Experimental Evolution of Cellular Response Networks in Yeast

Permanent link

Terms of Use

Share Your Story

Experimental Evolution of

Cellular Response Networks in Yeast

A dissertation presented

by

Nichole Wespe

to

The Department of Molecular and Cellular Biology

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Biochemistry

Harvard University

Cambridge, Massachusetts

February 2017

Dissertation Advisor   Author

Andrew Murray   **Nichole Wespe**

Experimental Evolution of
Cellular Response Networks in Yeast

Abstract

Organisms respond to environmental changes using complex signaling networks, whose evolution has been studied by comparing their structure and function in related organisms. We complemented this approach by using experimental evolution in the budding yeast, *Saccharomyces cerevisiae*, to explore how existing response networks change under selection. We selected for cell cycle arrest in response to an environmental change by alternating between an environment that favors cell proliferation and one that contains high salt (0.5 M NaCl) and hydroxyurea, a replication inhibitor that kills a genetically engineered ancestor if it attempts DNA replication. This selection yielded strains that arrest division in high NaCl. Causal mutations were identified by whole-genome sequencing and confirmed by genetic reconstruction. We identified four different forms of rewiring: mutations that alter three different hexose transporters, allowing them to admit sodium ions into cells; inactivation of *HAL5*, which encodes a protein kinase that regulates potassium channels; inactivation of *HOG1*, which encodes a MAP kinase that detects high external osmolarity; and loss-of-function mutations in *IMP2'*, a poorly characterized transcriptional activator also implicated in ion homeostasis.

Contents

Acknowledgments

The following work was conducted with the guidance of Andrew Murray, to whom I owe my deepest gratitude. Thank you for being an excellent advisor: for treating me as a peer, with respect and kindness, and for providing a true apprenticeship. I have learned much from you about research as both a scientific and a human endeavor. Thank you to my advisory committee, Rachelle Gaudet, Vlad Denic, and Michael Laub, for guidance on which of the many paths to pursue in my research. Your strong suggestion to do more replicates of the same selection likely helped me avoid some serious quicksand. Thank you also to previous committee members Sharad Ramanathan, Chris Marx, and Erin O'Shea, for advice in the early stages of this project.

Thank you to all of the members of the Murray lab with whom I shared this journey. To those who came before me – Erik Hom, Lori Huberman, Edel Hyland, Quincey Justman, John Koschwanez, Liedewij Laan, Derek Lau, Natalie Nannas, Mary Wahl, and Gregg Wildenberg – your collective intellectual rigor, creativity, and openness are a large part of why I joined the lab. Thank you especially to Erik and Quincey for your generosity in helping me navigate the intellectual formulation and experimental instantiation of my project. Thank you again to Quincey for teaching me microscopy techniques and demonstrating every day how good experiments are accomplished; your adage "the data are the data" not only cured me of noun-verb disagreement in my scientific language but

also saved my emotional health many times. Thank you to Liedewij for many supportive conversations about being a scientist; I credit my conference talk to you.

To those who have come after me – Zain Ali, David Armenta, Laura Bagamery, Felix Barber, Miguel Coelho, Marco Fumasoni, Phoebe Hsieh, Sriram Srikant, Patrick Stoddard, and Cara Weisman – thank you for continuing to make the Murray lab a great environment in which to do science. Thank you to Miguel for help with library prep and improving the sequence analysis pipeline; thank you to Marco and Phoebe for developing the growth curve setup and providing feedback on the analysis program.

My scientific education has benefitted in many ways from the Murray-Nelson symbiosis: thank you to the Nelson lab members Wolfram Möbius, Melanie Müller, and Bryan Weinstein, and particular thanks to David Nelson for lessons in how to structure presentation slides. Enormous thanks to Sri and Bryan for their help writing scripts for the Odyssey cluster and developing the programs described in Appendices B and C. You were both selfless with your time and enthusiastic about helping, and you made my first major foray into coding less intimidating and more fun. Thank you also to Bryan for helping me set up LaTeX for writing my dissertation (much to Andrew's chagrin).

I am so fortunate to have been supported in my work by the highly capable team of Beverly Neugeboren and Linda Kefalas. Thank you to Beverly for everything you have done for me – providing training, resources, and counsel – while somehow doing everything for everyone else, too, with patience and caring. Thank you to Linda for always wrangling a time to meet in Andrew's schedule when I needed one and for helping manage the behind-the-scenes support that research requires. Thank you to Sara Amaral for support with lab materials that made my work easier: I used a lot of media and a lot of plates, and you made them all.

Thank you to members of the Desai lab for helpful discussions and guidance, in

List of Figures

List of Tables

Chapter 1

Introduction

Organisms respond to environmental changes

The ability to respond to changes in the environment is a defining feature of all organisms. In order to survive and reproduce, organisms must adjust their physiology to match external conditions, and they have evolved complex methods to do so in regulated and defined ways. Such responses include restoring balance after a stress, adjusting metabolism to the presence of a new nutrient, or defending against a detected predator. For example, the *lac* operon in the bacterium *Escherichia coli* is regulated to express its gene products, enzymes required for lactose utilization, in the presence of lactose and absence of glucose. Information about the status of lactose and glucose in the environment is relayed by a network of interactions among sugars and other small molecules, proteins, and DNA, the architecture of which has been studied for over half a century [1, 2].

A large part of research in biology has been focused on determining the structure of such modern-day cellular response networks to address the fundamental question of

how organisms function. At the same time, scientists have also worked to understand its immediate corollary: in the words of scientist Barry Hall, "How did they get to be that way?" [3]. Information in different biological networks is physically transduced using a common language of interactions among proteins, DNA, RNA, and small molecules, including such activities as phosphorylation, localization, steric hindrance, synthesis and degradation. While the downstream targets of pathways provide the physical changes or components required for the response, the upstream components are messengers that may not have structural characteristics specific to the stimulus. This language of interactions makes it possible to imagine how networks evolve: components of networks are altered by mutations to form new or different interactions. In this chapter, I will discuss how response networks and their evolution have been investigated, recent work identifying "predictive" networks in microorganisms, and finally, how this work inspired the approach I took to study the evolution of response networks in the budding yeast, *Saccharomyces cerevisiae*.

Interaction networks have evolved specific responses

An important characteristic of cellular responses is specificity: organisms have distinct responses to different environmental changes. Specificity is important because alterations to cellular activity have a cost, whether it be the energy expenditure from gene transcription and protein translation or the time spent in adjustment, during which other cellular processes may be paused and competing organisms consume valuable resources [4, 5]. Unnecessary or inappropriate responses could reduce fitness, leading to the refinement of responses over generations by mutation and selection. Specificity in network structure comes from limiting promiscuous interactions with components outside of the

network. For example, Zarrinpar and colleagues investigated the specificity of the yeast kinase Pbs2 for the protein-binding SH3 domain of its biological partner Sho1 [6]. The binding of a non-yeast SH3 domain to Pbs2 *in vitro* was highly correlated with its ability to functionally substitute for the Sho1 SH3 domain *in vivo*. None of the 26 other SH3 domains found in the yeast proteome could bind Pbs2 or functionally replace the Sho1 SH3 domain, and altering the amino acid sequence of Pbs2 to generate non-specific interactions reduced strain fitness, providing evidence for evolutionary selection against promiscuous interactions, rather than just for increased affinity within the network.

This specificity can be achieved by inherent structural characteristics of two interacting partners. For example, bacterial two-component signaling systems consist of pairs of interacting proteins: a histidine kinase (HK) and its target response regulator (RR). Skerker and colleagues identified amino acid residues responsible for conferring the phosphotransfer specificity of an HK for its RR and demonstrated that mutating these residues can switch specificity to a different RR [7]. In other network interactions, specificity is achieved by controlling context via scaffolding or localization. Protein phosphatases with promiscuous catalytic subunits are given substrate specificity *in vivo* through complexing with diverse regulatory subunits [8]. In another example from bacterial two-component signaling systems, hybrid kinases, which contain both kinase and receiver domains in the same protein, rely on this proximity rather than binding-site specificity to limit cross-talk. In a phospho-relay system in *Myxococcus xanthus*, the full-length version of a hybrid HK shows phosphotransfer preference for its internal domain, overriding the kinetic preference of its isolated kinase domain for a separate RR [9]. The same phenomenon was observed for many of the hybrid kinases from *Caulobacter crescentus*, whose isolated kinase domains showed much broader substrate activity *in vitro* than typical, non-hybrid HKs [10].

3

Not only can non-specific interactions produce unfit responses, but they can affect regulatory behavior of the network itself, such as by reducing the free concentrations of protein components. But because networks have evolved their functions in the milieu of other cellular components, the presence of non-specific interactions can be as important to the behavior of a network as the absence. For example, binding of the repressor LacI to the *lac* operator sequence is reduced by its non-specific interactions with the rest of the genome; without these sites, modeling predicts that derepression of the operon would not occur [11]. As organisms adapt to the challenges of their environment, they necessarily do so within their existing system of interactions.

Research approaches to network evolution

A variety of research approaches have been taken to explore cellular response networks, including genetic analysis, the use of "omics" datasets, evolutionary comparisons, synthetic biology, experimental evolution, and combinations thereof. In particular, *Saccharomyces cerevisiae*, as a lab-adapted, genetically-tractable eukaryotic microorganism, has been a workhorse in systems-level analysis of response networks [12]. Profiling of the yeast transcriptome has surveyed the genes expressed under different environmental conditions [13, 14, 15, 16]. Identification of physical protein-protein interactions, first from yeast two-hybrid screens and then from affinity purification and mass spectrometry, has placed proteins into interaction networks [17, 18, 19, 20]. Components of response networks have been identified from data collected over a time period after an environmental change, and regulatory relationships have been inferred from comparisons between wild-type and defined mutants [14, 21, 22]. Alongside the massive increase in the size

of datasets, sophisticated computational analyses have been required to derive network information from gene expression patterns [23, 24].

The "omics" revolution in molecular biology has enabled the collection of vast amounts of data not only for select model organisms but also for many others with a wide range of evolutionary relationships, enabling analysis of how networks differ between species and giving insight into their evolution [25]. Comparisons between organisms has revealed instances of "rewiring" events, in which features of a defined interaction, like binding specificity, differ between homologous proteins [26, 27, 28, 29]. They have also enabled calculations of the relative rates of rewiring in different types of networks [30, 31, 32]. While comparisons between organisms can yield clues to the evolutionary history of these networks and the genetic events that have altered them, they create hypotheses about a single evolutionary trajectory and the selective forces that favored it rather than directly revealing the diversity of potential solutions to a particular selective pressure.

Experimental evolution to study biology

Experimental evolution provides a method to explore the genetic basis of adaptation to specific environmental challenges. Microorganisms can be passaged for hundreds or thousands of generations, such as in the Long Term Evolution Experiment (LTEE) being conducted by Lenski and associates since 1988, currently at over 66,000 generations [33]. Such experiments have described population-level characteristics such as the rate of mean fitness increase and the dynamics of selective sweeps and clonal interference [34, 35, 36]. They have also examined the genetic changes underlying adaptation in a variety of selection conditions, which can be identified by directly comparing the genome sequences of

ancestral and evolved populations [37, 38]. Analysis of replicate cultures can reveal the potential diversity of genetic mechanisms for creating the same phenotype [39].

Most laboratory evolution experiments have focused on adaptation in constant environments and generally use a single metric, competitive fitness, to evaluate evolved strains. Some experiments have directly tackled the evolution of complex behaviors, such as multicellularity and periodic oscillation [40, 41, 42]. Scientists have also combined synthetic biology with experimental evolution to test predictions about the adaptation of regulatory networks to defined selection regimes [43]. In one of many examples, a synthetic gene circuit, producing both fitness costs and benefits in a manner dependent on the addition of drugs to the environment, was engineered into *S. cerevisiae* and evolved under different drug combinations to examine the mutational pathways to improved fitness [44]. Similar work using a synthetic gene circuit in *E. coli* characterized the evolution of its regulatory behavior in variable environments with trade-offs in fitness [45].

Predictive behavior in response networks

Signaling requires sensing an environmental variable and then mounting an appropriate response. Because the molecular events in signal detection, processing, and response all take time, there is an inevitable delay between signal and response. If one stimulus repeatedly occurs before another, organisms can eliminate the delay by adapting so that the first stimulus induces the response normally elicited by the second. In the lifetime of an organism, this change is classical or Pavlovian conditioning, and a similar change over evolutionary time has been called "adaptive prediction" due to the fitness benefit of responding before rather than after an environmental change [46, 47]. Classical condition-

ing in animals involves changes in the wiring of the nervous system, whereas prediction requires rewiring signaling networks to form new connections between a specific stimulus and response. The difference between predictive and direct responses is the temporal relationship between the organism's response and the fitness gained by that response. A fitness benefit from prediction accrues when a second change occurs; it is also contingent on the second change occurring.

I will also define here a difference between prediction and non-predictive cross-protection. For both of these, an organism's fitness in a second environment is increased if it is exposed to a first, different environment. For example, yeast exposed to a heat shock exhibit resistance to otherwise lethal doses of ionizing radiation and ultraviolet light [48, 49]. In non-predictive cross-protection, a response increases fitness in both the environment eliciting the response and the second environment; therefore, loss of the response would decrease fitness in both environments. In contrast, a predictive behavior exists when the response in the first environment yields a neutral or negative fitness effect if the organism does not experience the second environment; loss of the response would have no effect on fitness in the first environment.

Distinguishing between these requires the ability to test the relative fitness of an organism lacking the response in each environment, a challenging feat when the response involves expression changes of several hundred genes. The measurement of gene expression changes in *S. cerevisiae* upon transfer to many different stresses identified a block of 500 to 900 genes which undergo similar changes in all tested stress conditions [14, 16]. This group of gene induction and repression activities was termed the environmental stress response (ESR) and includes the previously-identified "general" stress response of about 50 genes controlled by the transcription factors Msn2 and Msn4. However, subsequent

7

profiling of the yeast deletion collection showed very little overlap between genes induced in a stress condition and genes whose deletions conferred a fitness defect in that condition [50]. Work by Berry and Gasch demonstrated that neither protein synthesis nor Msn2/4 was required for tolerance of an initial stress but was required for acquired protection against a subsequent stress, indicating that much of the environmental stress response may be preparation for future conditions [51].

As one example of predictive behavior in microorganisms, Schild and colleagues investigated the role of genes expressed by the bacterial pathogen *Vibrio cholerae* during late stages of infection [52]. In this work, they describe a class of genes that are not required for infection but contribute to fitness in the post-infection aquatic environment. Strains lacking one of these late-induced genes show no fitness defect during host infection but are out-competed in an aquatic environment similar to that which bacteria experience subsequent to expulsion from the host. The fitness defect is eliminated if the tested strains (both mutant and the wild-type competitor) are not passaged through a host mouse before incubation in the aquatic environment. The authors speculate that gene induction in the host may allow the bacteria to gather resources from the host that are scarce in the aquatic environment. This could also indicate that gene expression in direct response to the environmental change is insufficient or does not occur. In either case, the gene expression profile described is predictive as defined above, in that the fitness benefit accrued from the late-induced gene expression is dependent on the bacteria entering an aquatic environment.

This pattern of regulatory behavior was first described as predictive by Tagkopolous and colleagues in 2008, who presented an *in silico* demonstration of the capacity for networks to evolve prediction in correlated variable environments [46]. Based on knowledge of

E. coli life history, they then looked for and found correlations in gene expression changes between responses to increased temperature and decreased oxygen, which co-occur upon entry to the mammalian gastrointestinal tract. In order to distinguish between cross-protection and prediction, they evolved *E. coli* in varying conditions with the opposite correlation: an increase in temperature was followed by increased oxygen. They observed large fitness gains in evolved strains and a "decoupling" of the original transcriptional responses as measured by transcriptome profiles.

A similar study was conducted by Mitchell and colleagues to look specifically for asymmetric anticipation, in which change A induces preparation for change B but not vice versa [47]. They examined a different correlation in *E. coli* ecology, that of the stereotypical order of carbon sources in the mammalian intestine: lactose precedes maltose in the digestive tract. *E. coli* induced expression of maltose genes after lactose exposure and pre-exposure to lactose improved fitness in maltose, demonstrating characteristics of adaptive prediction in the regulation of maltose. Also, evolution in constant lactose resulted in reduced expression of maltose genes in response to lactose, indicating a fitness cost of the original response. They also tackled the complexities of the yeast stress response by quantifying survival in pairs of sequential stresses. They found that most stress pairs are not symmetrically cross-protective, and the strength of cross-protection correlated with the typical order of stresses experienced by yeast during wine fermentation. Neither Tagkopolous et al. nor Mitchell et al. sequenced the evolved *E. coli* strains to identify mutations responsible for the decoupling.

More recently, Dhar and colleagues examined adaptation to changing environments by experimentally evolving *S. cerevisiae* in alternating exposures to salt stress and oxidative stress every 10 generations [53]. They also evolved yeast in constant environments of

each stress and in no stress. They found fitness increases of all evolved populations when tested in all conditions (salt stress, oxidative stress and no stress), with their relative increases only partly related to their evolved condition. Based on the relative magnitudes of fitness increases, they identify stress-specific adaptation in the salt-evolved (S) and oxidation-evolved (O) populations (i.e., salt-evolved strains are fitter than oxidation-evolved strains in salt) as well as cross-stress adaptation of the O, but not the S, populations. Based on their fitness data, they conclude that the O and alternating-evolved (SO) populations evolved cross-protection against salt stress. Unfortunately, their analysis conflates the asymmetry of the cross-stress adaptation with asymmetric temporal cross-protection (like that seen by Mitchell et al.), and they do not perform the experiments required to clearly demonstrate temporal cross-protection.

Dhar et al. also measured gene expression changes of the evolved populations after 20 minutes in each of the three conditions, and use the relative numbers of overlapping genes with expression changes to claim evidence for asymmetric anticipatory regulation. Their argument is as follows: the SO populations change some number of genes in response to salt that overlap with the genes the O populations change in response to oxidative stress. If this overlap is greater than that between O populations in oxidative stress and S populations in salt, the difference represents O-specific genes that SO populations change in response to salt. In this analysis, they only include genes that changed in regulation (e.g., have increased induction) relative to the ancestor, not all genes that change expression in these conditions. They do find a difference in the number of genes that changed regulation and conclude that the SO populations evolved salt-induced anticipation of oxidative stress. They concede that they "have not been able to disentangle the effects of cross-protection and anticipation on fitness in the cycling SO populations."

10

Experimental evolution of predictive behavior

To directly probe the evolution of response networks, I exploited the logic of predictive behavior to select for a new connection between an environmental stimulus and a cellular response. I devised a selection for a connection between cell cycle arrest and an environmental change. Cell cycle arrest is a physiological state characterized by the absence of progression through the processes necessary to replicate a cell. This was chosen because it is a phenotype that can be selected both for and against. Selection against arrest occurs in an environment that is permissive for proliferation, as arrested cells are outcompeted by replicating cells. Selection for arrest occurs by making replication lethal to the organism. I engineered budding yeast so that I could use a small molecule (hydroxyurea) to kill replicating cells (described in detail in Chapter 2).

Yeast were alternately exposed to two selective environments: replication-permissive and replication-lethal. As a signal of the upcoming change to replication-lethal, sodium chloride was added to the media. In each round of selection, cells were allowed to proliferate in standard media, then placed in high-salt media before being exposed to hydroxyurea, selecting for cells that arrested their cell cycle in response to high salt. I characterized clones from the evolved populations by hydroxyurea survival and population growth in various media. Using genetic tricks, whole-genome sequencing and the extensive database of genomic data available for budding yeast, I identified mutations in the evolved strains and verified causality by engineering the mutations into the ancestral strain. I also sequenced the final populations to characterize their genetic diversity. I assayed the hydroxyurea survival and population growth of the engineered strains to determine the genetic contributions to the evolved phenotypes.

In Chapter 2, I describe preliminary testing of the selection with defined mutants, the execution of the evolution, and the phenotypes of clones from replicate evolved populations. In Chapter 3, I describe the causal mutations identified by whole-genome sequencing and confirmed by strain engineering and the phenotypes conferred by these mutations. Chapter 4 contains details on the experimental procedures used in Chapters 2 and 3. Chapter 5 presents a summary of the results, analysis of the approach and suggestions for refinement, and future applications of this selection in studying the evolution of response networks.

References

[1] J. Monod and M. Cohn, "[Biosynthesis induced by enzymes; enzymatic adaptation].," *Advances in Enzymology and Related Areas of Molecular Biology*, vol. 13, pp. 67–119, 1952.

[2] F. Jacob and J. Monod, "Genetic regulatory mechanisms in the synthesis of proteins.," *Journal of Molecular Biology*, vol. 3, pp. 318–356, June 1961.

[3] B. G. Hall, "Evolution of a regulated operon in the laboratory.," *Genetics*, vol. 101, pp. 335–344, July 1982.

[4] D. I. Andersson and B. R. Levin, "The biological cost of antibiotic resistance.," *Current Opinion in Microbiology*, vol. 2, pp. 489–493, Oct. 1999.

[5] G. I. Lang, A. W. Murray, and D. Botstein, "The cost of gene expression underlies a fitness trade-off in yeast," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 106, pp. 5755–5760, Apr. 2009.

[6] A. Zarrinpar, S.-H. Park, and W. A. Lim, "Optimization of specificity in a cellular protein interaction network by negative selection.," *Nature*, vol. 426, pp. 676–680, Dec. 2003.

[7] J. M. Skerker, B. S. Perchuk, A. Siryaporn, E. A. Lubin, O. Ashenberg, M. Goulian, and M. T. Laub, "Rewiring the specificity of two-component signal transduction systems," *Cell*, vol. 133, pp. 1043–1054, June 2008.

[8] D. M. Virshup and S. Shenolikar, "From Promiscuity to Precision: Protein Phosphatases Get a Makeover," *Molecular Cell*, vol. 33, pp. 537–545, Mar. 2009.

[9] S. Wegener-Feldbrügge and L. Søgaard-Andersen, "The atypical hybrid histidine protein kinase RodK in Myxococcus xanthus: spatial proximity supersedes kinetic preference in phosphotransfer reactions.," *Journal of Bacteriology*, vol. 191, pp. 1765–1776, Mar. 2009.

[10] E. J. Capra, B. S. Perchuk, O. Ashenberg, C. A. Seid, H. R. Snow, J. M. Skerker, and M. T. Laub, "Spatial tethering of kinases to their substrates relaxes evolutionary constraints on specificity.," *Molecular Microbiology*, vol. 86, pp. 1393–1403, Dec. 2012.

[11] P. H. von Hippel, A. Revzin, C. A. Gross, and A. C. Wang, "Non-specific DNA binding of genome regulating proteins as a biological control mechanism: I. The lac operon: equilibrium aspects.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 71, pp. 4808–4812, Dec. 1974.

[12] G. D. Bader, A. Heilbut, B. Andrews, M. Tyers, T. Hughes, and C. Boone, "Functional genomics and proteomics: charting a multidimensional map of the yeast cell.," *Trends in Cell Biology*, vol. 13, pp. 344–356, July 2003.

[13] S. Chu, J. DeRisi, M. Eisen, J. Mulholland, D. Botstein, P. O. Brown, and I. Herskowitz, "The transcriptional program of sporulation in budding yeast.," *Science*, vol. 282, pp. 699–705, Oct. 1998.

[14] A. P. Gasch, P. T. Spellman, C. M. Kao, O. Carmel-Harel, M. B. Eisen, G. Storz, D. Botstein, and P. O. Brown, "Genomic expression programs in the response of yeast cells to environmental changes.," *Molecular Biology of the Cell*, vol. 11, pp. 4241–4257, Dec. 2000.

[15] T. R. Hughes, M. J. Marton, A. R. Jones, C. J. Roberts, R. Stoughton, C. D. Armour, H. A. Bennett, E. Coffey, H. Dai, Y. D. He, M. J. Kidd, A. M. King, M. R. Meyer, D. Slade, P. Y. Lum, S. B. Stepaniants, D. D. Shoemaker, D. Gachotte, K. Chakraburtty, J. Simon, M. Bard, and S. H. Friend, "Functional discovery via a compendium of expression profiles.," *Cell*, vol. 102, pp. 109–126, July 2000.

[16] H. C. Causton, B. Ren, S. S. Koh, C. T. Harbison, E. Kanin, E. G. Jennings, T. I. Lee, H. L. True, E. S. Lander, and R. A. Young, "Remodeling of Yeast Genome Expression in Response to Environmental Changes," *Molecular Biology of the Cell*, vol. 12, pp. 323–337, Feb. 2001.

[17] P. Uetz, L. Giot, G. Cagney, T. A. Mansfield, R. S. Judson, J. R. Knight, D. Lockshon, V. Narayan, M. Srinivasan, P. Pochart, A. Qureshi-Emili, Y. Li, B. Godwin, D. Conover, T. Kalbfleisch, G. Vijayadamodar, M. Yang, M. Johnston, S. Fields, and J. M. Rothberg, "A comprehensive analysis of protein-protein interactions in Saccharomyces cerevisiae.," *Nature*, vol. 403, pp. 623–627, Feb. 2000.

[18] A.-C. Gavin, P. Aloy, P. Grandi, R. Krause, M. Boesche, M. Marzioch, C. Rau, L. J. Jensen, S. Bastuck, B. Dümpelfeld, A. Edelmann, M.-A. Heurtier, V. Hoffman, C. Hoefert, K. Klein, M. Hudak, A.-M. Michon, M. Schelder, M. Schirle, M. Remor, T. Rudi, S. Hooper, A. Bauer, T. Bouwmeester, G. Casari, G. Drewes, G. Neubauer, J. M. Rick, B. Kuster, P. Bork, R. B. Russell, and G. Superti-Furga, "Proteome survey reveals modularity of the yeast cell machinery.," *Nature*, vol. 440, pp. 631–636, Mar. 2006.

[19] N. J. Krogan, G. Cagney, H. Yu, G. Zhong, X. Guo, A. Ignatchenko, J. Li, S. Pu, N. Datta, A. P. Tikuisis, T. Punna, J. M. Peregrín-Alvarez, M. Shales, X. Zhang, M. Davey, M. D. Robinson, A. Paccanaro, J. E. Bray, A. Sheung, B. Beattie, D. P. Richards, V. Canadien, A. Lalev, F. Mena, P. Wong, A. Starostine, M. M. Canete, J. Vlasblom, S. Wu, C. Orsi, S. R. Collins, S. Chandran, R. Haw, J. J. Rilstone, K. Gandi, N. J. Thompson, G. Musso, P. St Onge, S. Ghanny, M. H. Y. Lam, G. Butland, A. M. Altaf-Ul, S. Kanaya, A. Shilatifard, E. O'Shea, J. S. Weissman, C. J. Ingles, T. R. Hughes, J. Parkinson, M. Gerstein, S. J. Wodak, A. Emili, and J. F. Greenblatt, "Global landscape of protein complexes in the yeast Saccharomyces cerevisiae." *Nature*, vol. 440, pp. 637–643, Mar. 2006.

[20] A. Breitkreutz, H. Choi, J. R. Sharom, L. Boucher, V. Neduva, B. Larsen, Z.-Y. Lin, B.-J. Breitkreutz, C. Stark, G. Liu, J. Ahn, D. Dewar-Darch, T. Reguly, X. Tang, R. Almeida, Z. S. Qin, T. Pawson, A.-C. Gingras, A. I. Nesvizhskii, and M. Tyers, "A global protein kinase and phosphatase interaction network in yeast.," *Science*, vol. 328, pp. 1043–1046, May 2010.

[21] K. Natarajan, M. R. Meyer, B. M. Jackson, D. Slade, C. Roberts, A. G. Hinnebusch, and M. J. Marton, "Transcriptional profiling shows that Gcn4p is a master regulator of gene expression during amino acid starvation in yeast.," *Molecular and Cellular Biology*, vol. 21, pp. 4347–4368, July 2001.

[22] A. Szopinska, H. Degand, J.-F. Hochstenbach, J. Nader, and P. Morsomme, "Rapid response of the yeast plasma membrane proteome to salt stress.," *Molecular & Cellular Proteomics*, vol. 10, p. M111.009589, Nov. 2011.

[23] J. Ihmels, S. Bergmann, and N. Barkai, "Defining transcription modules using large-scale gene expression data.," *Bioinformatics*, vol. 20, pp. 1993–2003, Sept. 2004.

[24] A. Gitter, M. Carmi, N. Barkai, and Z. Bar-Joseph, "Linking the signaling cascades and dynamic regulatory networks controlling stress responses.," *Genome Research*, vol. 23, pp. 365–376, Feb. 2013.

[25] D. J. Wohlbach, D. A. Thompson, A. P. Gasch, and A. Regev, "From elements to modules: regulatory evolution in Ascomycota fungi.," *Current Opinion in Genetics & Development*, vol. 19, pp. 571–578, Dec. 2009.

[26] A. E. Tsong, B. B. Tuch, H. Li, and A. D. Johnson, "Evolution of alternative transcriptional circuits with identical logic," *Nature*, vol. 443, pp. 415–420, Sept. 2006.

[27] C. R. Baker, B. B. Tuch, and A. D. Johnson, "Extensive DNA-binding specificity divergence of a conserved transcription regulator.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 108, pp. 7493–7498, May 2011.

[28] E. K. Kruzel, S. S. Giles, and C. M. Hull, "Analysis of Cryptococcus neoformans sexual development reveals rewiring of the pheromone-response network by a change in transcription factor identity.," *Genetics*, vol. 191, pp. 435–449, June 2012.

[29] C. K. Dalal, I. A. Zuleta, K. F. Mitchell, D. R. Andes, H. El-Samad, and A. D. Johnson, "Transcriptional rewiring over evolutionary timescales changes quantitative and qualitative properties of gene expression.," *eLife*, vol. 5, p. 403, Sept. 2016.

[30] P. Beltrao and L. Serrano, "Specificity and evolvability in eukaryotic protein interaction networks.," *PLoS Computational Biology*, vol. 3, p. e25, Feb. 2007.

[31] P. Beltrao, J. C. Trinidad, D. Fiedler, A. Roguev, W. A. Lim, K. M. Shokat, A. L. Burlingame, and N. J. Krogan, "Evolution of phosphoregulation: comparison of phosphorylation patterns across yeast species.," *PLoS Biology*, vol. 7, p. e1000134, June 2009.

[32] C. Shou, N. Bhardwaj, H. Y. K. Lam, K.-K. Yan, P. M. Kim, M. Snyder, and M. B. Gerstein, "Measuring the evolutionary rewiring of biological networks.," *PLoS Computational Biology*, vol. 7, no. 1, p. e1001050, 2011.

[33] R. E. Lenski, M. R. Rose, S. C. Simpson, and S. C. Tadler, "Long-term experimental evolution in Escherichia coli. I. Adaptation and divergence during 2,000 generations," *The American Naturalist*, 1991.

[34] M. M. Desai, D. S. Fisher, and A. W. Murray, "The speed of evolution and maintenance of variation in asexual populations," *Current Biology*, vol. 17, pp. 385–394, Mar. 2007.

[35] J. E. Barrick, D. S. Yu, S. H. Yoon, H. Jeong, T. K. Oh, D. Schneider, R. E. Lenski, and J. F. Kim, "Genome evolution and adaptation in a long-term experiment with Escherichia coli," *Nature*, vol. 461, pp. 1243–1247, Oct. 2009.

[36] G. I. Lang, D. Botstein, and M. M. Desai, "Genetic variation and the fate of beneficial mutations in asexual populations.," *Genetics*, vol. 188, pp. 647–661, July 2011.

[37] D. Gresham, M. M. Desai, C. M. Tucker, H. T. Jenq, D. A. Pai, A. Ward, C. G. DeSevo, D. Botstein, and M. J. Dunham, "The Repertoire and Dynamics of Evolutionary Adaptations to Controlled Nutrient-Limited Environments in Yeast," *PLoS Genetics*, vol. 4, p. e1000303, Dec. 2008.

[38] C. L. Araya, C. Payen, M. J. Dunham, and S. Fields, "Whole-genome sequencing of a laboratory-evolved yeast strain.," *BMC Genomics*, vol. 11, p. 88, Feb. 2010.

[39] H.-H. Chou and C. J. Marx, "Optimization of gene expression through divergent mutational paths.," *CellReports*, vol. 1, pp. 133–140, Feb. 2012.

[40] W. C. Ratcliff, R. F. Denison, M. Borrello, and M. Travisano, "Experimental evolution of multicellularity.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 109, pp. 1595–1600, Jan. 2012.

[41] J. H. Koschwanez, K. R. Foster, and A. W. Murray, "Improved use of a public good selects for the evolution of undifferentiated multicellularity.," *eLife*, vol. 2, p. e00367, Apr. 2013.

[42] G. A. Wildenberg and A. W. Murray, "Evolving a 24-hr oscillator in budding yeast," *eLife*, vol. 3, pp. 1082–17, Nov. 2014.

[43] S. G. Peisajovich, "Evolutionary synthetic biology.," *ACS Synthetic Biology*, vol. 1, pp. 199–210, June 2012.

[44] C. González, J. C. J. Ray, M. Manhart, R. M. Adams, D. Nevozhay, A. V. Morozov, and G. Balázsi, "Stress-response balance drives the evolution of a network module and its host genome.," *Molecular Systems Biology*, vol. 11, p. 827, Aug. 2015.

[45] F. J. Poelwijk, M. G. J. de Vos, and S. J. Tans, "Tradeoffs and optimality in the evolution of gene regulation.," *Cell*, vol. 146, pp. 462–470, Aug. 2011.

[46] I. Tagkopoulos, Y.-C. Liu, and S. Tavazoie, "Predictive behavior within microbial genetic networks.," *Science*, vol. 320, pp. 1313–1317, June 2008.

[47] A. Mitchell, G. H. Romano, B. Groisman, A. Yona, E. Dekel, M. Kupiec, O. Dahan, and Y. Pilpel, "Adaptive prediction of environmental changes by microorganisms.," *Nature*, vol. 460, pp. 220–224, July 2009.

[48] R. E. Mitchel and D. P. Morrison, "Heat-shock induction of ionizing radiation resistance in Saccharomyces cerevisiae, and correlation with stationary growth phase.," *Radiation Research*, vol. 90, pp. 284–291, May 1982.

[49] R. E. Mitchel and D. P. Morrison, "Heat-shock induction of ultraviolet light resistance in Saccharomyces cerevisiae.," *Radiation Research*, vol. 96, pp. 95–99, Oct. 1983.

[50] G. Giaever, A. M. Chu, L. Ni, C. Connelly, L. Riles, S. Véronneau, S. Dow, A. Lucau-Danila, K. Anderson, B. André, A. P. Arkin, A. Astromoff, M. El-Bakkoury, R. Bangham, R. Benito, S. Brachat, S. Campanaro, M. Curtiss, K. Davis, A. Deutschbauer, K.-D. Entian, P. Flaherty, F. Foury, D. J. Garfinkel, M. Gerstein, D. Gotte, U. Güldener, J. H. Hegemann, S. Hempel, Z. Herman, D. F. Jaramillo, D. E. Kelly, S. L. Kelly, P. Kötter, D. LaBonte, D. C. Lamb, N. Lan, H. Liang, H. Liao, L. Liu, C. Luo, M. Lussier, R. Mao, P. Menard, S. L. Ooi, J. L. Revuelta, C. J. Roberts, M. Rose, P. Ross-Macdonald, B. Scherens, G. Schimmack, B. Shafer, D. D. Shoemaker, S. Sookhai-Mahadeo, R. K. Storms, J. N. Strathern, G. Valle, M. Voet, G. Volckaert, C.-y. Wang, T. R. Ward, J. Wilhelmy, E. A. Winzeler, Y. Yang, G. Yen, E. Youngman, K. Yu, H. Bussey, J. D. Boeke, M. Snyder, P. Philippsen, R. W. Davis, and M. Johnston, "Functional profiling of the Saccharomyces cerevisiae genome.," *Nature*, vol. 418, pp. 387–391, July 2002.

[51] D. B. Berry and A. P. Gasch, "Stress-activated genomic expression changes serve a preparative role for impending stress in yeast.," *Molecular Biology of the Cell*, vol. 19, pp. 4580–4587, Nov. 2008.

[52] S. Schild, R. Tamayo, E. J. Nelson, F. Qadri, S. B. Calderwood, and A. Camilli, "Genes induced late in infection increase fitness of Vibrio cholerae after release into the environment.," *Cell Host & Microbe*, vol. 2, pp. 264–277, Oct. 2007.

[53] R. Dhar, R. Sägesser, C. Weikert, and A. Wagner, "Yeast adapts to a changing stressful environment by evolving cross-protection and anticipatory gene regulation.," *Molecular Biology and Evolution*, vol. 30, pp. 573–588, Mar. 2013.

Chapter 2

Selection for Altered Regulation of Cellular Response Networks

2.1 Abstract

I set out to create a selection for a new or altered connection between an environmental change and cellular response using the budding yeast, *Saccharomyces cerevisiae*. The scheme exploits the ability to select on two different states of the cell cycle, replicating and arrested. In the selection, yeast are exposed to two alternating environments: the first selects against cell cycle arrest by being permissive for growth, and the second selects for cell cycle arrest by killing cells that are replicating their DNA. I added 0.5 M NaCl as an environmental change to signal the upcoming switch. Cells that arrest the cell cycle during exposure to high salt are protected from death in the second environment. To kill replicating cells, I used *mec1Δ sml1Δ* yeast, which are deficient in the DNA replication checkpoint and thus susceptible to the drug hydroxyurea. Eleven populations

were passaged through the selection cycle, and clones from each final population were assayed for their ability to survive hydroxyurea. Evolved clones have differing specificity for environmental signals and distinct population growth dynamics in high salt despite having similar levels of survival, implying diversity in the underlying genetic causes.

2.2 Introduction

The selective pressure for organisms to match their activities to environmental conditions has resulted in complex regulatory behaviors. Direct regulation involves an organism first sensing a stimulus or a change in the environment and then mounting an appropriate response. However, the time required for sensing and responding results in a delay. Environmental variation at regular intervals selects for organisms that match their physiological regulation to the cycle, thereby reducing the delay. An example is the circadian clock, which regulates physiology in sync with the 24-hour day/night cycle. Recent work has demonstrated that, in environments where changes in two or more variables are coupled, a change in one variable can induce a transcriptional response for the other variable in microorganisms [1, 2]. In other words, single-celled organisms can encode information about correlations in their environment within their gene regulatory networks. This phenomenon has been termed "adaptive prediction" due to the fitness benefit of responding before an environmental challenge.

The research described here seeks to better understand the genetic events that enable this information about the environment to become encoded in the genome by directly evolving a new connection between an environmental stimulus and a cellular response.

Here, I describe an experimental evolution setup that uses the logic of adaptive prediction to select for altered regulation of cellular processes in the single-celled eukaryote *Saccharomyces cerevisiae* (budding yeast). The budding yeast is well-suited for experimental evolution: it combines a short generation time, ease of culturing, well-characterized cell biology, and sophisticated tools for genetic analysis [3, 4].

To select for a connection between a stimulus and a response, I needed a response phenotype that provided a fitness benefit in an manner dependent on the environment, such as gene transcription to enable utilization of a carbon source or resistance to a noxious agent. To be selected against, the response would also have to create a fitness burden in a different environment. I was skeptical that the fitness effect of such behaviors may not be large enough to effectively select against a constitutive response in a short laboratory evolution experiment, and so I chose to select on the state of the cell division cycle. The cell cycle is a well-studied cellular behavior that has many points of regulation, responding to both internal and external factors [5, 6]. A growth-permissive environment selects strongly against cell cycle arrest, as arrested cells are outcompeted by replicating cells[1]. Cell cycle arrest can be selected for by killing replicating cells; this requires using known yeast mutants that are drug-sensitive during replication.

To kill replicating cells, we took advantage of the importance of cell division cycle checkpoints in maintaining viability when crucial cellular processes are perturbed [7]. Cells with mutations in a cell cycle checkpoint will proceed through cell division in inappropriate conditions. The failure to stop can be fatal, creating a strong selection for cell cycle arrest by another mechanism. For example, the DNA replication checkpoint prevents cells from

---

[1]Indeed, a permanent cell cycle arrest is equivalent to inviability.
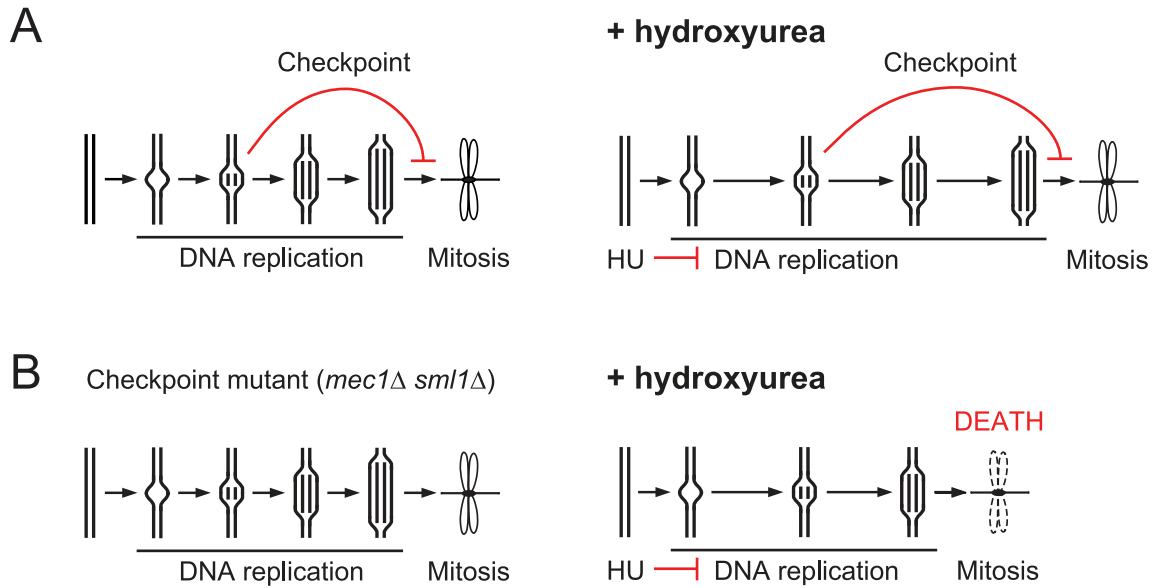
Figure 2.1 Mechanism of hydroxyurea lethality in $mec1^-$ $sml1^-$ yeast.
(A) The DNA replication checkpoint blocks progression to M phase until the genome is fully replication. Checkpoint-competent yeast extend S phase in the presence of hydroxyurea, a replication inhibitor, to complete replication before entering mitosis.
(B) The DNA replication checkpoint protein Mec1 is dispensible in the absence of Sml1 for most replication cycles. In the presence of hydroxyurea, checkpoint-deficient yeast will proceed to mitosis without a fully replication genome, resulting in death of the cell.

entering mitosis until the genome is fully replicated (Figure 2.1A) [8]. Wild-type cells treated with hydroxyurea (HU), a DNA replication inhibitor, trigger this checkpoint to stabilize stalled replication forks and extend S phase. HU exerts its inhibitory effect on DNA replication by inactivating ribonucleotide reductase (RNR), the enzyme complex that generates deoxyribonucleotides from ribonucleotides, thereby interfering with the production of raw material for DNA synthesis [9]. Mec1 is an essential protein involved in activating the checkpoint and preventing the collapse of stalled replication forks [10, 11, 12, 13]. Yeast lacking Mec1 function are viable if Sml1, a repressor of RNR, is also absent; however, $mec1^-$ $sml1^-$ yeast die if DNA replication is perturbed by treatment with HU (Figure 2.1B) [11, 14]. Lethality results from the irreversible collapse of stalled

A Selection against and for arrest
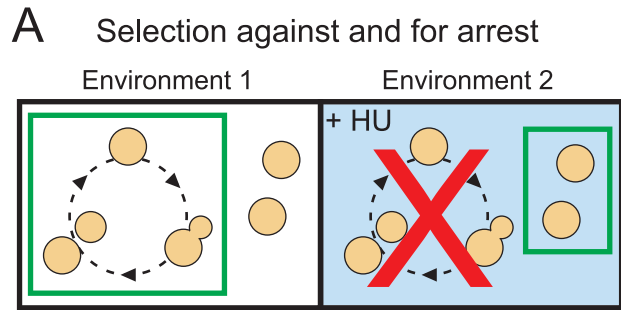
Figure 2.2 Alternating selective environments to evolve a regulated response.

