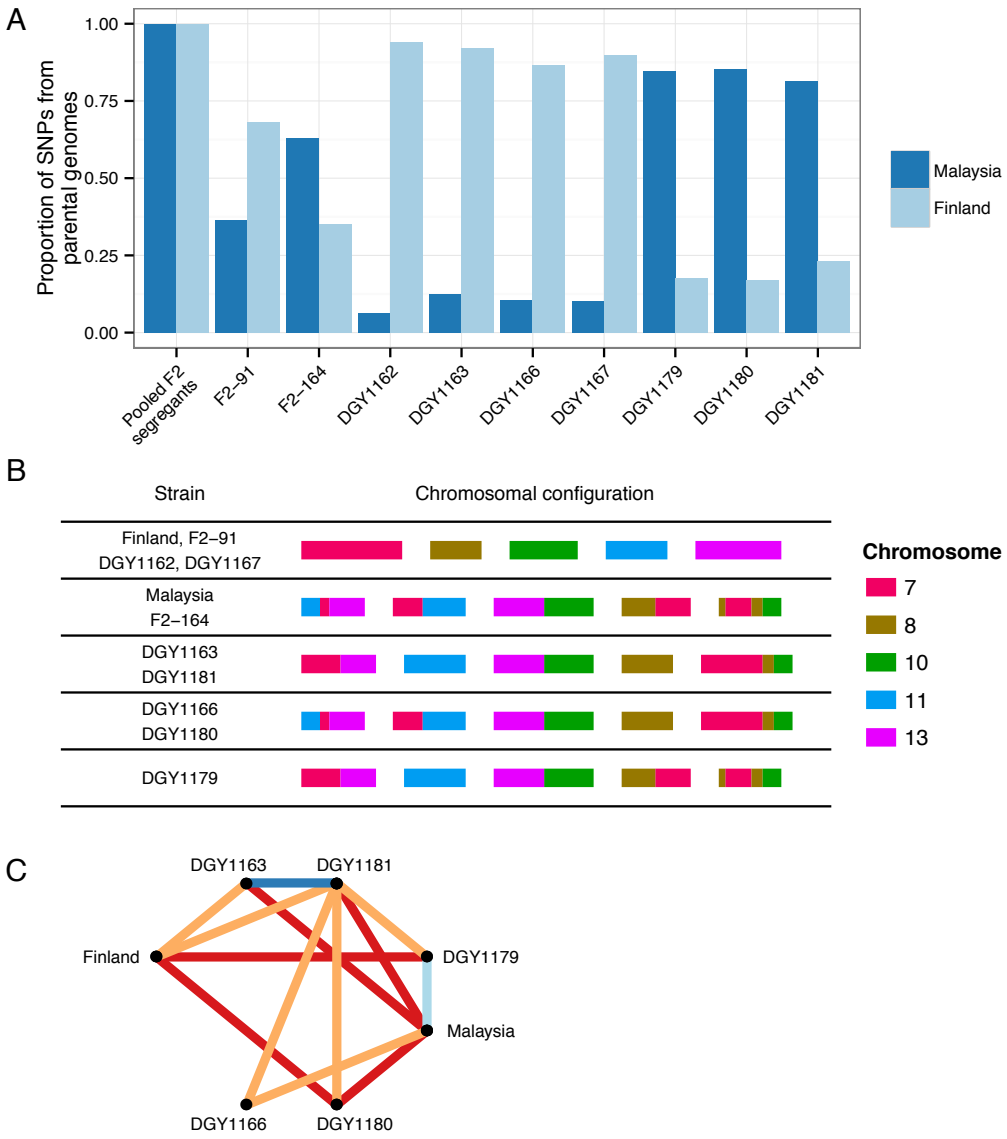


Figure 4.6: Characterization of backcrossed strains

A) Proportion of SNPs originating from the Finland or Malaysia strain, normalized to the F2 segregant pool. Sequenced backcrossed strains are progenitors of strains shown in **Figure 4.4**. B) Diagram of chromosomal configurations. Only major Malaysian translocations are shown, which comprise re-arrangements in 5 chromosomes. Colors represent reference chromosome identifier. C) Representation of crosses performed. For clarity, strains DGY1162 and DGY1167 are

excluded from the diagram; both strains had the same viability pattern as the Finland strain. Line colors represent spore viability of resulting hybrids (red: 0% – 25%, orange: 25% – 50%, light blue: 50% – 75% and blue: 75% – 100%).

4.4: Discussion

A number of studies have demonstrated that there is significant phenotypic variation in genetically identical individuals (Raser and O'Shea 2005; Levy et al. 2012). Despite its prevalence, the genetic basis of phenotypic variability is not well understood, particularly in natural populations. The main limitation in studying variability is the necessity of experimental and statistical methods that permit accurate estimation and analysis (Geiler-Samerotte et al. 2013). We have used a high-throughput microcolony growth assay (Levy et al. 2012; Ziv, Siegal, et al. 2013) to explore the genetic basis of growth rate variability in natural isolates of *S. cerevisiae*.

This study has focused on two strains, isolated from Finland and Malaysia, which differ in the variance of microcolony growth rate distributions. These two strains are divergent at both the nucleotide and karyotype level. The unique Malaysian karyotype may be the product of genomic instability. This genomic instability may be associated with an increased mutation rate leading to high variance in the measured growth rate distributions. However, two observations argue against this. First, the Malaysian lineage has not accumulated mutations as the level of sequence divergence is similar to other lineages (Liti et al. 2009). Second, higher variance in Malaysia growth rate distributions is due to equal proportions of fast and slow growing cells (Ziv, Siegal, et al. 2013), whereas random mutations tend to have deleterious effects (Hall and Joseph 2010).