(A) Environment 1, without HU, selects for cell growth and division. Environment 2, containing HU (blue shading), selects for cell cycle arrest.
(B) Cells are passaged through multiple cycles of alternating selection for and against cell cycle arrest; selection for arrest is preceded by the addition of a signal.
(C) The goal of the selection is for evolved cells to survive HU by arrest only in the presence of the signal.
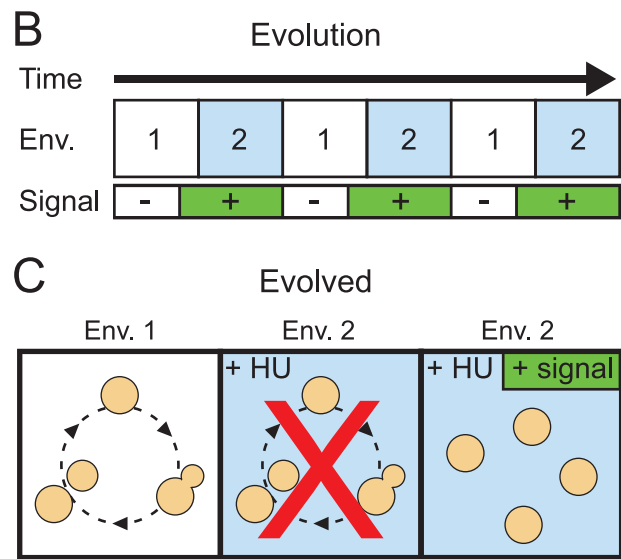
replication forks followed by entering mitosis with unreplicated DNA. I used the lethal effects of HU on *mec1⁻ sml1⁻* yeast to select for cells with the cell cycle arrested.

With a selection both for and against a given phenotype, I can subject yeast to two environments with opposite selective pressures: a replication-permissive environment selecting for cell growth and division, and a replication-lethal environment selecting for cell cycle arrest (Figure 2.2A). The selection cycle couples the alternation of selective pressure with the addition of a signal before the change in environment (Figure 2.2B). The selection cycle is designed to evolve cells that arrest in the presence of the signal (Figure 2.2C). Another option that would also pass the selection is for evolved cells to

22

survive HU regardless of the signal, either by arresting directly in response to HU or by attaining resistance in a way that enables continued replication. However, no resistance of checkpoint-deficient cells to HU has been previously described. A third option would be for evolved cells to stochastically switch between replicating and arrested states, thereby increasing population heterogeneity to always have a subset of cells that could survive HU. This option, known as "bet-hedging," was described to explain the phenomenon of bacterial persistance, in which part of the population is slow-growing and resistant to antibiotics [15, 16].

As a stimulus, I chose 0.5 M NaCl, an osmotic and ionic stress that yeast respond to using known response networks. In yeast, an increase in osmolarity is detected and transduced via the HOG (High Osmolarity Glycerol) pathway, one of several mitogen-activated protein kinase (MAPK) pathways [17]. Mutations in *HOG1* and *PBS2*, two components of the osmolarity response network, were previously shown to yield cross talk between the osmolarity and mating response signaling pathways due to a shared upstream component [18]. Because part of the mating response is cell cycle arrest in G1 phase, loss-of-function mutations in either *HOG1* or *PBS2* were predicted to be selected in the evolution.

Replicate cultures of *mec1Δ sml1Δ* yeast were passaged through multiple cycles alternating selection for and against cell cycle arrest; selection for arrest was preceded by the addition of NaCl to 0.5 M. The evolving populations had rapid increases in fitness, as monitored by cell number after each cycle. Clones characterized from the evolved populations produced different solutions to this selection, evidenced in this chapter by variety in the stimulus specificity of the phenotype and in the population growth dynamics of the clones in high salt.

2.3 Design and tests of the selection cycle

Selection against replicating cells

To select against replicating cells, I constructed a DNA replication checkpoint-deficient strain susceptible to hydroxyurea (HU) treatment by deleting *MEC1* and *SML1*. Survival after HU treatment was determined by counting colony-forming units in a culture before and after exposure (see 4.3 for details). A crucial aspect of killing cells by HU is that cells that are genetically dead at the end of treatment may not be physiologically dead. This precludes use of a live/dead cell staining method to measure survival, instead requiring a colony growth assay. The *mec1Δ sml1Δ* strain had similar levels of survival after HU treatment as an existing *mec1-1 sml1-1* strain (Figure 2.3). This same treatment had no effect on the viability of checkpoint-competent cells. All hydroxyurea survival experiments were performed on the *mec1Δ sml1Δ* strain or a derivative of it. Survival was determined for a range of concentrations and exposure times (Figure 2.3). Exposure to 50 mM HU for four hours, which killed over 99% of the starting population, was used in the evolution experiment. Experiments without error bars were typically performed once; data represent averages of technical duplicates. The level of survival was consistent for similar conditions across experiments testing different variables (e.g., concentration versus time); because these were considered pilot experiments to decide parameters for the evolution, they were not repeated.

Arresting the cell cycle in G1 phase or mitosis before exposure to HU enabled survival of the checkpoint-deficient strain (Figure 2.4). Yeast were arrested in G1 by exposure to the pheromone alpha factor for two hours before adding HU, then plated for survival four
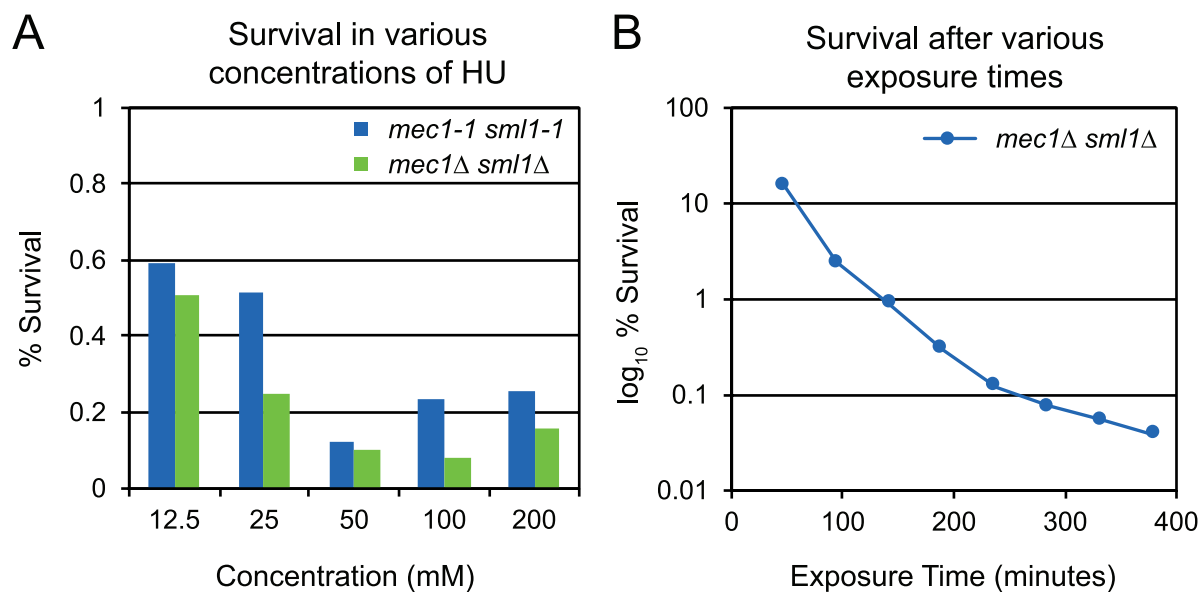
Figure 2.3 Hydroxyurea survival of *mec1⁻ sml1⁻* yeast.
(A) Cultures of each strain were plated after incubating in media with different concentrations of HU for 4 hours. Survival was determined by dividing the number of CFUs after treatment by the number of cells before treatment. Both *mec1-1 sml1-1* and *mec1Δ sml1Δ* strains had survival of less than 1% for all concentrations shown here.
(B) Cultures of *mec1Δ sml1Δ* yeast were plated after incubating in media with 50 mM HU for different times and survival was determined as in (A). Survival is graphed on a log scale to better display low values. Survival rapidly decreases to below 1% after 2 hours of exposure. See section 4.3.1 for more experimental details.

hours later. To arrest cells in mitosis, I used a strain lacking *CDC26*, which displays a temperature-sensitive mitotic arrest. *CDC26* is a gene whose function in the anaphase-promoting complex is essential only at high temperatures; shifting *cdc26Δ* cells to 37°C causes arrest at metaphase [19]. Yeast were arrested in mitosis by incubating at 37°C for two hours before adding HU, then assayed for survival four hours later. Both experiments were internally controlled by co-culturing with a strain that could not arrest, either a *ste2Δ* strain, which lacks the alpha factor receptor, or a *CDC26* strain, which continues growth at 37°C.

Figure 2.4 Hydroxyurea survival with cell cycle arrest.

For G1 arrest, *mec1Δ sml1Δ* yeast were exposed to alpha factor for two hours before adding HU, then plated for survival four hours later. For mitotic arrest, *mec1Δ sml1Δ cdc26Δ* yeast were incubated at 37°C for two hours before adding HU, then plated for survival four hours later. Arrest in either stage greatly increased survival.

Survival with cell cycle arrest
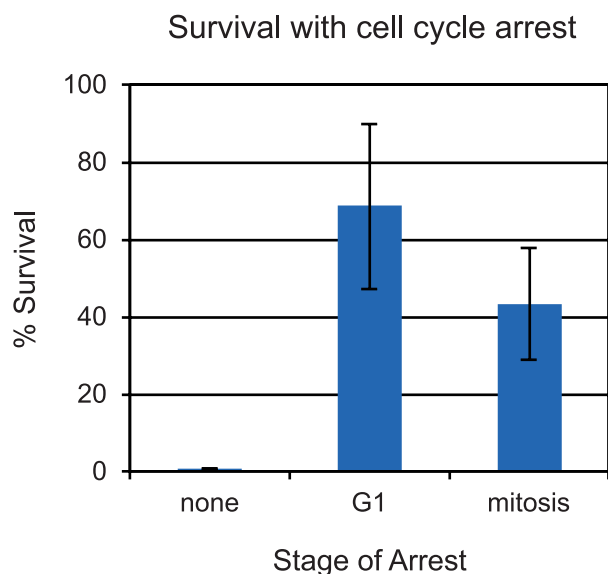
Testing sodium chloride as a stimulus

The stimulus chosen for the first evolution experiment was 0.5 M sodium chloride (NaCl). This creates a moderate level of osmotic and ionic stress and is known to induce a transient pause in the cell cycle while cells adjust their physiology. Mutations in *HOG1* and *PBS2*, two components of the osmolarity response network, were previously shown to yield cross talk between the osmolarity and mating response signaling pathways [18]. One potential solution to our selection is loss of signaling through the osmolarity response network leading to induction of the mating response, which includes cell cycle arrest in G1, and therefore to increased survival in hydroxyurea. I tested whether a *hog1Δ* mutant had increased HU survival when exposed to an increase in osmolarity. Survival did increase in a manner dependent on the presence of NaCl, for a range of NaCl concentrations and pre-exposure times (Figure 2.5A-C). Sorbitol, a nonionic osmolyte, was also tested as a stimulus and resulted in even higher survival than NaCl for *hog1Δ* yeast (Figure 2.5C). However, a stock solution of sorbitol in standard media could only be made to 2 M, forcing

a 1/2 dilution of cultures upon adding the stimulus, whereas NaCl stock solution could be made to 4 M, resulting in a 7/8 dilution of cultures for the equivalent osmotic stress. In order to avoid yeast responding to the possible repletion of nutrients along with the stimulus, I wanted to add the lowest volume of media as possible; this was the primary reason for choosing NaCl as a stimulus instead of sorbitol. The preliminary experiments in Figure 2.5 aided in determining the parameters to be used in the evolution experiment. I aimed to maximize the fold-increase in survival of the *hog1Δ* mutants over the wild-type, so I chose to use a stimulus of 0.5 M NaCl with a pre-exposure time of two hours. The full selection cycle is depicted in Figure 2.6.

I tested whether the full selection cycle, including growth in the absence of salt, would enrich for *hog1Δ* mutants by selecting on a large population of wild-type cells containing a small number of mutant cells. The ratio was measured by plating a sample of the culture after each NaCl/HU treatment, followed by replica-plating to determine the fraction of *hog1Δ* cells. As expected, the relative number of *hog1Δ* cells in the population increased after treatment with NaCl and HU (Figure 2.5D). However, *hog1Δ* cells did not completely take over the population over several rounds of selection. This is probably due to the slower growth rate of *hog1Δ* cells in rich media (see section 3.6). During the growth phase of each cycle, the relative number of *hog1Δ* cells in the population would decrease. For the survival and time parameters of the selection cycle, the increase in survival was roughly balanced by the slower growth over 18 hours. This indicated that while loss of Hog1 function may be one solution to the selection, it was not an "optimal" solution due to the fitness tradeoff between the two selection phases.
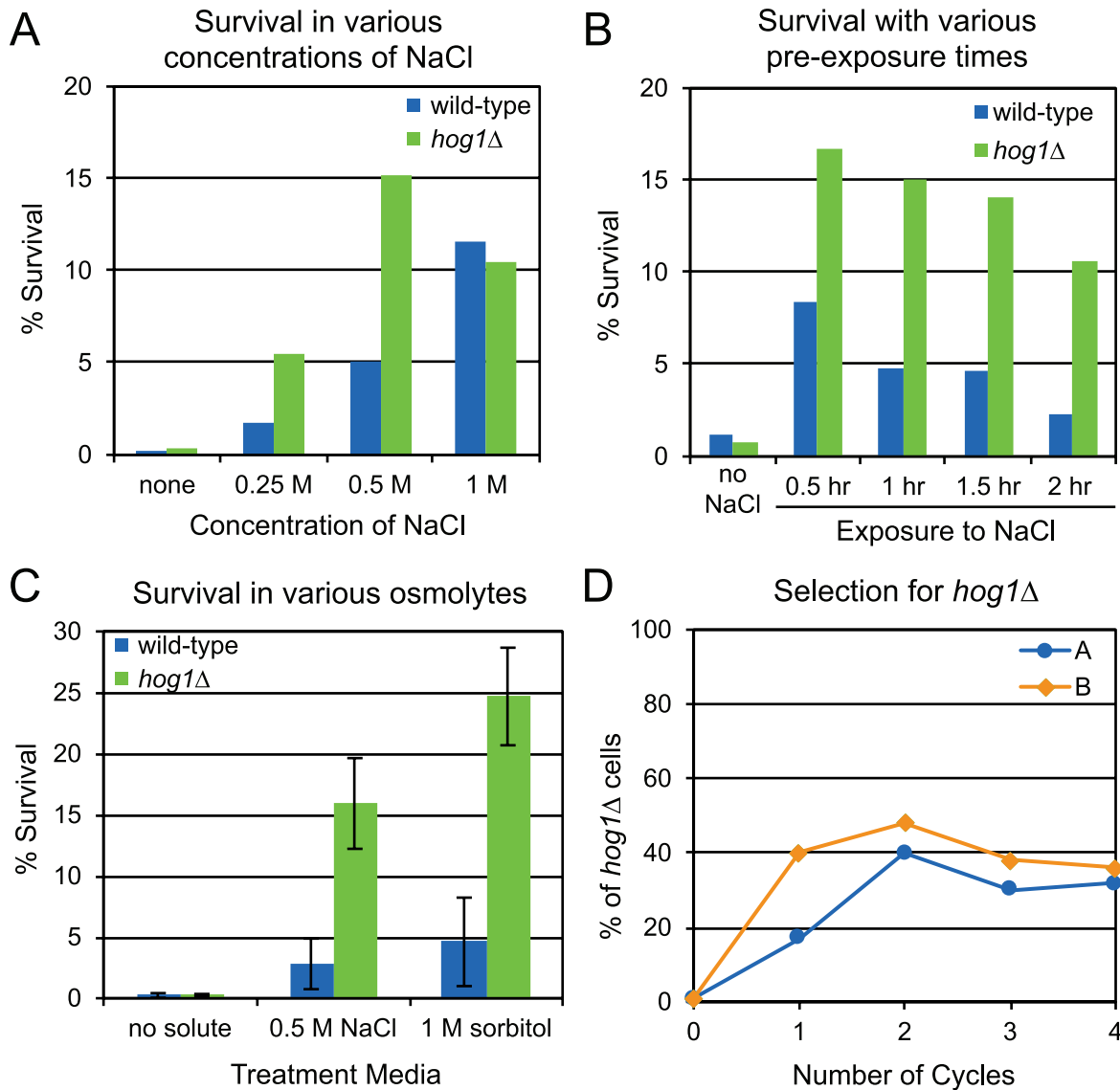
Figure 2.5 Hydroxyurea survival with osmotic stimuli.

(A) Cultures of *mec1Δ sml1Δ* ("wild-type") and *mec1Δ sml1Δ hog1Δ* yeast were placed in media with the indicated concentration of NaCl for 1.5 hours before adding HU.

(B) Cultures were placed in YPD media and NaCl was added to 0.5 M at the indicated times before adding HU. Data for (A) and (B) are averages of technical duplicates.

(C) Cultures were placed in media with either 0.5 M NaCl, 1 M sorbitol, or no additional solute for 2 hours before adding HU. Data are averages of two experiments with technical replicates; error is the sample standard deviation.

(D) *mec1Δ sml1Δ hog1Δ* cells were mixed with *mec1Δ sml1Δ* cells at 1:100 and passaged through multiple NaCl/HU cycles, using 0.5 M NaCl with a 2-hour pre-exposure, separated by overnight growth periods. The fraction of *hog1Δ* cells was determined by plating after each NaCl/HU treatment. Data are shown separately for two technical replicates, A and B. All samples in (A)-(D) were plated for survival four hours after adding HU.

Figure 2.6 Selection cycle using NaCl and hydroxyurea.

For each passage, cells growing in exponential phase in standard media (YPD) were exposed to 0.5 M NaCl for two hours. HU was then added to a final concentration of 50 mM. After four hours in NaCl/HU media, cells were transferred into fresh YPD, diluted and cultured for 18 hours.
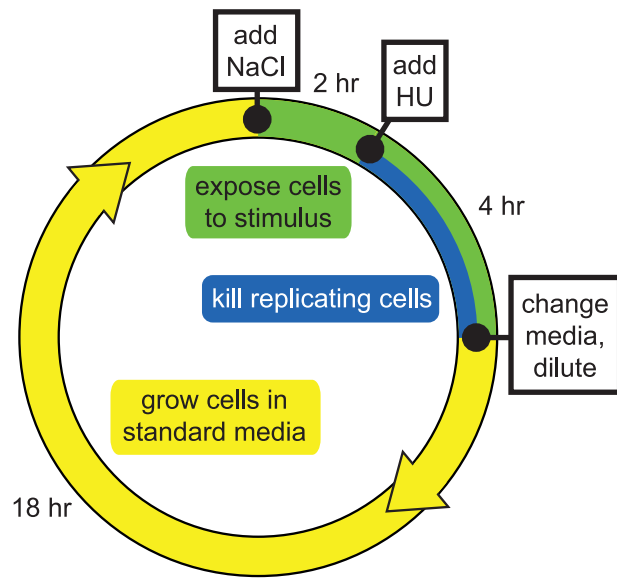
Testing other possible solutions to the selection

Before beginning the selection, I tested two other potential mechanisms of increasing survival. The first was to increase the length of G1 phase relative to the entire cell cycle by deleting *CLN3*, a cyclin governing the G1/S transition [20]. Deletion of *CLN1* and *CLN2*, the two other main G1 cyclins, had been reported to rescue *mec1* lethality in cells with *SML1* [21]. This was posited to be due to prolonged expression of *RNR1*, one of many genes that are upregulated at the G1/S transition, leading to an increase in dNTP levels before the onset of S phase. The *cln3Δ* strain had increased HU survival relative to the *CLN3* strain in both no-solute and NaCl treatment conditions, and its survival in NaCl was almost as high as the *hog1Δ* strain (Figure 2.7A).

The second mechanism tested was to disrupt the ability to physiologically adapt to the osmotic stress by deleting *GPD1*, a glycerol-3-phosphate dehydrogenase important in restoring osmotic balance by producing glycerol [22]. *GPD1* expression is upregulated by Hog1 upon osmotic shock and its activity is important for growth in high osmolarity. The
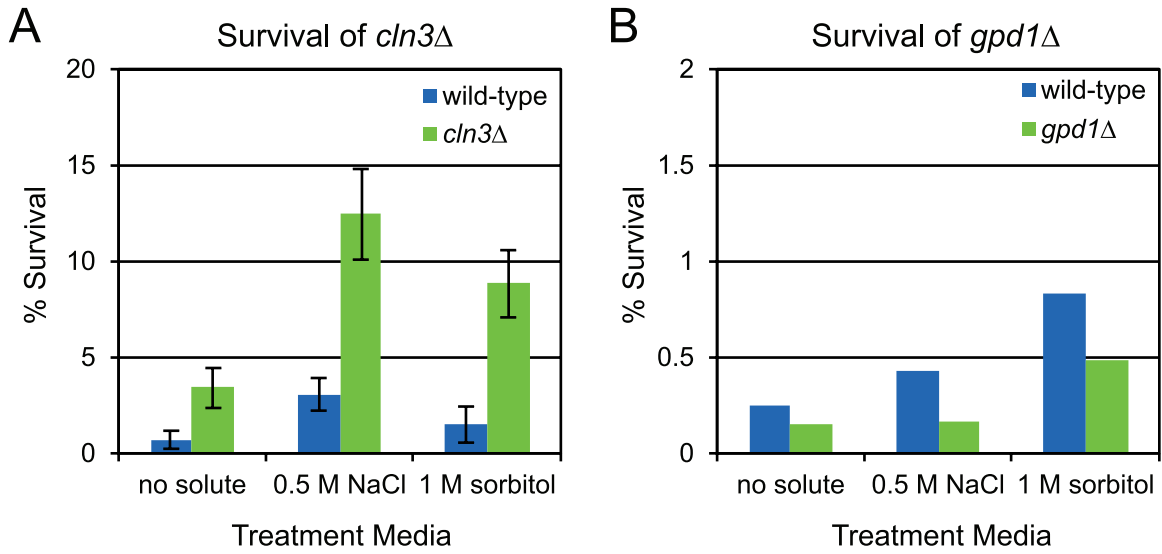
Figure 2.7 Testing *cln3Δ* and *gpd1Δ* mutants in hydroxyurea survival.
(A) Cultures of *mec1Δ sml1Δ* and *mec1Δ sml1Δ cln3Δ* yeast were placed in media with the indicated solute for 2 hours before adding HU. Data are averages of two experiments with technical replicates; error is the sample standard deviation.
(B) Cultures of *mec1Δ sml1Δ* and *mec1Δ sml1Δ gpd1Δ* yeast were placed in media with the indicated solute for 2 hours before adding HU. Data are averages of technical duplicates. Note that the scales for (A) and (B) are different.

gpd1Δ strain had no discernable increase in HU survival in treatment conditions with either no solute or an added osmolyte (Figure 2.7B).

Use of mutator ancestral strain

The strain of yeast used here has a basal mutation rate of about 5×10^{-3} mutation per genome per division, or one mutation every 200 divisions [23]. To increase the genetic diversity of the populations undergoing selection, I constructed a version of the *mec1Δ sml1Δ* strain with a "mutator" phenotype. I used an allele of *POL3*, the catalytic subunit of DNA polymerase δ, which was reported to elevate the mutation rate 100-fold, to about 0.5 mutation per genome per cell division [24]. The Pol3-L523D mutant has

reduced $3' \rightarrow 5'$ exonuclease activity, involved in proofreading, without an impact on its polymerase activity. However, this polymerase's processivity is strongly inhibited by nucleotide imbalances, which result in increased misincorporation. The *mec1Δ sml1Δ pol3-L523D* strain had a strong growth defect and often had variable overnight growth after back-dilution. This is consistent with synthetic effects from a possible imbalance in nucleotide pools (due to reduced regulation of RNR activity) leading to reduced polymerase activity (due to increased misinsertion) combined with an inability to compensate for slower DNA synthesis (due to the absence of the replication checkpoint). However, this strain retained its sensitivity to HU.

Because of its growth defect, the mutator ancestor was pre-cultured for several days in standard medium to improve growth before being used in the experimental evolution. The consequences of using such an impaired strain were not appreciated at the time; in future experiments, the mutator ancestor proved to be ill-suited for use as an experimental control because of difficulty in achieving consistent overnight culture growth and high variability in HU survival. It was also refractory to further genetic modifications and was not used in later testing of causal mutations by strain engineering. However, as will be discussed in Chapter 3, mutator populations yielded similar genetic solutions as non-mutator populations, and the evolved phenotypes of mutator clones were not dependent on the mutator allele. Subsequent to the evolution experiments, additional populations of the mutator ancestor were passaged in standard medium and found to have decreased HU survival concomitant with increased growth rate, indicating that the strain's survival in some assays may be due to its slow growth rate and/or variability in growth rate.

2.4　Evolution of salt-dependent arrest

Eleven cultures, each with an effective population size of 5 x 10^5 cells, were passaged through the selection cycle (Figure 2.6, see section 4.2 for more details). For each passage, cells in standard media were exposed to 0.5 M NaCl for two hours. HU was then added to a final concentration of 50 mM. After four hours in NaCl/HU media, cells were transferred into standard media, diluted to permit exponential growth and cultured for 18 hours. This procedure was repeated for a total of 24 NaCl/HU exposures for three populations (Evolved 1 to 3) and 10 exposures for eight populations (Evolved 4 to 11), with roughly 18-hour growth periods between each exposure to NaCl and HU. Populations 1 through 6 were of the *mec1Δ sml1Δ pol3-L532D* mutator strain, and populations 7 through 11 were of the *mec1Δ sml1Δ* strain.

The levels of HU survival in the evolving populations were tracked by extrapolation from the amount of overnight growth following each NaCl/HU treatment (Figure 2.8). The number of cells in the culture each morning results from the number of cells before treatment multiplied by the (unknown) survival rate, the dilution level after treatment, and the exponential growth factor dependent on the number of doublings (based on time elapsed). Given a known number of cells before treatment, the survival rate can be determined. This calculation requires several assumptions: overnight growth is exponential, the doubling time is known and constant, and the period of exponential growth is known. The estimations of survival in Figure 2.8 assume that no growth occurs during the 6-hour NaCl/HU period and use a doubling time of 120 minutes for the mutator populations and 100 minutes for the non-mutator populations. Based on these estimations, all evolving cultures quickly reached moderate-to-high levels of HU survival. The drop in survival in
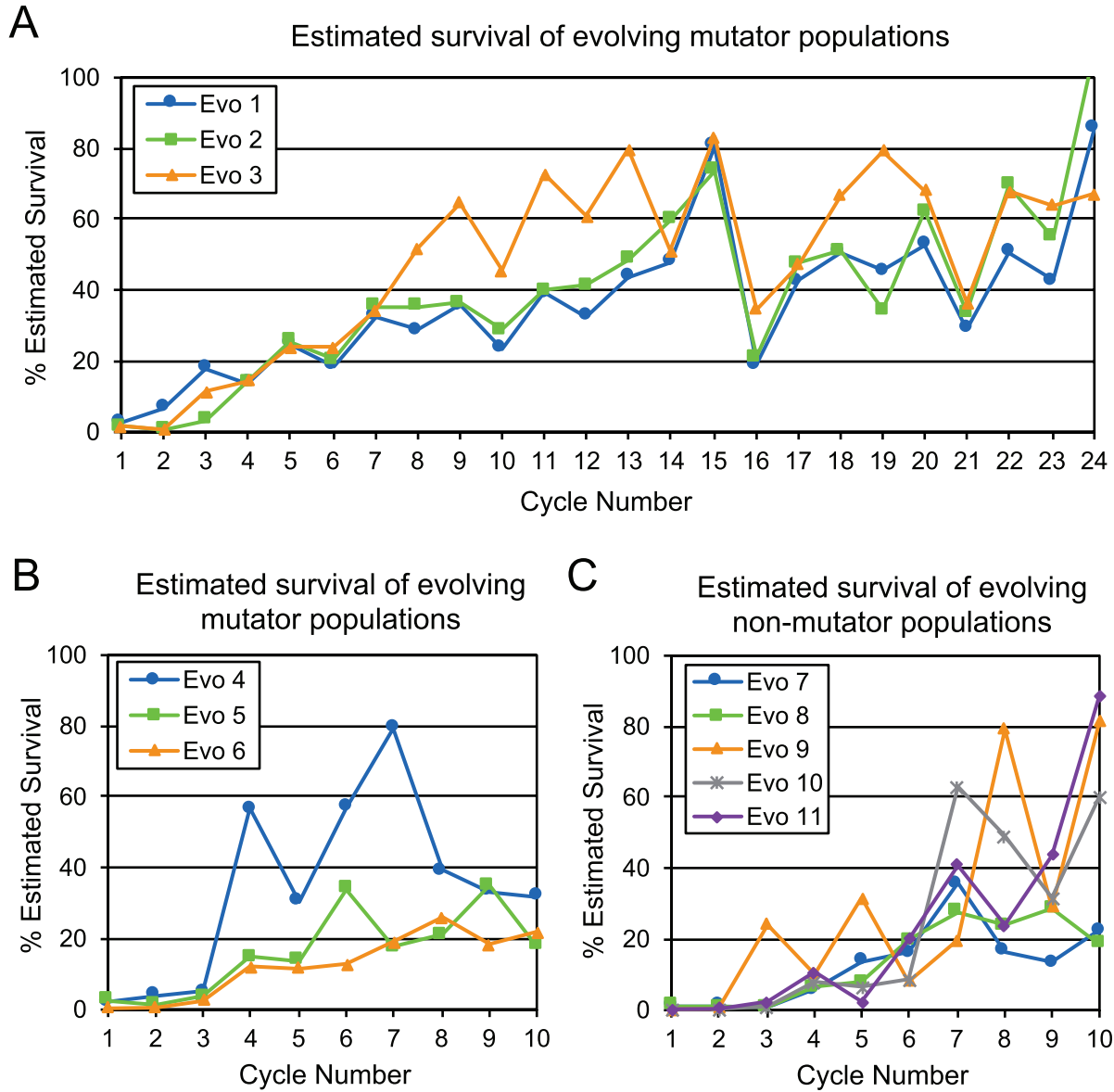
Figure 2.8 Estimated survival of evolving populations.

A rough estimate of survival was extrapolated from cell counts before and after each cycle by dividing the actual cell number after a cycle by the cell number expected from 100% survival, calculated from the starting cell count, the number of doublings in the growth phase, and any dilutions made. See section 4.2 for more details. Populations 1-3 were passaged through 24 cycles (A), and populations 4-11 were passaged through 10 cycles (B)-(C).

round 16 for Evolved 1 to 3 was due to a failure to promptly return treated cultures to the roller drum at 30°C; luckily, the cultures and the project withstood this error.

In most evolution experiments, fitness is a direct outcome of an organism's ability to proliferate, and an increase in fitness occurs by increasing the number of replication cycles in a given period of time. In this selection cycle, the rate of proliferation contributes to fitness in only part of the cycle, and drug survival determines fitness in the other part. To measure fitness of the evolved clones, I will primarily focus on their survival in HU, since this is the part of the selection cycle in which a new stimulus-response connection is expected to provide a fitness benefit.

Clones were picked by plating samples of each evolved population after the final selection cycle. Three or more clones from each final population were assayed for their ability to survive HU by plating culture samples before and after HU treatment. Clones were evaluated for survival on a scale from - to +++, and a clone was considered to survive if it had a score of ++ or +++; see section 4.3 for details on how the survival score was determined. Phenotypic and genetic analyses focused primarily on one clone with salt-dependent survival from each population (Figure 2.9 and Table 2.1). Subsequent to the genetic characterization of these clones, more clones were assayed for survival and sequenced to better understand the diversity of the final evolved populations (Table 2.2).[2] All populations contained clones whose increased survival depends on exposure to NaCl, and populations 9 and 11 also contained clones that survive HU regardless of NaCl.

To characterize the specificity of the response, HU survival was also assayed in the

---

[2]No additional clones were sequenced from populations 1 and 10, because none were isolated that lacked the known causal mutation. Only sequenced clones have phenotypic data listed in Tables 2.1 and 2.2.
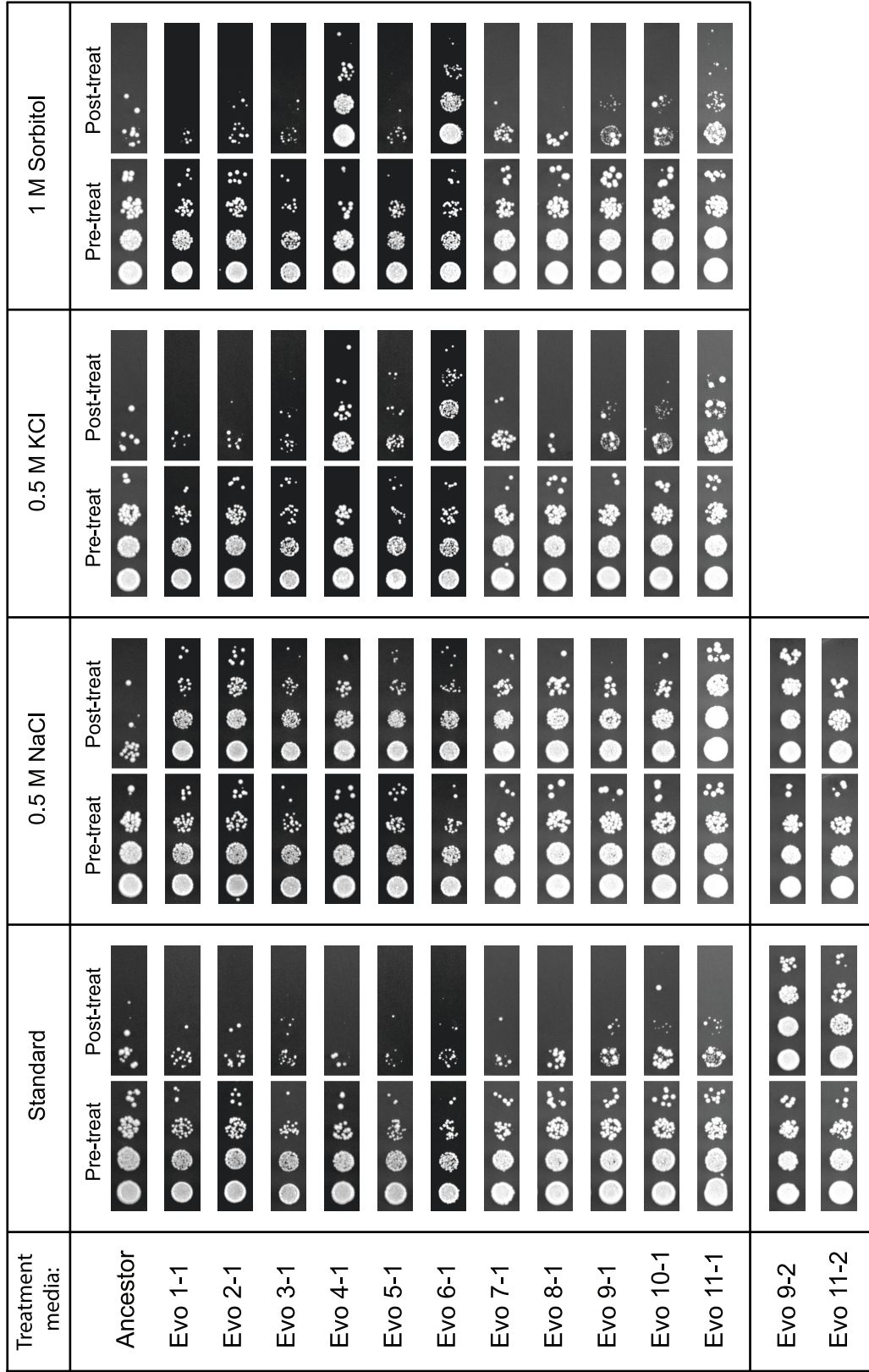
Figure 2.9 Hydroxyurea survival of evolved clones.

Clones were assayed for survival by transferring exponentially growing cultures to the treatment media, YPD or YPD plus either 0.5 M NaCl, 0.5 M KCl, or 1 M sorbitol. A sample of each culture was diluted and spotted onto standard media agar plates immediately after transfer ("Pre-treat"). Two hours later, HU was added to 50 mM, and four hours later cultures were diluted and spotted onto plates ("Post-treat").

Table 2.1 Hydroxyurea survival[1] of evolved clones.

Strain	Standard Media	0.5 M NaCl	0.5 M KCl	1 M sorbitol
Non-mutator Ancestor	-	-	-	-
Evolved 1-1	-	+ + +	-	-
Evolved 2-1	-	+ + +	-	-
Evolved 3-1	-	+ + +	-	-
Evolved 4-1	-	+ + +	+ +	+ + +
Evolved 5-1	-	+ + +	-	-
Evolved 6-1	+	+ + +	+ + +	+ + +
Evolved 7-1	-	+ + +	-	-
Evolved 8-1	-	+ + +	-	-
Evolved 9-1	-	+ +	-	+
Evolved 10-1	-	+ +	+	+
Evolved 11-1	-	+ + +	+	+

[1]Survival was determined on a scale of -/+/+ +/+ + + as described in section 4.3. Examples of raw data are in Figure 2.9. Data are averaged from six or more total replicates, from at least three separate experiments.

presence of either 0.5 M KCl or 1 M sorbitol, a non-ionic osmolyte. These solutes increased survival for the first clones assayed from populations 4 and 6, and for one clone later assayed from population 5. However, not all clones from populations 4, 5, and 6 survived in the other osmolytes, demonstrating phenotypic diversity in these populations (Figure 2.9 and Tables 2.1 and 2.2). The specificity of survival of the rest of the clones indicates that their survival is not based on a response to hyperosmolarity, but rather to an increase in the level of sodium ions.

Table 2.2 Hydroxyurea survival[1] of more evolved clones.

Strain	Standard Media	0.5 M NaCl	0.5 M KCl	1 M sorbitol
Evolved 2-2	-	+++	-	-
Evolved 2-3	-	+++	-	-
Evolved 2-4	-	+++	-	-
Evolved 3-2	+	+++	+	+
Evolved 3-3	-	+++	+	+
Evolved 3-4	-	+++	-	-
Evolved 4-2	-	+++	-	-
Evolved 4-3	-	+++	-	+
Evolved 5-3	-	+++	-	-
Evolved 5-4	-	+++	+++	+++
Evolved 5-5	-	+++	+	-
Evolved 6-2	-	+++	+	+
Evolved 6-3	-	++	-	-
Evolved 6-4	-	+++	-	-
Evolved 7-2	-	++	-	-
Evolved 7-3	-	+++	-	-
Evolved 7-4	-	+++	-	-
Evolved 8-2	-	+++	-	-
Evolved 8-3	-	+++	-	-
Evolved 9-2	+++	+++	n.d.[2]	n.d.
Evolved 9-3	-	++	+	+
Evolved 11-2	+++	+++	n.d.	n.d.
Evolved 11-3	-	+++	n.d.	n.d.
Evolved 11-4	+++	++	n.d.	n.d.

[1]Survival was determined on a scale of $-/+/+ +/+ + +$ as described in section 4.3. Examples of raw data are in Figure 2.9. Data are averaged from six or more total replicates, from at least three separate experiments.
[2]n.d.: No data were collected for these strains and conditions.

2.5 Phenotypes of evolved clones

Population growth rates

Population growth rates of the ancestor strains and evolved clones were measured from changes in optical density (OD) after transfer into either standard media or 0.5 M NaCl. Cultures growing in exponential phase in standard media (YPD) were divided and combined 1:1 with either YPD or YPD plus 1 M NaCl (final concentration of 0.5 M). The OD600 was measured every ten minutes, and an "effective" growth rate[3] for the time period of two to six hours after transfer, equivalent to the timing of HU treatment in the selection,[4] was calculated by fitting an exponential curve to the data (see section 4.4 for more details). If the population reached saturation before six hours, the rate was calculated for data between hour two and the saturation time. Analysis of this time period also removed the confounding factor of the initial decrease in OD upon transfer to NaCl media, which reverses within the first hour after NaCl addition[5] (for example, see plots in Figure 2.10C). This decrease is most likely due to a decrease in individual cell size, as OD is dependent on cell size as well as cell number.

All ancestor strains have a reduced growth rate in NaCl media relative to standard media (Figure 2.10A). The ancestor strains also show some differences from each other in their growth rates in NaCl, which correlate with their relative growth rates in stan-

---

[3]I use the term "effective" to distinguish from other methods of calculating growth rates.

[4]No HU was added to the media for experiments measuring population growth.

[5]I refer to the time before increasing OD as a "lag" period, but it is not meant to imply equivalence to the lag period typically seen in microbial growth curves of cells coming out of saturation phase.
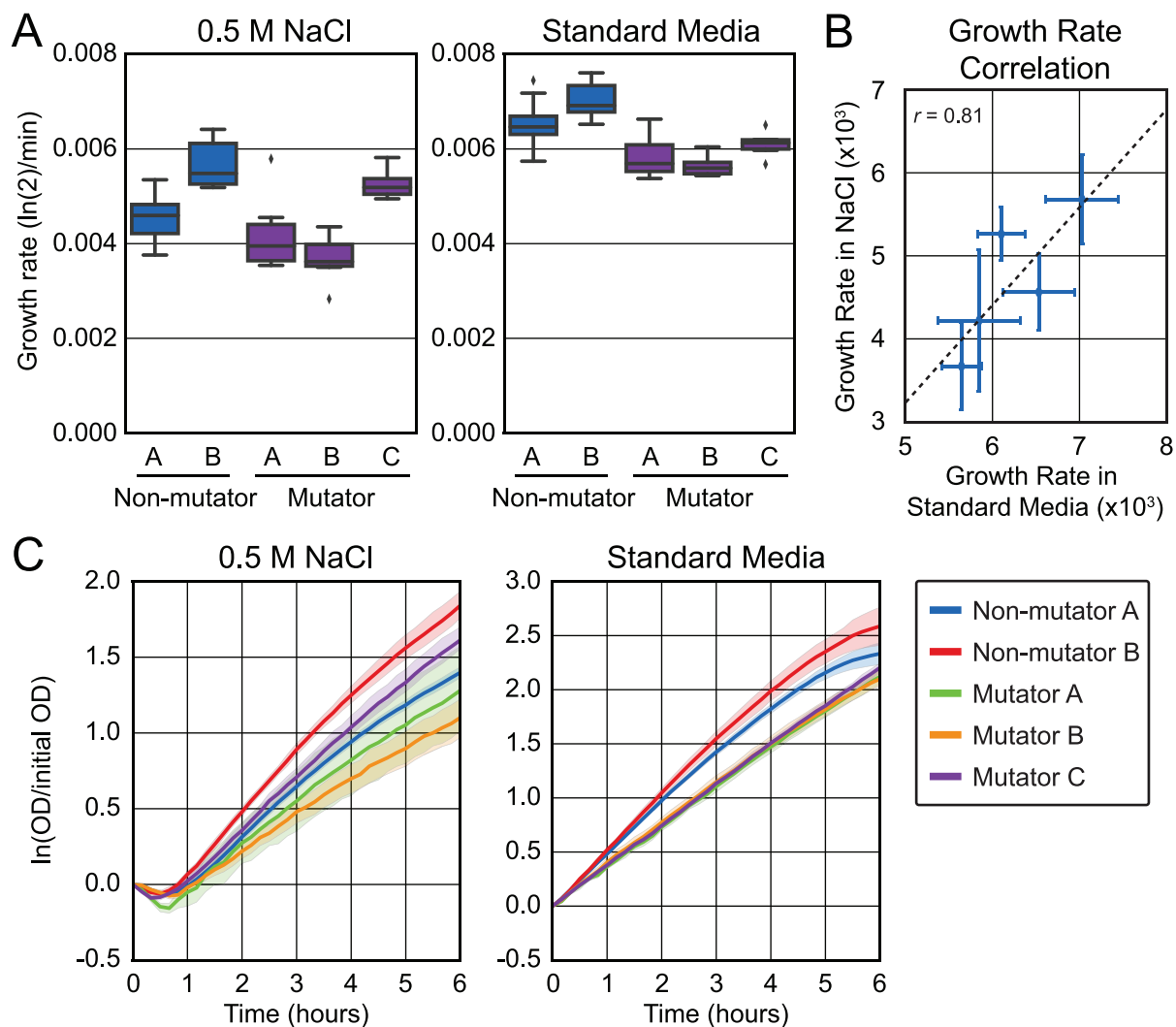
Figure 2.10 Growth rates and dynamics of ancestor strains.

(A) The starting strains for the evolved populations differed from each other in growth rate. The effective growth rate of each strain was calculated for the period of 2 to 6 hours after transfer into 0.5 M NaCl media or standard media, or from hour 2 until saturation if reached before hour 6. Rates were determined from six or more biological replicates. All mutator strains were derived from non-mutator strain A.

(B) The growth rates of the ancestor strains in NaCl media correlated with their rates in standard media. Points are mean values of the same experimental data plotted in (A) and error bars are standard deviations. The dotted line is a linear least-squares regression. The correlation coefficient (Pearson's r) is 0.81.

(C) Averaged growth curves show details of population growth dynamics not captured by the effective growth rate. Raw data were calibrated by a blank value and transformed to account for the non-linearity above OD 0.6 before normalizing by initial OD. Lines are mean values from six or more replicates and shaded bands are bootstrap 95% confidence intervals. All strains decreased in OD from the initial value, which reversed within the first hour after NaCl addition.

39

Figure 2.11 Growth rates of evolved clones in NaCl.

All clones have a decreased rate of growth two to six hours after transfer into media with 0.5 M NaCl, but three clones (Evolved 8, 9, and 10) have slight decreases while the other seven have large decreases. Rates plotted are the effective growth rates as described in Figure 2.10A. No HU was in the media. The ancestor strain shown is non-mutator A.

Growth rate after NaCl addition
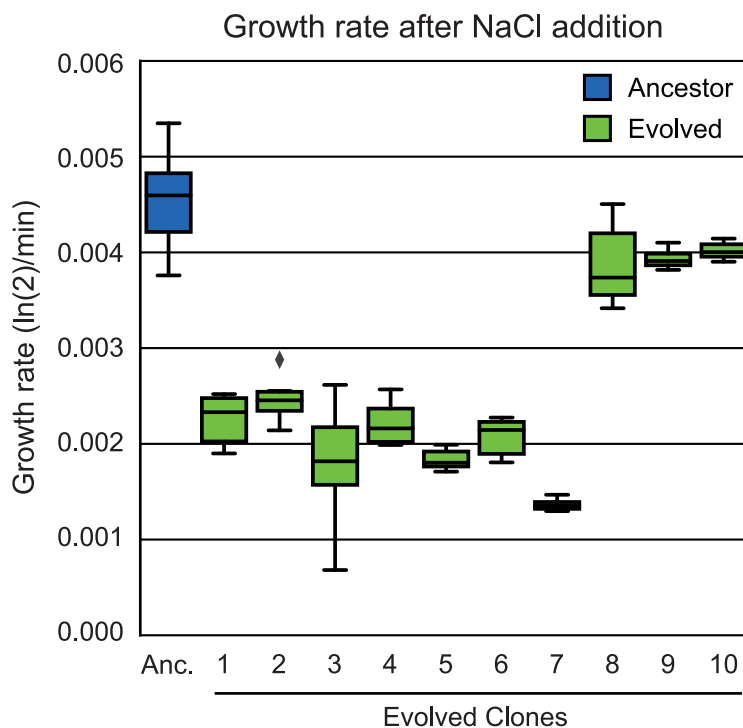
dard media (r=0.81, Figure 2.10B). Multiple experiments were averaged together after normalization to initial OD to display the growth curve traces over time (Figure 2.10C). All growth curves are shown with log-transformed data, in which a straight line indicates exponential growth. These curves show that the population dynamics for the ancestor strains differ primarily in the rate of growth after adaptation to the NaCl, when OD begins increasing. There may also be slight differences in the time to increasing OD, but all strains show steady increases after 40 minutes.

The same population growth measurements were made for the primary clone from each evolved population (e.g., Evolved 1-1) except 11. Growth of Evolved 11-1 could not be assayed by OD600 due to its flocculation phenotype, which alters the relationship between cell number and optical density and results in significant settling of the cells in culture during the experiment. Slower population growth in NaCl media was observed for all clones to varying degrees (Figure 2.11). All mutator clones and one non-mutator

40

clone (Evolved 7) had population growth at a doubling time of greater than four hours (rate equal to 0.0029), while three of the non-mutator clones slowed to a doubling time of about three hours (rate equal to 0.0039).

Growth rates of the evolved clones are compared to their direct ancestors in Figure 2.12. The growth rates in NaCl from Figure 2.11 and in both media from Figure 2.10A are reproduced in Figure 2.12A and B with the data arranged by ancestor strain. The three mutator clones from the first evolution experiment show no growth rate changes in standard media, in contrast to the mutator clones from the second experiment: Evolved 4 and 6 have lower growth rates in standard media than their ancestors, while Evolved 5 has a higher growth rate than its ancestor (Figure 2.12C). The non-mutator evolved clones have no growth rate changes from their ancestors in standard media, and Evolved 9 and 10 maintained the elevated rate of their ancestor, non-mutator B, relative to the non-mutator A strain (Figure 2.12D).

Averaged growth curves after transfer to NaCl show that all evolved clones have altered population dynamics from their ancestors to some degree (Figure 2.13). Mutator evolved clones show differences in the decrease from initial OD as well as a slower rate of growth after the decrease (Figure 2.13A). Evolved 4, 5, and 6 are particularly divergent from their ancestors in the first hour. In contrast, the non-mutator clones all resemble their ancestors in the initial decrease, but have different growth dynamics afterwards (Figure 2.13B). Evolved 7 shows a longer "lag" period before increasing in OD, and then shows growth similar to the mutator evolved clones. Evolved 8 shows only a slight difference from its ancestor. Evolved 9 and 10 have growth curves identical to each other but distinct from their ancestor and from all other clones: they appear to have faster than, as opposed to slower than, exponential growth, as indicated by an upward bend in

41
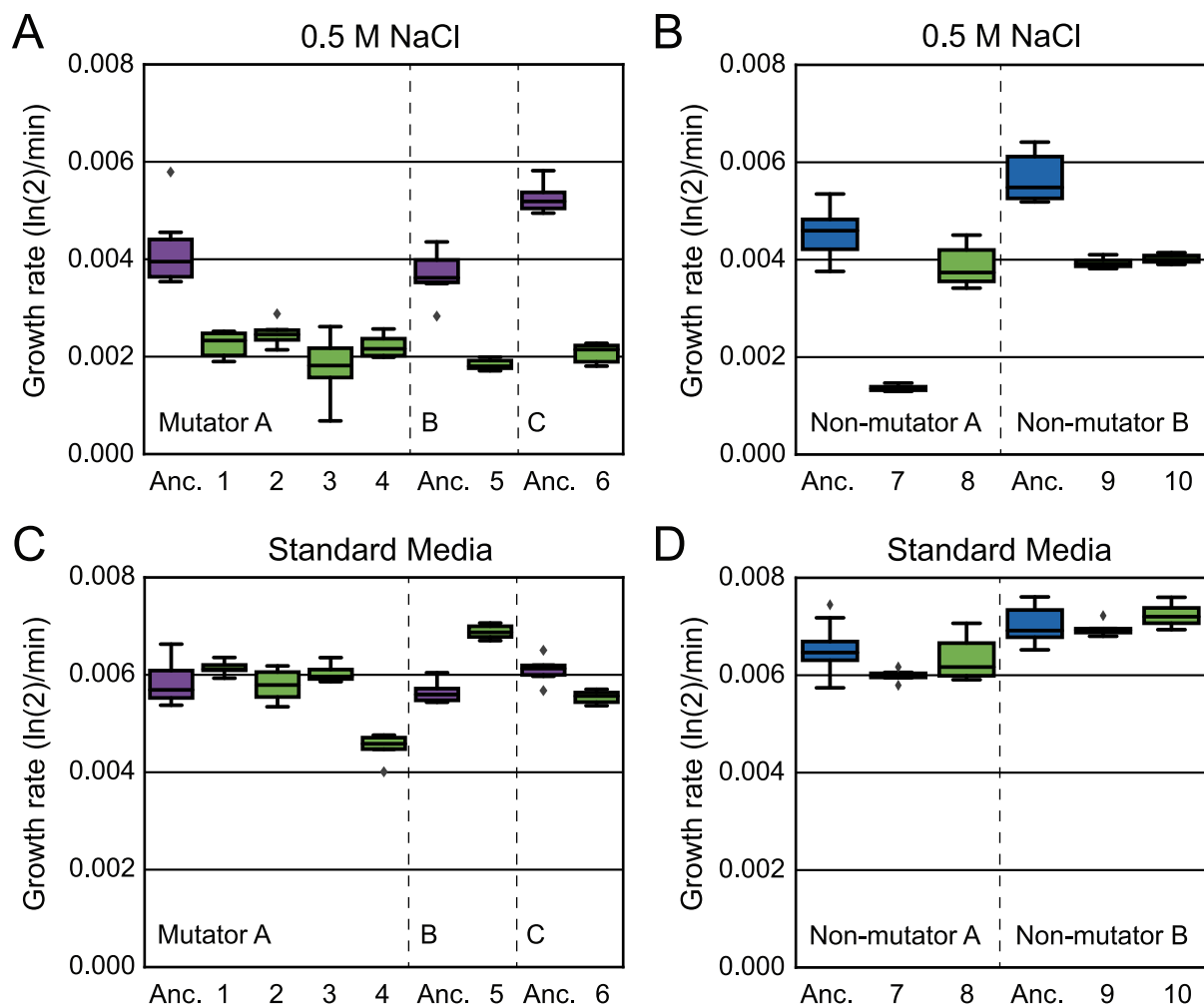
Figure 2.12 Growth rates of evolved clones and their direct ancestors.

(A) All mutator clones evolved to have a similarly low growth rate in NaCl despite deriving from different immediate ancestors. Rates plotted are the effective growth rates for 2 to 6 hours after media transfer, as described in Figure 2.10A.

(B) Only one of the four non-mutator clones has a growth rate in NaCl as low as the mutator clones; the other three have smaller decreases in growth rate.

(C) Mutator evolved clones from the first experiment (1, 2, and 3) have no change from their ancestor in standard media. The clones from the second experiment (4, 5, and 6) do show differences from their respective ancestors.

(D) The growth rates of non-mutator clones in standard media do not differ from their respective ancestors.
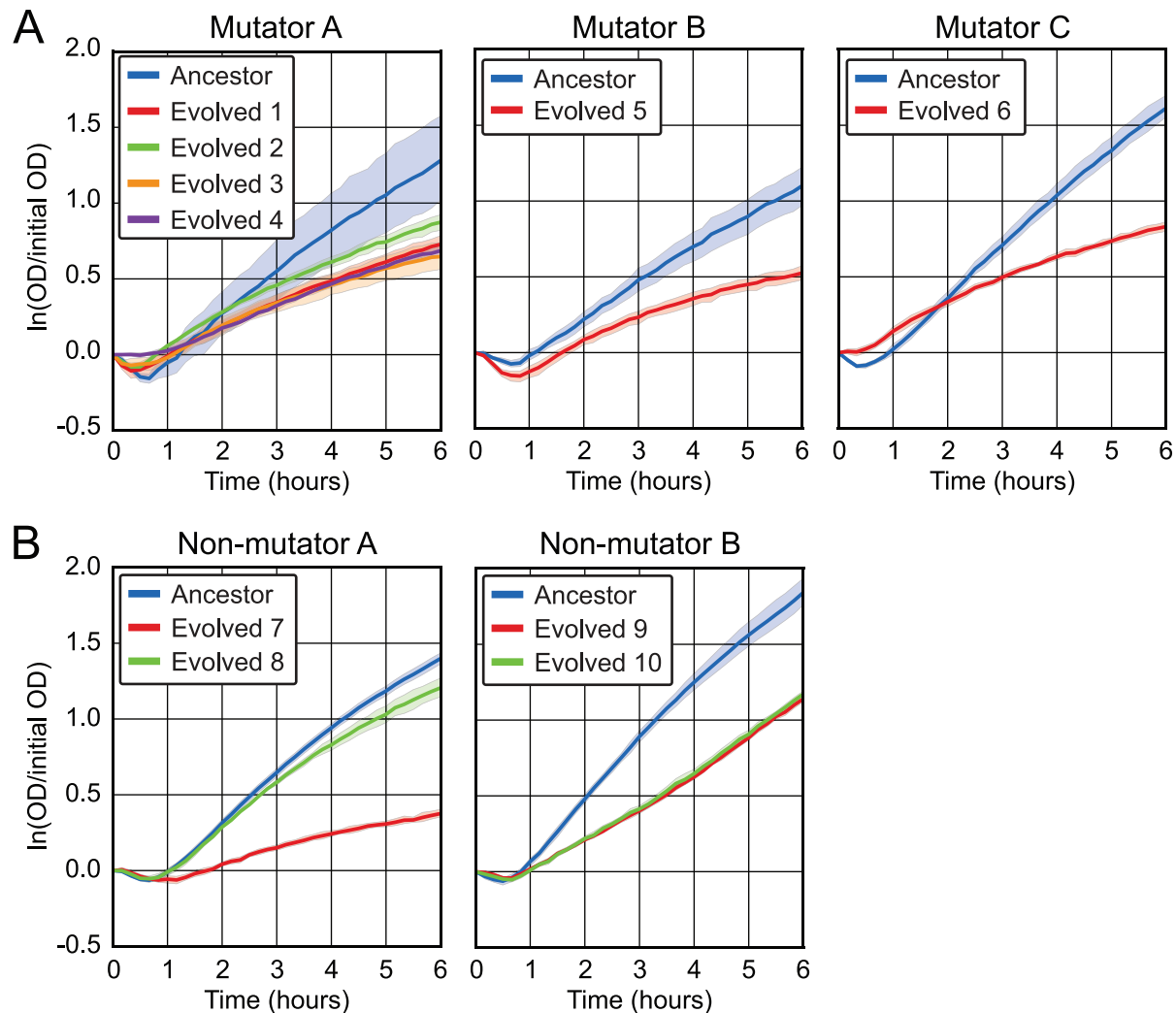
Figure 2.13 Growth curves of evolved clones and their direct ancestors in NaCl.

(A) All mutator clones evolved lower growth rates in NaCl despite deriving from different immediate ancestors. Evolved clones 4, 5 and 6 also show differences from their ancestors in the initial decrease in OD within the first hour. Growth curves from multiple experiments were aggregated after normalizing by initial OD, as described in Figure 2.10C. Lines are mean values from six or more replicates and shaded bands are bootstrap 95% confidence intervals. Clones from mutator ancestor A also have greater experimental reproducibility, as evidenced by the thinner shaded bands.

(B) Evolved 7 has the lowest level of growth of any evolved clone and the greatest difference from its ancestor, and Evolved 8 has the least difference in growth from its ancestor as measured at hour 6. Evolved 9 and 10 have identical population growth profiles that are distinct from their ancestor and from the other evolved clones.

Figure 2.14 Growth curves of evolved clones in NaCl and standard media.

Growth curves were aggregated after normalizing by initial OD, as described in Figure 2.10C. Parts A, C, D and E display the same data as Figure 2.13, but without the ancestor data and with the clones grouped into low (A, C, D) vs. high (E) growth rate. Parts C and D are enlargements of A. Lines are mean values from six or more replicates. Shaded bands represent 68% confidence intervals instead of 95% to more easily distinguish between the clones, and axis scales were varied to highlight intra-group differences.

(A) The low-growth rate clones have slower than exponential growth in NaCl media, evidenced by downward arcs of the log-transformed data. Evolved 7 has the lowest OD at hour 6 and the longest "lag" time before increasing OD after transfer to NaCl media.
(B) Evolved 1 through 7 have steady exponential growth in standard media, though at a range of rates.
(C) A closer look at the first two hours better illustrates the variety in dynamics among the low-growth rate clones after transfer to NaCl media, and reveals that Evolved 4 and 6 do not experience a decrease in OD, although they do have a "lag" period before OD increases.
(D) The trajectories look similar after the first 2 hours; most of the difference in the clones' ODs at hour 6 is due to differences in their early trajectories, with slight differences in growth rates as reflected by the data in Figure 2.11.
(E) Evolved 8 has slower than exponential growth, similar to the low-growth rate clones but differing in degree. In contrast, Evolved 9 and 10 have qualitatively different dynamics: their curves show faster than exponential growth.
(F) Evolved 8, 9 and 10 have steady exponential growth in standard media until saturation is reached.
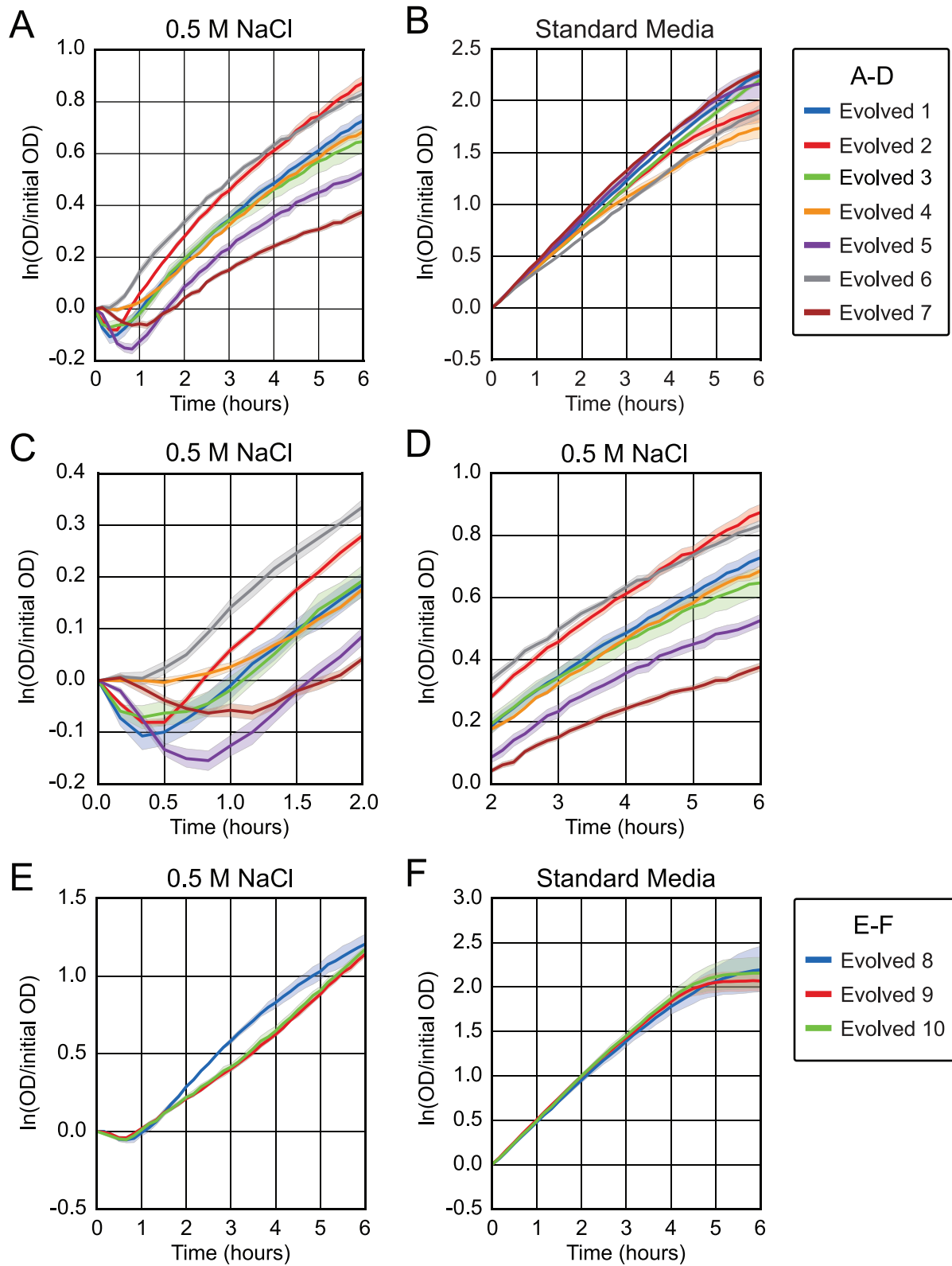
Figure 2.14 (continued)
Growth curves of evolved clones in NaCl and standard media.

the log-transformed data.

The growth curves of the evolved clones are compared to each other in Figure 2.14, grouped by effective growth rate. All mutator clones have slower than exponential growth, evidenced by the downward arcs of the log-transformed data, but have steady exponential growth in standard media (Figure 2.14A, B). The non-mutator strain Evolved 7 has the most severe growth phenotype, with the lowest OD at hour 6 and the longest "lag" time before increasing OD after transfer to NaCl media. The mutator clones show greater differences from each other in their trajectories in the first two hours (Figure 2.14C) than from hours two to six (D). The slight differences in their growth after hour 2 are accurately reflected by the calculated effective growth rates (Figure 2.11). Evolved 4 and 6 do not have a decrease in OD upon transfer to NaCl media, in contrast to their ancestors and all other evolved clones, although they still have a "lag" period before OD increases.

Among the higher growth rate strains, Evolved 9 and 10 have very different growth dynamics from Evolved 8, despite have nearly identical effective growth rates from hours two to six (Figure 2.14E). Evolved 8 has a downward arc, like Evolved 1-7, but Evolved 9 and 10 have upward arcs, indicating an increasing rate of population growth over time. No differences are seen in the growth of Evolved 8, 9, and 10 in standard media (Figure 2.14F).

The downward arcs of the log-transformed data for Evolved 1 to 8 indicate that their culture densities are not increasing exponentially. In plots of averaged growth curves without log transformation, the increase in OD appears to be linear for these strain, or possibly even sub-linear for Evolved 5 and 7; exponential growth is seen after an initial period of linear growth for Evolved 9 and 10 (Figure 2.15). The implications of this are examined in the discussion.
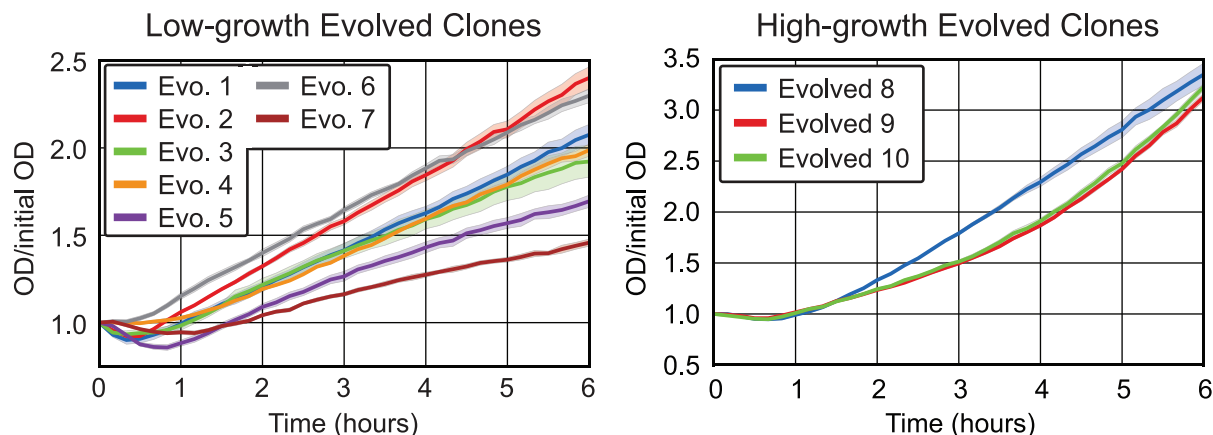
Figure 2.15 Linear growth curves of evolved clones in NaCl media.
Averaged growth curves without log transformation show linear, or possibly sub-linear, growth for Evolved 1 through 8, in contrast to the exponential growth seen for Evolved 9 and 10. Clones are divided into two plots according to their effective growth rates.

2.6   Discussion

In this chapter, I described how I created a selection scheme for adaptive prediction and executed it for parallel populations of *Saccharomyces cerevisiae*. By using the genetically-altered *mec1Δ sml1Δ* strain lacking the DNA replication checkpoint, I administered the replication inhibitor hydroxyurea (HU) as a selection for cell cycle arrest. I demonstrated that replicating cells died at a very high level when treated with HU, and cells arrested in G1 or M phase survived. With the ability to select both for and against cell cycle arrest, I could then add a cue to alert cells to the change in selection. The addition of sodium chloride was chosen as an environmental change to which yeast could respond. I also created a version of the *mec1Δ sml1Δ* strain with an elevated mutation rate to select on populations with greater genetic diversity.

I passaged eleven separate populations of yeast through cycles of HU treatment and standard media, adding 0.5 M NaCl before and during the time HU was present. Each

population had rapid increases in their survival of HU. From the final populations, I selected clones which individually showed high survival when treated with HU plus NaCl. All eleven populations contained clones with survival dependent on the presence of NaCl. Two populations also had clones with NaCl-independent survival, demonstrating heterogeneity within the final populations. Investigating the specificity of the survival phenotype revealed more diversity among the clones, as clones from three populations survived with any osmolyte present, whereas the remainder survived only with NaCl. These three populations all contained clones with both specificity phenotypes, further indicating population heterogeneity.

After characterization of HU survival phenotypes, the clones can be grouped into three classes. Class I clones, which have sodium-specific survival, were found in every population, and from six of the eleven populations (three mutator, three non-mutator) only Class I clones were isolated.[6] Class II clones, with osmolyte-dependent survival, were found in three mutator populations. Class III clones, which survive HU regardless of osmolyte presence, were found in two non-mutator populations. Although this is a small number of populations from which to generalize, I can speculate that NaCl-specific survival is the phenotype with the largest target size, and that the relative fitness of other survival phenotypes may have some dependence on the background of the ancestor (mutator versus non-mutator).

The selection using hydroxyurea was intended to select for cell cycle arrest upon exposure to NaCl, and survival phenotypes were tested as a proxy for arrest. To examine whether the evolved strains replicate during exposure to NaCl, I measured population

---

[6]It is possible that other survival phenotypes were present but not identified; fewer than ten clones were assayed from each population.

growth by optical density of exponentially growing cultures after the addition of NaCl. All mutator clones and one non-mutator clone have a reduced rate of population growth equal to a doubling time of four hours or longer during the period two to six hours after NaCl addition. In other words, their populations undergo one doubling or less than one doubling during the time period equivalent to hydroxyurea treatment in the selection. The culture ODs increase linearly, indicating a constant number of replicating cells and therefore a decreasing fraction of the population that is replicating.

Yeast cells arrested in the cell cycle can continue to grow in size, depending on the nature of the arrest. It is unknown to what extent an increase in individual cell size contributes to the increase in culture OD of the evolved strains in NaCl media. Cells with greater volume have fewer cells per OD unit [25], which means that cell growth without division could account for some or all of the OD increase. More precise data on the increase in cell number could be collected using a particle counter instead of a spectrophotometer; this data is more challenging to collect, as it is destructive to the measured sample and is not currently in a high-throughput format. Growth curves conducted by particle counts of Evolved 1 clones showed an increase of less than 50% in the cell number over six hours in NaCl media, versus an average 110% increase in OD; these experiments were not replicated in favor of the OD method with higher throughput and higher time resolution. Therefore, it remains unclear to what extent the OD increases for Evolved 1 through 7 represent cell division versus cell size increases.

The relationship of culture OD to individual cell size could also explain the phenomenon of decreasing OD at early times after transfer to NaCl media. Within the first hour, the ancestor strains and most of the evolved strains show decreases of up to 20% from the initial OD. The two exceptions, Evolved 4 and 6, are also two evolved clones

49

that demonstrate osmolyte-dependent survival (Class II); all other assayed clones are Class I. This will be explored further in Chapter 3, when the genetic causes of the evolved phenotypes are elucidated.

Three of the non-mutator clones have an effective growth rate in NaCl roughly equal to that of the mutator ancestors A and B, but have higher survival in hydroxyurea. This indicates that this metric is not sufficient to predict higher survival nor to explain it for these evolved clones. The clear differences in population growth dynamics of Evolved 9 and 10 versus Evolved 8 illustrate the inadequacy of a single parameter, effective growth rate, for comparing growth phenotypes. This inadequacy is also true of the experimental method in general: a measurement of a population does not necessarily reflect single-cell behavior. To visualize growth dynamics at a single-cell level, time-lapse microscopy experiments were performed for the evolved clones 1, 2, and 3, in which individual cells appeared to arrest. However, the small sample size and lack of automated analysis made it difficult to discern whether cells were proceeding very slowly or arresting, and if arresting, at what stage in the cell cycle. Efforts were instead focused on elucidating the genetic basis of the evolved phenotypes, detailed in Chapter 3. Alternative methods to molecularly characterize the cell cycle arrest of the evolved clones will be discussed in Chapters 3 and 5.

References

[1] I. Tagkopoulos, Y.-C. Liu, and S. Tavazoie, "Predictive behavior within microbial genetic networks.," *Science*, vol. 320, pp. 1313–1317, June 2008.

[2] A. Mitchell, G. H. Romano, B. Groisman, A. Yona, E. Dekel, M. Kupiec, O. Dahan, and Y. Pilpel, "Adaptive prediction of environmental changes by microorganisms.," *Nature*, vol. 460, pp. 220–224, July 2009.

[3] S. F. Elena and R. E. Lenski, "Microbial genetics: Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation," *Nature Reviews Genetics*, vol. 4, pp. 457–469, June 2003.

[4] M. J. Dunham, "Experimental evolution in yeast: a practical guide.," *Methods in Enzymology*, vol. 470, pp. 487–507, 2010.

[5] A. M. Murray and T. Hunt, *The Cell Cycle: An Introduction.* 1993.

[6] L. Alberghina, C. Smeraldi, B. M. Ranzi, and D. Porro, "Control by nutrients of growth and cell cycle progression in budding yeast, analyzed by double-tag flow cytometry.," *Journal of Bacteriology*, vol. 180, pp. 3864–3872, Aug. 1998.

[7] L. H. Hartwell and T. A. Weinert, "Checkpoints: controls that ensure the order of cell cycle events.," *Science*, vol. 246, pp. 629–634, Nov. 1989.

[8] P. Zegerman and J. F. X. Diffley, "DNA replication as a target of the DNA damage checkpoint.," *DNA Repair*, vol. 8, pp. 1077–1088, Sept. 2009.

[9] S. J. Elledge, Z. Zhou, J. B. Allen, and T. A. Navas, "DNA damage and cell cycle regulation of ribonucleotide reductase," *BioEssays*, vol. 15, no. 5, pp. 333–339, 1993.

[10] T. A. Weinert, G. L. Kiser, and L. H. Hartwell, "Mitotic checkpoint genes in budding yeast and the dependence of mitosis on DNA replication and repair.," *Genes & Development*, vol. 8, pp. 652–665, Mar. 1994.

[11] X. Zhao, E. G. D. Muller, and R. Rothstein, "A suppressor of two essential checkpoint genes identifies a novel protein that negatively affects dNTP pools.," *Molecular Cell*, vol. 2, pp. 329–340, Sept. 1998.

[12] M. Lopes, C. Cotta-Ramusino, A. Pellicioli, G. Liberi, P. Plevani, M. Muzi-Falconi, C. S. Newlon, and M. Foiani, "The DNA replication checkpoint response stabilizes stalled replication forks.," *Nature*, vol. 412, pp. 557–561, Aug. 2001.

[13] J. A. Tercero and J. F. X. Diffley, "Regulation of DNA replication fork progression through damaged DNA by the Mec1/Rad53 checkpoint.," *Nature*, vol. 412, pp. 553–557, Aug. 2001.

[14] B. A. Desany, A. A. Alcasabas, J. B. Bachant, and S. J. Elledge, "Recovery from DNA replicational stress is the essential function of the S-phase checkpoint pathway," *Genes & Development*, vol. 12, pp. 2956–2970, Sept. 1998.

[15] N. Q. Balaban, J. Merrin, R. Chait, L. Kowalik, and S. Leibler, "Bacterial persistence as a phenotypic switch.," *Science*, vol. 305, pp. 1622–1625, Sept. 2004.

[16] E. Kussell, R. Kishony, N. Q. Balaban, and S. Leibler, "Bacterial persistence: a model of survival in changing environments.," *Genetics*, vol. 169, pp. 1807–1814, Apr. 2005.

[17] M. A. Schwartz and H. D. Madhani, "Principles of MAP kinase signaling specificity in Saccharomyces cerevisiae.," *Annual Review of Genetics*, vol. 38, pp. 725–748, 2004.

[18] S. M. O'Rourke and I. Herskowitz, "The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in *Saccharomyces cerevisiae*," *Genes & Development*, vol. 12, pp. 2874–2886, Sept. 1998.

[19] W. Zachariae, T. H. Shin, M. Galova, B. Obermaier, and K. Nasmyth, "Identification of subunits of the anaphase-promoting complex of Saccharomyces cerevisiae.," *Science*, vol. 274, pp. 1201–1204, Nov. 1996.

[20] F. R. Cross, "DAF1, a mutant gene affecting size control, pheromone arrest, and cell cycle kinetics of Saccharomyces cerevisiae.," *Molecular and Cellular Biology*, vol. 8, pp. 4675–4684, Nov. 1988.

[21] E. A. Vallen and F. R. Cross, "Interaction between the MEC1-dependent DNA synthesis checkpoint and G1 cyclin function in Saccharomyces cerevisiae.," *Genetics*, vol. 151, pp. 459–471, Feb. 1999.

[22] J. Albertyn, S. Hohmann, J. M. Thevelein, and B. A. Prior, "GPD1, which encodes glycerol-3-phosphate dehydrogenase, is essential for growth under osmotic stress in Saccharomyces cerevisiae, and its expression is regulated by the high-osmolarity glycerol response pathway.," *Molecular and Cellular Biology*, vol. 14, pp. 4135–4144, June 1994.

[23] G. I. Lang and A. W. Murray, "Estimating the per-base-pair mutation rate in the yeast Saccharomyces cerevisiae," *Genetics*, vol. 178, no. 1, pp. 67–82, 2008.

[24] Y. H. Jin, P. Garg, C. M. W. Stith, H. Al-Refai, J. F. Sterling, L. J. W. Murray, T. A. Kunkel, M. A. Resnick, P. M. Burgers, and D. A. Gordenin, "The multiple biological roles of the 3 5 exonuclease of Saccharomyces cerevisiae DNA polymerase δ require switching between the polymerase and exonuclease domains," *Molecular and Cellular Biology*, vol. 25, no. 1, pp. 461–471, 2005.

[25] B. Volkmer and M. Heinemann, "Condition-dependent cell volume and concentration of Escherichia coli to facilitate data conversion for systems biology modeling.," *PLoS ONE*, vol. 6, no. 7, p. e23126, 2011.

Chapter 3

Identification and Confirmation of Causal Mutations

3.1 Abstract

To identify mutations causing the evolved phenotype, clones with NaCl-dependent HU survival from each population were analyzed by whole-genome sequencing. Additionally, bulk segregant analysis was performed for one clone from each mutator population. Putative causal mutations in the mutator strains were identified based on segregation frequency, the annotated function of the affected gene and any notable mutant phenotypes. Candidate causal mutations were tested by allele replacement, both by engineering the mutation into the ancestor background and by reversion of the mutation to wild-type in the evolved clone. We confirmed causality of nine mutations in six different genes that encode the following proteins: two protein kinases (Hal5 and Hog1), a transcriptional regulator (Imp2'), and three hexose transporters (Hxt1, Hxt3, and Hxt7). The muta-

tions in the hexose transporters are gain-of-function mutations whereas those in the other genes inactivate their protein products, demonstrating that both altering and eliminating protein function can rewire the cellular response to environmental stimuli in a way that produces adaptive prediction.

3.2 Introduction

In this chapter, I describe genetic characterization of the evolved strains presented in Chapter 2 and the survival and growth phenotypes caused by evolved mutations. Laboratory evolution enables identification of the genetic changes causing a new phenotype by (1) comparing the genomes of ancestral and evolved populations and (2) testing the effect of candidate mutations. The development of "next-generation" sequencing and its analysis have made the first part both precise and affordable [1, 2]. With sufficient sequencing depth, heterogeneous populations can also be sequenced and analyzed for mutation frequency [3]. The second part, confirmation of causality through strain engineering, is possible in *S. cerevisiae* due to its native DNA repair machinery: an exact mutation, including deletion or addition of DNA, can be engineered at the endogenous locus via homologous recombination [4, 5, 6]. In my evolution experiments, a mutator strain was used as the ancestor for six populations. Sequencing the evolved clones would yield dozens or hundreds of mutations, so the more sophisticated "bulk segregant" analysis was also done for these strains.

Bulk segregant analysis assists the identification of causal mutations in a genetic background with neutral or deleterious mutations (Figure 3.1A) [7, 8]. This method takes

Figure 3.1 Diagram of bulk segregant analysis.
(A) Bulk segregant analysis identifies causal mutations by mating the evolved strains to the ancestor, sporulating, and selecting for haploid recombinants that demonstrate the evolved phenotype. Non-causal mutations are reduced to an average frequency of 50% in the selected recombinants, whereas mutations required for the phenotype are present at high frequency.
(B) After mating and sporulation, the population of haploid recombinant spores was split into three pools. One was immediately prepared for sequencing and two were passaged in media before sequencing: one in standard media and one through five rounds of the NaCl/HU selection cycle.

advantage of laboratory control over the yeast life cycle and the ability to collect a great number of offspring from the same set of parents. In this method, a haploid strain containing unknown mutations of interest is mated to a haploid strain of the opposite mating type lacking those mutations to generate a heterozygous diploid cell. In experimental evolution, this pair typically consists of an evolved haploid strain and an ancestor strain that has had its mating type switched or a similar non-evolved strain. Mutations in the diploid parent are randomly distributed to offspring by the processes of chromosomal recombination and independent segregation that occur during meiosis. Four spores result from the meiosis of a single diploid cell. By first growing the parental diploid cell to a large

population, spores are obtained from many independent sporulation events of the same parent strain. These haploid recombinant spores are then passaged through selection for the phenotype of interest. Population frequencies of mutations change during the selection based on the fitness benefit or deficit they confer. The frequency of a mutation in the selected pool is determined by preparing genomic DNA from the pool of cells, sequencing it by next-generation technology, and calculating the fraction of independent reads containing the mutation. Mutations required for the phenotype are present at high frequency in the selected pool while neutral mutations are reduced to an average frequency of 50% and deleterious mutations are at low frequency.