Moreover, any mechanism creating genetic variation between microcolonies will have to be reconciled with the low growth rate variance of the F1 hybrid.

The association between Malaysian chromosomal configuration and high variability when cells are grown in low glucose can be either direct or indirect. An indirect link may be the chance genetic linkage of loci determining variability and the translocation breakpoints. On the other hand, the chromosomal rearrangements may have changed the relative position and chromatin environment of growth related genes. This in turn could have caused lower and more variable expression of many or few genes. Change in expression variation of key genes (such as transcription factors) could propagate through cellular networks resulting in increased growth rate variability (Stewart-Ornstein et al. 2012).

This study contributes to the growing body of work that aims to investigate phenotypic variability. It demonstrates the feasibility of treating variability as a quantitative trait and provides evidence that chromosomal structure may contribute to phenotypic variability. The connection between phenotypic variability and the structure and associated epigenetic landscape of the genome warrants further investigation.

4.5: Materials and methods

4.5.1: Yeast mating

The Finland and Malaysia strains are capable of mating type switching and hence do not exist as stable haploids. Crosses in this study were performed either between two diploids, between a

diploid cell and a haploid cell or between haploids. For diploid crosses (original Finland/Malaysia cross and initial intercrosses/backcrosses shown in **Figure 4.2B**), parents were sporulated and single spores from each strain were dissected and placed in adjacent positions on a YPD plate. If the spores are of opposite mating types (expectation: 50%), they will form a hybrid. In the absence of any selectable markers, hybrids were identified using restriction fragment length polymorphisms (RFLPs).

Subsequently, the Finland and Malaysia strains were transformed, replacing one copy of the HO endonuclease with a construct containing the G418 resistance marker (kanMX). The heterozygote strains were sporulated and colonies derived from dissected tetrads contained both haploid and diploid cells. Haploid cells of both mating types in both genetic backgrounds were isolated. For the backcrossing strategy, F2-91 and F2-164 were crossed to Malaysia and Finland haploids. Crosses between diploids and haploids were similar to diploid-diploid mating. These crosses also produced both haploids and diploid progeny. Only haploid cells were used for the second backcrossing round. Haploids were crossed by mixing the two strains on an agar plate, incubating for ~4 hours at 30°, streaking out the mixture and physically "pulling zygotes" out of the mating mixture using a dissecting microscope.

4.5.2: Random spore isolation

A sporulated culture of the Finland/Malaysia hybrid was resuspended in equal amounts of water and ether and vortexed for 10 min to kill unsporulated cells. Spores were separated using centrifugation, washed with water and incubated in 1 mg/ml Zymolase for 1 hour at 30°C. 0.15

grams of sterile 0.5mm glass beads were added per ml of sporulated culture, incubated for an additional hour at 30°C and vortexed for 2 minutes. Cells were subsequently diluted in 0.1% triton and plated on YPD plates. 168 colonies were picked at random and arrayed in 2 96-well plates. Preliminary micro-colony growth rate analysis revealed bimodal distributions in a subset of wells. When wells were subsequently streaked for single colonies and growth rate distributions were remeasured, no strains retained bimodal distributions, suggesting that some of the original colonies picked were a mix of two different segregants. This may be a result of incomplete separation of spores during random spore analysis.

4.5.3: Growth rate analysis

All growth conditions, microscopy and analysis of microcolony growth profiles are as described (Ziv, Siegal, et al. 2013). Where multiple replicates were available, estimates for strain growth rate and variability were estimated as in (Ziv, Siegal, et al. 2013) using mixed effect modeling. Otherwise, distributions of mean or standard deviation per well were compared between strains and conditions. For the F2-91/F2-36 cross, some of the segregants were phenotyped across 6 wells (3 wells X 2 plates) and some were phenoptyped across 3 wells (1 plate) using a different microscope. All plates had 15 wells containing the same strains (3 wells X 5 strains). In order to correct for plate effects, the average growth rate standard deviation (or average mean growth rate) for the common strains was subtracted from the segregant estimates. Standard deviations were estimated based on 300-3500 microcolonies.

4.5.4: Whole genome sequencing and analysis

Libraries for sequencing were prepared and multiplexed using standard protocols and sequenced using an Illumina HiSeq. Reads were aligned to the reference genome (Ref.SGD020311.fasta) and single nucleotide polymorphisms (SNPs) were identified using BWA (Li and Durbin 2009) and SAMtools (Li et al. 2009). SNP alleles, position, quality and the number of high-quality reads mapping to reference or alternate alleles (DP4) were extracted from VCF files and analyzed using R (Team 2012). Read depth was calculated as the sum of reads mapping to reference and alternate alleles. Finland and Malaysia specific alleles were identified in each sample by comparing to SNPs found by sequencing the Finalnd and Malaysia strains. Data was formatted to reflect Malaysia allele frequencies.

4.6: Supplemental figures

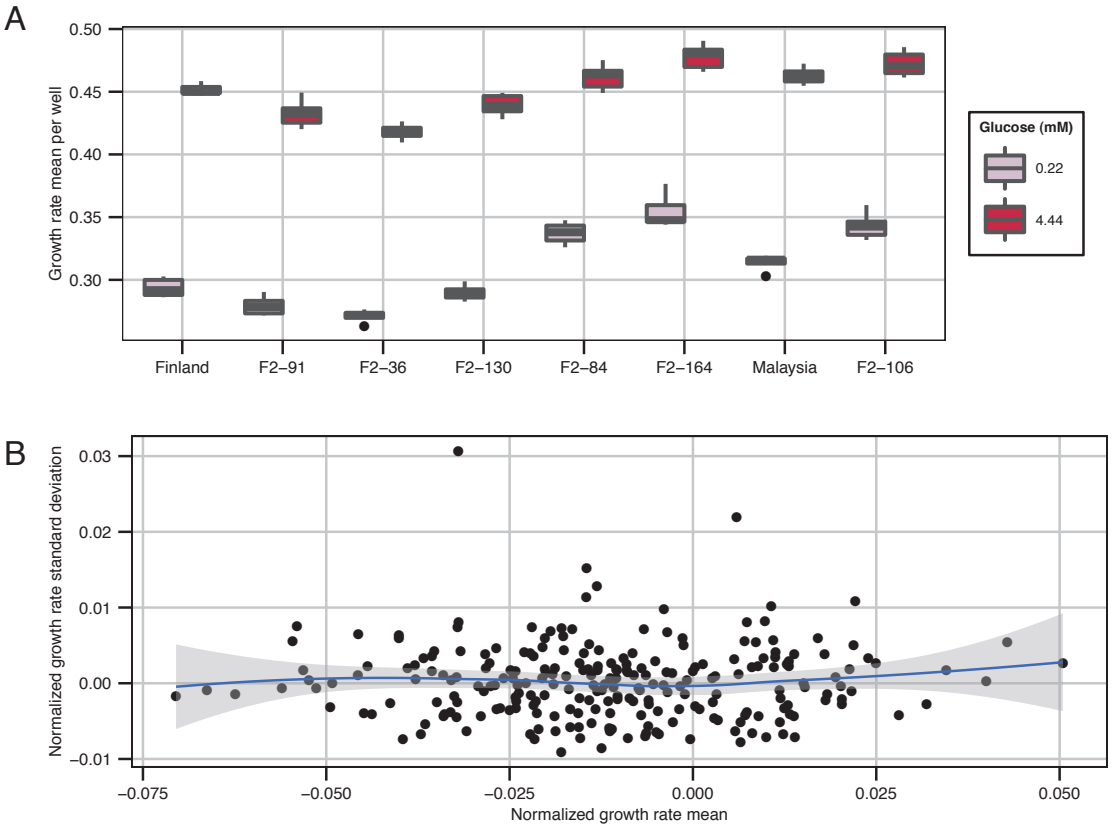


Figure 4.S1: Mean growth rate in Finland/Malaysia segregants

A) Growth rate means (estimated for 6 wells per strain/media combination) for the Finland and Malaysia strains and six viable segregants. Strains are in same order as **Figure 4.2A**. Colors represent media. B) Normalized growth rate standard deviation verses normalized growth rate mean for all growth rate distributions (3-6 per strain) representing 54 segregants from the F2-91/F2-36 cross. The line depicts loess regression.

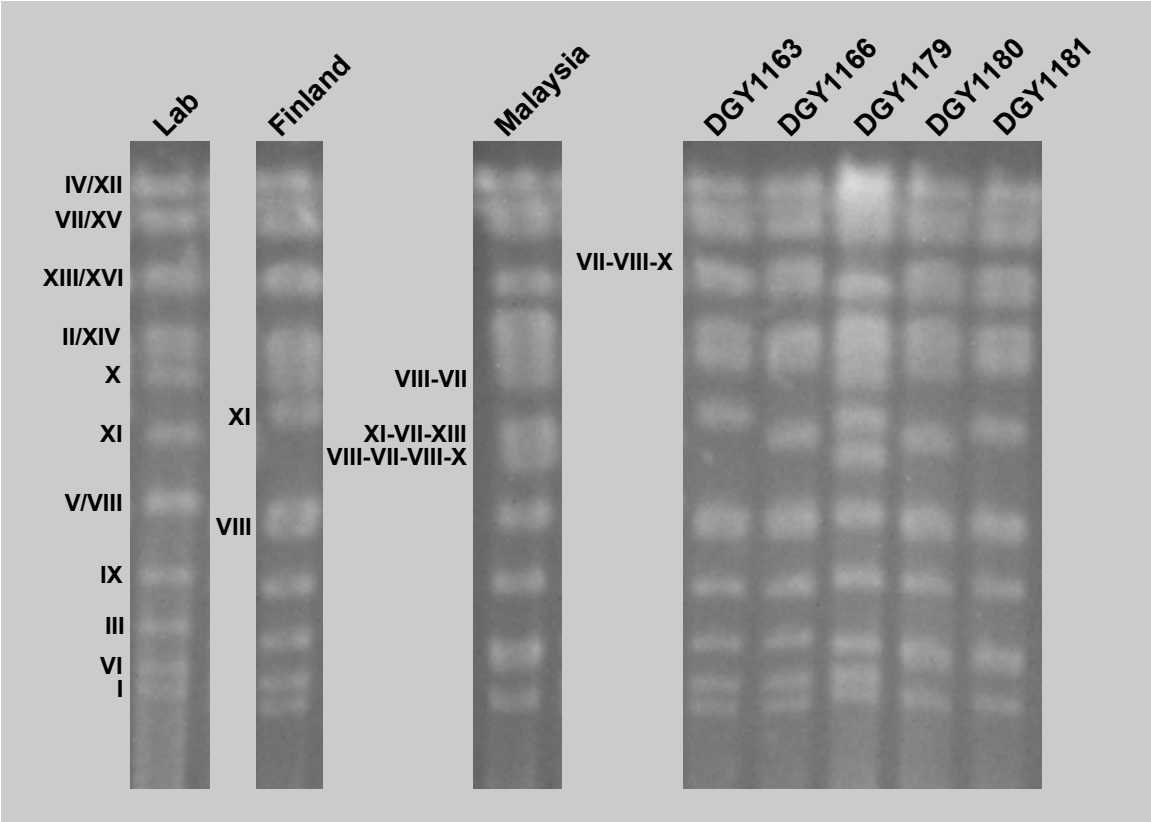


Figure 4.S2: Pulse-field gel supporting inferred chromosome configurations

Whole chromosome preparations for eight strains: the reference (Lab strain), Finland, Malaysia and five backcrossed strains. Chromosome annotation is based on expected size.