In this work, the method of bulk segregant analysis was extended by sequencing two additional pools: the initial pool of spores before selection, and a pool subjected to a separate selection by culturing in standard media (Figure 3.1B). These additional pools aided in determining whether a mutation's high frequency in the NaCl/HU pool was due to a high initial frequency, a contribution to fitness in standard culture conditions, or a contribution to the evolved survival phenotype. These are not mutually exclusive options; however, the additional evidence was used to prioritize the list of top candidate mutations. Candidate mutations were then tested for causality by genetically altering the ancestor and evolved strains in several ways. Allele replacements, both putting the mutation in the ancestor at the endogenous locus and reverting it in the evolved clone, tested for sufficiency and necessity. Deleting the mutated gene in both backgrounds determined whether the mutation was loss- or gain-of-function, and testing the phenotype of heterozygous diploid strains determined dominance. The population growth curves of reconstructed strains were measured as in Chapter 2 to test the relationship between genotype and the "extended" phenotype of growth dynamics.

3.3 Sequencing for putative causal mutations

Sequencing of evolved clones and populations

To identify mutations causing the evolved phenotypes, clones with NaCl-dependent HU survival from each population were analyzed by whole-genome sequencing (see section 4.6 for method details). Mutator clones were mated to a non-mutator *mec1Δ sml1Δ* strain and subjected to bulk segregant analysis to aide in distinguishing causal from non-causal mutations (Figure 3.1). After initially sequencing one clone from each population and engineering strains to test the causality of identified mutations, Sanger sequencing revealed that most final populations were heterogeneous for the causal mutations (discussed in section 3.4.1). Therefore, additional clones from each population that did not contain the known causal mutations were sequenced along with the frozen glycerol stocks of the final evolved populations to identify other potential causal mutations; no strain engineering was done to test causality of these mutations.

Three separate rounds of whole-genome sequencing were performed. The target level of coverage was 30x for clones and 50x for pools (i.e., heterogeneous populations), and the total number of bases sequenced for each sample was generally close to or above this target (Figure 3.2A, Table 3.1). However, coverage by the number of aligned bases was lower than expected due to short insert sizes in the library preparations. If the insert size is less than twice the read length (e.g., <300 bp for 150-bp reads), the second read of paired-end sequencing will overlap the first read, resulting in some bases being read twice. This double-reading can reduce the rate of false positives arising from sequencing errors [9]. However, it also creates false coverage because the information is not from
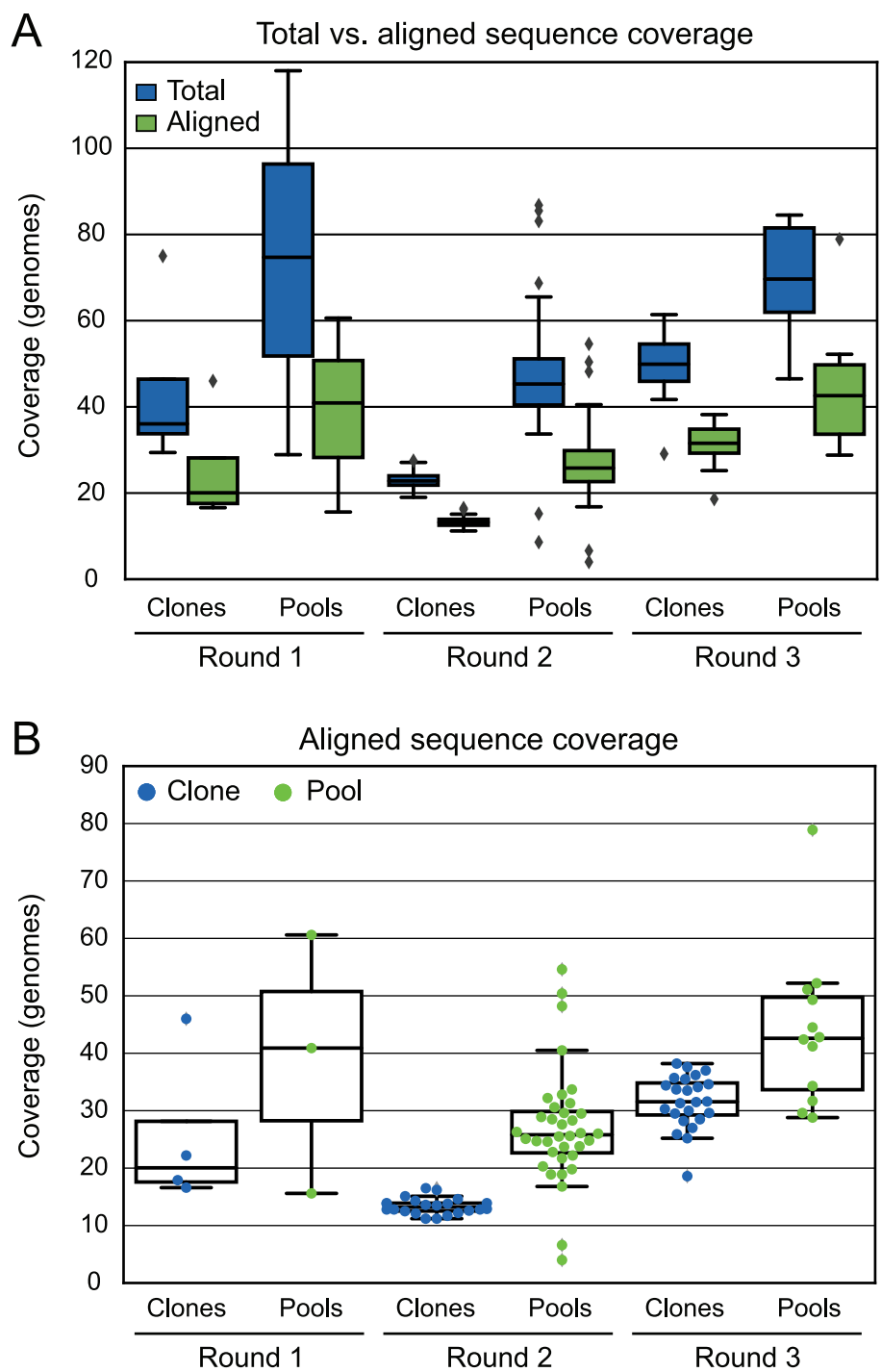
Figure 3.2 Sequencing coverage of clones and pools.

(A) Target coverage was 25-30x for clones and 50x for pools; however, short insert sizes in the library preparations resulted in discrepancies between the total bases sequenced and non-overlapping aligned bases for each sample.

(B) Circles show data for individual samples of clones and pools. Almost all samples had coverage sufficient for analysis despite being lower than desired in rounds 1 and 2.

two independent pieces of DNA. In a haploid or homozygous clonal sample, all readings of the same location in the genome are expected to be the same, and variants present at low frequencies are assumed to be from errors in the sequencing process or mutations acquired during culturing of the clone in preparation for sequencing. In a heterozygous or mixed population sample, readings of the same location can have variants as a result of mutations. Higher coverage – more readings of the same location – increases the confidence that the proportion of a mutation in a population has been determined accurately, and that a variant present at low frequency represents a real mutation instead of a sequencing error. This confidence depends on independent random sampling of the DNA; because overlapping reads of the same insert are not independent, the overlaps were excluded from analysis.

The first round included one clone from each of the first three evolved populations, their mutator ancestor clone, and their respective NaCl/HU-selected pools for bulk segregant analysis. Sequencing coverage averaged 25.7 for the four clones and 39.0 for the three pools but with wide variation (Figure 3.2B, Table 3.1). This first round of sequencing yielded an obvious top candidate mutation for Evolved 1 but inconclusive results for Evolved 2, due to low coverage, and Evolved 3, due to a large number of mutations linked to the mating locus. These clones were resequenced in the second round, and the bulk segregant analysis was entirely redone with new diploid strains and more rounds of selection. The analysis presented below is based on the sequencing results from the second bulk segregant analysis.

The second round of sequencing included one clone from each of the eleven evolved populations, five different ancestor clones, three pools for each bulk segregant analysis of six mutator clones, and additional clones and pools. Coverage averaged 13.4 for 22

Table 3.1 Sequencing coverage of clones and pools.

Round 1 Clones	
Sample	Coverage
Mutator ancestor A	22.2
Evolved 1-1	46.0
Evolved 2-1	16.6
Evolved 3-1	17.9
Average	25.7

Round 2 Clones	
Sample	Coverage
Ancestor A	12.8
Ancestor B	13.8
Mutator ancestor A	16.2
Mutator ancestor B	12.8
Mutator ancestor C	14.3
Evolved 1-1	14.6
Evolved 2-1	13.9
Evolved 3-1	16.5
Evolved 4-1	12.6
Evolved 5-1	15.1
Evolved 5-2	12.5
Evolved 6-1a	13.8
Evolved 6-1b	13.9
Evolved 7-1	12.9
Evolved 8-1	12.8
Evolved 9-1	12.2
Evolved 10-1	13.6
Evolved 11-1	11.2
Evolved 11-2	11.2
Mut. anc. A evolved in YPD	13.5
Mut. anc. B evolved in YPD	11.7
Mut. anc. C evolved in YPD	12.1
Average	13.4

Round 3 Clones	
Sample	Coverage
Evolved 2-2	29.6
Evolved 2-3	35.7
Evolved 2-4	37.6
Evolved 3-2	28.5
Evolved 3-3	33.7
Evolved 3-4	34.6
Evolved 4-2	33.5
Evolved 4-3	34.4
Evolved 4-4	18.6
Evolved 5-3	30.0
Evolved 5-4	31.5
Evolved 5-5	25.2
Evolved 6-2	28.2
Evolved 6-3	30.3
Evolved 6-4	29.5
Evolved 7-2	31.3
Evolved 7-3	25.9
Evolved 7-4	34.1
Evolved 8-2	38.2
Evolved 8-3	27.0
Evolved 9-2	36.2
Evolved 9-3	37.0
Evolved 11-3	31.6
Evolved 11-4	35.5
Average	31.6

Table 3.1 (continued) Sequencing coverage of clones and pools.

Round 1 Pools

Sample	Coverage
Evolved 1 NaCl/HU	40.9
Evolved 2 NaCl/HU	15.6
Evolved 3 NaCl/HU	60.6
Average	39.0

Round 2 Pools

Sample	Pool	Coverage
Evolved 1 spores	Initial	26.0
	NaCl/HU	30.6
	YPD	18.9
Evolved 2a spores	Initial	24.6
	NaCl/HU	29.6
	YPD	6.6
Evolved 2b spores	Initial	23.8
	NaCl/HU	54.6
	YPD	31.3
Evolved 3 spores	Initial	40.5
	NaCl/HU	32.2
	YPD	4.0
Evolved 4 spores	Initial	23.7
	NaCl/HU	28.3
	YPD	20.3
Evolved 5-1 spores	Initial	24.7
	NaCl/HU	32.8
	YPD	26.1
Evolved 5-2 spores	Initial	21.7
	NaCl/HU	28.5
	YPD	16.8
Evolved 6 spores	Initial	26.3
	NaCl/HU	29.5
	YPD	50.4

Round 2 Pools (continued)

Sample	Pool	Coverage
Mut. anc. A spores	Initial	18.9
	NaCl/HU	28.9
	YPD	25.5
Mut. anc. C spores	Initial	25.6
	NaCl/HU	33.7
	YPD	48.2
Mut. anc. A (evolved in YPD) spores	Initial	22.2
	NaCl/HU	25.1
	YPD	19.8
Mut. anc. C (evolved in YPD) spores	Initial	22.8
	NaCl/HU	24.8
	YPD	27.6
Average		27.1

Round 3 Pools

Sample	Coverage
Evolved 3 spores YPD (redo)	31.7
Final Population 1	28.8
Final Population 2	41.2
Final Population 3	44.5
Final Population 4	34.3
Final Population 5	42.8
Final Population 6	51.1
Final Population 7	29.6
Final Population 8	49.3
Final Population 9	42.4
Final Population 10	78.9
Final Population 11	52.2
Average	43.9

clones and 27.1 for 36 pools (Figure 3.2B). Two pools in the second round were too low for analysis; one was resequenced in round 3 and the other was redundant. The final evolved populations were sequenced in the third round, along with additional clones from each population that did not contain the known causal mutations identified by round 2. Coverage averaged 31.6 for 24 clones and 43.9 for 12 pools (Figure 3.2B).

Bulk segregant analysis of mutator clones

Bulk segregant analysis was performed for one clone from each mutator population, Evolved 1 to 6 (Figure 3.1). The clones used for bulk segregant analysis are those denoted as clone 1 in Chapter 2. To identify causal mutations for NaCl-dependent survival as well as mutations improving growth in standard media, pooled haploid cells were split into two cultures and subjected to five rounds of NaCl/HU selection or to continuous growth in standard media (Figure 3.1B). Genomic DNA was prepared from the final cultures and sequenced by Illumina sequencing (see section 4.6 for methods). The frequency of each mutation in the pool was determined from the numbers of independent reads with and without the mutation. Mutations were only considered for causality if they were present in the evolved clone but not in the direct ancestor. The mutations segregating above 80% in the NaCl/HU-selected pool for each clone are listed in Table 3.2. Mutations are listed by their position in the genome to better illustrate linkage. Only mutations that occur in open reading frames (ORFs) or promoter regions are listed; promoter regions were defined as the 500 bp upstream of an ORF. Regions with mutations not listed include terminators, autonomous replicating sequences (ARS), and small nucleolar RNAs (snoRNAs). Only one affected region is listed per mutation; where a mutation occurs in more than one

region (e.g., the ORF of one gene and the promoter of an adjacent gene), the region to be listed was determined by the following order of priority: ORF (nonsynonymous), Promoter, ORF (synonymous).

Genomic DNA from the initial pool of spores was also sequenced to account for biases in initial frequencies, which can result from close linkage to a locus used in the selection for haploid spores (see section 4.5 for more details on spore selection). Our method of selecting for haploid spores used genetic markers, specifically the mating locus from the evolved clone and the *URA3* locus from the (mating-type switched) ancestor clone, to select for *MAT***a** haploids. Therefore, any mutations in the evolved clone linked to the *MAT***a** locus (chromosome 3, position 198671 - 201177) would be enriched and those linked to the *URA3* locus (chromosome 5, position 116167 - 116970) would be depleted in the initial pool. Indeed, all six evolved clones have one or more mutations whose presence in the top 80% of segregants could be explained by linkage to the mating locus (Table 3.2). The top four segregating mutations in the analysis of Evolved 3-1 are above 85% in all three pools and are located within 27 kb of the *MAT***a** locus.[1] I further evaluated the mutations in Evolved 3-1 by testing linkage of the evolved phenotype to mating type. Spores from dissected tetrads were genotyped for mating type and tested for HU survival. The absence of linkage eliminated these mutations from consideration.

---

[1]One centiMorgan (cM) in yeast is 3 kb on average, i.e., mutations within 3 kb are separated in 1% of meioses.

Table 3.2 Mutations segregating above 80% in NaCl/HU-selected pools.

Evolved 1-1

Chrom.	Position	Type	Anc.	Evo.	Initial	NaCl/HU	YPD	Gene	Region
3	111334	SNP	T	C	75	80	68	*RER1*	Promoter
4	88103	SNP	T	G	100	100	100	*GLE1*	Promoter
4	125696	Insertion	C	+A	100	96	100	*PPH22*	Promoter
4	133160	Deletion	T	-C	92	100	100	*LYS20*	Promoter
4	140535	SNP	C	A	100	100	100	*DLD2*	ORF (nonsynonymous)
10	99025	SNP	A	C	14	92	8	*CPS1*	ORF (nonsynonymous)
10[1]	**107918**	**SNP**	**C**	**A**	**23**	**100**	**18**	***HAL5***	**ORF (nonsynonymous)**
12	138436	SNP	A	G	53	80	80	*SPO75*	ORF (nonsynonymous)
12	422828	SNP	A	G	61	80	81	*SLS1*	ORF (synonymous)
12	525560	SNP	T	A	100	100	100	*SKG3*	ORF (nonsynonymous)
12	739964	SNP	G	T	100	100	100	*STT4*	ORF (nonsynonymous)
12	787331	SNP	A	G	100	100	100	*REC102*	ORF (nonsynonymous)
12	798118	SNP	T	A	100	100	100	*NUP2*	ORF (nonsynonymous)
13	495276	SNP	A	C	100	100	100	*FOL3*	ORF (synonymous)
13	887055	SNP	C	T	100	100	100	*GAS1*	ORF (synonymous)
15	429280	SNP	C	T	44	80	69	*VHS3*	ORF (nonsynonymous)
16	76257	SNP	C	A	100	100	100	*RPL36B*	Promoter
16	80486	SNP	C	T	100	100	100	*GAL4*	ORF (nonsynonymous)

[1]Mutations listed in bold were tested for causality by strain engineering.

64

Table 3.2 (continued) Mutations segregating above 80% in NaCl/HU-selected pools.

Evolved 2-1a[2]

Chrom.	Position	Type	Anc.	Evo.	Initial	NaCl/HU	YPD	Gene	Region
3	200335	SNP	A	G	n.d.	100	n.d.[3]	*MATALPHA2*	Promoter
3	201933	SNP	A	G	93	100	n.d.	*TAF2*	ORF (nonsynonymous)
8	**291627**	**SNP**	**G**	**T**	**53**	**100**	**n.d.**	***HXT1***	**ORF (nonsynonymous)**
14	**18700**	**SNP**	**T**	**A**	**56**	**97**	**50**	***RPD3***	**ORF (nonsynonymous)**
14	84416	SNP	G	A	37	96	n.d.	*MID1*	ORF (nonsynonymous)
14	88557	SNP	T	A	47	85	66	*PCL1*	ORF (nonsynonymous)
14	125101	SNP	A	G	53	80	50	*TOF1*	ORF (synonymous)
16	516778	SNP	A	G	40	80	40	*VTC3*	ORF (synonymous)

Evolved 2-1b[2]

Chrom.	Position	Type	Anc.	Evo.	Initial	NaCl/HU	YPD	Gene	Region
3	200335	SNP	A	G	100	100	n.d.	*MATALPHA2*	Promoter
3	201933	SNP	A	G	92	95	100	*TAF2*	ORF (nonsynonymous)
8	**291627**	**SNP**	**G**	**T**	**45**	**90**	**9**	***HXT1***	**ORF (nonsynonymous)**
14	**18700**	**SNP**	**T**	**A**	**46**	**83**	**44**	***RPD3***	**ORF (nonsynonymous)**
14	84416	SNP	G	A	45	88	69	*MID1*	ORF (nonsynonymous)

[2]Evolved 2-1a and 2-1b are separate sporulations and bulk segregant analyses of the same evolved clone, 2-1.

[3]n.d. (no data): Position had fewer than 4 mapped reads. The YPD pool for Evolved 2-1a had very low coverage (6.6x).

65

Table **3.2 (continued)** Mutations segregating above 80% in NaCl/HU-selected pools.

Evolved 3-1

Chrom.	Position	Type	Anc.	Evo.	Initial	NaCl/HU	YPD	Gene	Region
3	171817	SNP	T	C	85	95	85	FEN2	ORF (nonsynonymous)
3	183278	SNP	G	T	100	90	96	BPH1	ORF (nonsynonymous)
3	196461	SNP	C	A	100	97	91	PHO87	ORF (nonsynonymous)
3	208906	Deletion	T	-C	87	100	89	RRT12	ORF (nonsynonymous)
4[4]	**1155677**	**SNP**	C	**T**	**10**	**39**	**7**	**HXT7**	**ORF (nonsynonymous)**
4[4]	**1161069**	**SNP**	C	**T**	**15**	**54**	**5**	**HXT6**	**ORF (nonsynonymous)**
4	1222686	SNP	A	C	40	93	19	VPS74	Promoter
5	436706	SNP	A	C	52	81	100	YER133W-A	Promoter
5	489297	SNP	A	T	40	83	59	YER158C	ORF (synonymous)
5	**524109**	**SNP**	C	**T**	**61**	**95**	**70**	**RPH1**	**ORF (nonsynonymous)**
5	547076	SNP	T	C	52	88	70	PDA1	ORF (synonymous)
9	83849	SNP	A	T	30	80	58	CCT2	ORF (nonsynonymous)

Evolved 4-1

Chrom.	Position	Type	Anc.	Evo.	Initial	NaCl/HU	YPD	Gene	Region
3	67527	SNP	A	C	50	81	47	HIS4	ORF (nonsynonymous)
3	211873	SNP	C	T	88	80	75	BUD23	Promoter
12	368807	SNP	C	A	80	100	12	AHP1	ORF (nonsynonymous)
12	**372115**	**SNP**	C	**T**	**66**	**100**	**9**	**HOG1**	**ORF (nonsynonymous)**

[4]The data for mutations in HXT6 and HXT7 represent a single mutation in HXT7; see text for details.

Table 3.2 (continued) Mutations segregating above 80% in NaCl/HU-selected pools.

Evolved 5-1

Chrom.	Position	Type	Anc.	Evo.	Initial	NaCl/HU	YPD	Gene	Region
3	121490	SNP	A	G	68	80	72	*CIT2*	ORF (nonsynonymous)
8	**291627**	**SNP**	**G**	**T**	**27**	**96**	**0**	***HXT1***	**ORF (nonsynonymous)**

Evolved 5-2[5]

Chrom.	Position	Type	Anc.	Evo.	Initial	NaCl/HU	YPD	Gene	Region
4	892312	SNP	C	A	15	82	18	*UPC2*	ORF (nonsynonymous)

Evolved 6-1

Chrom.	Position	Type	Anc.	Evo.	Initial	NaCl/HU	YPD	Gene	Region
3	183804	SNP	T	C	91	93	88	*BPH1*	ORF (nonsynonymous)
12	370038	SNP	A	G	42	94	23	*YLR111W*	Promoter
12	**372142**	**SNP**	**G**	**A**	**55**	**100**	**8**	***HOG1***	**ORF (nonsynonymous)**
16	203667	SNP	C	T	45	81	64	*CTI6*	ORF (nonsynonymous)

[5]A second clone from population 5 was evaluated because it had a different survival phenotype than Evolved 5-1 (see text for details).

3.4 Determining candidate causal mutations from sequence data

Putative causal mutations in the mutator strains were identified based on segregation frequencies in the three pools, the annotated function of the affected gene and any notable mutant phenotypes. In Evolved 1-1, 14 mutations segregate above 90% in the NaCl/HU pool. However, *HAL5* and *CPS1* are the only mutations segregating above 90% in the NaCl/HU pool that are also at initial frequencies below 90%. The other 12 mutations, present at or close to 100% in all three pools, likely owe their high frequencies to mitotic gene conversion events in the diploid prior to expansion of the culture for sporulation. *HAL5* encodes a putative protein kinase with a previously-described role in salt tolerance, and *CPS1* encodes a vacuolar carboxypeptidase. These genes lie within 10 kb of each other in the genome, and the mutation in *HAL5* was determined more likely to be causal due to its reported NaCl-related mutant phenotypes. This analysis process was repeated for each mutator clone; the results are summarized in Figure 3.3.

All mutations considered here had high representation in the bulk segregant analysis NaCl/HU pool. For Evolved 2-1, two diploid cells were separately sporulated and subjected to bulk segregant analysis, providing experimental replicates for these data. The mutation in *RPD3*, encoding a histone deacetylase, was considered for causality due to the previously described involvement of Rpd3 in potassium transport. The *HXT1* mutation in Evolved 2-1 is identical to that in Evolved 5-1 and therefore was considered very likely to be causal for both clones, despite no clear role for Hxt1, a hexose transporter, in salt or osmotic stress response. Evolved 4-1 and 6-1 each contain a mutation in *HOG1*,
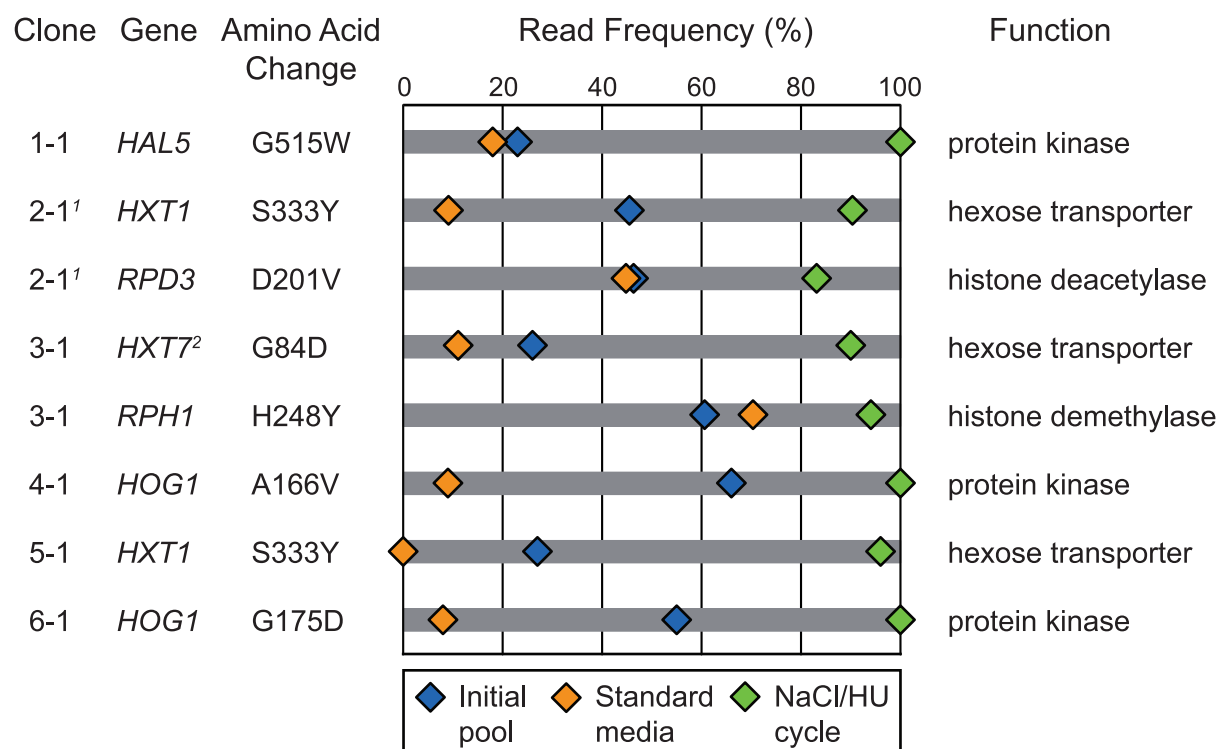
Figure 3.3 Top candidate mutations from bulk segregant analysis.
The top candidate mutations from the bulk segregant analysis of each evolved clone were chosen based on having frequencies above 90% after passage through the NaCl/HU selection cycle and initial frequencies around 50%. While not a criterion for candidacy, almost all showed decreases in standard media from their initial frequencies.
[1]Frequencies are from Evolved 2-1b analysis because it had higher coverage than 2-1a.
[2]Frequencies for *HXT7* are estimated from reads mapping to both *HXT6* and *HXT7*.

which encodes a protein kinase that is a key transducer of the osmotic stress response. The survival of a *hog1Δ* strain was previously described in section 2.3 (see Figure 2.5 for data). Consistent with a *hog1* loss-of-function phenotype, clones Evolved 4-1 and 6-1 also survive HU treatment with the osmolytes KCl and sorbitol (Figure 2.9, Table 2.1), providing further evidence that the *HOG1* mutations are causal.

In selecting clones from population 5, the clones displayed varying survival phenotypes: Evolved 5-1 had NaCl-dependent survival, while 5-2 had NaCl-independent survival of HU. However, Evolved 5-2 was not retested to confirm this result, and no other clones

analyzed from this population showed NaCl-independent survival. The only mutation segregating at above 80% in the NaCl/HU pool of Evolved 5-2 is in *UPC2*, which encodes a transcriptional activator of sterol biosynthesis genes. The Evolved 5-2 clone could not be retested because the original evolved clones used in the bulk segregant analysis for Evolved 4, 5, and 6 were unfortunately not saved for future work. Genomic DNA from the Evolved 5 clones was sequenced, but DNA samples for the Evolved 4 and 6 clones were unavailable. Instead, the diploid strains used for their bulk segregant analysis were sequenced and analyzed using different parameters to account for heterozygosity of the evolved mutations present. These analyses still gave clear results for the causality of the *HOG1* mutations.

Besides the mutations linked to the mating locus, the top segregants in Evolved 3-1 are a nonsynonymous mutation in *RPH1*, encoding a histone demethylase, and a mutation in the promoter of *VPS74*. The promoter mutation was considered unlikely to be causal, and the search for a causal mutation in Evolved 3-1 was renewed after strain engineering showed that *RPH1* was not causal (results described in section 3.5). The search focused on the region around *VPS74*, with the reasoning that the high frequency of its promoter mutation could be due to a linked causal mutation. After much consternation, a mutation in *HXT7*, another hexose transporter gene, was discovered. This mutation was undetected in earlier analyses because of the uncertainty inherent in the process of mapping very short reads to duplicated regions in the genome. *HXT7* has 99% sequence identity with the adjacent[2] ORF *HXT6*: there are only 3 mismatches over their coding regions of

---

[2]The adjacency of the ORFs did not contribute to the uncertainty in mapping, but it did cause Sanger sequencing to be required for determining which ORF was affected, since a mutation in either would show similar linkage to *VPS74*. It also, however, made it easier to identify the clonal heterogeneity at both sites in the sequence files.

Figure 3.4 Linkage to *HXT7* mutation.

The mutation in *HXT7* (at position 0) was calculated to be at a frequency of about 50% in the evolved clone and thus went undetected in initial analyses. The high frequency of the mutation in the *VPS74* promoter region in NaCl/HU pool 2 (arrow) led to closer examination of this region in the variant detection step of the sequencing pipeline. NaCl/HU pools 1 and 2 are from the first and second bulk segregant analyses of Evolved 3-1, respectively. The Initial and Standard media pools were done concurrently with NaCl/HU pool 2.

Linkage to *HXT7* mutation
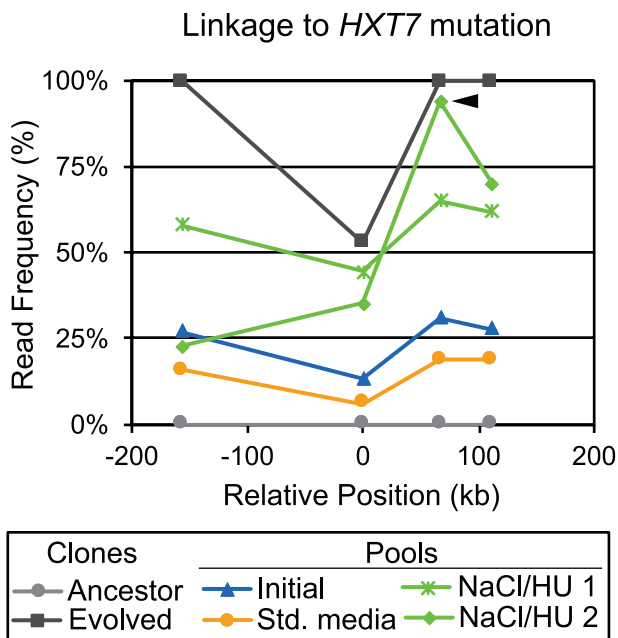
1713 bp plus 83 bp upstream. Therefore, Illumina sequencing reads (150 bp) cannot be unambiguously assigned to one region over the other. In our sequence analysis pipeline, reads with multiple matches were randomly assigned to one matching location. This resulted not only in the reported mutation frequency being lower than its true frequency in the sequenced pools, but also in the mutation not being detected in the evolved clone: our analysis used a threshold of 90% read frequencies for mutations in a clone, whereas the *HXT7* mutation was at a frequency close to 50% (Figure 3.4). Examination of the variant detection files revealed mutations present at heterogeneous levels in the evolved clone, and running the analysis with a lower threshold for clone mutations yielded the information on mutations in *HXT6* and *HXT7* listed in Table 3.2. Sanger sequencing with primers outside of the identical regions determined that this mutation is in *HXT7*. The frequencies for this mutation depicted in Figure 3.3 were estimated by doubling the proportion of variant reads out of all reads mapped to both of the *HXT6* and *HXT7* loci.

Clones from the non-mutator evolved populations were sequenced without back-

crossing; each contains fewer than ten mutations. All mutations in coding regions resulting in nonsynonymous changes and in promoter regions are listed in Table 3.3. Parallelism in the genes mutated aided in identifying mutations as potentially causal. The clones from populations 7 and 8 contain mutations in *HXT* genes: for Evolved 7, a mutation in *HXT3*, and for Evolved 8, a mutation in *HXT1* different from the mutation seen in Evolved 2 and 5. Evolved 9 and 10 each have a mutation in *IMP2'*, which encodes a protein characterized as a transcriptional activator. Initial characterization of population 11 yielded clones with two different survival phenotypes, and one clone of each type was sequenced. Evolved 11-1 has NaCl-dependent survival, and the nonsynonymous mutation in *TUP1*, encoding a transcriptional repressor, was judged the most likely to be causal. Evolved 11-2 is a clone with NaCl-independent survival of HU, and its single mutation is in *ERG10*, an essential gene that encodes an enzyme in the ergosterol biosynthesis pathway. Candidate mutations were confirmed by Sanger sequencing before commencing with strain construction.

Genetic diversity of final populations

The frequencies of candidate causal[3] mutations in the final populations were measured by Sanger sequencing of the final population and of ten newly-selected clones from the final population (Figure 3.5). Allele frequencies in the final populations were determined from relative peak heights of the Sanger traces from the population sample. In most but not all cases, the number of the ten sampled clones with the mutant allele approximated the frequencies calculated by peak heights. Surprisingly, known causal mutations were

---

[3]By this point, most of the causal mutations had been confirmed by strain construction.

Table 3.3 Mutations in non-mutator evolved clones.

Evolved 7

Chrom.	Position	Type	Anc.	Evo.	Gene	Region
4[1]	**1163347**	**SNP**	**G**	**A**	***HXT3***	**ORF (nonsynonymous)**
7	1039999	SNP	C	A	*TAF1*	ORF (nonsynonymous)
10	**56922**	**SNP**	**C**	**T**	***ACO2***	**ORF (nonsynonymous)**

Evolved 8

Chrom.	Position	Type	Anc.	Evo.	Gene	Region
8	**291541**	**SNP**	**T**	**C**	***HXT1***	**ORF (nonsynonymous)**
15	574117	SNP	A	C	*VPS17*	ORF (nonsynonymous)

Evolved 9

Chrom.	Position	Type	Anc.	Evo.	Gene	Region
9	**54095**	**SNP**	**A**	**C**	***IMP2′***	**ORF (nonsynonymous)**
10	356959	SNP	G	C	*SDH1b*	ORF (nonsynonymous)

Evolved 10

Chrom.	Position	Type	Anc.	Evo.	Gene	Region
9	**54651**	**SNP**	**C**	**A**	***IMP2′***	**ORF (nonsynonymous)**

Evolved 11-1

Chrom.	Position	Type	Anc.	Evo.	Gene	Region
2	473142	SNP	G	C	*LYS2*	ORF (nonsynonymous)
3	262417	SNP	A	T	*TUP1*	ORF (nonsynonymous)
10	585961	SNP	C	G	*STE18*	Promoter
15	424744	SNP	T	C	*YOR050C*	Promoter
16	292186	SNP	T	C	*SPP1*	ORF (nonsynonymous)

Evolved 11-2

Chrom.	Position	Type	Anc.	Evo.	Gene	Region
16	498488	SNP	C	G	*ERG10*	ORF (nonsynonymous)

[1]Mutations listed in bold were tested for causality by strain engineering.

Figure 3.5 Allele frequencies in final populations.

Allele frequencies were determined by Sanger sequencing (blue), sampling ten clones (green) and whole-genome sequencing (WGS, orange). Peak heights of Sanger sequence traces were compared using QSV Analyzer. Ten clones from the final population were selected from a plated sample and genotyped by Sanger sequencing. Read frequencies from whole-genome sequencing were determined from numbers of independent reads.
x: clone genotyping was not performed for *HXT7-G84D* in population 3 clones.

Frequencies in Final Populations
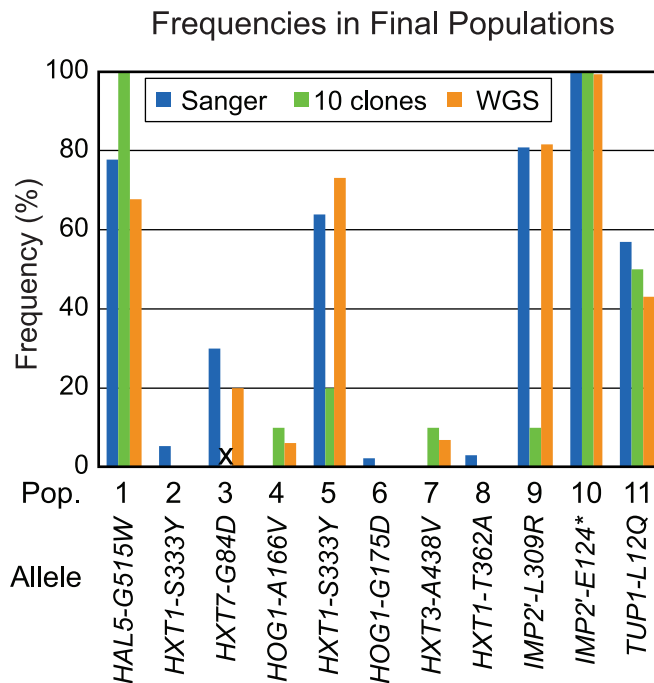
at 10% or less in five of the eleven final populations. Only one population had a causal mutation at 100%.

Due to this heterogeneity, all final populations were analyzed by whole-genome sequencing, along with clones not containing the candidate causal mutations. The frequencies determined by whole-genome sequencing closely matched those from the Sanger sequencing (Figures 3.5). Sequencing clones that lacked the initially identified mutations revealed other instances of *HXT1-S333Y*, *HXT7-G84D*, and *TUP1-L12Q*, early stop codon mutations in *HAL5* and *HOG1*, and three mutations in other *HXT* genes (Figure 3.6, Table 3.4).

The multiple instances of the same mutations arose independently, rather than as a result of contamination between populations, evidenced by each clone sharing mutations with other members of its population that are not found in other populations (data
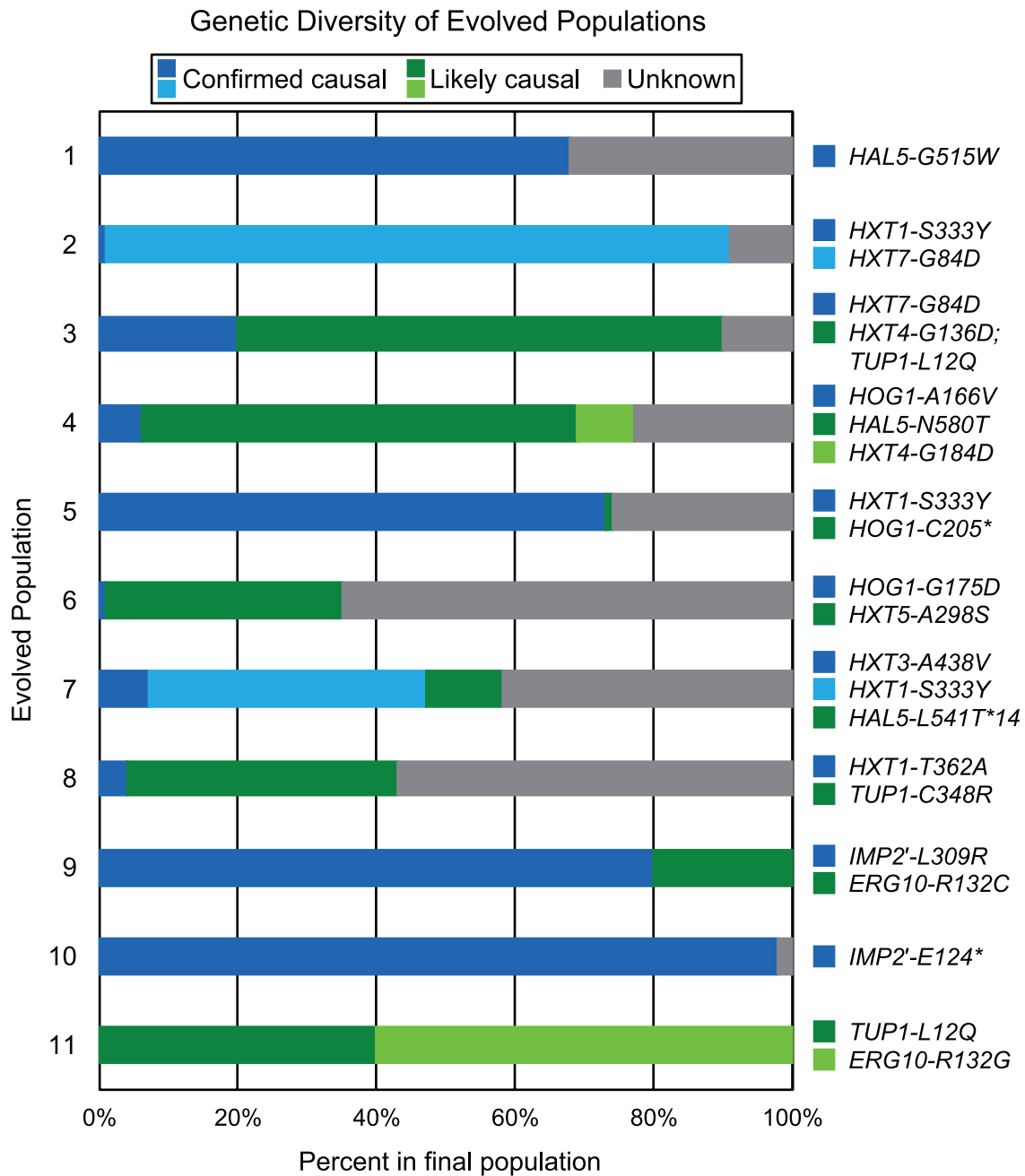
74

Figure 3.6 Genetic diversity in final evolved populations.

In five of eleven final populations, mutations that were confirmed to be causal (blues) account for 50% or more of the population. Greens represent "likely causal" mutations based on similarity to causal mutations or other parallelism. Including these still leaves most populations with a substantial fraction of cells with unidentified causal mutations (gray). Capitalization of gene names is not meant to indicate functional effect of mutation.

Table 3.4 Candidate causal mutations in all sequenced clones.