CHAPTER 5: CONCLUSION

Accurately predicting quantitative phenotypes based on the identity of underlying genetic variants is a central goal of basic research in genetics and biology with important medical and industrial applications (Nelson, Pettersson, and Carlborg 2013). However, mapping genotype to phenotype is a challenge even in model organisms due to the prevalence of gene-environment interactions, epistasis and variance among isogenic individuals (Lehner 2013). In this thesis, I have dissected the genetic basis of natural variation in cell growth, a quintessential quantitative trait, using natural isolates of the budding yeast (*Saccharomyces cerevisiae*). Using a high-throughput microscopy based assay, I defined the genetic architectures determining two growth traits, the exponential growth rate and time to growth initiation, in two ecologically relevant environments.

Studies presented in this thesis support the importance of quantitative trait loci (QTL) context dependency (Mackay et al. 2009; Chandler et al. 2013). Genome-wide analysis uncovered a multitude of loci with additive effects. However, at least in the case of *HXT7,* functional confirmation revealed extensive genetic background effects. Furthermore, epistatic interactions were also found, particularly determining variation in growth variability (variation in the absence of genetic and environmental variance). Gene-environment interactions were common both for mean growth rate and growth rate variability. These studies have set the stage for further functional characterization of natural alleles determining cell growth.

One of the major focuses of this thesis has been elucidating the genetic determinants of phenotypic variability. Phenotypic variability comprises variation between genetically identical individuals in the same environment (Geiler-Samerotte et al. 2013). Variability is a relative term; low variability implies homogeneity among individuals while high variability corresponds to heterogeneity among individuals (Levy and Siegal 2012). From an evolutionary perspective, theory suggests stabilizing selection on traits will decrease variability while strong directional selection and disruptive selection will select for increased variability (Mulder et al. 2007). Analysis of loci determining growth variability in the Oak/Vineyard cross did not reveal clear evidence of directional selection on variability, as combinations of oak or vineyard alleles could result in both high and low variability. However, due to limitations and detection bias, evolutionary inferences based on analysis of a small sample of QTL can be misleading (Rockman 2012).

Beyond the ultimate evolutionary significance of variability, the proximate causes of variability are also of interest. In many cases, variability is thought of in terms of bistability, which creates bimodal phenotypic distributions. Bistability is mediated by specific configurations of positive and negative feedback (Losick and Desplan 2008; Eldar and Elowitz 2010). Examples of such systems include bacterial persistence, the well studied *lac* operon and cell fate determination during development in multicellular organisms (Losick and Desplan 2008; Eldar and Elowitz 2010). In contrast, variability differences in continuous mRNA expression can be modulated by the frequency and magnitude of transcriptional bursts (Zenklusen et al. 2008). The system properties that enable differences in variability of more complex phenotypes are largely

unexplored. In this respect, cell growth is particularly appealing, as many cellular components required for cell growth regulation are known. Further functional analysis of the loci determining growth variability will lead to insights into the molecular mechanisms that can promote or reduce variability in continuous complex traits.

Valuable information is missed when only considering distribution averages. By studying phenotypic variability and analyzing the effect of QTL over different environments and genetic backgrounds, we can start to unravel the complexity of genotype to phenotype mapping. This thesis has laid the groundwork for the molecular dissection of genetic and non-genetic determinants of cell growth. In particular, it has provided a wealth of information on gene-gene and gene-environment interactions determining the central tendency and variability in cell growth distributions.

REFERENCES

Ansel J, Bottin H, Rodriguez-Beltran C, Damon C, Nagarajan M, Fehrmann S, François J, Yvert G. 2008. Cell-to-Cell Stochastic Variation in Gene Expression Is a Complex Genetic Trait.Flint J, editor. PLoS Genet. 4:e1000049.

Bates D, Maechler M, Bolker B. 2011. lme4: Linear mixed-effects models using S4 classes. Available from: http://CRAN.R-project.org/package=lme4

Blomberg A. 2011. Measuring growth rate in high-throughput growth phenotyping. Curr. Opin. Biotechnol. 22:94–102.

Bloom J, Cross FR. 2007. Multiple levels of cyclin specificity in cell-cycle control. Nat. Rev. Mol. Cell Biol. 8:149–160.

Bloom JS, Ehrenreich IM, Loo WT, Lite T-LV, Kruglyak L. 2013. Finding the sources of missing heritability in a yeast cross. Nature 494:234–237.

Boer VM, Crutchfield CA, Bradley PH, Botstein D, Rabinowitz JD. 2009. Growth-limiting Intracellular Metabolites in Yeast Growing under Diverse Nutrient Limitations. Mol. Biol. Cell 21:198–211.

Botstein D, White RL, Skolnick M, Davis RW. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32:314–331.