Pop.	Ancestor	Clone	Causal Mutations Confirmed	Likely	Other Mutations
1	Mut. ancestor A	1	HAL5-G515W		
2	Mut. ancestor A	1	HXT1-S333Y		RPD3-D201V
		2	HXT7-G84D		
		3			
		4			
3	Mut. ancestor A	1	HXT7-G84D		RPH1-H248Y
		2		HXT4-G136D, TUP1-L12Q	
		3			
		4			
4	Mut. ancestor A	1	HOG1-A166V		
		2		HXT4-G184D	
		3		HAL5-N580T, HAL5-Q376*	
		4	HOG1-A166V		
5	Mut. ancestor B	1	HXT1-S333Y		
		2			UPC2-L855I
		3			GRS2-L263P
		4		HOG1-C205*	
		5	HXT1-S333Y		
6	Mut. ancestor C	1	HOG1-G175D		CTI6-P83L
		2		HXT5-A298S	PPZ1-D597V
		3			
		4			PSR2-L166P, CMK1-R331S

Asterisks (*) denote early stop codons. Capitalization of gene names is not meant to indicate the functional effect of the mutation; it is used for consistency to avoid assumptions about untested mutations.

Table 3.4 (continued) Candidate causal mutations in all sequenced clones.

| Pop. | Ancestor | Clone | Causal Mutations | | Other Mutations |
			Confirmed	Likely	
7	Ancestor A	1	HXT3-A438V		
		2		HAL5-L541T*14	
		3	HXT1-S333Y		BZZ1-A385T
		4	HXT3-A438V		
8	Ancestor A	1	HXT1-T362A		VPS17-K315T
		2			SNQ2-A1233D, EGT2-delC246
		3		TUP1-C348R	
9	Ancestor B	1	IMP2′-L309R		
		2		ERG10-R132C	
		3	IMP2′-L309R		
10	Ancestor B	1	IMP2′-E124*		
11	Ancestor B	1		TUP1-L12Q	
		2		ERG10-R132G	
		3		TUP1-L12Q	
		4		ERG10-R132G	

Asterisks (*) denote early stop codons. Capitalization of gene names is not meant to indicate the functional effect of the mutation; it is used for consistency to avoid assumptions about untested mutations.

not shown). Also notable are the *ERG10-R132C* and *ERG10-R132G* mutations from populations 10 and 11, respectively (Table 3.4). All clones with one of these mutations had NaCl-independent HU survival (9-2, 11-2 and 11-4 in Table 2.2). The population frequencies of these "likely causal" mutations are depicted in Figure 3.6 as well. Despite including these, most populations still have a substantial fraction (10 to 65%) of cells with unidentified causal mutations. Some best guesses are included in Table 3.4 under "Other Mutations" based on gene associations with stress response and/or salt homeostasis (*PPZ1-D597V* in Evolved 6-1, *PSR2* and *CMK1* in Evolved 6-4) or their prominence in the final population (*SNQ2* and *EGT2* in Evolved 8-2). These guesses in mutator clones should be taken with many grains of salt, as these clones typically had over 100 mutations identified by the sequencing pipeline.

3.5 Strain engineering to test causality

Candidate causal mutations were tested by allele replacement, both by engineering the mutation into the ancestor background ("reconstruction") and by reversion of the mutation to wild-type in the evolved clone (Figure 3.7). By these methods, I confirmed mutations in six different genes that individually confer increased salt-dependent HU survival (Table 3.5). Mutations in the genes *RPD3* from Evolved 2-1, *RPH1* from Evolved 3-1, and *ACO2* from Evolved 7-1 were also tested and found to be non-causal (data not shown). All of the mutations tested were single nucleotide changes resulting in either an amino acid substitution or an early stop codon.

I tested the survival phenotypes of gene deletions in both the ancestor and evolved

Figure 3.7 Strain constructions to test causality of mutations.
Mutations identified in the evolved clones were engineered into an ancestral strain (non-mutator A) to test their sufficiency for conferring survival in hydroxyurea. Mutations in the evolved clones were reverted to the wild-type allele to test necessity. Whether the mutations constituted a loss or gain of function of the encoded protein was determined by comparing the phenotype of deletion strains of both backgrounds to the evolved phenotype. Dominance was determined from the phenotype of the heterozygous diploid.

backgrounds (Table 3.6). Gene deletions mimic the phenotype of the evolved mutations in *HAL5*, *HOG1*, and *IMP2′*, indicating loss-of-function mutations. For mutations in *HXT1*, *HXT3* and *HXT7*, gene deletions do not confer survival and deleting the gene in the evolved clones results in loss of survival, indicating that these mutations are gain-of-function. Dominance of the mutations was evaluated by testing HU survival of heterozygous diploids made from mating evolved clones with a non-evolved strain[4] (Table 3.6). The loss-of-function mutations in *HAL5*, *HOG1*, and *IMP2′* are recessive; the gain/change-of-function mutations in *HXT* genes are dominant except for *HXT1-T362A*.

Deletion of *TUP1* confers salt-independent HU survival (Table 3.6). This deletion also results in robust flocculation; the evolved clones with the *TUP1-L12Q* mutation show mild flocculation, indicating that this mutation is likely hypomorphic (partial loss-

---

[4]This actually tests whether the evolved *phenotype* is dominant or recessive. A direct test of the mutation would be to use a diploid made from the reconstructed strain and the ancestor, with no other evolved mutations present. Ancestor-Evolved diploid strains were used for expediency.

Table 3.5 Hydroxyurea survival of allele-replaced strains.

			Evolved Clone		Reconstruction		Reversion	
Gene	Mutation	Strain	Std.	NaCl	Std.	NaCl	Std.	NaCl
HAL5	G515W	1-1	-	+ + +	-	+ +	-	-
HOG1	A166V	4-1	-	+ + +	-	+ +	-	+
	G175D	6-1	-	+ + +	-	+	-	-
IMP2'	E124*	10-1	-	+ +	+	+ + +	-	-
	L309R	9-1	-	+ +	-	+ + +	-	-
HXT1	S333Y	2-1	-	+ + +	-	+ + +	-	-
	S333Y	5-1	-	+ + +			-	+
	T362A	8-1	-	+ + +	-	+ +	-	+
HXT3	A438V	7-1	-	+ + +	-	+ + +	-	+
HXT7	G84D	3-1	-	+ + +	-	+ + +	-	-

Table 3.6 Hydroxyurea survival of deletion and diploid strains.

			Anc. Deletion		Evo. Deletion		Diploid	
Gene	Mutation	Strain	Std.	NaCl	Std.	NaCl	Std.	NaCl
HAL5	G515W	1-1	-	+ +	-	+ + +	-	-
HOG1	A166V	4-1	-	+ + +	-	+ + +	-	-
	G175D	6-1			-	+ + +	-	-
IMP2'	E124*	10-1	+	+ + +	+	+ +	-	-
	L309R	9-1			+ +	+ + +	-	-
TUP1	L12Q	11-1	+ + +	+ + +	n.t.[1]	n.t.[1]	-	-
HXT1	S333Y	2-1	-	-	-	-	-	+ +
	S333Y	5-1			-	+	-	+ + +
	T362A	8-1			+	+	-	-
HXT3	A438V	7-1	-	-	-	+	-	+ +
HXT7	G84D	3-1	-	-	-	-	-	+ + +

[1]n.t.: Not tested; the evolved clone 11-1 with *TUP1* deleted was not constructed.

of-function). The diploid did not have any survival, so it is recessive. The mutation was not tested by reconstruction due to the salt-independent survival phenotype of the deletion; early assays with the evolved clone also showed variable results, with some salt-independent survival.

Five of the nine reconstructed strains survived in HU as well as or better than the evolved clone. Oddly, the reconstructed strains containing *IMP2'* mutations survived better than Evolved 9-1 and 10-1. The deletion strains also showed some increase in survival without NaCl. These differences imply the presence of phenotype-modifying mutations in the evolved clones, possibly acquired by their common ancestor, Non-mutator Ancestor B. The four strains whose reconstructions survived worse than their evolved clones contain mutations in *HAL5*, *HOG1*, and *HXT1*. The *hog1-G175D* mutation was found to be necessary but not sufficient; the reconstructed strain had only a slight increase in survival whereas Evolved 6-1 had no survival when the mutation was reverted. The *hog1-A166V* mutation also conferred a survival level lower than its evolved clone, Evolved 4-1. Because the ancestor deletion strain shows a stronger survival phenotype, these mutations are probably hypomorphic instead of a full loss of function, and Evolved 4-1 and 6-1 may have secondary mutations contributing to their survival.

Both the *hal5-G515W* mutation and deletion of *HAL5* conferred survival higher than the ancestor but lower than the evolved clone, indicating the potential presence of secondary mutations. Evolved 1-1 also contains a mutation in *VHS3*, which has a reported salt-sensitive null phenotype, that segregated at 80% in the bulk segregant analysis. The *HXT1-T362A* reconstructed strain also has lower survival than Evolved 8-1. The only other mutation in Evolved 8-1 is in *VPS17*; many *VPS* genes, involved in *v*acuolar *p*rotein *s*orting, have salt-sensitive null phenotypes but none has been reported for *VPS17* [10].

3.6 Phenotypes of reconstructed strains

Population growth phenotypes of reconstructed strains were measured as in section 2.5 (see section 4.4 for experimental details). These are compared to the phenotypes of the evolved clones from Chapter 2, grouped by the mutated gene. All allele replacements and deletions in an ancestor background were constructed using the same ancestor strain, non-mutator A, in order to compare phenotypes betweeen mutations. Thus, every "Ancestor" in the following charts is the non-mutator ancestor A. For comparisons between evolved clones and their direct ancestors, see section 2.5. Because population growth is a phenotype not dependent on the *mec1Δ sml1Δ* background, this was also assayed for reconstructed strains in a *MEC1 SML1* background (i.e., the typical lab strain). This is presented for all genes at the end to compare the mutant phenotypes to each other.

HAL5

The growth phenotypes of *HAL5* constructed strains closely match their levels of HU survival: the ancestor reconstruction and the ancestor deletion strains both have lower survival and higher growth rates in NaCl media than the Evolved 1-1 clone (Figure 3.8A). This again indicates the likely contribution of a second mutation to the evolved phenotype. However, reversion of *HAL5* completely restores the growth rate of the evolved clone in NaCl. The *HAL5* mutation appears to not greatly impact growth in standard media; the lower growth of Evolved 1 and the reverted strain is likely due to their mutator ancestor background (Figure 3.8B). The main function described for Hal5 is upregulating activity of the K^+ transporters Trk1 and Trk2 in response to salt stress [11, 12]. I
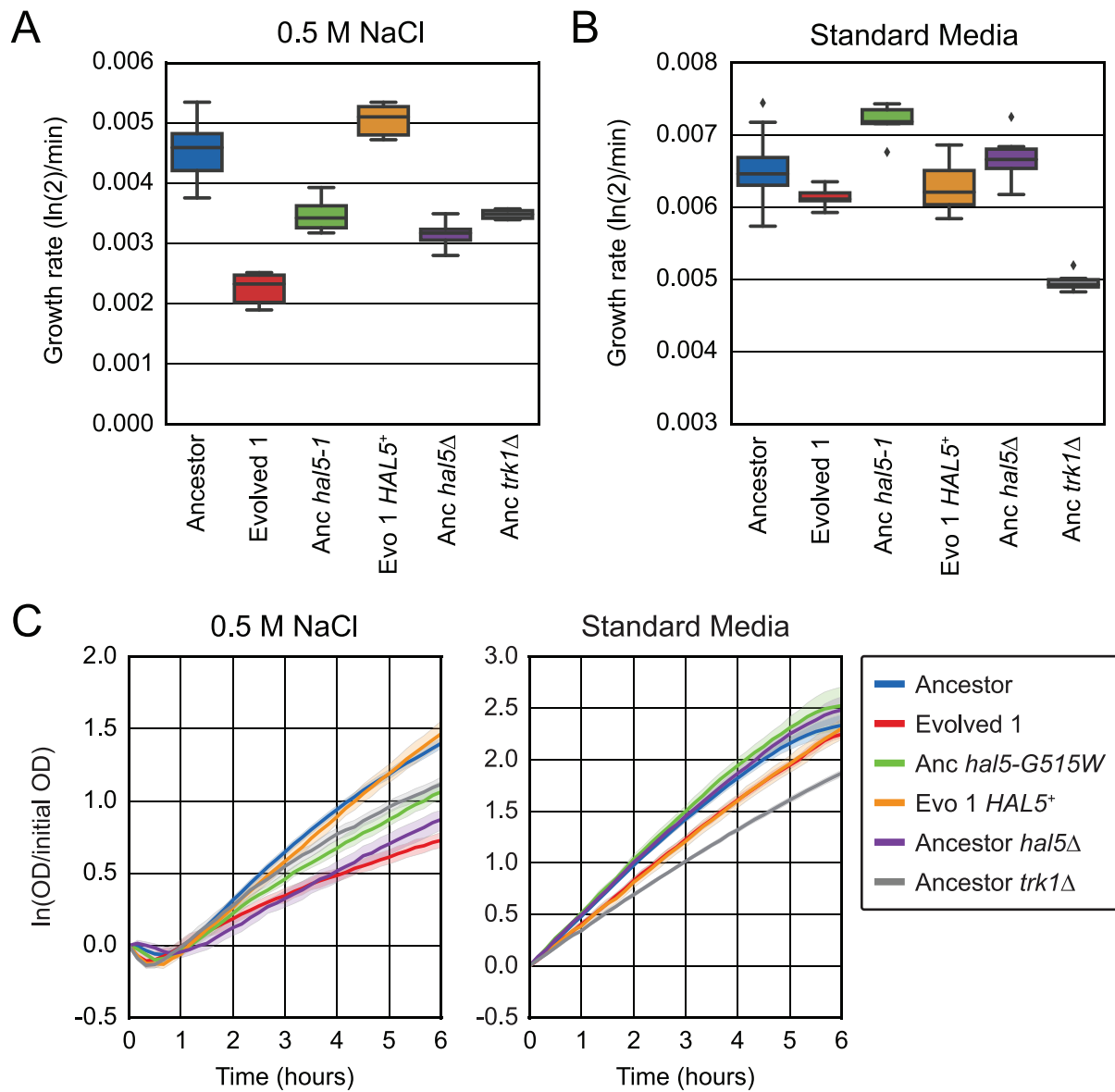
Figure 3.8 Growth rates of *HAL5* constructed strains.

(A) Reconstruction and deletion strains have reduced growth in NaCl media, but not to the level of Evolved 1-1. Deletion of *TRK1* matches the rate of the *hal5-G515W* and *hal5*Δ strains. The effective growth rate of each strain was calculated for the period of 2 to 6 hours after transfer into 0.5 M NaCl media, or from hour 2 until saturation if reached before hour 6. Rates were determined from six or more biological replicates.
(B) Deletion of *TRK1* confers a large growth defect in standard media, but deletion or mutation of *HAL5* does not. Growth rates are as in (A) except with standard media.
(C) Growth dynamics reflect the rate differences seen in (A) and (B). Raw data were calibrated by a blank value and transformed to account for the non-linearity above OD 0.6 before normalizing by initial OD. Lines are mean values from six or more replicates and shaded bands are bootstrap 95% confidence intervals.

tested HU survival of a strain lacking *TRK1*, the dominant transporter, and found it to be indistinguishable from the survival of strains containing the *hal5Δ* and *hal5-G515W* mutations in the ancestor background (- in standard media and ++ in NaCl), indicating that the *hal5* phenotypes are likely due to reduced activity of Trk1. This also indicates that the unknown secondary mutation in Evolved 1-1 is not exerting its effect through further suppression of Trk1 activity. A major difference between the *hal5* and *trkΔ* strains is seen in their growth rates in standard media (Figure 3.8B). Deletion of *TRK1* confers a strong defect in standard media, implying a need for Trk1 function in standard media conditions that does not depend on activation by Hal5.

HXT genes

Of the four different *HXT* mutations tested for causality by strain construction, three conferred a high level of hydroxyurea survival, and *HXT1-T362A* conferred a moderate level of survival. It is also the only recessive *HXT* mutation, despite all *HXT* mutations being gain-of-function. The growth phenotypes of the three *HXT* genes with mutations tested by strain construction are considered separately. The two alleles of *HXT1* have very different effects on growth rate in NaCl media; each matches the growth rate of its corresponding evolved clone(s) (Figure 3.9A). The *HXT1-S333Y* evolved clones and reconstructed strain have much lower growth rates than the *HXT1-T362A* strains. Growth deficits of the evolved clones in NaCl are eliminated by reverting the *HXT1* mutation. There is no strong defect in growth rates in standard media (Figure 3.9B).

For strains testing the *HXT1-S333Y* allele, the dynamics in NaCl media after the first hour segregate according to the strains' *HXT1* genotypes, matching the effective growth

rate calculations (Figure 3.9C). However, Evolved 2-1 shows different early dynamics that cannot be attributed its *HXT1* mutation; both it and the reverted strain have a slightly faster resumption of OD increase than Evolved 5-1 and the *HXT1-S333Y* reconstruction strain, indicating the possible contribution of a second mutation. Growth dynamics of these strains in standard media segregate by background, not the *HXT1* allele; this is more apparent from the aggregated growth curves than the effective growth rate (Figure 3.9B, C). For strains testing the *HXT1-T362A* allele, growth dynamics in NaCl media are initially identical for all strains, with a small degree of separation between hours 2 and 6 (Figure 3.9D). No differences are seen in standard media.

Evolved 7-1, which contains the *HXT3-A438V* allele, had the lowest growth rate in NaCl of the characterized evolved clones; this low rate is precisely matched by the reconstructed strain (Figure 3.10). The growth deficit of Evolved 3-1 in NaCl is also matched by its reconstructed strain containing the *HXT7-G84D* mutation (Figure 3.11). Two features of the growth dynamics of the *HXT7* strains suggest the influence of another mutation on the evolved clone's phenotype. The first is the faster resumption of OD increase in the evolved strain relative to the reconstructed strain. The second is the growth defect in standard media conferred by reverting the *HXT7* mutation, a result not seen in any other reverted strain.

HOG1

Two mutations in *HOG1* were tested for causality: *A166V* conferred a moderate level of HU survival, and *G175D* conferred a low level. This difference is reflected in the growth rates of the reconstructed strains in NaCl media: the lower-survival *hog1-G175D*

Figure 3.9 Growth rates of *HXT1* constructed strains.

(A) The *HXT1-S333Y* evolved and reconstructed strains have very low growth rates in NaCl media, in contrast to the *HXT1-T362A* strains. For the ancestor reconstruction strains, *HXT1-1* indicates the *HXT1-S333Y* allele and *HXT1-2* indicates the *HXT1-T362A* allele. The effective growth rate of each strain was calculated as described in Figure 3.8A. Rates were determined from six or more biological replicates.
(B) No strong growth defects are seen in standard media. Growth rates are as in (A) except with standard media.
(C) Growth dynamics in NaCl media show early recovery of Evolved 2-1 independent of its *HXT1* mutation; growth after the first hour segregates by *HXT1* genotype. Growth in standard media segregates by strain background, not the *HXT1* genotype. Growth curves are as described in Figure 3.8C. Lines are mean values from six or more replicates and shaded bands are bootstrap 95% confidence intervals.
(D) Growth dynamics reflect the small but reproducible and *HXT1*-dependent growth deficit of *HXT1-T362A* strains in NaCl media. This deficit is not seen in standard media.
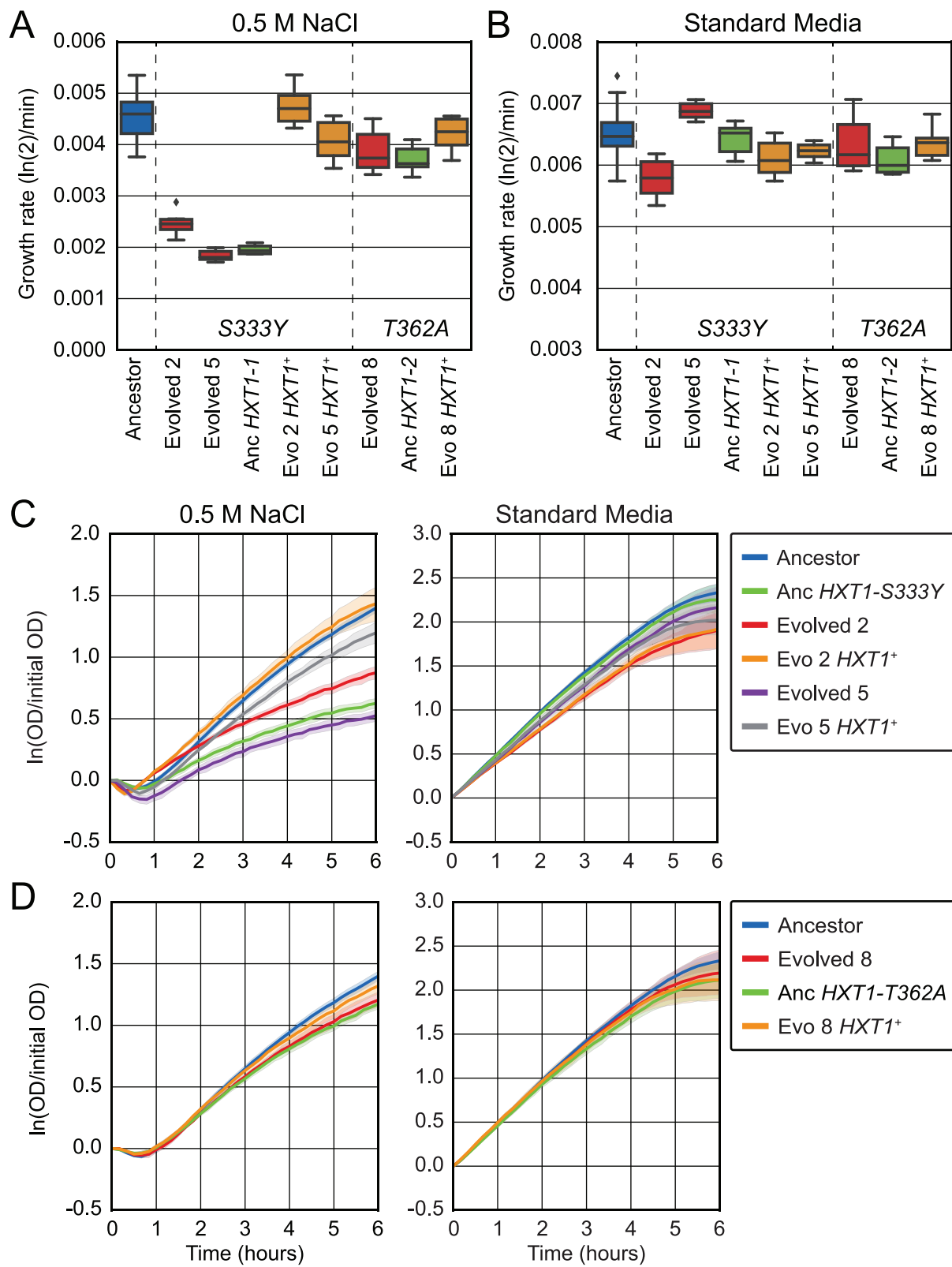
Figure 3.9 (continued) Growth rates of *HXT1* constructed strains.
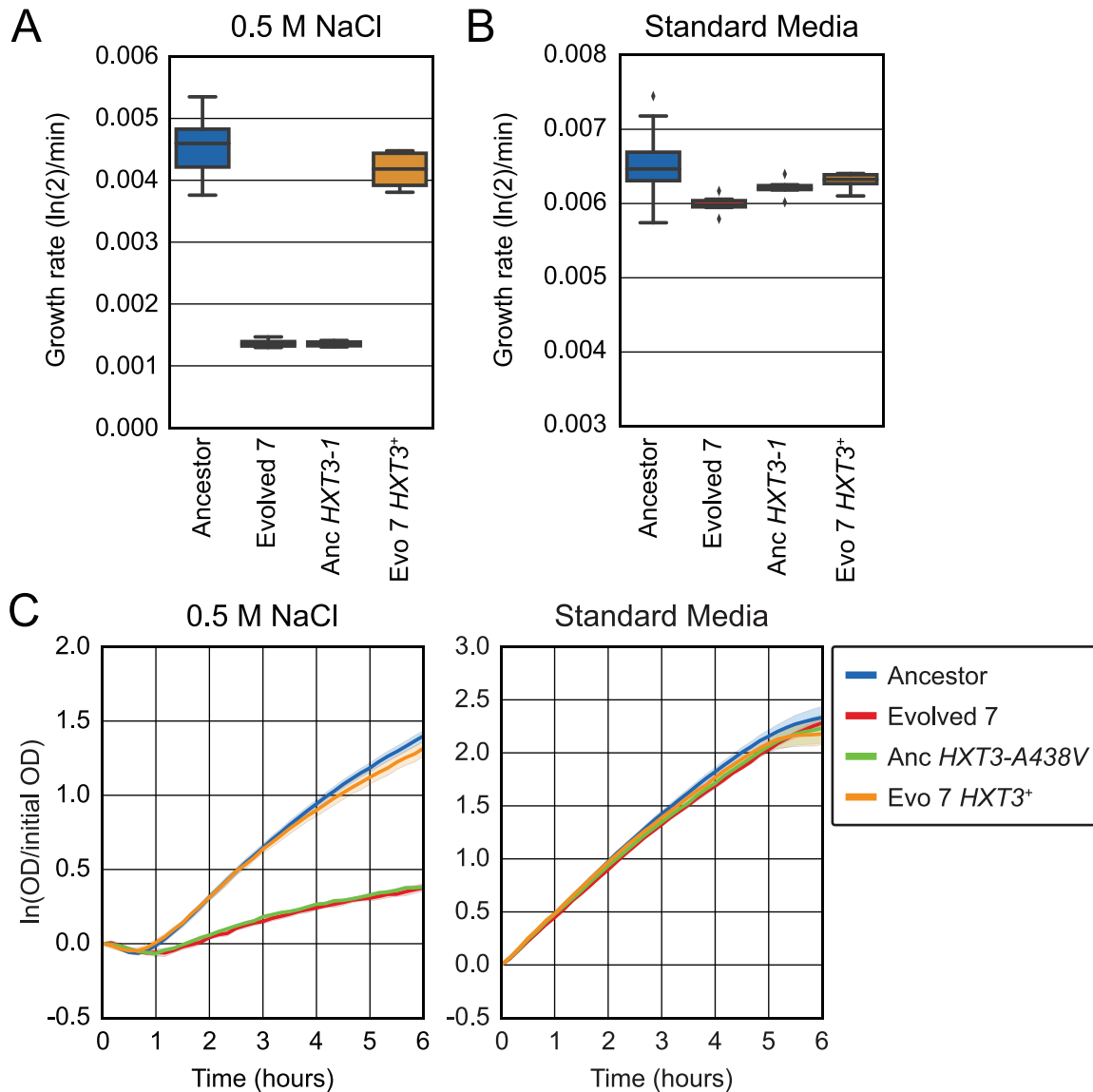
Figure 3.10 Growth rates of *HXT3* constructed strains.

The ancestor containing the *HXT3-A438V* allele perfectly matches both the growth rate and dynamics of the evolved clone in NaCl media. No differences are seen in standard media. Growth rates in (A) and (B) and aggregated growth curves in (C) are as described in Figure 3.8.
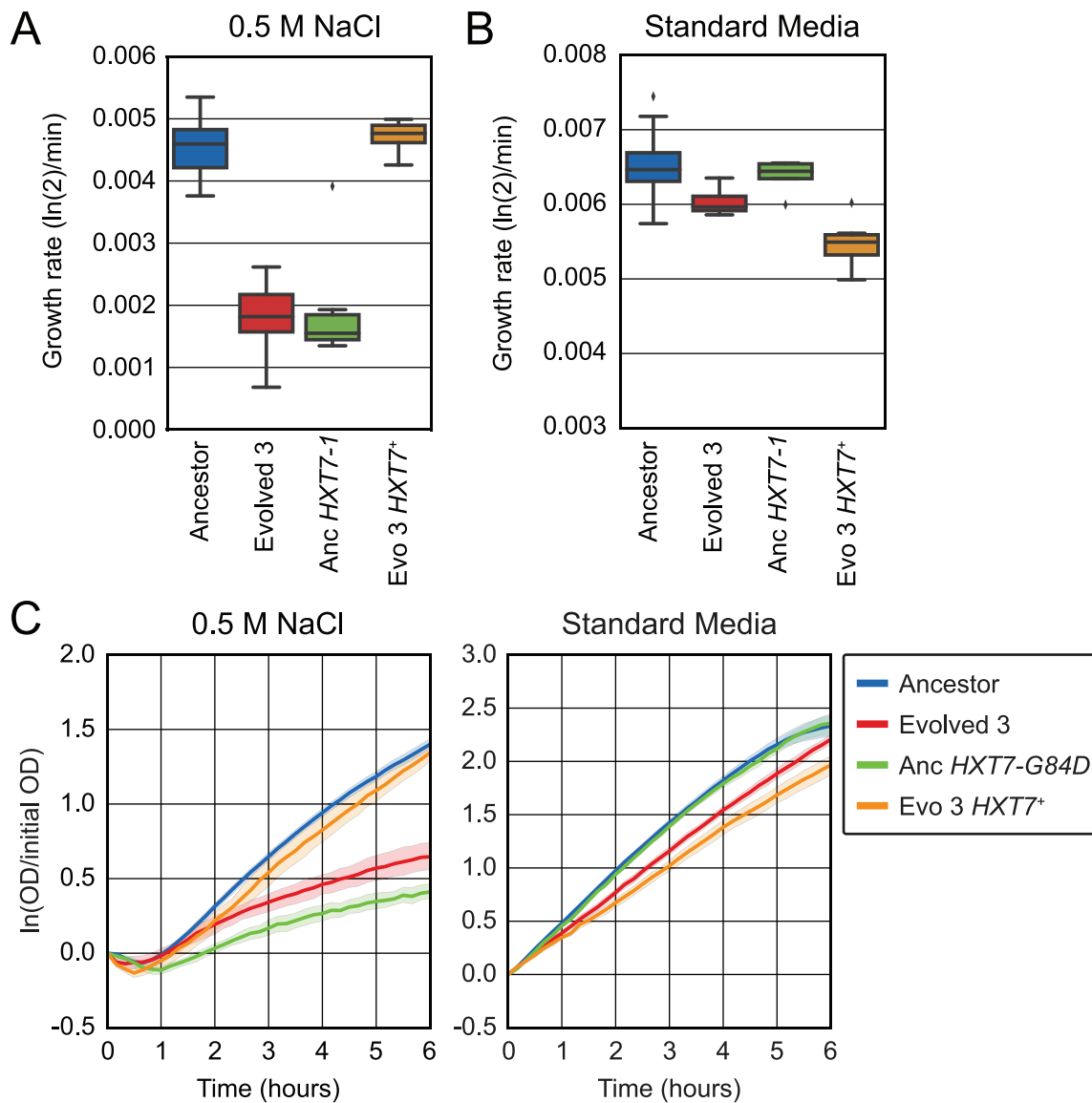
Figure 3.11 Growth rates of *HXT7* constructed strains.

(A) The ancestor containing the *HXT7-G84D* allele matches the decreased growth rate of the evolved clone in NaCl media. Reversion of the mutation eliminates the deficit.
(B) The *HXT7* mutation does not impact growth rate in standard media, but reversion of the mutation confers a growth defect.
(C) The reconstructed strain has a longer lag before increasing OD than the evolved strain. After the first hour, dynamics correlate with *HXT7* genotype. The reverted strain's growth defect in standard media is clearly seen in the growth curves.
Growth rates in (A) and (B) and aggregated growth curves in (C) are as described in Figure 3.8.

Figure 3.12 Growth rates of *HOG1* constructed strains.

(A) The *A166V* but not the *G175D* mutation confers a growth rate as low as its evolved clone of origin in NaCl media; none are as low as a *hog1Δ* strain. For the ancestor reconstruction strains, *hog1-1* indicates the *hog1-A166V* allele and *hog1-2* indicates the *hog1-G175D* allele. The effective growth rate of each strain was calculated as described in Figure 3.8A. Rates were determined from six or more biological replicates.
(B) Loss of *HOG1* function imparts a small growth deficit, but growth defects of the evolved clones in standard media are not greatly improved by reverting the *HOG1* mutations. Growth rates are as in (A) except with standard media.
(C) Growth dynamics in NaCl of strains testing *A166V* segregate by *HOG1* genotype but show small differences at later times. Growth curves are as described in Figure 3.8C. Lines are mean values from six or more replicates and shaded bands are bootstrap 95% confidence intervals.
(D) Growth dynamics in NaCl media show early recovery of Evolved 6-1 independent of its *HOG1* mutation; growth of Evolved 6-1 after the first hour resembles the *hog1Δ* but not the reconstructed *hog1-G175D* strain.
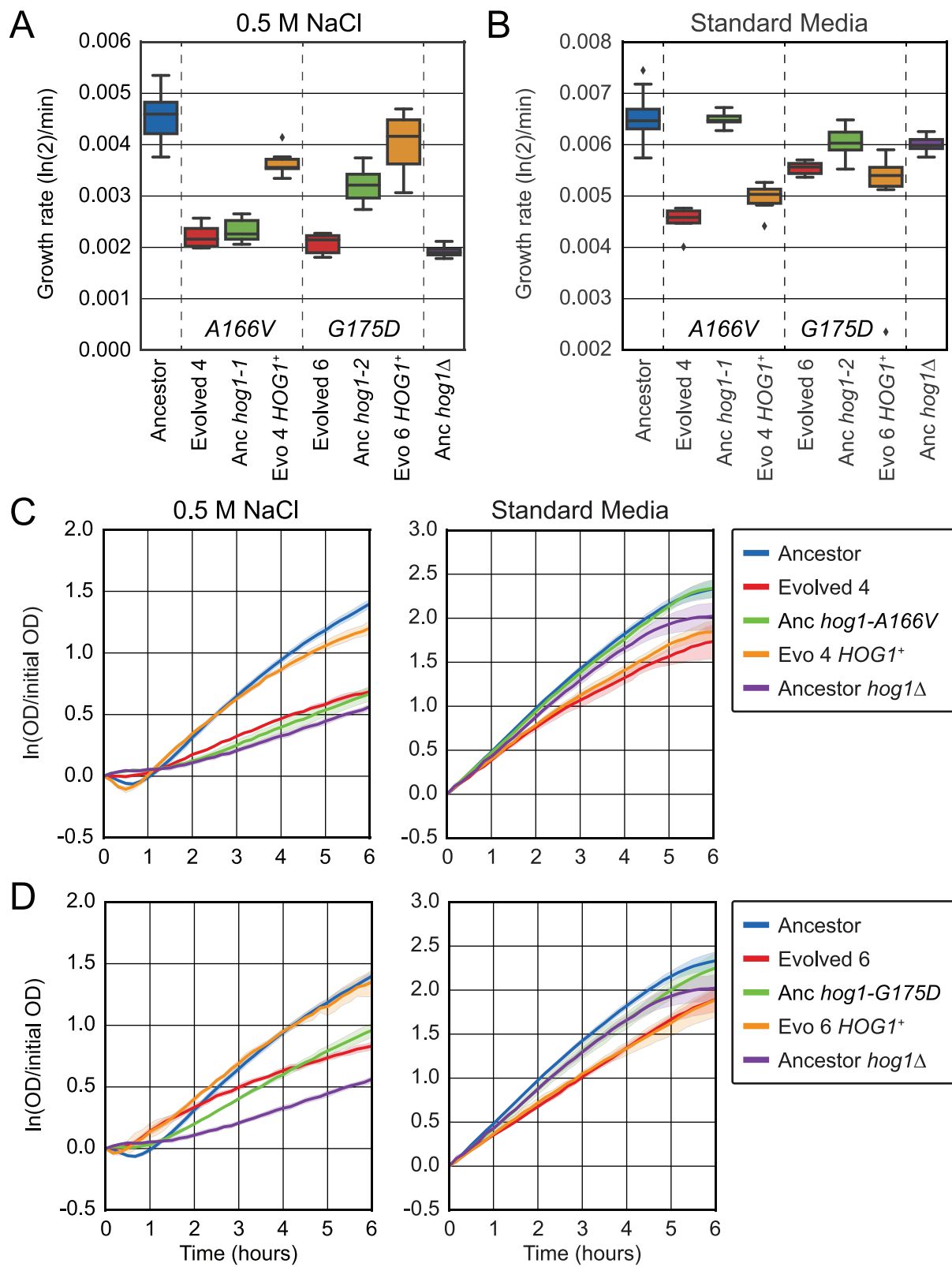
Figure 3.12 (continued) Growth rates of *HOG1* constructed strains.

reconstructed strain has a higher growth rate (Figure 3.12A). However, both Evolved 4-1 and Evolved 6-1 had a high level of survival. Evolved 6-1 has a much lower growth rate than the *hog1-G175D* reconstructed strain, but Evolved 4-1 does not have a big difference from the *hog1-A166V* strain. Oddly, the *G175D* allele confers a larger growth deficit in standard media than *A166V* and has about the same rate as the *hog1Δ* strain (Figure 3.12B). The growth rates of the reverted strains in standard media remain low; they are closer to, but still lower than, the growth rates of their mutator ancestor strains.

Evolved 4-1 and 6-1 were the only evolved strains to not show a decrease in OD after the addition of NaCl. Examining the growth curves of *HOG1* constructed strains demonstrates that the absence of a dip in OD right after NaCl addition is caused by *hog1* loss of function. The early growth dynamics of Evolved 4-1 in NaCl media can be accounted for by its *hog1-A166V* mutation; the difference at later times between the ancestor and the *HOG1*-reverted strain indicates the existence of a second mutation affecting growth in NaCl (Figure 3.12C). By contrast, Evolved 6-1 shows a distinct early trajectory that is not related to the *HOG1* mutation (Figure 3.12D). Both the evolved clone and the reverted strain have a faster resumption of OD increase than the ancestor reconstruction and deletion. Its rate at later times matches that of the *hog1Δ* strain but not the reconstructed strain, suggesting a continued influence of a secondary mutation on growth dynamics.

IMP2'

The reconstructed strains containing evolved alleles of *IMP2'* had higher survival than the evolved clones with those mutations, a trend not seen with any other mutation

tested. The growth dynamics of Evolved 9-1 and 10-1 were also different from all other evolved clones: they showed exponential rather than linear growth in NaCl media. The growth phenotypes of the strains testing the *imp2′-L309R* allele closely match those of the corresponding strain testing the *imp2′-E124** allele (Figure 3.13). The reconstructed strains have substantial growth defects in NaCl media, much lower than the evolved clones and close to the *imp2′Δ* strain (Figure 3.13A). Loss of *IMP2′* function has no effect in standard media; differences in growth rates are likely related to the higher growth rate of non-mutator ancestor B, the direct ancestor of Evolved 9-1 and 10-1 (Figure 3.13B). *IMP2′* loss-of-function strains have exponential growth at a low rate in NaCl media; Evolved 9-1 and 10-1 show an initially similar rate but increase their exponential growth rate over the first six hours in NaCl media (Figure 3.13C, D).

3.7 Discussion

Finding causal mutations

In this work, I sequenced and analyzed whole-genome sequencing data to determine the genetic basis of NaCl-dependent HU survival in evolved strains. Evolved strains had been passaged through either ten or twenty-four rounds of selection. Assuming growth only in standard media, evolving strains grew for about ten to twelve generations per cycle, and 100 to 120 (Evolved 4-11) or 240 to 288 (Evolved 1-3) generations total. Consistent with these estimates and assumed mutation rates, non-mutator evolved strains had only a few mutations per clone while mutator strains typically had over 100 mutations per clone. Therefore, I used bulk segregant analysis to determine causal mutations in mutator

Figure 3.13 Growth rates of *IMP2′* constructed strains.

(A) The reconstructed strains have substantial growth defects in NaCl media, much lower than the evolved clones and close to the *imp2′Δ* strain. The effective growth rate of each strain was calculated as described in Figure 3.8A. Rates were determined from six or more biological replicates.

(B) Growth rates in standard media segregate by background; strains derived from non-mutator ancestor A (Ancestor, reconstructions, and deletion) have lower rates than strains derived from ancestor B (Evolved 9 and 10 and their reversions). Growth rates are as in (A) except with standard media.

(C) Growth dynamics in NaCl media differ based on both the *IMP2′* genotype and the strain background. Evolved 9 shows increasing exponential growth over the first six hours. Growth curves are as described in Figure 3.8C. Lines are mean values from six or more replicates and shaded bands are bootstrap 95% confidence intervals.