Brauer MJ, Huttenhower C, Airoldi EM, Rosenstein R, Matese JC, Gresham D, Boer VM, Troyanskaya OG, Botstein D. 2007. Coordination of Growth Rate, Cell Cycle, Stress Response, and Metabolic Activity in Yeast. Mol. Biol. Cell 19:352–367.

Brauer MJ, Saldanha AJ, Dolinski K, Botstein D. 2005. Homeostatic Adjustment and Metabolic Remodeling in Glucose-limited Yeast Cultures. Mol. Biol. Cell 16:2503–2517.

Broach JR. 2012. Nutritional Control of Growth and Development in Yeast. Genetics 192:73–105.

Broman KW, Wu H, Sen S, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. Bioinforma. Oxf. Engl. 19:889–890.

Brown CJ, Todd KM, Rosenzweig RF. 1998. Multiple duplications of yeast hexose transport genes in response to selection in a glucose-limited environment. Mol. Biol. Evol. 15:931–942.

Buchanan R., Whiting R., Damert W. 1997. When is simple good enough: a comparison of the Gompertz, Baranyi, and three-phase linear models for fitting bacterial growth curves. Food Microbiol. 14:313–326.

Chandler CH, Chari S, Dworkin I. 2013. Does your gene need a background check? How genetic background impacts the analysis of mutations, genes, and evolution. Trends Genet. 29:358–366.

Cheverud JM, Routman EJ. 1995. Epistasis and its contribution to genetic variance components. Genetics 139:1455–1461.

Cross FR, Buchler NE, Skotheim JM. 2011. Evolution of networks and sequences in eukaryotic cell cycle control. Philos. Trans. R. Soc. B Biol. Sci. 366:3532–3544.

Cubillos FA, Billi E, ZöRgö E, Parts L, Fargier P, Omholt S, Blomberg A, Warringer J, Louis EJ, Liti G. 2011. Assessing the complex architecture of polygenic traits in diverged yeast populations. Mol. Ecol. 20:1401–1413.

Cubillos FA, Louis EJ, Liti G. 2009. Generation of a large set of genetically tractable haploid and diploid Saccharomyces strains. FEMS Yeast Res. 9:1217–1225.

Cubillos FA, Parts L, Salinas F, Bergstrom A, Scovacricchi E, Zia A, Illingworth CJR, Mustonen V, Ibstedt S, Warringer J, et al. 2013. High-Resolution Mapping of Complex Traits with a Four-Parent Advanced Intercross Yeast Population. Genetics 195:1141–1155.

Darvasi A, Soller M. 1995. Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199–1207.

Edwards MD, Gifford DK. 2012. High-resolution genetic mapping with pooled sequencing. BMC Bioinformatics 13 Suppl 6:S8.

Ehrenreich IM, Torabi N, Jia Y, Kent J, Martis S, Shapiro JA, Gresham D, Caudy AA, Kruglyak L. 2010. Dissection of genetically complex traits with extremely large pools of yeast segregants. Nature 464:1039–1042.

Eldar A, Elowitz MB. 2010. Functional roles for noise in genetic circuits. Nature 467:167–173.

Falconer DS, Mackay TFC. 1996. Introduction to quantitative genetics. Essex, England: Longman

Fendt S-M, Sauer U. 2010. Transcriptional regulation of respiration in yeast metabolizing differently repressive carbon substrates. BMC Syst. Biol. 4:12.

Ferenci T. 1999. “Growth of bacterial cultures” 50 years on: towards an uncertainty principle instead of constants in bacterial growth kinetics. Res. Microbiol. 150:431–438.

Ferguson ML, Le Coq D, Jules M, Aymerich S, Radulescu O, Declerck N, Royer CA. 2012. Reconciling molecular regulatory mechanisms with noise patterns of bacterial metabolic promoters in induced and repressed states. Proc. Natl. Acad. Sci. U. S. A. 109:155–160.

Ferretti AC, Larocca MC, Favre C. 2012. Nutritional stress in eukaryotic cells: Oxidative species and regulation of survival in time of scarceness. Mol. Genet. Metab. 105:186–192.

Fischer G, James SA, Roberts IN, Oliver SG, Louis EJ. 2000. Chromosomal evolution in Saccharomyces. Nature 405:451–454.

Fontana L, Partridge L, Longo VD. 2010. Extending Healthy Life Span--From Yeast to Humans. Science 328:321–326.

Frank SA. 2011. Natural selection. II. Developmental variability and evolutionary rate*. J. Evol. Biol. 24:2310–2320.

Frank SA, Rosner MR. 2012. Nonheritable Cellular Variability Accelerates the Evolutionary Processes of Cancer. PLoS Biol. 10:e1001296.

Fraser HB, Schadt EE. 2010. The Quantitative Genetics of Phenotypic Robustness.Gibson G, editor. PLoS ONE 5:e8635.

Geiler-Samerotte K, Bauer C, Li S, Ziv N, Gresham D, Siegal M. 2013. The details in the distributions: why and how to study phenotypic variability. Curr. Opin. Biotechnol.

Gerke J, Lorenz K, Cohen B. 2009. Genetic Interactions Between Transcription Factors Cause Natural Variation in Yeast. Science 323:498–501.

Gerke J, Lorenz K, Ramnarine S, Cohen B. 2010. Gene–Environment Interactions at Nucleotide Resolution.Mackay TFC, editor. PLoS Genet. 6:e1001144.

Gerke JP, Chen CTL, Cohen BA. 2006. Natural Isolates of Saccharomyces cerevisiae Display Complex Genetic Variation in Sporulation Efficiency. Genetics 174:985–997.