(D) Growth dynamics of Evolved 10 and the strains testing *imp2′-E124** closely match the corresponding strain in (C) testing the *imp2′-L309R* allele.
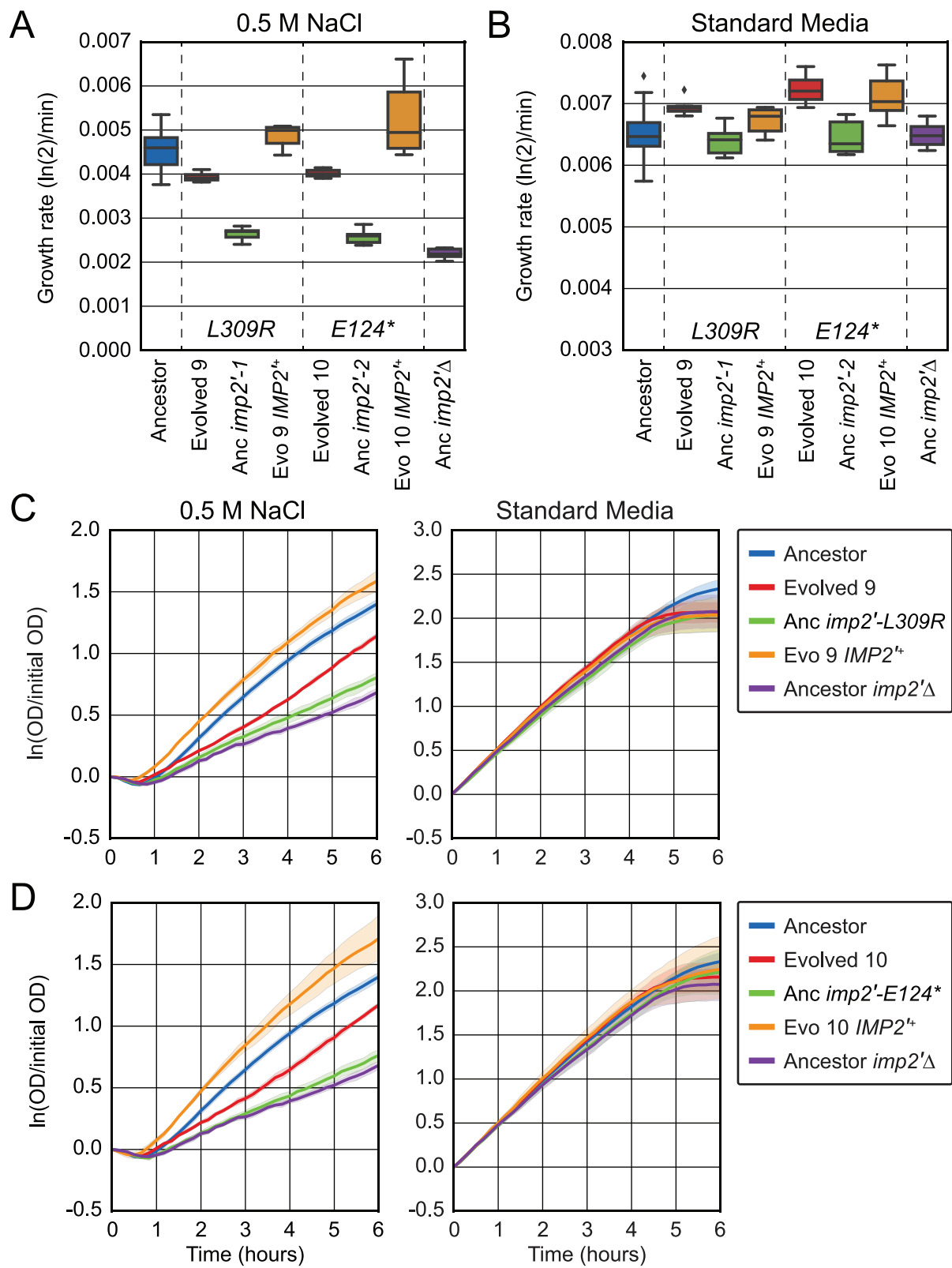
Figure 3.13 (continued) Growth rates of *IMP2′* constructed strains.

evolved strains. I extended the typical analysis to account for initial frequency in the pool of spores and for mutations that improved or diminished growth in the absence of HU selection. This method was very successful at identifying single causal mutations of large effect for the six mutator clones analyzed. Whole-genome sequencing without back-crossing was sufficient to identify causal mutations in the non-mutator clones, and was aided by a high degree of parallelism between independent populations.

Additional evolved clones and the final populations were sequenced after preliminary work showed low final frequencies for several confirmed causal mutations. This work confirmed the genetic basis for the phenotypic heterogeneity described in Chapter 2: most final populations contained multiple independent causal mutations. Often, evolution experiments involve sequencing end populations but not individual clones to identify causal mutations. Experiments that take advantage of the "fossil record" - frozen samples of the population at time points - have identified subpopulations by grouping together mutations that have correlated changes in frequency over time; they also use this data to distinguish low-frequency mutations from sequencing errors [3]. In the absence of time series information, determining whether mutations are in the same genome is impossible without additional manipulations [13]. The approach I took to identify causal mutations, sequencing clones with the desired phenotype, was because of my interest in the genetic basis of the phenotype rather than the dynamics of evolution. This approach turned out to be essential to identifying causal mutations. Sequencing the final populations revealed that only two populations (2 and 10) had causal mutations at over 90%. Notably, the *HOG1* mutations found in populations 4, 5, and 6 were all at very low frequencies ($<10\%$) and might not have been identified by sequencing populations.

Sequencing both populations and multiple clones from each population not only re-

vealed the genetic heterogeneity of final populations but also uncovered a high level of genetic parallelism between populations. From these two characteristics, I infer that there is a relatively small number of single, large-effect mutations, and that these were selected in populations with high initial diversity. The high degree of parallelism is surprising in light of the many, many genes that have been reported to have salt-sensitive phenotypes. In the *Saccharomyces* Genome Database, 449 genes have been annotated with "decreased hyperosmotic stress resistance" and 172 genes with "decreased ionic stress resistance," with an overlap of 67 genes [14, 15]. *HAL5*, *HOG1*, and *IMP2′* all appear in this overlapping set.

Among eight evolved populations, there were ten independent instances of *HXT* mutations, seven of which were confirmed to be causal. The large number of such mutations is reminiscent of the "duplication and divergence" model of gene evolution, except that many duplicate copies already exist in genome. Redundancy and feedback regulation of expression means that one non-glucose-transporting *HXT* gene would likely not impose a fitness cost, and the large number of *HXT* genes increases the target size for mutations that would otherwise seem rare.

Even with the high level of parallelism among evolved populations, there were differences between populations evolved from different ancestors. Surprisingly, the differences in the types of genes mutated were not between mutator and non-mutator strains, but between those derived from different non-mutator ancestors. *HAL5* and *HOG1* mutations account for about 20%[5] of the eight final populations evolved from the non-mutator ancestor A or its derived mutators (Evolved 1-8), and change-of-function mutations in

---

[5]Percents were calculated by adding the percents from each population and dividing by the number of populations.

HXT genes account for nearly half (43%) of these final populations. Conversely, no mutations in any of these genes were found in the non-mutator B populations (Evolved 9-11). Instead, *IMP2′* mutations account for about 60% of Evolved 9 through 11, and *ERG10* mutations conferring NaCl-independent survival account for another 27%. The only gene found mutated in populations from both ancestors is *TUP1*, with the *L12Q* allele found in both Evolved 3 and 11, and *C348R* found in Evolved 8. Analysis of the final populations is of course biased by knowledge of causal mutations in the subset of analyzed clones; roughly 25% of the final populations is composed of clones without a mutation in one of these genes and whose causal mutations remain unknown.

Connecting genotype to phenotype

In the discussion in Chapter 2, clones from the evolved populations were grouped into three classes based on their HU survival phenotype. Considering these classes after sequencing, I can now assign genetic causes to the different phenotypes. The class I clones, which demonstrate NaCl-specific survival, are dominated by mutations in *HXT* genes, *HAL5*, and *IMP2′*. All of the class II clones (osmolyte-dependent survival) have mutations in *HOG1*, which matches with its described role in osmotic stress response independent of the nature of the osmolyte. The two class III clones, which have NaCl-independent survival, both have mutations in the same residue of *ERG10*. *ERG10* is an essential gene encoding an enzyme required for *erg*osterol biosynthesis, and heterozygous null mutants show increased sensitivity to hydroxyurea [16, 17]. These are therefore likely to be hypomorphic alleles that result in cell cycle arrest in the presence of HU in a manner independent of the DNA replication checkpoint, through inhibition of ergosterol

biosynthesis.

Genetic reconstructions demonstrated that most clones contained a single mutation with a large effect on HU survival and that these mutations decrease population growth rate in NaCl media. However, differences were found in either HU survival level, growth dynamics, or both between most clones and their reconstructed strains, providing evidence for secondary mutations. For example, the reversion of the *hog1-G175D* mutation in Evolved 6-1 revealed its contribution to the late, but not early, growth dynamics in NaCl media. Some differences could also be due to background effects: a single non-mutator ancestor, not each direct ancestor, was used to test reconstructions of mutations. This is most likely the cause of differences between Evolved 9-1 and 10-1 and the *IMP2′* reconstructed strains. Both Evolved 7-1 and 8-1 were evolved from non-mutator ancestor A, the exact strain used for reconstruction, which could account for the relative precision in matching growth dynamics between the evolved clones and their reconstructed strains, and between the reverted clones and the ancestor. The remaining difference between the *HXT1*-reverted Evolved 8-1 and its ancestor is additional evidence for a second mutation.

Mechanisms of action

The mutations described above lead to the hypothesis that the evolved clones survive HU in NaCl media due to an inability to properly regulate or restore Na^+ ion balance and are therefore inhibited in passage through the cell cycle. Alkali cation regulation is important for many aspects of cellular physiology, including membrane potential, cell volume, intracellular pH, and signaling [18]. Cells maintain a high intracellular concentration of K^+ ions, necessary for several processes, and a low concentration of Na^+ ions, which

are toxic to cells. Measurements of intracellular ion concentrations in high NaCl indicate that the ratio of K^+ to Na^+ ions, rather than the absolute concentration, is important for cellular function [11]. In *S. cerevisiae*, Trk1 and Trk2 are channels responsible for uptake of K^+ ions, and their activity is upregulated by Hal5 [11]. The membrane channels Ena1 and Nha1 pump Na^+ ions out of the cell, powered by ATP hydrolysis and H^+ antiport respectively [19, 20].

The simplest interpretation of our results is that cell cycle progression is dependent on a proper ion balance, and the evolved strains are defective in creating this balance in the presence of high NaCl. While many genes in *S. cerevisiae* have been implicated in ion homeostasis and salt-sensitive mutants often having overlapping defects with other cellular processes, little is known about the regulatory connection between ion homeostasis and the cell cycle. Some evidence for the dependence of the cell cycle on ion homeostasis comes from the complete G1/S blockage in a *sit4⁻ hal3⁻* mutant background that is rescued by overexpression of genes involved in ion transport and regulation [**?**]. Also, mutants with increased salt tolerance (*ppz1,2⁻*) show accelerated adaptation to α-factor pheromone arrest in G1 in a manner dependent on the Trk1,2 K^+ transporters [21]. This latter evidence argues for a role for ion regulation in a typical cell cycle, not just for the sensitivity of cellular processes to ion imbalance.

Loss of Hal5 function causes salt sensitivity by removing the upregulation of Trk1,2 K^+ ion channels [11, 12]. Loss of Imp2′ function has been found to cause defects in many cellular processes[6], among which is monovalent and divalent cation homeostasis

---

[6]To quote directly from the *IMP2′* SGD webpage: "null mutant exhibits slow growth, impaired endocytosis, large cell size, abnormally round buds, decreased life span, and is sensitive to osmotic stress, metals, heat, radiation, dessication, antimalarial chloroquine, various mutagens, and various antibiotics."

[22]. Imp2′ has been characterized as a transcriptional regulator but its targets remain unknown.

In contrast to the *hal5* and *imp2′* mutations, the *HXT* mutations conferred salt sensitivity through gain/change of function. The *HXT* genes encode hexose transporters with high homology to each other and to several other hexose transporter genes [23]. All of the evolved mutations are single amino acid changes that affect putative transmembrane residues. Three of the confirmed causal mutations (*HXT7-G84D*, *HXT1-S333Y*, *HXT3-A438V*) and one untested mutation (*HXT4-G184D*) are analogous to mutations in *HXT1* and *HXT3* that were found in a selection for mutations that enabled *trk1,2⁻* cells to grow in K$^+$-limited medium [?]. The mutations enabled growth by allowing the ion to enter through the altered hexose transporter. These mutations also caused sodium sensitivity, most likely by being permeable to Na$^+$ as well as K$^+$ ions and thus increasing the internal sodium concentration. Mutations in *HXT* genes created some of the strongest growth defects in NaCl media found in this study. The difference in growth rates between the *HXT1-S333Y* and *HXT1-T362A* mutants is probably due to differences in the ion permeability created by the amino acid change.

Mutations in *TUP1* were found in clones from three evolved populations. Tup1 complexes with Cyc8 to repress transcription of over 180 genes and of large subtelomeric regions; the repression of subsets of these genes is mediated by different sequence-specific repressor partners [24]. This complex is converted from repressive to activating in the Hog1-mediated transcriptional response to osmotic stress [25]. Genes repressed by Tup1-Cyc8 include osmotic stress response genes, such as the Na$^+$ exporter *ENA1*, but also DNA damage response genes, including *RNR2* and *RNR3*. Because the *tup1*Δ strain showed NaCl-independent survival and early assays with Evolved 11-1 produced variable

survival results, its evolved allele *TUP1-L12Q* was not tested by reconstruction. All *TUP1* mutant clones also displayed some flocculation, a phenotype that has been demonstrated to protect cells from various stresses in a non-specific way [26].

As an osmostress-activated MAP kinase, Hog1 has been extensively studied in its capacity to effect broad transcriptional changes and impinge on cell cycle checkpoints [27, 28, 29]. Hog1 has been reported to phosphorylate the cyclin-dependent kinase (CDK) inhibitor Sic1 and repress cyclin transcription in G1 and to inhibit CDK activity in G2 after transfer to 0.4 M NaCl [30, 28, 31]. However, cell cycle delays still occur in *hog1*Δ strains [32]. Hog1 also mediates an immediate, non-transcriptional response to NaCl stress by activating ion channel Nha1 [27]. Sudden NaCl stress causes dissociation of proteins from chromatin, and the absence of the Hog1/Nha1 initial response delays the resumption of transcription for over 10 minutes [27]. However, a delay in adaptation does not account for the sustained growth defect of *hog1*Δ yeast in NaCl media.

While a persistant imbalance in ion levels might explain the salt sensitivity of the *hog1* mutants, it does not account for their survival in sorbitol. This survival could be due to either a similar inhibition by osmotic stress or cross talk with the pheromone response, another MAPK pathway [33]. This cross talk arises from the common function of Ste11, the MAPKKK, in both pathways: Ste11 is activated by Ste20 in response to either stimulus, and phosphorylates either Pbs2 or Ste7, depending on the stimulus [34]. The downstream specificity of these signals despite having a shared component has been extensively examined, and is thought to occur through "insulation" of the signal by scaffolding proteins and cross-inhibition [35, 36]. Osmotic stress applied to yeast lacking Hog1 or Pbs2 function induce transcription of a pheromone-responsive gene [33]. Whether or not they also arrest the cell cycle via the pheromone-specific CDK inhibitor Far1 has,

to my knowledge, not been characterized; studies of the cell cycle in $hog1\Delta$ yeast have focused on identifying the interactions of Hog1 itself with cell cycle machinery. Testing of the $hog1^-$ phenotype in a strain also lacking $FAR1$ is planned.

References

[1] D. R. Smith, A. R. Quinlan, H. E. Peckham, K. Makowsky, W. Tao, B. Woolf, L. Shen, W. F. Donahue, N. Tusneem, M. P. Stromberg, D. A. Stewart, L. Zhang, S. S. Ranade, J. B. Warner, C. C. Lee, B. E. Coleman, Z. Zhang, S. F. McLaughlin, J. A. Malek, J. M. Sorenson, A. P. Blanchard, J. Chapman, D. Hillman, F. Chen, D. S. Rokhsar, K. J. McKernan, T. W. Jeffries, G. T. Marth, and P. M. Richardson, "Rapid whole-genome mutational profiling using next-generation sequencing technologies.," *Genome Research*, vol. 18, pp. 1638–1642, Oct. 2008.

[2] C. L. Araya, C. Payen, M. J. Dunham, and S. Fields, "Whole-genome sequencing of a laboratory-evolved yeast strain.," *BMC Genomics*, vol. 11, p. 88, Feb. 2010.

[3] G. I. Lang, D. P. Rice, M. J. Hickman, E. Sodergren, G. M. Weinstock, D. Botstein, and M. M. Desai, "Pervasive genetic hitchhiking and clonal interference in forty evolving yeast populations.," *Nature*, vol. 500, pp. 571–574, Aug. 2013.

[4] A. Hinnen, J. B. Hicks, and G. R. Fink, "Transformation of yeast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 75, pp. 1929–1933, Apr. 1978.

[5] S. Scherer and R. W. Davis, "Replacement of chromosome segments with altered DNA sequences constructed in vitro.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, pp. 4951–4955, Oct. 1979.

[6] Z. Moqtaderi and J. V. Geisberg, "Construction of mutant alleles in Saccharomyces cerevisiae without cloning: overview and the delitto perfetto method.," *Current Protocols in Molecular Biology*, vol. 104, p. Unit 13.10C, Oct. 2013.

[7] M. J. Brauer, C. M. Christianson, D. A. Pai, and M. J. Dunham, "Mapping novel traits by array-assisted bulk segregant analysis in Saccharomyces cerevisiae.," *Genetics*, vol. 173, pp. 1813–1816, July 2006.

[8] A. V. Segrè, A. W. Murray, and J.-Y. Leu, "High-resolution mutation mapping reveals parallel experimental evolution in yeast," *PLoS Biology*, vol. 4, p. e256, July 2006.

[9] D. J. Kvitek and G. Sherlock, "Whole genome, whole population sequencing reveals that loss of signaling networks is the major adaptive strategy in a constant environment.," *PLoS Genetics*, vol. 9, p. e1003972, Nov. 2013.

[10] J. Zhao, W. Lin, X. Ma, Q. Lu, X. Ma, G. Bian, and L. Jiang, "The protein kinase Hal5p is the high-copy suppressor of lithium-sensitive mutations of genes involved in

the sporulation and meiosis as well as the ergosterol biosynthesis in Saccharomyces cerevisiae," *Genomics*, vol. 95, pp. 290–298, May 2010.

[11] J. M. Mulet, M. P. Leube, S. J. Kron, G. Rios, G. R. Fink, and R. Serrano, "A novel mechanism of ion homeostasis and salt tolerance in yeast: the Hal4 and Hal5 protein kinases modulate the Trk1-Trk2 potassium transporter.," *Molecular and Cellular Biology*, vol. 19, pp. 3328–3337, May 1999.

[12] C. Casado, L. Yenush, C. Melero, M. del Carmen Ruiz, R. Serrano, J. Pérez-Valle, J. Ariño, and J. Ramos, "Regulation of Trk-dependent potassium transport by the calcineurin pathway involves the Hal5 kinase," *FEBS Letters*, vol. 584, pp. 2415–2420, June 2010.

[13] J. N. Burton, I. Liachko, M. J. Dunham, and J. Shendure, "Species-level deconvolution of metagenome assemblies with Hi-C-based contact probability maps.," *Genes Genomes Genetics*, vol. 4, pp. 1339–1346, May 2014.

[14] *hyperosmotic stress resistance: decreased.* Accessed: 2017-01-29.

[15] "ionic stress resistance: decreased." http://www.yeastgenome.org/phenotype/decreased_ionic_stress_resistance/overview. Accessed: 2017-01-29.

[16] L. Hiser, M. E. Basson, and J. Rine, "ERG10 from Saccharomyces cerevisiae encodes acetoacetyl-CoA thiolase.," *The Journal of Biological Chemistry*, vol. 269, pp. 31383–31389, Dec. 1994.

[17] D. Hoepfner, S. B. Helliwell, H. Sadlish, S. Schuierer, I. Filipuzzi, S. Brachat, B. Bhullar, U. Plikat, Y. Abraham, M. Altorfer, T. Aust, L. Baeriswyl, R. Cerino, L. Chang, D. Estoppey, J. Eichenberger, M. Frederiksen, N. Hartmann, A. Hohendahl, B. Knapp, P. Krastel, N. Melin, F. Nigsch, E. J. Oakeley, V. Petitjean, F. Petersen, R. Riedl, E. K. Schmitt, F. Staedtler, C. Studer, J. A. Tallarico, S. Wetzel, M. C. Fishman, J. A. Porter, and N. R. Movva, "High-resolution chemical dissection of a model eukaryote reveals targets, pathways and gene functions.," *Microbiological Research*, vol. 169, pp. 107–120, Feb. 2014.

[18] J. Ariño, J. Ramos, and H. Sychrova, "Alkali metal cation transport and homeostasis in yeasts.," *Microbiology and Molecular Biology Reviews*, vol. 74, pp. 95–120, Mar. 2010.

[19] R. Haro, B. Garciadeblás, and A. Rodríguez-Navarro, "A novel P-type ATPase from yeast involved in sodium transport," *FEBS Letters*, vol. 291, no. 2, pp. 189–191, 1991.

[20] M. A. Bañuelos, H. Sychrová, C. Bleykasten-Grosshans, J.-L. Souciet, and S. Potier, "The Nha1 antiporter of Saccharomyces cerevisiae mediates sodium and potassium efflux.," *Microbiology*, vol. 144 (Pt 10), pp. 2749–2758, Oct. 1998.

[21] L. Yenush, J. M. Mulet, J. Ariño, and R. Serrano, "The Ppz protein phosphatases are key regulators of K+ and pH homeostasis: implications for salt tolerance, cell wall integrity and cell cycle progression.," *The EMBO Journal*, vol. 21, pp. 920–929, Mar. 2002.

[22] J.-Y. Masson and D. Ramotar, "The transcriptional activator Imp2p maintains ion homeostasis in Saccharomyces cerevisiae," *Genetics*, vol. 149, no. 2, pp. 893–901, 1998.

[23] S. Ozcan and M. Johnston, "Function and regulation of yeast hexose transporters.," *Microbiology and Molecular Biology Reviews*, vol. 63, pp. 554–569, Sept. 1999.

[24] T. M. Malavé and S. Y. R. Dent, "Transcriptional repression by Tup1-Ssn6.," *Biochemistry and Cell Biology*, vol. 84, pp. 437–443, Aug. 2006.

[25] M. Proft and K. Struhl, "Hog1 kinase converts the Sko1-Cyc8-Tup1 repressor complex into an activator that recruits SAGA and SWI/SNF in response to osmotic stress.," *Molecular Cell*, vol. 9, pp. 1307–1317, June 2002.

[26] S. Smukalla, M. Caldara, N. Pochet, A. Beauvais, S. Guadagnini, C. Yan, M. D. Vinces, A. Jansen, M. C. Prevost, J.-P. Latgé, G. R. Fink, K. R. Foster, and K. J. Verstrepen, "FLO1 Is a Variable Green Beard Gene that Drives Biofilm-like Cooperation in Budding Yeast," *Cell*, vol. 135, pp. 726–737, Nov. 2008.

[27] M. Proft and K. Struhl, "MAP kinase-mediated stress relief that precedes and regulates the timing of transcriptional induction.," *Cell*, vol. 118, pp. 351–361, Aug. 2004.

[28] X. Escoté, M. Zapater, J. Clotet, and F. Posas, "Hog1 mediates cell-cycle arrest in G1 phase by the dual targeting of Sic1," *Nature Cell Biology*, vol. 6, pp. 997–1002, Sept. 2004.

[29] K. E. Cook and E. K. O'Shea, "Hog1 controls global reallocation of RNA Pol II upon osmotic shock in Saccharomyces cerevisiae.," *Genes Genomes Genetics*, vol. 2, pp. 1129–1136, Sept. 2012.

[30] M. R. Alexander, M. Tyers, M. Perret, B. M. Craig, K. S. Fang, and M. C. Gustin, "Regulation of cell cycle progression by Swe1p and Hog1p following hypertonic stress.," *Molecular Biology of the Cell*, vol. 12, pp. 53–62, Jan. 2001.

[31] M. À. Adrover, Z. Zi, A. Duch, J. Schaber, A. González-Novo, J. Jiménez, M. Nadal-Ribelles, J. Clotet, E. Klipp, and F. Posas, "Time-dependent quantitative multicomponent control of the G-S network by the stress-activated protein kinase Hog1 upon osmostress.," *Science Signaling*, vol. 4, p. ra63, Sept. 2011.

[32] G. Bellí, E. Garí, M. Aldea, and E. Herrero, "Osmotic stress causes a G1 cell cycle delay and downregulation of Cln3/Cdc28 activity in Saccharomyces cerevisiae.," *Molecular Microbiology*, vol. 39, pp. 1022–1035, Feb. 2001.

[33] S. M. O'Rourke and I. Herskowitz, "The Hog1 MAPK prevents cross talk between the HOG and pheromone response MAPK pathways in *Saccharomyces cerevisiae*," *Genes & Development*, vol. 12, pp. 2874–2886, Sept. 1998.

[34] H. Saito, "Regulation of cross-talk in yeast MAPK signaling pathways," *Current Opinion in Microbiology*, vol. 13, pp. 677–683, Dec. 2010.

[35] A. Zarrinpar, R. P. Bhattacharyya, M. P. Nittler, and W. A. Lim, "Sho1 and Pbs2 act as coscaffolds linking components in the yeast high osmolarity MAP kinase pathway," *Molecular Cell*, 2004.

[36] J. G. Zalatan, S. M. Coyle, S. Rajan, S. S. Sidhu, and W. A. Lim, "Conformational control of the Ste5 scaffold protein insulates against MAP kinase misactivation.," *Science*, vol. 337, pp. 1218–1222, Sept. 2012.

Chapter 4

Materials and Methods

4.1 Yeast culturing and strain construction

Media and materials

Yeast were grown and handled using standard techniques [1, 2]. Yeast were cultured in the complete medium YPD (1% yeast extract, 2% peptone and 2% dextrose). Liquid cultures were grown at 30°C on a roller drum. Glycerol stocks were made by combining a culture 1:1 with 30% (w/v) glycerol in ultrapure water and stored at -80°C. Strains were sporulated by culturing overnight in YEPA (1% yeast extract, 2% peptone and 2% potassium acetate), then washing twice and resuspending in ultrapure water containing 2% potassium acetate. Sporulating cultures were incubated at 25°C on a roller drum for 2-7 days. Spores were treated with zymolyase and separated by micromanipulator for tetrad analysis.

Stock solutions of sodium chloride, potassium chloride and sorbitol were made by

dissolving the solute in YPD and filtering with 0.22 μm pore vacuum filters (Corning). Hydroxyurea was dissolved in ultrapure water at 2 M, filtered with a 0.22 μm pore syringe filter (Pall), aliquoted and stored at -20°C. To use, aliquots were thawed at 25 to 65°C, then vortexed to ensure full dissolution. All media components and chemicals were purchased from Sigma-Aldrich (`www.sigmaaldrich.com`) unless otherwise indicated. Zymolyase was purchased from Zymo Research.

Strain construction

Strain genotypes are listed in Tables 4.1, 4.4, and 4.3 in section 4.8. All strains used in this study were derived from the W303 background (*ura3-1 ade2-1 his3-11,15 leu2-3,112 trp1-1 can1-100*) and are *MAT***a** haploid unless otherwise indicated. Genetic modifications were made using standard techniques [3]. Gene deletions were constructed by PCR-mediated one-step gene disruption (Figure 4.1A). DNA products used in transformations, consisting of a selectable gene and 100 to 300 bp of flanking homology, were constructed using fusion PCR. Allele replacements were constructed by transformation with a linear DNA fragment containing a region of the gene to be replaced, including the mutation and continuing through the terminator, fused with the *Kluyveromyces lactis URA3* cassette including its native promoter and terminator regions (Figure 4.1B). The *K. lactis URA3* locus was amplified from the plasmid pBS1369 [4]. Gene deletion and allele replacement transformants were screened by colony PCR and verified by PCR from genomic DNA. Allele replacements were also verified by Sanger sequencing of the mutation region.

The *mec1Δ sml1Δ* strain (yNMC006) was constructed by first deleting *sml1Δ* by
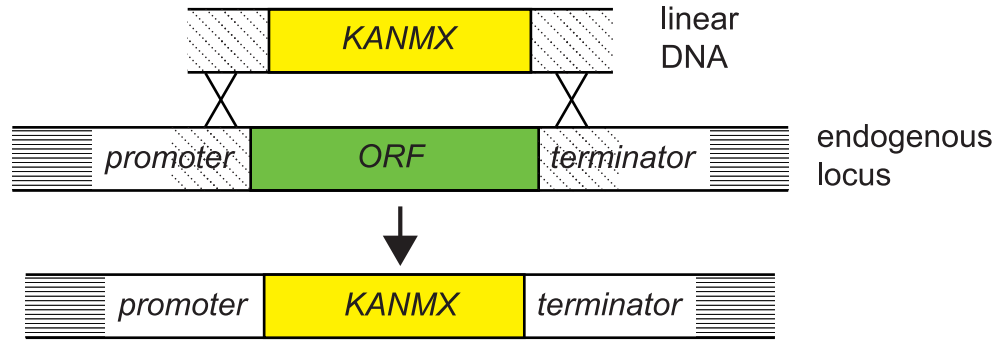
A Gene Deletion

B Allelic Replacement
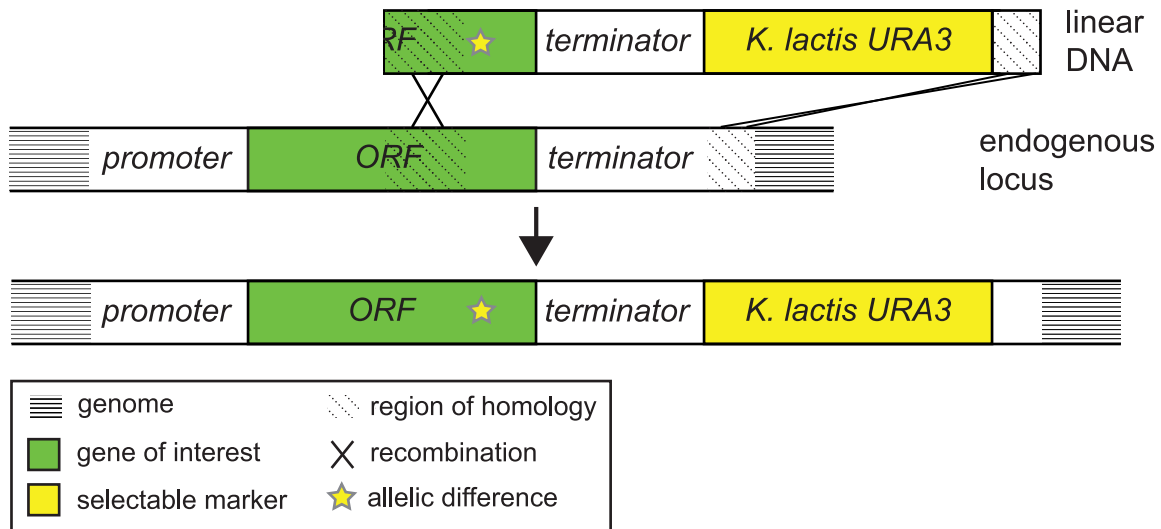
Figure 4.1 Diagrams of strain construction.

(A) Genes were deleted by one-step replacement with a selectable gene cassette (typically *KANMX*) using a linear construct containing the cassette flanked by 100-300 bases of homology. The cassette contains its own promoter and terminator.

(B) Allele replacements were done by one-step replacement of a region of the gene of interest, mediated by a downstream selectable gene cassette containing its own promoter and terminator. This technique was chosen for expediency over the more commonly used two-step methods, as there was no need to reuse the selectable marker.

replacement with *KANMX6*, followed by deleting *mec1Δ* by replacement with *HIS3MX6*. The *POL3-L523D* mutator strain (yNMC028) was constructed using the *URA3* loopout strategy with plasmid PY24D. The plasmid was linearized before transformation with restriction enzyme Eag1 (New England Biolabs). Both the initial plasmid integration and the nucleotide substitutions after loopout were confirmed by Sanger sequencing.

The strain yNMC047, the *MATα* version of yNMC006, was constructed by mating-type switching. Strain yNMC006 was transformed with plasmid pJH1912 carrying P_{GAL1}-*HO*, cultured in galactose media to induce *HO* expression and plated onto YPD agar [5]. The haploid MATα strain was isolated from the resulting colonies by sporulation and tetrad dissection. Strain yNMC063 was created to enable selection of haploid spores in bulk segregant analysis (see section 4.5). This strain was derived from yNMC047 by integrating *NATMX6-P_{STE2}-URA3* at the endogenous *URA3* locus.

Diploid strains were constructed by mixing together strains of opposite mating types on an agar plate, incubating at 30°C, and selecting zygotes by micromanipulation. Diploids were confirmed by complementation mating test and by sporulation.

4.2 Experimental evolution

Two clones of yNMC006 (*mec1Δ sml1Δ*) and three clones of yNMC028 (*mec1Δ sml1Δ POL3-L523D*, all derived from yNMC006A) were used as ancestors for one or more evolved populations. See Table ?? for the specific ancestor for each evolved population. Clones A and B of strain yNMC006 are independent transformants of the initial *sml1Δ* gene deletion.

Each population was inoculated from a single colony of the ancestral strain into a small 3-ml culture, then diluted for overnight growth. From the 8-ml overnight culture, 1 ml was removed, of which 100 μl was used for cell counts and the remainder frozen as a glycerol stock. Then, 1 ml of 4 M NaCl in YPD was added to 7 ml of culture for a final concentration of 0.5 M. Two hours after adding NaCl, 205 μl of 2 M hydroxyurea (HU) was added for a final concentration of 50 mM. Four hours after adding HU, the entire culture was transferred to a plastic 14-ml conical tube and centrifuged at 1800 RPM for 3 minutes. The supernatant was decanted and the pellet resuspended in 800 μl of YPD. A varying amount (2-400 μl) from the ten-fold concentrated cells was diluted into 8 ml of YPD for overnight growth. A glycerol stock was made of 500 μl of the remaining concentrated cells. The goal of overnight dilution was to achieve a mid-log phase culture (i.e., not exceed 10^7 cells/ml) 18 hours later from an unknown number of viable cells. Thus, the dilution level was varied between cultures and between rounds of selection. The level of overnight dilution for each culture was determined based on the growth of the culture from the previous round's dilution. In early rounds, the dilution level was from 1/2 to 1/20, and in late rounds, the dilution level was from 1/100 to 1/400.

The estimated survival of a culture was determined after each round by comparing the number of cells after the overnight growth phase with the expected number based on the number of cells before NaCl/HU treatment. The equation used is

$$\text{survival} = \frac{\text{actual cell count}}{\text{expected cell count}} = \frac{C_E}{C_S \times \text{dilutions} \times 2^{t/\tau}}$$

where C_S and C_E are the starting and ending cell counts, respectively, t is the length of time the culture was in the growth phase, and τ is the doubling time. Dilutions are the

dilution factors used after the NaCl/HU treatment (e.g., 1/400) and occasionally again in the morning to keep the cultures in exponential growth. For the estimated values in Figure 2.8, $\tau =$120 minutes for the mutator populations and 100 minutes for the non-mutator populations. The result was multiplied by 100 for percent survival.

In the first evolution (February-March 2013), three cultures were passaged through a total of 24 NaCl/HU exposures with overnight growth periods in between each exposure to HU. These became strains yNMC033, 034 and 035. In the second evolution (August 2014), the cultures were passaged through five rounds of selection, placed at 4°C for about 24 hours, then passaged through another five rounds for a total of ten rounds. These cultures became strains yNMC110 to 117. Concurrently with the second evolution, three cultures of the yNMC028 ancestors (one each of clones A, B, and C) were passaged through growth in YPD without exposure to NaCl or HU. Cultures were diluted once per day in the evening, 1/400 to 1/4000, and mostly reached late-log phase or saturation. These became strains yNMC107, 108, and 109.

4.3   Hydroxyurea survival and analysis

The general scheme for a hydroxyurea survival assay is depicted in Figure 4.2. For each experiment, strains were inoculated in duplicate into a starter culture 24 hours in advance of the experiment, and diluted for overnight growth. The growth phase of the culture was important to achieve accurate and consistent results; a culture close to stationary phase would have higher survival due to having fewer growing cells in the population. Details specific to either the colony-counting version or the high-throughput

dilution spot version are below.

Colony-counting survival assay

Cultures were grown in YPD liquid media at 30°C on a roller drum. Each strain to be tested was inoculated in duplicate into 2 to 5 ml YPD one day before the experiment, then back-diluted (1/100 to 1/2000 depending on strain) into 5 to 10 ml for overnight growth. For the experiment, cultures were diluted to 2×10^6 cells/ml in 4 ml of Start and end samples were plated in triplicate onto YPD agar and colonies were counted two days later. An ImageJ macro was written to aid in counting colonies on agar plates. See Appendix A for a more detailed description.

The first executions of the colony-counting assay, in which a single strain was tested in a range of conditions, did not include plating a start sample; instead, survival was calculated based on the number of cells in the starting culture as determined by Coulter counter. In most other colony-counting assays, a knock-out strain containing a drug selection gene was combined with a control strain in the same tube, treated together, then distinguished in the start and end samples by replica-plating. However, a large difference in survival made it difficult to accurately assay survival for both strains in a tube. Later assays were done with strains cultured individually.

High-throughput survival assay

Cultures were grown in a 2-ml deep 96-well plate ("deep-well" plate, VWR Cat. 89237-526) covered with a breathable rayon sealing film (VWR Cat. 60941-086). Each
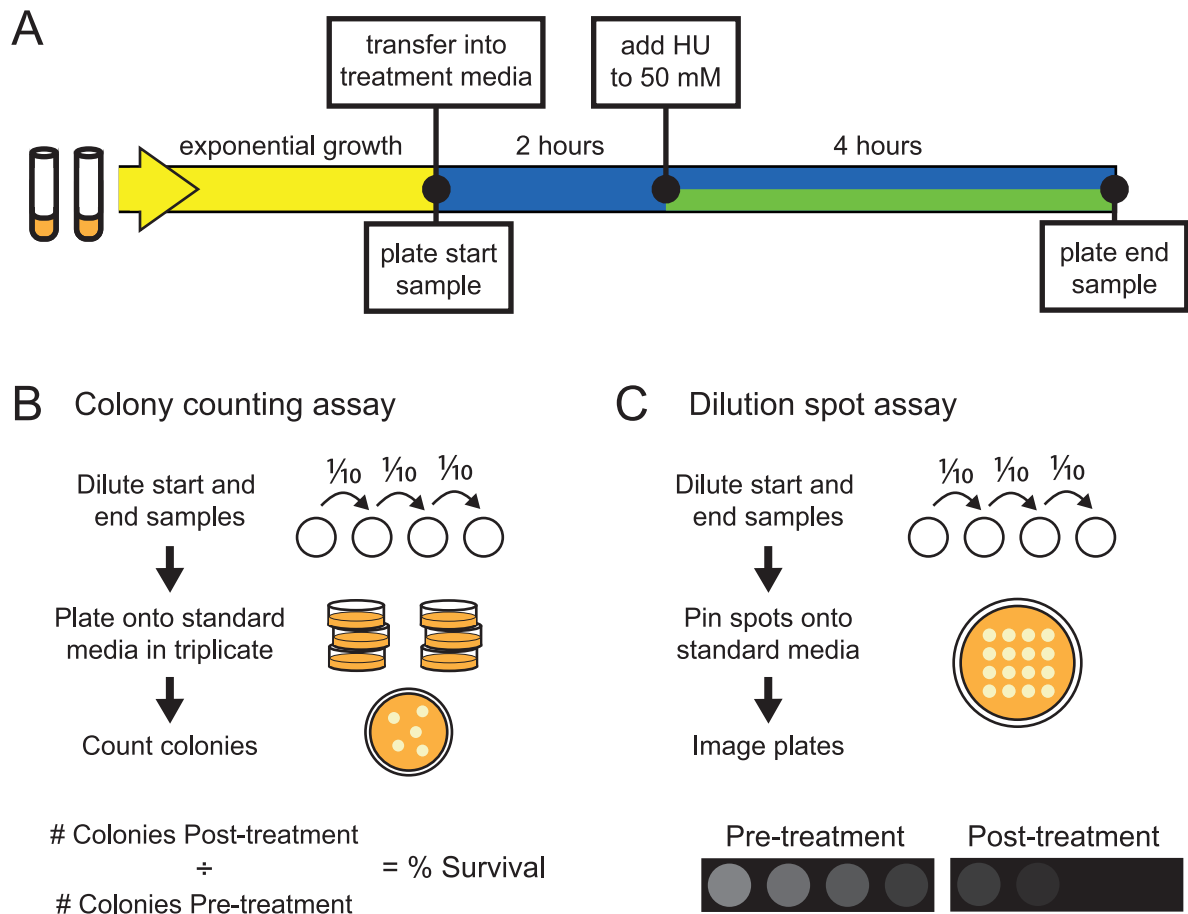
Figure 4.2 Diagram of hydroxyurea survival assay.

(A) General scheme of survival assay. Each strain to be assayed is inoculated in duplicate, grown to exponential phase, then passaged through the NaCl/HU selection analogous to that used in the evolution experiment. A no-stimulus control for each strain is always included.

(B) To quantify percent survival, start and end samples are plated in triplicate onto YPD agar and colonies are counted two days later. In some assays, two strains differing by a single gene deletion were co-cultured; the YPD plates were counted for the total number of cells, then replica-plated to a drug selection plate to count the number of deletion cells.

(C) To determine the order of magnitude of survival, start and end samples are serially diluted and spotted onto YPD agar. Spots from start and end are compared as described in section 4.3.3.

strain to be tested was inoculated in duplicate into 400 μl YPD one day before the experiment, then back-diluted (1/100 to 1/2000 depending on strain) into 400 μl for overnight growth. For the experiment, each culture was diluted into 300 μl YPD plus 300 μl of 2x test media in a deep-well plate. Then, 200 μl of each sample was transferred to a 96-well plate for the "start" plating, leaving 400 μl in each well. The plate was incubated at 30ºC on a roller drum for two hours, then 10 μl of 2 M HU (final concentration 50 mM) was added to each well and the plate was incubated for another four hours. 200 μl of each sample was transferred to a 96-well plate for the "end" plating. Each 200-μl sample culture was serially diluted 1/10 three times by transferring 20 μl into wells containing 180 μl YPD. Cultures were spotted onto YPD agar plates using a replica plater (Sigma-Aldrich). Each plate contained 48 spots: 12 samples with 4 serial dilution spots per sample (undiluted, 1/10, 1/100, and 1/1000). Plates were grown at 30ºC for two days before imaging. Each agar plate was imaged with a gel imager (AlphaImager HP, AlphaInnotech) using reflective white light with no filter, at the following settings: aperture 8.00, zoom 25, exposure 75 ms.

Hydroxyurea survival analysis

A suite of ImageJ macros and Python programs was written to automatically analyze the plate images acquired in the HU survival assays. See Appendix A for more detailed description of scripts. First, images were adjusted, cropped and split into series of spots for individual samples using batch_plate_cropper.py. Then the start and end images for each sample are combined into one file using batch_montage.py. Montage images for a single experiment were placed into one PDF document using combine_experiment.py.

The images were checked by eye for contamination and low starting growth, and these samples were excluded from analysis. If the negative control strain (yNMC006) showed high survival, the entire experiment was excluded from analysis.

Each experimental sample was matched to the strain and media used and placed into a folder for that strain using strain_sorter.py. Each image of serial dilution spots was converted to a binary image and the pixel area of each spot measured and recorded using batch_survival_strain.py. Next, the end/start pixel area ratio for each dilution level was calculated and the highest dilution level determined for which this ratio was at least 0.5 using HU_strain_analysis.py. The dilution levels were assigned values of 0, 1, 2, 3, and 4 for calculating a mean. A survival category (-/+/++/+++) for the strain and condition was assigned based on the mean maximum dilution level from at least six replicates (three separate experiments with two replicates each).

4.4 Growth curve assays and analysis

Growth curve acquisition

Experimental growth curves were conducted using the BioTek PowerWave 340 with shaking at 30ºC. The optical density at a wavelength of 600 nm (OD600) was recorded every 10 minutes. The first time point was recorded 10 minutes after the start of the experiment to account for settling during plate setup. Each strain to be tested was inoculated in duplicate into 400 μl YPD in a deep-well one day before the experiment, then back-diluted (1/100 to 1/2000 depending on strain) into 400 μl for overnight growth. For the experiment, strains were diluted to OD600 0.1 in 100 μl YPD in a flat-bottom 96-

well plate (Falcon Cat. 353072) and the plate was covered with a permeable, transparent film (Breathe-Easy, Diversified Biotech). A growth curve was recorded for 2 to 3 hours to assure exponential growth ("pre-incubation"). Then, the plate was removed, 20 μl of YPD was added to each well, and 50 μl was transferred from each well to a well containing 50 μl of either YPD or YPD with 1 M NaCl. A film was applied and growth was recorded for 18 hours.

Data calibration

The raw OD600 values were calibrated by subtracting a value from each well to account for the optical density of the plate, culture media and cover film. Two factors affected the accuracy of this "blank" value (Figure 4.3). It was determined that the machine in use had consistent between-row variation in measurement, and this was accounted for by using separate blank values for each row. The permeable film used to cover the wells had an inconsistent contribution to the optical density, and this was accounted for by using an average value from several plates with films. Because the optical density contributed by the plate and culture media were highly consistent, the same blank values were used for every experiment.

The blank-subtracted OD600 values were also calibrated to account for the nonlinearity of optical density reads above 0.6 (Figure 4.4). The equation was determined by comparing measured OD600 with the expected OD600 of a dilution series. I created separate dilution series of a log phase culture and a saturated culture[1], then measured OD600

---

[1]The saturated culture had lower OD600 values than the log phase culture for the same number of cells, probably due to smaller cell size.

(without film) of all samples in a 96-well plate. I plotted the relative dilution versus OD600 (minus blank OD600) for the saturation phase cultures below OD 0.6 to get the linear regression: y = 1.28x + 0.0162 (R^2 = 0.985) (Figure 4.4A). I used this equation to determine the expected OD for values greater than 0.6. I did the equivalent analysis for the log phase data, using the fit line for low OD log phase to determine expected OD for high OD log phase (Figure 4.4B). I combined the expected versus measured OD data for both phases above 0.6 (Figure 4.4C). I used the exponential fit equation y = $0.2141\,e^{1.7935x}$ to adjust OD values above 0.6 during data analysis.

Growth analysis and plots

The Python program growth_curve_analysis.py was written in collaboration with Bryan Weinstein, with input from Yu-Ying Phoebe Hsieh and Marco Fumasoni, to analyse parameters of growth curves. For a description of the script and how to use it, see Apppendix B. The effective growth rate for a defined time interval (i.e., 0 to 6 hours) was calculated by fitting a linear regression to all of the data points for the time interval. If the culture reached saturation during the time interval, only the data points prior to saturation were used for the linear regression. Saturation points were determined by growth_curve_analysis.py using an approximation of the second derivative.

Aggregated growth curves were created by plotting the mean values of six or more normalized growth curves for each strain. Data were normalized by calibrating the raw OD600 values as described above and dividing by the initial value. Natural log transformations of the data were applied after normalization. Plots were created using the Seaborn (version 0.7.1) data visualization package for Python (seaborn.pydata.org). See

Figure 4.3 Blank calculation for growth curve assay.

(A) Measurement of only media in a 96-well plate on three different experiment dates showed row-to-row variation but date-to-date consistency in media absorbance values.

(B) Measurement of media in plates sealed with a permeable, transparent film showed that the film creates noise in the measurement in addition to the row-to-row variation. (C) Row averages were calculated from the five plate measurements in (B); the gray lines indicate the standard deviation of the 60 measurements for each row (5 measurements of 12 wells per row). These row averages were used as the "blank" values in the analysis of all growth curve data.
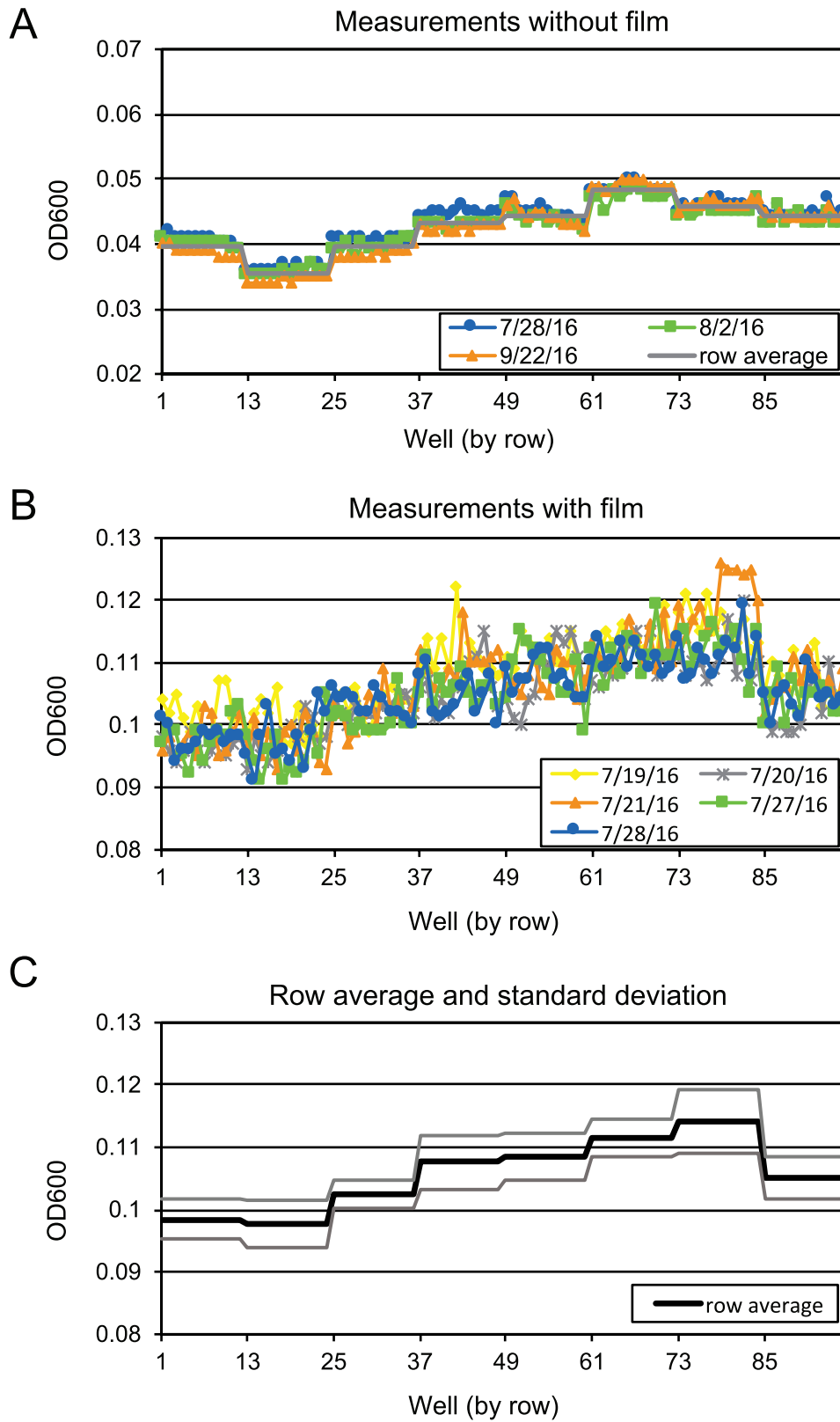
Figure 4.3 (continued) Blank calculation for growth curve assay.
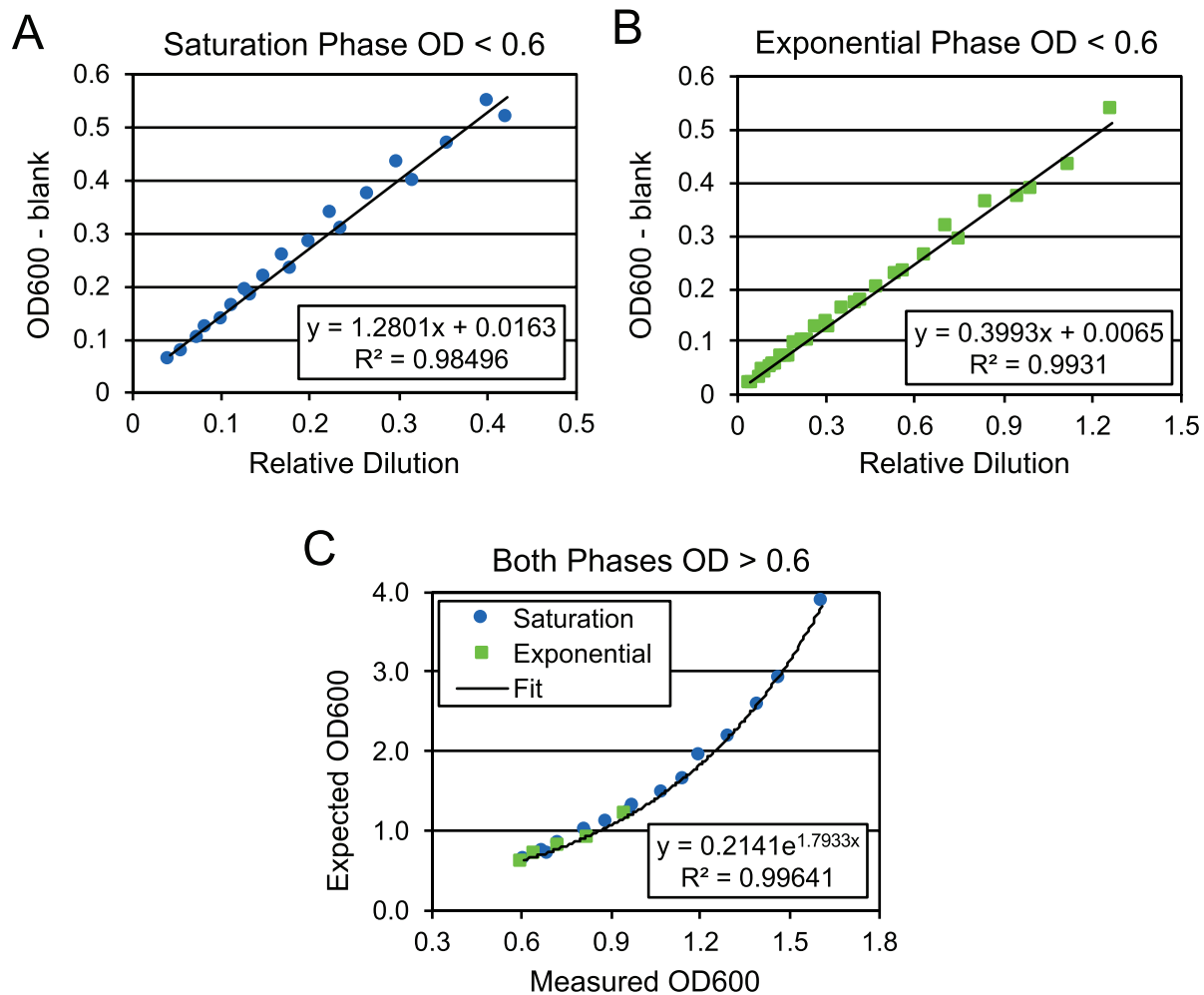
Figure 4.4 Calibration for high OD values.

(A) and (B) Separate dilution series of an exponential phase culture and a saturated culture were measured without film. The relative dilution vs. OD - blank OD for the saturation phase (A) and exponential phase (B) cultures below OD 0.6 were plotted and a linear regression equation determined for each.

(C) These equations were used to calculate the expected OD for values above OD 0.6. The resulting values were plotted against the measured OD and used to determine an exponential fit equation, which was used to adjust experimental OD values above 0.6 in the analysis of all growth curve data.

Appendix B for a description of the Python code used to generate boxplots and aggregate growth curve plots.

4.5 Bulk segregant analysis

One or more clones from each evolved population was mated to yNMC063 ($MAT\alpha$ $mec1\Delta$ $sml1\Delta$ $NATMX6$-P_{STE2}-$URA3$) and sporulated for several days as described in section 4.1 until spores were observed. The sporulation culture was pelleted, resuspended in 50 μl of zymolyase (diluted 1/10 in ultrapure water), and incubated at 30°C for one hour. Then, 400 μl water and 50 μl 10% Triton X-100 was added. The mixture was sonicated for five seconds then pelleted at 6000 RPM for one minute. The cells were resuspended in 100 μl water and plated onto synthetic complete medium minus uracil to select for haploid $MAT\boldsymbol{a}$ cells. Once colonies were visible, the plate was rinsed with media to pool cells. The pooled cells were split into three parts. The "initial" pool was used directly to prepare genomic DNA (see section 4.7.2). The "YPD" pool was cultured in YPD with dilution once per day for 5 days, then processed for genomic DNA. The "NaCl/HU" pool was passaged through 5 rounds of the full selection cycle (see Figure 2.6), then processed for genomic DNA.

4.6 Whole-genome sequencing and analysis

Libraries were prepared for sequencing using the Illumina Nextera kit and a protocol modified for smaller volumes [6]. Library concentrations were quantified using a

Qubit Fluorimeter and Agilent TapeStation 2200. Sequencing was performed by the Harvard Bauer Core Facility on the Illumina HiSeq 2500 using a Rapid Run flow cell. All sequencing reactions were 150-bp paired-end reads.

Sequence data were analyzed by a modified version of the procedure described in [7]. The FASTQ files were first processed for adapter trimming using cutadapt 1.8.1 (cutadapt.readthedocs.io/en/stable/index.html) [8]. Reads were aligned to the S288c reference genome R64 (downloaded from the Saccharomyces Genome Database, accessible at www.yeastgenome.org) by the Burrows-Wheeler Alignment tool BWA-MEM (version 0.7.9a) (bio-bwa.sourceforge.net) [9, 10]. The resulting SAM (Sequence Alignment/Map) file was converted to a BAM (binary SAM) file, sorted and indexed with samtools 1.3.1 (www.htslib.org) [10]. Insertions and deletions (indels) were realigned using the IndelRealigner tool of the Genome Analysis Toolkit 3.3.0 (software.broadinstitute.org/gatk) [11]. Overlapping bases, which result from short inserts of paired-end reads, were removed from analysis using the clipOverlap function of bamUtil 1.0.13 (genome.sph.umich.edu/wiki/BamUtil:_clipOverlap) before creating the pileup file with samtools mpileup. All variant detection (SNP, indel and copy number) was performed using VarScan 2.4.1 (dkoboldt.github.io/varscan/) [12]. Sequencing coverage was determined using the function CollectAlignmentSummaryMetrics from Picard Tools 1.119 (broadinstitute.github.io/picard). Coverage reported is the number of aligned bases, not including overlapping bases, divided by the genome size (12,157,105 bp; nuclear and mitochondrial genome lengths of S288c reference genome, www.yeastgenome.org/genomesnapshot).

SNPs and indels detected by VarScan were further analyzed using the mutantanalysis.py program written by a previous Murray lab member, John Koschwanez (available at github.com/koschwanez/mutantanalysis) [7]. This program was modified slightly for

use with ClustalW 2.0.12. The exact version used for the analysis presented here was run with Python 2.7.6, pysam 0.6 and Biopython 1.63 and is available at github.com/nwespe/ sequence_analysis. Copy number variation determined by VarScan was visualized with a script written by Daniel Rice, with modifications by Phoebe Hsieh, Marco Fumasoni, and me, and expanded for multiple samples by me. These scripts (plot_copynumber.py and batch_plot_copynumber.py) are available at github.com/nwespe/sequence_analysis.

I wrote a suite of Python and Bash scripts to automate and parallelize execution of the sequence analysis pipeline on the Harvard Research Computing Odyssey cluster. See Appendix C for a full description of scripts, which are available at github.com/nwespe/ sequence_analysis.

4.7   Molecular biology protocols

Protocols used frequently in this work are presented here in full. The yeast transformation and genomic DNA preparation protocols were obtained from Murray lab members and have minor modifications. The colony PCR protocol was personally developed after intermittant success with other methods; it has proven successful in the hands of several rotation students and beginning bench scientists.

Yeast transformation

Protocol modified from Gottschling Lab.

1. Grow yeast to exponential phase in 5-10 ml YPD.

2. Spin down culture 3 min at 1800 rpm and discard supernatant.

3. Resuspend pellet in 1 ml PBS to wash, then spin down again.

4. Discard supernatant and resuspend pellet in 100 μl 0.1 M LiAc/1x TE. Transfer to Eppendorf tube.

5. Add 240 μl 50% PEG (MW 3350), 36 μl 1 M LiAc, 10 μl ssDNA, X μl insert DNA, and 74-X μl H_2O.

6. Vortex briefly to combine.

7. Place tube in 42°C water bath for 30 to 90 min.

8. Spin down tube 3 min at 1800 rpm.

9. Decant supernatant.

10. For auxotrophic selection, resuspend gently in 300 μl PBS and plate directly onto selection plate. For drug selection, resuspend gently in 300 μl YPD and plate onto YPD plate. Replica-plate 24 hours later onto selection plate.

Genomic DNA preparation

Protocol modified slightly from Gregg Wildenberg.

1. Grow yeast to saturation in 1-5 ml YPD. I generally use 1.5 ml; for larger volumes, I double the amount of the reagents in steps 2-6.

2. Spin down culture 3 min at 1800 rpm and discard supernatant.

3. Add 50 μl 0.5 M EDTA (pH 7.5 – 8.0), 200 μl dH2O, and 2.5 μl zymolyase. Mix by inversion and incubate 30-60 min at 37°C.

4. Add 50 μl of "miniprep mix" (0.2 M EDTA pH 8.0, 0.4 M Tris-Cl pH 8.0, 2% SDS). Mix by inversion and incubate 15-30 min at 65°C.

5. Add 63 μl 5 M KAc. Mix by inversion and incubate 15-30 min on ice.

6. Spin down 10 min maximum speed. Transfer supernatant to a new tube.

7. Add 720 μl ice-cold 100% ethanol. Mix by inversion and spin down 5-10 min.

8. Discard supernatant. Add 130 μl dH$_2$O or elution buffer and 1 μl RNase A to pellet. Incubate at 37°C for 60 min.

9. Add 2 μl Proteinase K (20 mg/ml) and incubate at 37°C for 2 hours (omit this step unless sample is for sequencing).

10. Add 130 μl 100% isopropanol. Mix by inversion and spin down 5-10 min. Discard supernatant.

11. Add 500 μl 70% ethanol. Mix by inversion and spin down 5-10 min. Discard supernatant.

12. Dry pellet 5 min at 65°C. Resuspend in 100-500 μl of 10 mM Tris-Cl pH 8.0. Incubate at 50-65°C to speed up dissolving.

Colony PCR

Nichole's protocol, adapted from freeze-thaw and boil-in-base methods. This protocol regularly yielded PCR products of up to 4.5 kb with enough product for Sanger sequencing and no gel purification required (used spin column or magnetic bead purification).

1. Aliquot 20 μl of 20 mM NaOH into a PCR tube for each colony.

2. Using a sterile toothpick, pick a colony and smear onto a selective plate. Swish the same toothpick around in a PCR tube containing NaOH to transfer cells. Stabbing up and down works well. The liquid should appear slightly cloudy. With a fresh toothpick, streak out from each smear for future colonies of the clone.

3. Vortex tubes well for 30 seconds (seriously, count to 30).

4. Heat at 100°C for 20 minutes using a PCR block.

5. Vortex tubes again for 30 seconds.

6. Place in a -80°C freezer for 10+ minutes.

7. Microwave for 2 minutes.

8. Vortex again. *Optional*: do another freeze-thaw cycle (steps 6-8).

9. Use 1 μl of the NaOH-colony solution in a 25 μl reaction (sufficient for checking for a band); use 2 μl in a 50 μl reaction if purifying product for further use.

10. Store at -20°C if you want to do PCR from the same prep again. It's important to not let the colony prep sit at room temperature. If not using immediately after step 8, keep at -20°C and thaw right before using.

Pro tip: Each toothpick has a thin end and a thick end. I use the thin end for picking the colony and swishing in the NaOH because the thick end wicks out too much of the liquid. I use the thick end (of a different toothpick!) for streaking out because the thin

end tears into the agar more easily.

PCR mix for 50 μl reaction:

    27.5 μl H$_2$O

      10 μl reaction buffer

     3.5 μl DMSO

       5 μl primers (2.5 μl each of 10 nM stock)

       2 μl colony prep

       1 μl dNTPs

       1 μl polymerase

Troubleshooting: If your PCR yields no product, try the following:

1. Test the reaction with a control DNA sample. If your PCR doesn't work on good genomic DNA, it's not going to work on a colony prep. Go redesign your primers.

2. Repeat the reaction in parallel with control DNA. Maybe you forgot something.

3. Use more or less of the colony prep in the PCR. There is a Goldilocks amount of template, particularly with colony PCR because other cellular matter can interfere with the reaction. The streaking of the colony before adding it to the NaOH tends to achieve this level more robustly but it's not perfect.

4. Redo the colony prep, from a fresh colony if possible.

5. If you did a 25-μl reaction, try a 50-μl reaction.

6. Did you use DMSO? Use DMSO.

7. Alter your PCR program parameters (annealing temperature, extension time).

8. Do you really need this product? If so, looks like you're just going to have to do a genomic DNA prep (see section 4.7.2).

4.8 Strains

All strains used in this study were derived from the W303 background (*ura3-1 ade2-1 his3-11,15 leu2-3,112 trp1-1 can1-100*) and are *MAT***a** haploid unless otherwise indicated.

Table 4.1 Strains used in Chapter 2.

Description in Text	Systematic Name	Relevant Genotype
mec1-1 sml1-1	yBS10[1]	*mec1-1 sml1-1*
mec1Δ sml1Δ	yNMC006A	*mec1Δ::HIS3 sml1Δ::KANMX*
cdc26Δ	yNMC025	yNMC006 *cdc26Δ::HPHMX*
ste2Δ	yNMC022	yNMC006 *ste2Δ::HPHMX*
hog1Δ	yNMC014	yNMC006 *hog1Δ::HPHMX*
cln3Δ	yNMC019	yNMC006 *cln3Δ::HPHMX*
gpd1Δ	yNMC016	yNMC006 *gpd1Δ::HPHMX*

Ancestors

Description in Text	Systematic Name	Relevant Genotype
Non-mutator ancestor A	yNMC006A	*mec1Δ::HIS3 sml1Δ::KANMX*
Non-mutator ancestor B	yNMC006B	*mec1Δ::HIS3 sml1Δ::KANMX*
Mutator ancestor A	yNMC028A	yNMC006A *pol3-L523D*
Mutator ancestor B	yNMC028B	yNMC006A *pol3-L523D*
Mutator ancestor C	yNMC028C	yNMC006A *pol3-L523D*

[1]This strain contains alleles referenced in [13].

<div align="center">**Table 4.2** Evolved clones.</div>

Mutator evolved clones

Description in Text	Systematic Name	Direct Ancestor
Evolved 1-1	yNMC033B	yNMC028A
Evolved 2-1	yNMC034A	yNMC028A
Evolved 2-2	evo 2-01	yNMC028A
Evolved 2-3	evo 2-02	yNMC028A
Evolved 2-4	evo 2-10	yNMC028A
Evolved 3-1	yNMC035C	yNMC028A
Evolved 3-2	evo 3-01	yNMC028A
Evolved 3-3	evo 3-04	yNMC028A
Evolved 3-4	evo 3-06	yNMC028A
Evolved 4-1	yNMC110-c4	yNMC028A
Evolved 4-2	evo 4-01	yNMC028A
Evolved 4-3	evo 4-02	yNMC028A
Evolved 4-4	evo 4-04	yNMC028A
Evolved 5-1	yNMC111-c3	yNMC028B
Evolved 5-2	yNMC111F	yNMC028B
Evolved 5-3	evo 5-05	yNMC028B
Evolved 5-4	evo 5-08	yNMC028B
Evolved 5-5	evo 5-09	yNMC028B
Evolved 6-1	yNMC112E	yNMC028C
Evolved 6-2	evo 6-02	yNMC028C
Evolved 6-3	evo 6-07	yNMC028C
Evolved 6-4	evo 6-08	yNMC028C
Mut. anc. A evolved in YPD	yNMC107	yNMC028A
Mut. anc. B evolved in YPD	yNMC108	yNMC028B
Mut. anc. C evolved in YPD	yNMC109	yNMC028C

Table 4.2 (continued) Evolved clones.

Non-mutator evolved clones

Description in Text	Systematic Name	Direct Ancestor
Evolved 7-1	yNMC113A	yNMC006A
Evolved 7-2	evo 7-02	yNMC006A
Evolved 7-3	evo 7-03	yNMC006A
Evolved 7-4	evo 7-05	yNMC006A
Evolved 8-1	yNMC114A	yNMC006A
Evolved 8-2	evo 9-01	yNMC006A
Evolved 8-3	evo 9-09	yNMC006A
Evolved 9-1	yNMC115A	yNMC006B
Evolved 9-2	evo 10-04	yNMC006B
Evolved 9-3	evo 10-08	yNMC006B
Evolved 10-1	yNMC116A	yNMC006B
Evolved 11-1	yNMC117D	yNMC006B
Evolved 11-2	yNMC117I	yNMC006B
Evolved 11-3	evo 12-04	yNMC006B
Evolved 11-4	evo 12-06	yNMC006B

Table 4.3 Strains used in Chapter 3.

Description in Text	Systematic Name	Relevant Genotype
n/a	yNMC047[1]	*MATα mec1Δ::HIS3 sml1Δ::KANMX*
MATα mating strain	yNMC063	yNMC047 *NATMX-pSTE2-URA3@URA3*
n/a	yNMC051[2]	*mec1Δ::HIS3 sml1Δ::NATMX*
Ancestor *trk1Δ*	yNMC130	yNMC051 *trk1Δ::KANMX*

Reconstructed strains

Description in Text	Systematic Name	Relevant Genotype
Ancestor *hal5-G515W*	yNMC213	yNMC006A *hal5-G515W; Kl-URA3*
Ancestor *hog1-A166V*	yNMC247	yNMC006A *hog1-A166V; Kl-URA3*
Ancestor *hog1-G175D*	yNMC212	yNMC006A *hog1-G175D; Kl-URA3*
Ancestor *imp2'-L309R*	yNMC191	yNMC006A *imp2'-L309R; Kl-URA3*
Ancestor *imp2'-E124**	yNMC219	yNMC006A *imp2'-E124*; Kl-URA3*
Ancestor *HXT1-S333Y*	yNMC169	yNMC006A *HXT1-S333Y; Kl-URA3*
Ancestor *HXT1-T362A*	yNMC170	yNMC006A *HXT1-T362A; Kl-URA3*
Ancestor *HXT3-A438V*	yNMC228	yNMC006A *HXT3-A438V; Kl-URA3*
Ancestor *HXT7-G84D*	yNMC258	yNMC006A *HXT7-G84D; Kl-URA3*

Ancestor deletions

Description in Text	Systematic Name	Relevant Genotype
Ancestor *hal5Δ*	yNMC096	yNMC006A *hal5Δ::URA3*
Ancestor *hog1Δ*	yNMC014	yNMC006A *hog1Δ::HPHMX*
Ancestor *imp2'Δ*	yNMC208	yNMC006A *imp2'Δ::NATMX*
Ancestor *hxt1Δ*	yNMC157	yNMC051 *hxt1Δ::KANMX*
Ancestor *hxt3Δ*	yNMC246	yNMC006A *hxt3Δ::NATMX*
Ancestor *hxt7Δ*	yNMC252	yNMC006A *hxt7Δ::NATMX*

[1]This strain was made by converting the mating type of yNMC006A.
[2]This strain was made from yNMC006A by swapping the *sml1Δ* selective marker. It was used for the *trk1Δ* and *hxt1Δ* strains.

Table 4.3 (continued) Strains used in Chapter 3.

Reverted strains

Description in Text	Systematic Name	Relevant Genotype
Evolved *1 HAL5$^+$*	yNMC188	yNMC033B *HAL5$^+$; Kl-URA3*
Evolved 2 *HXT1$^+$*	yNMC215	yNMC034A *HXT1$^+$; Kl-URA3*
Evolved 3 *HXT7$^+$*	yNMC256	yNMC035C *HXT7$^+$; Kl-URA3*
Evolved 4 *HOG1$^+$*	yNMC248	yNMC110-c4 *HOG1$^+$; Kl-URA3*
Evolved 5 *HXT1$^+$*	yNMC255	yNMC111-c3 *HXT1$^+$; Kl-URA3*
Evolved 6 *HOG1$^+$*	yNMC217	yNMC112E *HOG1$^+$; Kl-URA3*
Evolved 7 *HXT3$^+$*	yNMC229	yNMC113A *HXT3$^+$; Kl-URA3*
Evolved 8 *HXT1$^+$*	yNMC216	yNMC114A *HXT1$^+$; Kl-URA3*
Evolved 9 *IMP2$'^+$*	yNMC209	yNMC115A *IMP2$'^+$; Kl-URA3*
Evolved 10 *IMP2$'^+$*	yNMC210	yNMC116A *IMP2$'^+$; Kl-URA3*

Evolved deletions

Description in Text	Systematic Name	Relevant Genotype
Evolved 1 *hal5Δ*	yNMC097	yNMC033B *hal5Δ::URA3*
Evolved 2 *hxt1Δ*	yNMC194	yNMC034A *hxt1Δ::NATMX*
Evolved 3 *hxt7Δ*	yNMC253	yNMC035C *hxt7Δ::NATMX*
Evolved 4 *hog1Δ*	yNMC261	yNMC110-c4 *hog1Δ::HPHMX*
Evolved 5 *hxt1Δ*	yNMC254	yNMC111-c3 *hxt1Δ::NATMX*
Evolved 6 *hog1Δ*	yNMC262	yNMC112E *hog1Δ::HPHMX*
Evolved 7 *hxt3Δ*	yNMC245	yNMC113A *hxt3Δ::NATMX*
Evolved 8 *hxt1Δ*	yNMC193	yNMC114A *hxt1Δ::NATMX*
Evolved 9 *imp2$'$Δ*	yNMC205	yNMC115A *imp2$'$Δ::NATMX*
Evolved 10 *imp2$'$Δ*	yNMC206	yNMC116A *imp2$'$Δ::NATMX*

Table 4.4 Diploid strains.

Diploids of evolved clones

Heterozygous diploid of:	Systematic Name	yNMC063 mated with:
Ancestor A	yNMC266	yNMC006A
Mutator ancestor A	yNMC144	yNMC028A
Mutator ancestor B	yNMC145	yNMC028B
Mutator ancestor C	yNMC146	yNMC028C
Evolved 1-1	yNMC066	yNMC033B
Evolved 1-1	yNMC119	yNMC033B
Evolved 2-1	yNMC067	yNMC034A
Evolved 2-1a,b	yNMC152(A,B)	yNMC034A
Evolved 3-1	yNMC068	yNMC035C
Evolved 3-1	yNMC121	yNMC035C
Evolved 4-1	yNMC122	yNMC110 clone
Evolved 5-1	yNMC151	yNMC111G
Evolved 5-2	yNMC150	yNMC111F
Evolved 6-1	yNMC123	yNMC112 clone
Evolved 7-1	yNMC267	yNMC113A
Evolved 8-1	yNMC268	yNMC114A
Evolved 9-1	yNMC220	yNMC115A
Evolved 10-1	yNMC221	yNMC116A
Evolved 11-1	yNMC222	yNMC117D
Evolved 11-2	yNMC223	yNMC117I
Mut. anc. A evolved in YPD	yNMC147	yNMC107
Mut. anc. B evolved in YPD	yNMC148	yNMC108
Mut. anc. C evolved in YPD	yNMC149	yNMC109

All diploid strains above were constructed by mating with yNMC063: *MATα mec1Δ::HIS3 sml1Δ::KANMX NATMX-pSTE2-URA3@URA3*

References

[1] D. A. Treco and F. Winston, "Growth and manipulation of yeast.," *Current Protocols in Molecular Biology*, vol. Chapter 13, p. Unit 13.2, Apr. 2008.

[2] D. A. Treco and V. Lundblad, "Preparation of yeast media.," *Current Protocols in Molecular Biology*, vol. Chapter 13, p. Unit13.1, May 2001.

[3] V. Lundblad, G. Hartzog, and Z. Moqtaderi, "Manipulation of cloned yeast DNA.," *Current Protocols in Molecular Biology*, vol. Chapter 13, p. Unit13.10, May 2001.

[4] O. Puig, B. Rutz, B. G. Luukkonen, S. Kandels-Lewis, E. Bragado-Nilsson, and B. Séraphin, "New constructs and strategies for efficient PCR-based gene manipulations in yeast.," *Yeast*, vol. 14, pp. 1139–1146, Sept. 1998.

[5] I. Herskowitz and R. E. Jensen, "Putting the HO gene to work: practical uses for mating-type switching.," *Methods in Enzymology*, vol. 194, pp. 132–146, 1991.

[6] M. Baym, S. Kryazhimskiy, T. D. Lieberman, H. Chung, M. M. Desai, and R. Kishony, "Inexpensive Multiplexed Library Preparation for Megabase-Sized Genomes," *PLoS ONE*, vol. 10, pp. e0128036–15, May 2015.

[7] J. H. Koschwanez, K. R. Foster, and A. W. Murray, "Improved use of a public good selects for the evolution of undifferentiated multicellularity.," *eLife*, vol. 2, p. e00367, Apr. 2013.

[8] M. Martin, "Cutadapt removes adapter sequences from high-throughput sequencing reads," *EMBnet.journal*, vol. 17, p. 10, Aug. 2011.

[9] S. R. Engel, F. S. Dietrich, D. G. Fisk, G. Binkley, R. Balakrishnan, M. C. Costanzo, S. S. Dwight, B. C. Hitz, K. Karra, R. S. Nash, S. Weng, E. D. Wong, P. Lloyd, M. S. Skrzypek, S. R. Miyasato, M. Simison, and J. M. Cherry, "The reference genome sequence of Saccharomyces cerevisiae: then and now.," *Genes Genomes Genetics*, vol. 4, pp. 389–398, Mar. 2014.

[10] H. Li, B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin, and 1000 Genome Project Data Processing Subgroup, "The Sequence Alignment/Map format and SAMtools," *Bioinformatics*, vol. 25, pp. 2078–2079, Aug. 2009.

[11] A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly, and M. A. DePristo, "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data.," *Genome Research*, vol. 20, pp. 1297–1303, Sept. 2010.

[12] D. C. Koboldt, Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding, and R. K. Wilson, "VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing.," *Genome Research*, vol. 22, pp. 568–576, Mar. 2012.

[13] X. Zhao, E. G. D. Muller, and R. Rothstein, "A suppressor of two essential checkpoint genes identifies a novel protein that negatively affects dNTP pools.," *Molecular Cell*, vol. 2, pp. 329–340, Sept. 1998.

Chapter 5

Conclusions and Future Directions

5.1 Summary of major results

I designed a selection cycle to evolve NaCl-dependent cell cycle arrest, using hydroxyurea lethality as a selective pressure. Every evolved population yielded clones with NaCl-dependent survival of HU; of these clones, the majority have drastic reductions in population growth in NaCl media, consistent with some or all of the cells arresting the cell cycle. Growth defects in standard media were not observed in evolved clones relative to their direct ancestors with one exception, Evolved 4-1. Phenotyping of the evolved clones was extended beyond fitness in the selection conditions and included examining the specificity of HU survival and growth dynamics in both NaCl and standard media. The population growth measurements of evolved clones revealed several features of growth dynamics not captured by a single growth rate parameter, such as differences between early (0-2 hours) and late (2-6 hours) times after the addition of NaCl[1].

---

[1]Population growth by OD was recorded for 18 hours, but dynamics after 6 hours were not discussed.

Causal mutations were identified by whole-genome sequencing and confirmed by strain construction. Evolved clones contained single mutations with large effects on HU survival, and reconstruction provided evidence for secondary mutations that influence HU survival and/or dynamics of growth in NaCl. The final evolved populations contained subpopulations with different causal mutations and a high degree of parallelism between independent populations. The genes repeatedly mutated were *HAL5* (3), *HOG1* (3), *IMP2'* (2), *TUP1* (3), *ERG10* (2), and several *HXT* genes (10 in total). In section 3.7, I discuss hypotheses about the molecular basis of HU survival based on the identified mutations, evidence from prior literature, and the "extended" phenotypes (survival specificity, growth dynamics) of the evolved clones.

One prediction before our experiment was that cross talk in MAPK signaling pathways resulting from mutations in *HOG1* and *PBS1* could increase survival in hydroxyurea. Preliminary work using a *hog1Δ* strain demonstrated that loss of *HOG1* increases survival in hydroxyurea in a high-osmolarity environment, but a growth defect in standard media inhibited its enrichment in a cycling population. Mutations in *HOG1* were identified in evolved clones but were not dominant in the final populations, which aligns with expectations from preliminary work.

5.2 Further experiments with evolved strains

Molecular characterization of cell cycle arrest

As discussion in section 2.6, the measurement of population growth by OD can provide only a rough indication of the behavior of individual cells. Examination of single

cells can be done by time-lapse microscopy, which was performed for three evolved clones, and by flow cytometry to analyze certain cellular features. DNA content can be used as an indicator of cell cycle stage and measured by flow cytometry for large numbers of individual cells in a population. Early experiments showed that Evolved 1-1 maintained arrest in G1 when released from α-factor into NaCl media. Recent attempts to replicate and extend these experiments have been stymied by laboratory equipment issues, but these experiments are planned for a small number of strains.

As discussion in section 3.7, the sustained low growth rate of $hog1^-$ cells could be related to activation of Far1, the pheromone-responsive CDK inhibitor. Strains to test whether deletion of $FAR1$ alters the HU survival and growth phenotypes of $hog1^-$ mutant strains are under construction. The absence of a dip in initial OD upon transfer to NaCl media, a feature unique to the $hog1^-$ mutant strains, indicates a basal difference in the physiology of $hog1^-$ strains growing in standard media that impacts their response to salt stress. To my knowledge, this has not been described before, and deserves further investigation.

Identification of secondary mutations

Almost all of the evolved clones characterized had single mutations of large effect. However, I described evidence from both survival and growth phenotypes that some evolved clones may have secondary mutations contributing to their phenotype. I have listed the mutations I think are most likely to be contributors in Table 3.4, but they remain to be characterized as such. Not every sequenced evolved clone was characterized by growth curves; other evolved clones containing apparent loss-of-function mutations in

HAL5 could be assayed and compared to Evolved 1-1, which could help clarify whether the difference between Evolved 1-1 and the reconstructed strain is due to its initial genetic background or a secondary mutation. The *RPD3* mutation in Evolved 2-1 and the *RPH1* mutation in Evolved 3-1 were not found to affect HU survival, but could contribute to the differences in growth dynamics not explained by their causal mutations in *HXT* genes.

Assuming that replication continues to be lethal in the presence of HU, the selective pressure in environment 2 (with HU) has a cap of 100% survival. The selective pressure in the cycle would then shift to environment 1 (no HU) and additional gains in fitness would come from increased growth rate in standard media and/or delayed response to the signal, depending on the time parameters. In sequence analysis of the mutator evolved clones, the bulk segregant pool was also passaged in standard media; the results were not analyzed for this particular purpose but could provide evidence of adaptation to environment 1. However, adaptive mutations could be specifically mitigating the fitness defect of the *pol3-L523D* mutator allele; a non-mutator strain was used for the back-crossing, so these mutations would not necessarily be enriched in the pool.

5.3   Variations on the selection cycle

By using 50 mM hydroxyurea, the evolved cultures were subjected to very strong selective pressure to survive. Combined with the moderate stress from 0.5 M NaCl, the cultures repeatedly gave rise to clones with single, large-effect mutations. A lower level of salt stress could result in a diminished effect from the same mutations, leading to evolved clones with multiple beneficial mutations. A lower concentration of hydroxyurea

or shorter exposure time would decrease the selective pressure, thereby allowing mutations that would otherwise have a small effect to potentially be more competitive against those of large effect. The relative effect size of a mutation could be quantified by assaying survival under conditions more challenging than the evolution experiment.

One obvious extension of this experimental setup is the use of other environmental changes to signal the change in selective pressure. A summer undergraduate student, Frances Diehl, undertook one such selection using a shift in temperature to 37°C, with all other parameters being the same as the NaCl experiment. Cultures of both mutator and non-mutator *mec1Δ sml1Δ* strains achieved moderate increases in survival over ten rounds of selection. Clones with survival phenotypes were sequenced by whole-genome sequencing but not further analyzed; unfortunately, most samples had very low coverage. A more challenging evolution would be to add a substance that does not typically induce a stress response as the signal, such as a nutrient. The selection was designed with the idea that the signal would be something that a yeast cell already "detects" but this need not be a criterion.

A similar selective pressure for cell cycle arrest can be imposed by abolishing the spindle checkpoint through mutations in *MAD1*, *MAD2*, or *MAD3*. When treated with microtubule-destabilizing drugs such as benomyl, these mutants proceed through mitosis without properly segregating their chromosomes and die [1]. This could also be a strategy to avoid resistance to HU; by using a strain with defects in both checkpoints and alternating treatments of HU and benomyl, there is reduced selective pressure for drug-specific resistance/survival.

Another extension of this experimental evolution includes altering the time parame-

142

ters of the selection. Our experimental setup creates the ability to introduce a time gap between exposure to stimulus and treatment with HU. This strategy exploits the fact that the factor triggering the cellular response (salt stress) is different from that providing the selective pressure (HU). Because of this decoupling of the environmental change from selective pressure, we can administer them separately. With a time gap between the presence of stimulus and addition HU, the optimal strategy for a cell is to continue cell replication until the HU is present, using the earlier presence of the stimulus as an indicator of timing. This line of experiments would potentially evolve cells that exhibit a phenomenon more similar to "prediction". A version of this selection was performed, using a one-hour exposure to NaCl followed by one hour in standard media before adding HU. Cultures of the mutator ancestor strain and of Evolved 1-1, 1-2, and 1-3 (i.e., pre-evolved to arrest in NaCl) were passaged through rounds of this selection cycle. However, these experiments quickly yielded strains that survive HU with or without exposure to NaCl, implying the development of drug resistance. Further versions of this experiment were not pursued.