Giaever G, Chu AM, Ni L, Connelly C, Riles L, Véronneau S, Dow S, Lucau-Danila A, Anderson K, André B, et al. 2002. Functional profiling of the Saccharomyces cerevisiae genome. Nature 418:387–391.

Gresham D, Boer VM, Caudy A, Ziv N, Brandt NJ, Storey JD, Botstein D. 2011. System-Level Analysis of Genes and Functions Affecting Survival During Nutrient Starvation in Saccharomyces cerevisiae. Genetics 187:299–317.

Gresham D, Desai MM, Tucker CM, Jenq HT, Pai DA, Ward A, DeSevo CG, Botstein D, Dunham MJ. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. PLoS Genet. 4:e1000303.

Gutteridge A, Pir P, Castrillo JI, Charles PD, Lilley KS, Oliver SG. 2010. Nutrient control of eukaryote cell growth: a systems biology study in yeast. BMC Biol. 8:68.

Hall DW, Joseph SB. 2010. A High Frequency of Beneficial Mutations Across Multiple Fitness Components in Saccharomyces cerevisiae. Genetics 185:1397–1409.

Hall MC, Dworkin I, Ungerer MC, Purugganan M. 2007. Genetics of microenvironmental canalization in Arabidopsis thaliana. Proc. Natl. Acad. Sci. U. S. A. 104:13717–13722.

Hillenmeyer ME, Fung E, Wildenhain J, Pierce SE, Hoon S, Lee W, Proctor M, St.Onge RP, Tyers M, Koller D, et al. 2008. The Chemical Genomic Portrait of Yeast: Uncovering a Phenotype for All Genes. Science 320:362–365.

Hill WG, Mulder HA. 2010. Genetic analysis of environmental variation. Genet. Res. 92:381–395.

Hittinger CT. 2013. Saccharomyces diversity and evolution: a budding model genus. Trends Genet. TIG 29:309–317.

Hou J, Friedrich A, de Montigny J, Schacherer J. 2014. Chromosomal Rearrangements as a Major Mechanism in the Onset of Reproductive Isolation in Saccharomyces cerevisiae. Curr. Biol. 24:1153–1159.

Huh W-K, Falvo JV, Gerke LC, Carroll AS, Howson RW, Weissman JS, O'Shea EK. 2003. Global analysis of protein localization in budding yeast. Nature 425:686–691.

Illingworth CJR, Parts L, Bergström A, Liti G, Mustonen V. 2013. Inferring Genome-Wide Recombination Landscapes from Advanced Intercross Lines: Application to Yeast Crosses.Lichten M, editor. PLoS ONE 8:e62266.

Jimenez-Gomez JM, Corwin JA, Joseph B, Maloof JN, Kliebenstein DJ. 2011. Genomic Analysis of QTLs and Genes Altering Natural Variation in Stochastic Noise.Gibson G, editor. PLoS Genet. 7:e1002295.

Johnston GC, Pringle JR, Hartwell LH. 1977. Coordination of growth with cell division in the yeast Saccharomyces cerevisiae. Exp. Cell Res. 105:79–98.

Johnston IG, Gaal B, Neves RP das, Enver T, Iborra FJ, Jones NS. 2012. Mitochondrial Variability as a Source of Extrinsic Cellular Noise.Haugh JM, editor. PLoS Comput. Biol. 8:e1002416.

Kaniak A, Xue Z, Macool D, Kim J-H, Johnston M. 2004. Regulatory Network Connecting Two Glucose Signal Transduction Pathways in Saccharomyces cerevisiae. Eukaryot. Cell 3:221–231.

Kovárová-Kovar K, Egli T. 1998. Growth kinetics of suspended microbial cells: from single-substrate-controlled growth to mixed-substrate kinetics. Microbiol. Mol. Biol. Rev. MMBR 62:646–666.

Kussell E, Leibler S. 2005. Phenotypic Diversity, Population Growth, and Information in Fluctuating Environments. Science 309:2075–2078.

Lander ES, Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics 121:185–199.

Lee H-Y, Chou J-Y, Cheong L, Chang N-H, Yang S-Y, Leu J-Y. 2008. Incompatibility of Nuclear and Mitochondrial Genomes Causes Hybrid Sterility between Two Yeast Species. Cell 135:1065–1073.

Legendre P. 2011. lmodel2: Model II Regression. Available from: http://CRAN.R-project.org/package=lmodel2

Lehner B. 2011. Molecular mechanisms of epistasis within and between genes. Trends Genet. 27:323–331.

Lehner B. 2013. Genotype to phenotype: lessons from model organisms for human genetics. Nat. Rev. Genet. 14:168–178.

Levy SF, Siegal ML. 2008. Network Hubs Buffer Environmental Variation in Saccharomyces cerevisiae.Levchenko A, editor. PLoS Biol. 6:e264.

Levy SF, Siegal ML. 2012. The Robustness Continuum. In: Soyer OS, editor. Evolutionary Systems Biology. Vol. 751. New York, NY: Springer New York. p. 431–452. Available from: http://www.springerlink.com/index/10.1007/978-1-4614-3567-9_20

Levy SF, Ziv N, Siegal ML. 2012. Bet Hedging in Yeast by Heterogeneous, Age-Correlated Expression of a Stress Protectant.Hurst LD, editor. PLoS Biol. 10:e1001325.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinforma. Oxf. Engl. 25:1754–1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinforma. Oxf. Engl. 25:2078–2079.

Liti G, Carter DM, Moses AM, Warringer J, Parts L, James SA, Davey RP, Roberts IN, Burt A, Koufopanou V, et al. 2009. Population genomics of domestic and wild yeasts. Nature 458:337–341.

Liti G, Louis EJ. 2012. Advances in Quantitative Trait Analysis in Yeast.Fay JC, editor. PLoS Genet. 8:e1002912.

Liu Y. 2007. Overview of some theoretical approaches for derivation of the Monod equation. Appl. Microbiol. Biotechnol. 73:1241–1250.

Loewith R, Hall MN. 2011. Target of Rapamycin (TOR) in Nutrient Signaling and Growth Control. Genetics 189:1177–1201.

Losick R, Desplan C. 2008. Stochasticity and Cell Fate. Science 320:65–68.

Mackay TFC. 2013. Epistasis and quantitative traits: using model organisms to study gene–gene interactions. Nat. Rev. Genet. 15:22–33.

Mackay TFC, Lyman RF. 2005. Drosophila bristles and the nature of quantitative genetic variation. Philos. Trans. R. Soc. B Biol. Sci. 360:1513–1527.

Mackay TFC, Stone EA, Ayroles JF. 2009. The genetics of quantitative traits: challenges and prospects. Nat. Rev. Genet. 10:565–577.

Marie-Nelly H, Marbouty M, Cournac A, Flot J-F, Liti G, Parodi DP, Syan S, Guillén N, Margeot A, Zimmer C, et al. 2014. High-quality genome (re)assembly using chromosomal contact data. Nat. Commun. 5:5695.

Michelmore RW, Paran I, Kesseli RV. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proc. Natl. Acad. Sci. 88:9828–9832.

Monod J. 1949. The Growth of Bacterial Cultures. Annu. Rev. Microbiol. 3:371–394.

Monod J. 1950. La technique de culture continue, theorie et applications. Ann Inst Pasteur 79:390–410.

Mulder HA, Bijma P, Hill WG. 2007. Prediction of Breeding Values and Selection Responses With Genetic Heterogeneity of Environmental Variance. Genetics 175:1895–1910.

Naumov GI, Serpova EV, Naumova ES. 2006. A genetically isolated population of Saccharomyces cerevisiae in Malaysia. Mikrobiologiia 75:245–249.

Nelson RM, Pettersson ME, Carlborg Ö. 2013. A century after Fisher: time for a new paradigm in quantitative genetics. Trends Genet. 29:669–676.

Nelson RM, Pettersson ME, Li X, Carlborg Ã. 2013. Variance Heterogeneity in Saccharomyces cerevisiae Expression Data: Trans-Regulation and Epistasis.Peddada SD, editor. PLoS ONE 8:e79507.

Newcomb LL, Diderich JA, Slattery MG, Heideman W. 2003. Glucose Regulation of Saccharomyces cerevisiae Cell Cycle Genes. Eukaryot. Cell 2:143–149.

Novick A, Szilard L. 1950. Description of the Chemostat. Science 112:715–716.

Parts L, Cubillos FA, Warringer J, Jain K, Salinas F, Bumpstead SJ, Molin M, Zia A, Simpson JT, Quail MA, et al. 2011. Revealing the genetic structure of a trait by sequencing a population under selection. Genome Res. 21:1131–1138.

Peleg M, Corradini MG. 2011. Microbial Growth Curves: What the Models Tell Us and What They Cannot. Crit. Rev. Food Sci. Nutr. 51:917–945.

Pelkmans L. 2012. Using Cell-to-Cell Variability--A New Era in Molecular Biology. Science 336:425–426.

Ramachandran V, Herman PK. 2011. Antagonistic Interactions Between the cAMP-Dependent Protein Kinase and Tor Signaling Pathways Modulate Cell Growth in Saccharomyces cerevisiae. Genetics 187:441–454.

Raser JM, O'Shea EK. 2005. Noise in gene expression: origins, consequences, and control. Science 309:2010–2013.

Regenberg B, Grotkjaer T, Winther O, Fausbøll A, Akesson M, Bro C, Hansen LK, Brunak S, Nielsen J. 2006. Growth-rate regulated genes have profound impact on interpretation of transcriptome profiling in Saccharomyces cerevisiae. Genome Biol. 7:R107.

Reifenberger E, Boles E, Ciriacy M. 1997. Kinetic Characterization of Individual Hexose Transporters of Saccharomyces Cerevisiae and their Relation to the Triggering Mechanisms of Glucose Repression. Eur. J. Biochem. 245:324–333.

Rockman MV. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. Evolution 66:1–17.

Ronnegard L, Valdar W. 2011. Detecting Major Genetic Loci Controlling Phenotypic Variability in Experimental Crosses. Genetics 188:435–447.

Saldanha AJ, Brauer MJ, Botstein D. 2004. Nutritional Homeostasis in Batch and Steady-State Culture of Yeast. Mol. Biol. Cell 15:4089–4104.

Scheiner SM, Lyman RF. 1989. The genetics of phenotypic plasticity I. Heritability. J. Evol. Biol. 2:95–107.

Scott M, Hwa T. 2011. Bacterial growth laws and their applications. Curr. Opin. Biotechnol. 22:559–565.

Shen X, Pettersson M, Rönnegård L, Carlborg Ö. 2012. Inheritance Beyond Plain Heritability: Variance-Controlling Genes in Arabidopsis thaliana.Barsh GS, editor. PLoS Genet. 8:e1002839.

Shi L, Tu BP. 2013. Acetyl-CoA induces transcription of the key G1 cyclin CLN3 to promote entry into the cell division cycle in Saccharomyces cerevisiae. Proc. Natl. Acad. Sci. 110:7318–7323.

Silverman SJ, Petti AA, Slavov N, Parsons L, Briehof R, Thiberge SY, Zenklusen D, Gandhi SJ, Larson DR, Singer RH, et al. 2010. Metabolic cycling in single yeast cells from unsynchronized steady-state populations limited on glucose or phosphate. Proc. Natl. Acad. Sci. 107:6946–6951.

Smets B, Ghillebert R, De Snijder P, Binda M, Swinnen E, De Virgilio C, Winderickx J. 2010. Life in the midst of scarcity: adaptations to nutrient availability in Saccharomyces cerevisiae. Curr. Genet. 56:1–32.

Smith EN, Kruglyak L. 2008. Gene–Environment Interaction in Yeast Gene Expression.Mackay T, editor. PLoS Biol. 6:e83.

Snoep JL, Mrwebi M, Schuurmans JM, Rohwer JM, Teixeira de Mattos MJ. 2009. Control of specific growth rate in Saccharomyces cerevisiae. Microbiology 155:1699–1707.

Son S, Tzur A, Weng Y, Jorgensen P, Kim J, Kirschner MW, Manalis SR. 2012. Direct observation of mammalian cell growth and size regulation. Nat. Methods 9:910–912.

Steinmetz LM, Sinha H, Richards DR, Spiegelman JI, Oefner PJ, McCusker JH, Davis RW. 2002. Dissecting the architecture of a quantitative trait locus in yeast. Nature 416:326–330.

Stewart-Ornstein J, Weissman JS, El-Samad H. 2012. Cellular Noise Regulons Underlie Fluctuations in Saccharomyces cerevisiae. Mol. Cell 45:483–493.

Swinnen IAM, Bernaerts K, Dens EJJ, Geeraerd AH, Van Impe JF. 2004. Predictive modelling of the microbial lag phase: a review. Int. J. Food Microbiol. 94:137–159.

Swinnen S, Schaerlaekens K, Pais T, Claesen J, Hubmann G, Yang Y, Demeke M, Foulquie-Moreno MR, Goovaerts A, Souvereyns K, et al. 2012. Identification of novel causative genes determining the complex trait of high ethanol tolerance in yeast using pooled-segregant whole-genome sequence analysis. Genome Res. 22:975–984.

Di Talia S, Skotheim JM, Bean JM, Siggia ED, Cross FR. 2007. The effects of molecular noise and size control on variability in the budding yeast cell cycle. Nature 448:947–951.

Taylor MB, Ehrenreich IM. 2014. Genetic Interactions Involving Five or More Genes Contribute to a Complex Trait in Yeast.Fay JC, editor. PLoS Genet. 10:e1004324.

Taylor MB, Ehrenreich IM. 2015. Higher-order genetic interactions and their contribution to complex traits. Trends Genet. 31:34–40.

Team RDC. 2012. R: A Language and Environment for Statistical Computing. Vienna, Austria Available from: http://www.R-project.org/

Wang X, Kruglyak L. 2014. Genetic Basis of Haloperidol Resistance in Saccharomyces cerevisiae Is Complex and Dose Dependent.Copenhaver GP, editor. PLoS Genet. 10:e1004894.

Westerhoff HV, Lolkema JS, Otto R, Hellingwerf KJ. 1982. Thermodynamics of growth. Non-equilibrium thermodynamics of bacterial growth. The phenomenological and the mosaic approach. Biochim. Biophys. Acta 683:181–220.

Wilkening S, Lin G, Fritsch ES, Tekkedil MM, Anders S, Kuehn R, Nguyen M, Aiyar RS, Proctor M, Sakhanenko NA, et al. 2014. An evaluation of high-throughput approaches to QTL mapping in Saccharomyces cerevisiae. Genetics 196:853–865.

Wolfinger RD, Gibson G, Wolfinger ED, Bennett L, Hamadeh H, Bushel P, Afshari C, Paules RS. 2001. Assessing Gene Significance from cDNA Microarray Expression Data via Mixed Models. J. Comput. Biol. 8:625–637.

Yin Z, Wilson S, Hauser NC, Tournu H, Hoheisel JD, Brown AJP. 2003. Glucose triggers different global responses in yeast, depending on the strength of the signal, and transiently stabilizes ribosomal protein mRNAs. Mol. Microbiol. 48:713–724.

Zaman S, Lippman SI, Zhao X, Broach JR. 2008. How Saccharomyces responds to nutrients. Annu. Rev. Genet. 42:27–81.

Zenklusen D, Larson DR, Singer RH. 2008. Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. 38 Mol. Biol. 15:1263–1271.

Ziv N, Brandt NJ, Gresham D. 2013. The Use of Chemostats in Microbial Systems Biology. J. Vis. Exp. [Internet]. Available from: http://www.jove.com/video/50168/the-use-of-chemostats-in-microbial-systems-biology

Ziv N, Siegal ML, Gresham D. 2013. Genetic and non-genetic determinants of cell-growth variation assessed by high-throughput microscopy. Mol. Biol. Evol. [Internet]. Available from: http://mbe.oxfordjournals.org/cgi/doi/10.1093/molbev/mst138

Zorgo E, Gjuvsland A, Cubillos FA, Louis EJ, Liti G, Blomberg A, Omholt SW, Warringer J. 2012. Life History Shapes Trait Heredity by Accumulation of Loss-of-Function Alleles in Yeast. Mol. Biol. Evol. 29:1781–1789.