5.4   Closing remarks

This system has several features that make it well-suited to probing evolutionary dynamics: the ability to control the degree of selective pressure used during evolution, to determine the relative effect size of mutations by altering selective pressure in assays of evolved strains, and to separately measure contributions to fitness in the two selective environments. The success of a strain in the full selection cycle depends both on its survival in hydroxyurea and its proliferation in another environment. The physical features

of each environment can be varied without changing their basic selective pressures (arrest or growth), permitting the simultaneous evolution of different responses to environmental changes and imposing selection against trade-offs.

This execution of experimental evolution was not designed to determine the full target size of genetic adaptations to the selection nor their the relative likelihood. It was also not a goal to determine the "best" adaptation(s), although this could be tested for a subset of mutations by directly competing the evolved strains. The evolving populations were not passaged through hundreds of generations for an in-depth analysis of selective sweeps and clonal interference, although there are hints of both phenomena in the genetic profiles of the final populations. What this experiment did do was reveal change-of-function genetic adaptations that would not have been identified using a screen of the deletion collection or via transposon mutagenesis, which also predominately generates loss-of-function mutations.

In addition to the study of evolution as a process, experimental evolution provides a different way to approach the study of cellular physiology. Future experiments with the evolved strains described here could aid in clarifying the regulation of arrest under ionic stress and in characterizing the consequences of proceeding through the cell cycle with ion imbalance. This unique method of probing the cell cycle in general could reveal new characteristics of biological networks in a manner complementary to large-scale methods.

References

[1] R. Li and A. W. Murray, "Feedback control of mitosis in budding yeast.," *Cell*, vol. 66, pp. 519–531, Aug. 1991.

Appendices

Introduction

In the course of this project, it was frequently necessary to analyze very large datasets in a reproducible manner, and I wrote computer programs to automate these analyses. I created several ImageJ macros to aid in analyzing data from hydroxyurea survival experiments, including a colony-counting program with broad applications. They are described in Appendix A. In collaboration with Bryan Weinstein, I wrote a growth curve analysis program to calculate growth parameters from time courses of OD600 measurements, described in Appendix B. In order to efficiently analyze genomic sequencing data, I created programs to execute existing software functions for large numbers of files in parallel on the Harvard Odyssey research computing cluster. These are explained in Appendix C. These projects have so far enabled others in my research group to quickly analyze their data as well. The open-source code for these programs is currently available online at https://www.github.com/nwespe.

Appendix A

ImageJ Macros

The ImageJ macros and Python files described in this appendix are available at https://github.com/nwespe/ImageJ_functions. To install the macros, copy the files to the ImageJ or Fiji application plugins folder and restart the application.

A.1   Colony counting

1. Take images of plates facing up with trans-illuminating white light. Settings used on Gel-Doc: aperture = 8.00, zoom = 25, focus = 2.0, exposure = 75 ms.

2. In ImageJ/Fiji, run "auto count" from Plugins folder. Select folder containing plate images.

3. Run "manual count" from Plugins folder, selecting same folder as before. Click on unmarked colonies, typically around the edge of the plate. Colony counts (auto, manual, and total) are saved in a text file in the folder selected before.

Input image: "somecolonies.tif"
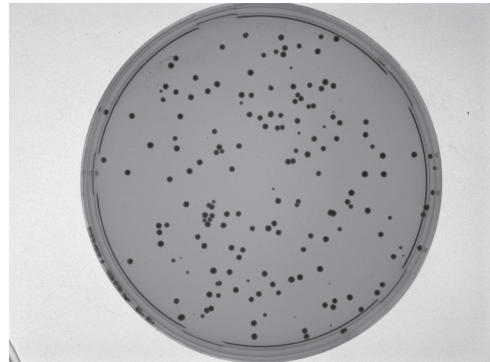
Convert to 8-bit, duplicate as mask and auto-threshold to find plate. Fit circle and shrink to define area for auto particle analysis.
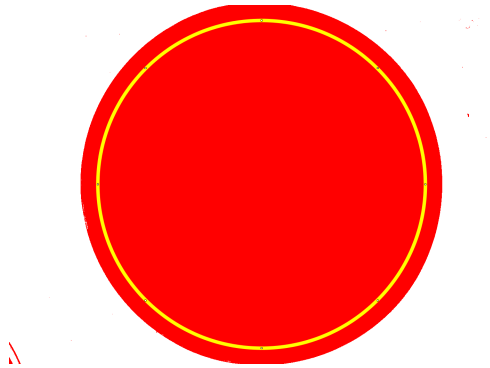
Subtract background of original image and add circle selection defined above.
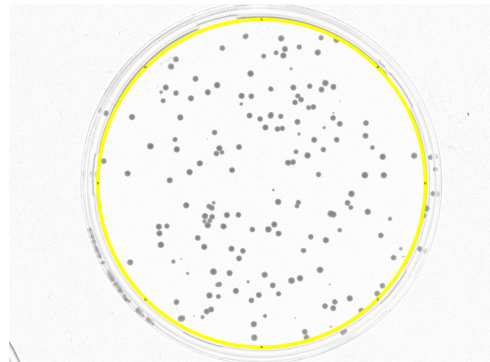
Clear outside and convert to binary. Run Watershed algorithm.
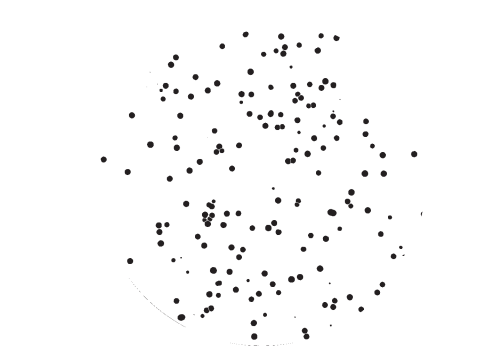
Run "Analyze particles."

Colony counting: "manual count" processing steps

Add outlines of auto-counted colonies to original image and ask user to click on uncounted colonies.

Save original image with overlay of auto-counted and manually counted colonies.
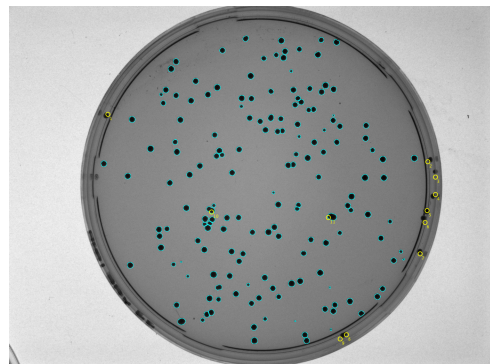
Save text file with auto, manual and total counts for all images: "Colony counts.txt"

| Image	Auto count	Manual count	Total
1	lotsofcolonies.tif	352	11	363
2	somecolonies.tif	162	11	173

A.2   Frogger sample splitting

1. Take images of plates facing up with trans-illuminating white light. Settings used on Gel-Doc: aperture = 8.00, zoom = 25, focus = 2.0, exposure = 75 ms. The next three steps are run directly from ImageJ/Fiji.

2. Place plate images in a folder. Select "batch plate cropper" from Plugins folder. Select folder with images. Move yellow box to be centered over samples.
   Subroutine information: batch_plate_cropper.py calls plate_cropper.py, which calls save_-roi.ijm.

3. Select "batch montage" from Plugins folder. Select "Individual" folder created by "batch plate cropper" in first step. Select destination folder.
   Subroutine information: batch_montage.py calls create_montage.ijm.

4. Select "combine experiment" from Plugins folder. Select "Montages" folder created by "batch montage" in second step. Select destination folder.
   Subroutine information: combine_experiment.py calls place_image.ijm.

   The strain sorting functions below are run from the command line or an ipython notebook and require a CSV file containing information about each sample.

5. strain_sorter.py: The montage files are copied to a folder created for each strain and named with the information provided. This program could be modified to sort by and include any information, not just strain and medium.

6. strain_report.py: All montage image files in a folder created by strain_sorter.py are compiled into a single PDF file.

Frogger sample splitting: "batch plate cropper" processing steps

Input image: "20161026 AB 1-6 Start.tif"
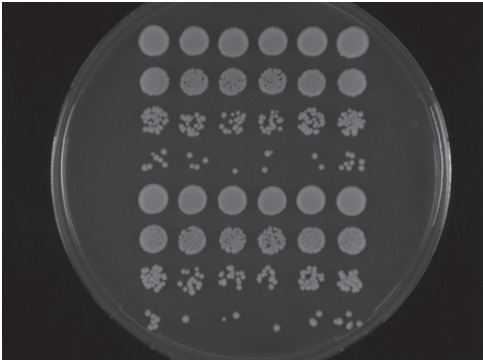
Adjust brightness and contrast.
Ask user to move selection box to be
centered over samples.
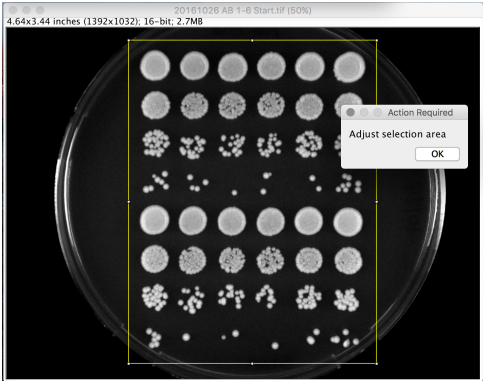
Crop image and save in "Adjusted" folder.
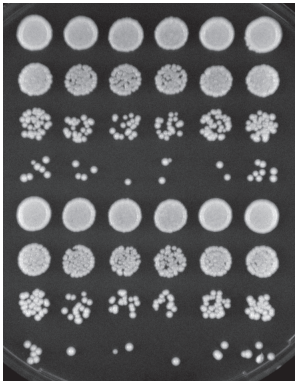
Select regions of individual dilution series
and save as separate files in "Individual"
folder.
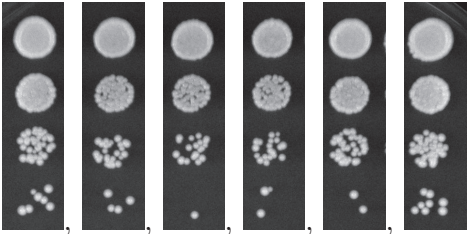
Frogger sample splitting: "batch montage" processing steps

Input images: "20161026 A1 Start.tif" and
"20161026 A1 End.tif." Program matches each start
image with its corresponding end image.

,
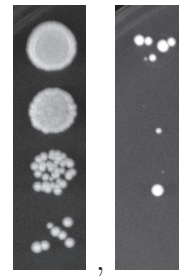
Rotate each image, combine into one image
file, and save in "Montages" folder.
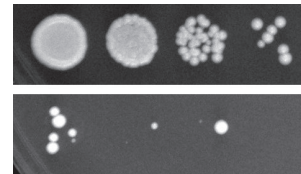
Frogger sample splitting: "combine experiment" processing steps

All montage images from
one experiment are
arranged in a single PDF
document.
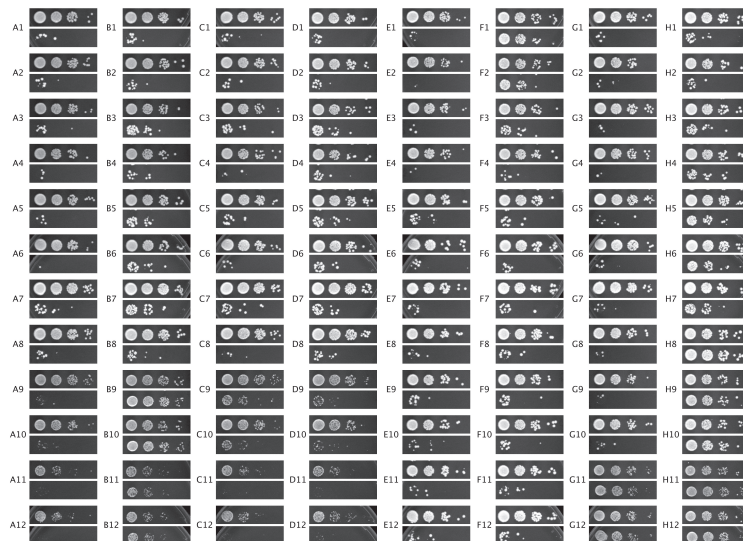
Sort montage image files into folders by strain or other characteristic.
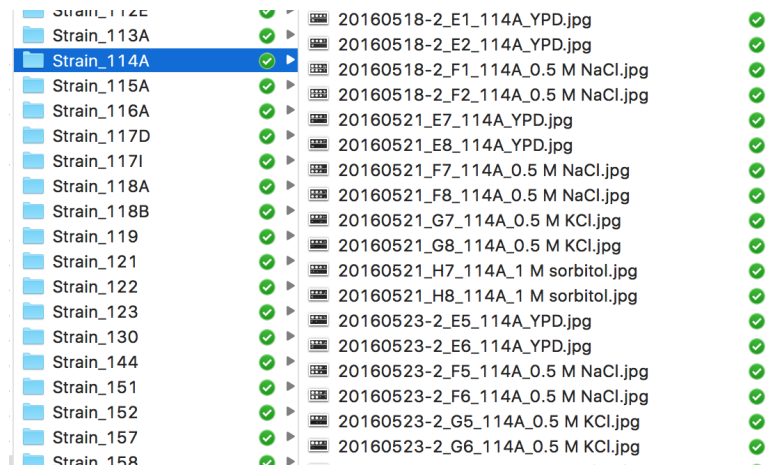
Compile report of samples for each strain.
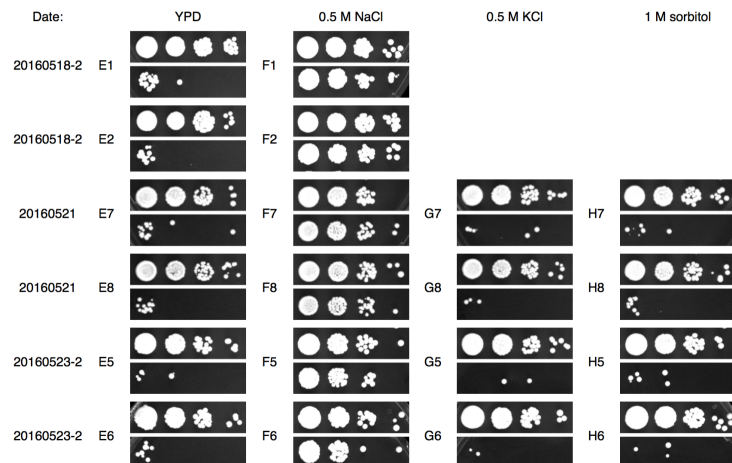
A.3 Evaluating hydroxyurea survival

The first step is run directly from ImageJ/Fiji and operates on the montage images created by the frogger splitting functions described above.

1.   a. Select "batch survival strain" from Plugins folder. Select folder containing subfolders with montage images for each strain.

   Subroutine information: batch_survival_strain.py calls analyze_pixels.ijm.

   Output: strainid_pixel_analysis.csv file for each strain; "Regions" folder in each strain folder containing copy of each montage image with regions marked.

       b. Alternatively, select "batch survival expt" to run analysis on a single experiment instead of on images grouped by strain.

   Output: one pixel_analysis.csv file; "Regions" folder containing copy of each montage image with regions marked.

2. HU_strain_analysis.py is run from the command line. This uses the pixel analysis files generated in step 1. This program computes end to start area ratios and determine the maximum dilution at which the pixel ratio is at least 0.5. This program also creates graphs displaying the maximum dilution results grouped by media condition.

   Output: one summary file plus three files for each strain: strainid_ratios.csv, strainid_max_dilutions.csv, and strainid_HU_survival.png.

Input image:
"20160518 E1 114 YPD.jpg"
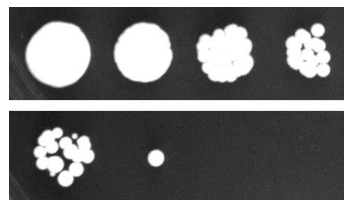
Select first dilution series in montage,
duplicate, threshold, and convert to mask.

Select region containing first spot,
duplicate, invert, and measure pixel area.

Repeat for each spot.

Repeat process for second dilution series in
montage. Mark spot selection regions on
dilution series image and save in "Regions"
folder.
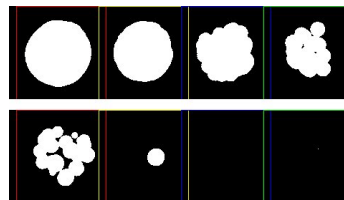
Save CSV file containing area
measurements.

Area	Dilution	Date	Well	Strain	Condition	Time
6682	1	20160518	E1	114	YPD	start
5501	2	20160518	E1	114	YPD	start
5053	3	20160518	E1	114	YPD	start
3565	4	20160518	E1	114	YPD	start
4202	1	20160518	E1	114	YPD	end
472	2	20160518	E1	114	YPD	end
0	3	20160518	E1	114	YPD	end
1	4	20160518	E1	114	YPD	end

The start and end area measurements for each sample are matched, and the end/start ratio is calculated.

Dilution	Area_start	Area_end	Ratio
1	6682	4202	0.629
2	5501	472	0.086
3	5053	0	0.000
4	3565	1	0.000

The maximum dilution level at which the ratio is still at least 0.5 is determined.

Max Dilution	Date	Well	Strain	Condition
1	20160518	E1	114A	YPD
1	20160518	E2	114A	YPD
3	20160518	F1	114A	0.5 M NaCl
4	20160518	F2	114A	0.5 M NaCl

Maximum dilutions are compiled by strain and condition, and a survival level is assigned based on the mean value.

Condition	N	Strain	mean	median	std	Survival
YPD	14	113A	0.14285714	0	0.53452248	-
0.5 M NaCl	13	113A	3.61538462	4	0.76794765	+++
0.5 M KCl	8	113A	0.5	0.5	0.53452248	-
1 M sorbitol	8	113A	0.25	0	0.46291005	-
YPD	12	114A	0.58333333	0.5	0.66855792	-
0.5 M NaCl	12	114A	3.25	3	0.62158156	+++
0.5 M KCl	8	114A	0.375	0	0.74402381	-
1 M sorbitol	8	114A	0.5	0	0.9258201	-

Maximum dilutions are also graphed by condition for each strain.

157
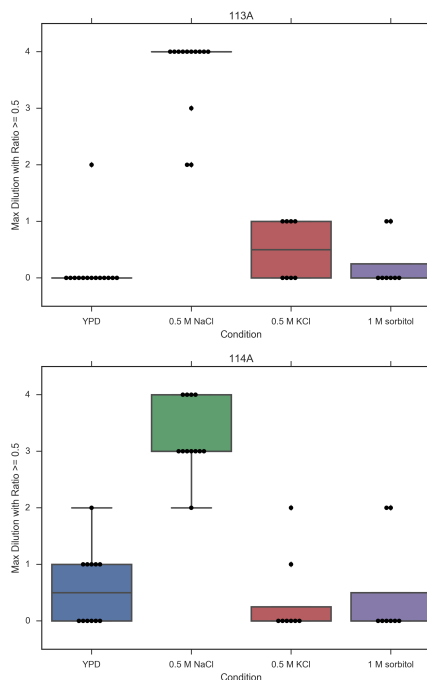
Appendix B

Growth Curve Analysis

A Python script was written to automate analysis of growth curves, and ipython notebooks were developed to plot the data and analysis results from multiple experiments. See Section 4.4 for details on how data are collected and calibrated. The files described in this appendix are available at https://github.com/nwespe/OD_growth_finder/.

B.1 Population growth rates from OD curves

This program was written to extract growth rate parameters from optical density time series readings of samples in a 96-well or 384-well plate. The program determines maximum growth rate, lag time and saturation time, as well as an "effective" growth rate for any specified time period. The program can be run from the command line or an ipython notebook; examples of running the program are shown in notebook format. Formatting of input files, program options, and examples of output are described below.

The raw data file contains timepoint information in the first column and optical density values in each subsequent column. The first row contains headings with well labels. Timepoints can be formatted as HH:MM:SS or as a number indicating minutes elapsed (e.g., 0, 10, 20). The input file can be an Excel (.xlsx), CSV or tab-delimited text file.

	A	B	C	D	E	F	G
1	Kinetic read	A1	A2	A3	A4	A5	A6
2	0:00:00	0.157	0.156	0.158	0.159	0.16	0.147
3	0:10:00	0.163	0.159	0.163	0.164	0.163	0.15
4	0:20:00	0.169	0.165	0.17	0.172	0.168	0.153
5	0:30:00	0.176	0.172	0.18	0.179	0.173	0.157
6	0:40:00	0.183	0.179	0.187	0.189	0.177	0.162
7	0:50:00	0.189	0.187	0.197	0.198	0.183	0.166
8	1:00:00	0.197	0.193	0.206	0.205	0.186	0.17
9	1:10:00	0.207	0.203	0.216	0.219	0.192	0.173
10	1:20:00	0.218	0.213	0.23	0.233	0.196	0.183

An optional plate layout Excel file can contain any metadata describing the well contents; these will be associated with each sample in the results output. The first column contains the well labels, and subsequent columns contain sample information.

	A	B	C	D	E	F
1	well	strain	media	replicate	expt_date	run
2	A1	006	YPD	1	10/5/16	YPD+/-NaCl
3	A2	006	YPD	2	10/5/16	YPD+/-NaCl
4	A3	003	YPD	1	10/5/16	YPD+/-NaCl
5	A4	003	YPD	2	10/5/16	YPD+/-NaCl
6	A5	035	YPD	1	10/5/16	YPD+/-NaCl
7	A6	035	YPD	2	10/5/16	YPD+/-NaCl

Blank values are essential for accurately calculating growth rate. These can be input in several ways: as a single value; as one or more wells, in which case the average value over all timepoints is used as the blank value; or as an Excel file containing a single value for each well. The last option enables the use of a different blank value for different wells, as in the case of using multiple media types with different background optical densities in the same experiment.

	A	B	C	D	E	F
1	A1	A2	A3	A4	A5	A6
2	0.098	0.098	0.098	1.020	1.020	1.020

Here is how to run the main analysis in an ipython/Jupyter notebook. The analyze_experiment function creates an output Excel file, a summary text file, and, if sample_plots = True, an image file for each sample with the fit parameters plotted onto the raw data. The make_plots function generates a histogram and a heatmap displaying a summary of the entire experiment.

```
import growth_curve_analysis as od
```

```
expt = od.analyze_experiment('/Users/nwespe/Desktop/GROC/ypd nacl data.txt',
                             plate_layout = '/Users/nwespe/Desktop/GROC/ypd nacl layout.xlsx',
                             blank_file = '/Users/nwespe/Desktop/GROC/gen_blank_file.xlsx',
                             out_dir = '/Users/nwespe/Desktop/GROC/test/',
                             sample_plots = True, correction = [0.6, 0.2141, 1.7935])
initialized experiment
analyzed samples
created output data table
```

```
od.make_plots(expt)
```

Below is an example of running the "effective growth rate" method, with input options for start and end times in minutes.

```
expt2 = od.analyze_experiment('/Users/nwespe/Desktop/GROC/ypd nacl data.txt',
                              plate_layout = '/Users/nwespe/Desktop/GROC/ypd nacl layout.xlsx',
                              blank_file = '/Users/nwespe/Desktop/GROC/gen_blank_file.xlsx',
                              out_dir = '/Users/nwespe/Desktop/GROC/',
                              method = 'effective_growth_rate',
                              start = 120, end = 360, saturation = True,
                              sample_plots = True)
od.make_plots(expt2, save = True,
              metric1 = 'growth rate 120-360', unit1 = 'ln(2)/minutes',
              metric2 = 'growth rate 120-360', unit2 = 'ln(2)/minutes')
initialized experiment
analyzed samples
created output data table
```

The output Excel file lists the calculated growth rate and details of this calculation for each sample.

	A	B	C	D	E	F	G	H
		well	growth rate	r-squared	doubling time	time of max growth rate	start of fit region	end of fit region
49		E1	0.00638313	0.99777214	108.590499	110	4	19
50		E2	0.00631378	0.99691896	109.783178	110	3	19
51		E3	0.00511805	0.99858381	135.431847	130	5	21
52		E4	0.00549975	0.99750496	126.032507	130	5	21
53		E5	0.00269513	0.99607891	257.185323	150	6	25
54		E6	0.00269228	0.99793209	257.457187	170	6	29

The output file also includes lag time, saturation time, and maximum OD reached.

	A	B	I	J	K	L	M	N
		well	lag time	OD at end of lag	saturation time	OD at saturation	max OD	time of max OD
49		E1	46.8515814	0.247	400	0.783	0.989	1140
50		E2	37.6644629	0.273	390	0.834	1.016	1140
51		E3	52.7314242	0.268	470	0.742	0.891	1180
52		E4	54.6872237	0.241	490	0.726	0.882	1140
53		E5	65.9894846	0.222	940	0.425	0.545	1140
54		E6	64.2155202	0.253	730	0.487	0.621	1140

The last columns in the output file include information from the optional plate layout file.

	A	B	O	P	Q	R	S	T	U
		well	row	column	name	media	replicate	expt_date	run
49		E1	5	1	003	0.5 M NaCl	1	Jul 27 2016	YPD+/-NaCl
50		E2	5	2	003	0.5 M NaCl	2	Jul 27 2016	YPD+/-NaCl
51		E3	5	3	006A	0.5 M NaCl	1	Jul 27 2016	YPD+/-NaCl
52		E4	5	4	006A	0.5 M NaCl	2	Jul 27 2016	YPD+/-NaCl
53		E5	5	5	110-c4	0.5 M NaCl	1	Jul 27 2016	YPD+/-NaCl
54		E6	5	6	110-c4	0.5 M NaCl	2	Jul 27 2016	YPD+/-NaCl

If sample_plots = True, an SVG file is created for each sample showing the calculated parameters plotted with the raw data.

161
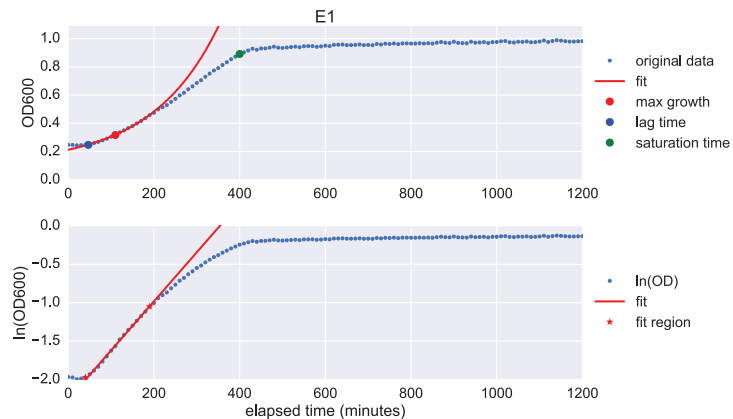
Any numeric column of the output file can be plotted in a histogram or heatmap using the function "make_plots." The histogram shows the distribution of values of a given output parameter for the experiment. The default parameter is doubling time. Other parameter options include growth rate, lag time, saturation time, and max OD.
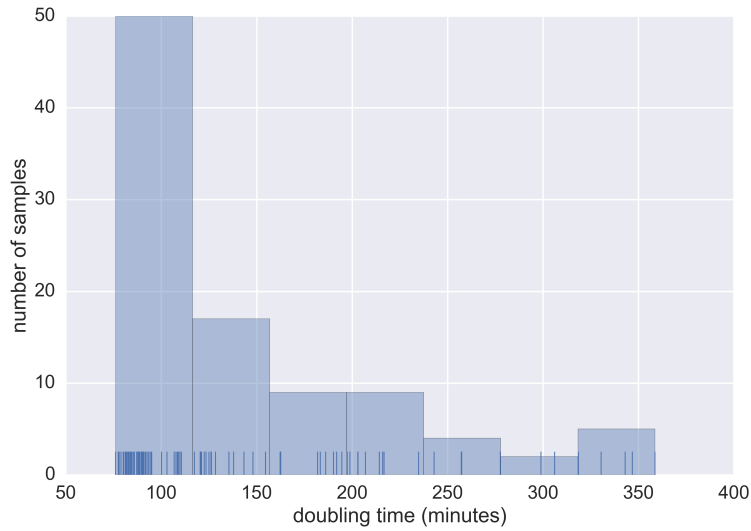
The heatmap shows values of a given output parameter for each well in the experiment. The default parameter is growth rate. Here, the usage of duplicate wells for each strain can be clearly seen, especially in rows E through H.
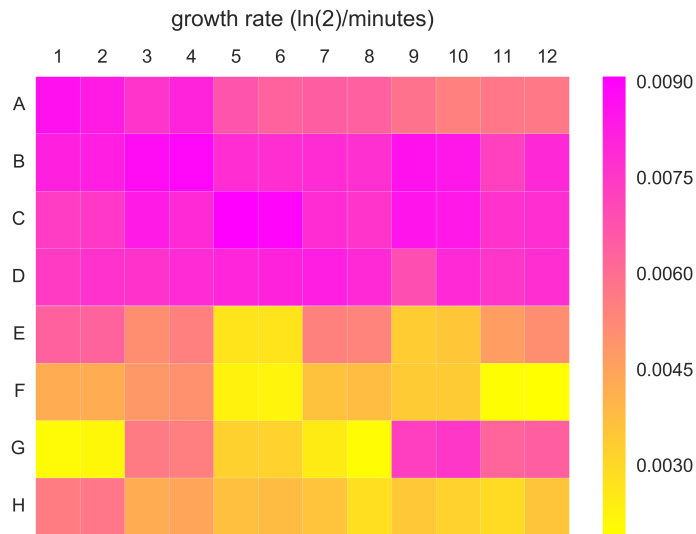
Growth rates: input options

Parameter	Description	Default value
data_file	file containing OD data	None (required)
plate_layout	file containing sample metadata	None
blank	blank value	0
blank_file	file containing blank value by well; overrides blank	None
method	see descriptions in separate table	'sliding_window'
out_dir	destination for output files	current directory './'
window_size	number of OD values used in sliding window to fit linear regression	9
sample_plots	create sample plots (adds significantly to runtime)	False
droplow	drop very low values from analysis (below -4.6 after calibration, equal to OD 0.01)	False
start	start time for "effective growth rate"	0
end	end time for "effective growth rate"	None (uses last point)
saturation	use saturation point as end time for "effective growth rate" if sample saturates before specified end (recommended)	False
correction	parameters for correcting non-linearity of OD readings; input as list [A, B, C] where A is OD value above which correction will be applied, and B and C are from the exponential fit y = B * exp(C*x) of measured vs. expected OD values, e.g., [0.6, 0.2141, 1.7935]	None

Growth rates: calculation method options

Method	Description
'sliding_window' (default)	Finds maximum growth rate. First uses a sliding window to find the maximum slope of the log-transformed data. Calculates maximum growth rate as slope of linear regression fit to all points whose sliding-window slopes were within 90% of the maximum.
'smooth_n_slide'	Fits a spline to all data points, then finds maximum growth rate using a sliding window as above. Used internally by 'sliding_window' method for samples whose r-squared value is below 0.9 in initial calculation.
'spline'	Fits a spline to all data points; maximum growth rate is the maximum derivative of spline. Faster but less accurate than sliding window methods.
'effective_growth_rate'	Calculates growth rate by fitting an exponential curve to all data points within specified start and end times.

B.2 Plotting growth rates results

An ipython notebook was developed to generate boxplots summarizing the growth rate results from multiple experiments. The input files for this notebook are the results Excel files created by the program growth_curve_analysis.py described in Section B.1. The key steps executed by this notebook are described below.

All results to be plotted are first imported and compiled into a single dataframe. User specifies a subset of the data to be plotted using a list of identifiers.

```
strain_list = ['006A', '006B', '028A', '028B', '028C']
group = 'all_ancestors'
```

The results for a given metric are grouped by identifier, segregated by condition, and plotted onto a boxplot.

```
g = sns.factorplot(x='name', y='growth rate 120-360',
                   col='media', col_order=['YPD', '0.5 M NaCl'],
                   data=gc, order=strain_list,
                   kind='box', legend=False, sharey=False)
plt.savefig('/Users/nwespe/Desktop/'+group+'_boxplot.svg', format='svg', bbox_inches='tight')
```
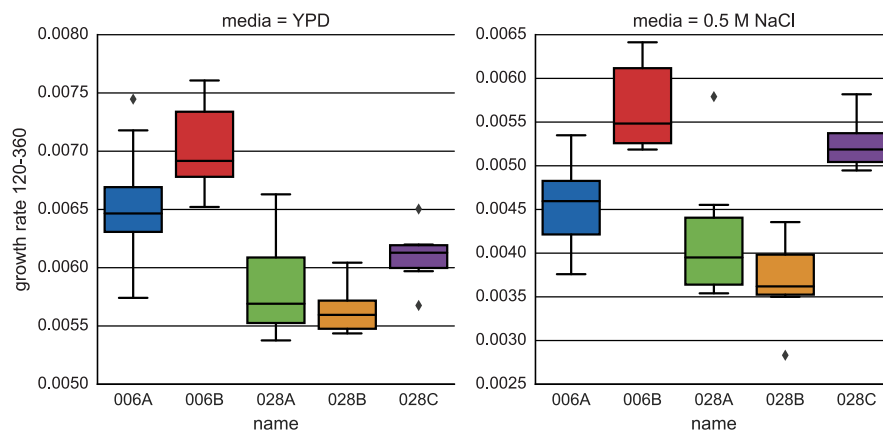

The metric (y) can be any numeric column of the results file (e.g., saturation time, lag time). The identifier (x) and condition (col) are descriptors from the plate layout file.

B.3 Aggregate growth curve plots

An ipython notebook was developed to create aggregate curves using the seaborn (version 0.7.1) data visualization package for python (seaborn.pydata.org). The input files for this notebook are the same as for growth_curve_analysis.py. See seaborn.tsplot for more details. The key steps executed by this notebook are described below.

The raw data and plate layout information are imported for each experiment.

To plot data from multiple wells and experiments together, the data are normalized by dividing each OD value by the initial OD. Dataframes are created of the normalized data and the log-transformed normalized data.

User specifies a subset of the data to be plotted using either a single identifier or a list of identifiers, e.g., strain names. The first example is a plot for a single strain with data separated by media type.

```
strain = '006A'
```

```
subset = trimmed_norm_data[trimmed_norm_data.name == strain]
lf_norm_data = get_long_form(subset, 'OD/init OD')

fig = plt.figure()
ax = sns.tsplot(data=lf_norm_data, time='time', value='OD/init OD', unit='sample',
                condition='media', err_style='unit_traces')
```

The data are plotted with a light line for each replicate and a dark line representing the mean value of the replicates at each timepoint with interpolation between the timepoints.
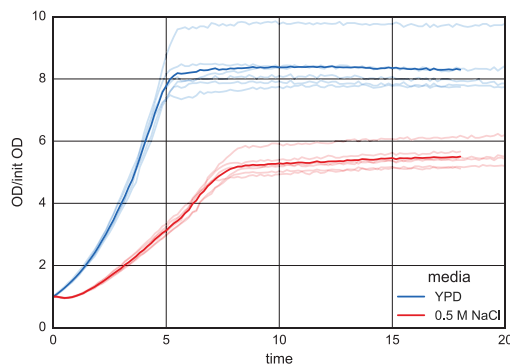
The following examples are plots of data for multiple strains. A time subset of the data can also be used; e.g., the first 6 hours of an 18-hour timecourse.

```
strain_list = ['006A', '114', '115', '116']
group = 'non-mutator evolved'
```

```
data = trimmed_log_norm_data   # or trimmed_norm_data
val = 'ln(OD/init OD)'   # or 'OD/init OD'
```

```
# to plot a time subset of the data
first_three_hrs = pd.concat([data.iloc[:, 0:19], data.iloc[:, -4:]], axis=1)
first_six_hrs = pd.concat([data.iloc[:, 0:37], data.iloc[:, -4:]], axis=1)
first_twelve_hrs = pd.concat([data.iloc[:, 0:73], data.iloc[:, -4:]], axis=1)

data = first_six_hrs
```
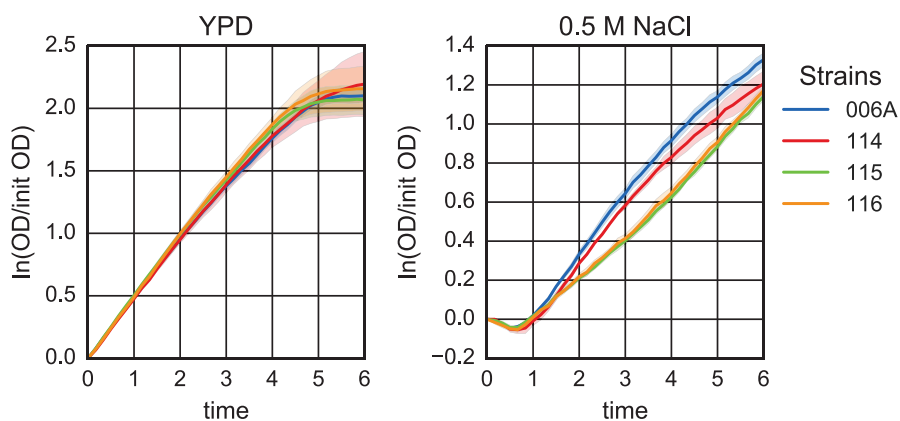

In the figure above, data for different media are separated onto two plots. These can also be plotted onto a single figure as shown below. Shaded regions around the lines show 95% confidence intervals.

167
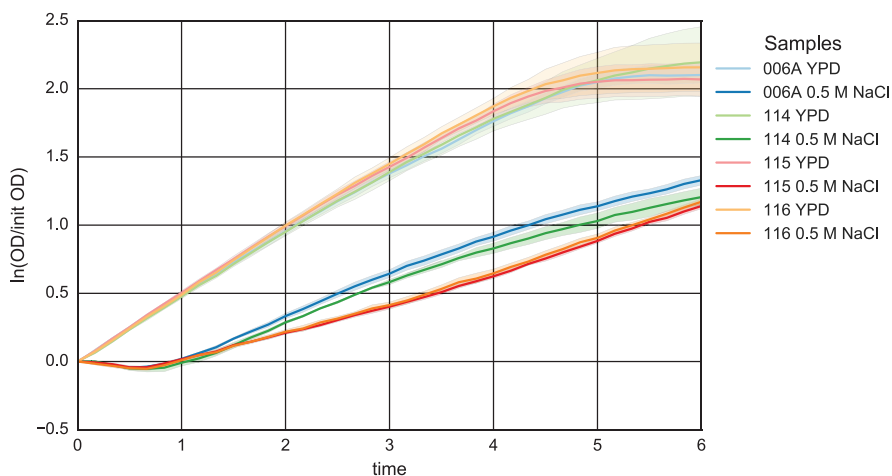
Appendix C

Sequence Analysis

Python and shell scripts were written to execute a sequence analysis pipeline on the Harvard University Odyssey research computing cluster. Scripts were also written to find genomic copy number variations and plot the results. The details of what these scripts do and how to run them are described below. The files described in this appendix are available at https://github.com/nwespe/sequence_analysis/.

C.1   Analysis pipeline

The pipeline is executed by a series of commands in the Bash script "segtools" written by John Koschwanez, which was modified to include additional processing steps and to accommodate parallel sample analysis. The execution of these commands was divided into three parts to enable parallel processing. The first part takes raw FASTQ files as received from the Bauer Sequencing Core Facility, i.e., already demultiplexed, and aligns

the sequence data to a reference genome, generating pileup files. The second part identifies mutations by comparing the pileup files to a reference genome. The third part locates the mutations relative to known features on the reference genome using a script written by John Koschwanez, "mutantanalysis.py." The software used, the input files, and the output files are described below for each part. The method of running this analysis pipeline using the Python scripts I wrote is described in Section C.2.

Potential future updates to the pipeline include updating the versions of the software programs used and adding a step to remove PCR duplicates from the analysis. There is room for improvement in the handling of sequences that map to more than one location in the reference genome; see Section 3.4 for the trouble that can arise from this. Detection of structural variation is absent from the pipeline described here; Miguel Coelho has used the program SoftSV to analyze chromosome rearrangements and this step could be added to the existing pipeline. Programs for running SoftSV and Breseq in parallel on the cluster were written and used by others but not extensively tested and are not described here.

I. Generate alignment

INPUT. The FASTQ data file contains four lines of information for each read; one example is rearranged into numbered lines below. Each read has a unique identifier including information on the machine used to generate the data and the adapter sequence used for multiplexing (lines 1-2). Following that is the sequencing read (lines 3-5) and the associated quality score for each base read (lines 7-9), separated by a line containing a '+' symbol. More information on the FASTQ format can be found in online resources.

```
1  @ILLUMINA-D00365:412:HCMTVADXX:2:1116:17461:37814 1:N:0:TAG
2  GCATGGTAAGGAG
3  CCAAAGACAATAGAAGCTTCAAAAGAGTCACTTAAACCGACAGCCTGGAAAACAATGGT
4  ACCGTAGTAGAAGAAATAGTTATCACCAGTTAATTGTTGTAGAGATTGAATCATGATAC
5  CCATCATAGTACGTTGAAACATGGCTGGTTTA
6  +
7  CCCFFFFFHHGHHJJJJJJJJJJJJIJJHHIJJJJJJJJJGIIIIIIIJJJJJJJJJJJJJ
8  HIJHHHGEFFFFFEEEEEDDEEEDEDDDDCDDEEEDCDDDDDACDDDDDDDCDDDDDDDC
9  CDDDDDDDDEEEDDDBDDDDDDCCDDDDDDBD
```

169

I. Generate alignment (continued)

STEP 1: ADAPTER TRIMMING. The cutadapt program searches each read for sequences matching the adapters used to generate the library and removes these bases, and any following bases, from the read. The adapter sequences are hard-coded into the segtools program and can be changed there if necessary. This step generates a new FASTQ file for each file analyzed and a report for each file or pair of files, in the case of paired-end reads; a part of the report is excerpted below.

```
Total read pairs processed:  1,315,728
  Read 1 with adapter:  728,181 (55.3%)
  Read 2 with adapter:  725,430 (55.1%)
Pairs written (passing filters):  1,315,728 (100.0%)

Total basepairs processed:  394,718,400 bp
  Read 1:  197,359,200 bp
  Read 2:  197,359,200 bp
Total written (filtered):  329,882,059 bp (83.6%)
  Read 1:  164,886,064 bp
  Read 2:  164,995,995 bp
```

STEP 2: QUALITY CONTROL. The FastQC program generates an HTML report providing an overview of several quality checks of the read data, such as per-base sequence quality, sequence length distribution and overrepresented sequences. A version of this report is also provided by the Bauer Core Facility with the FASTQ files, so it is possible to compare this information before and after adapter trimming.
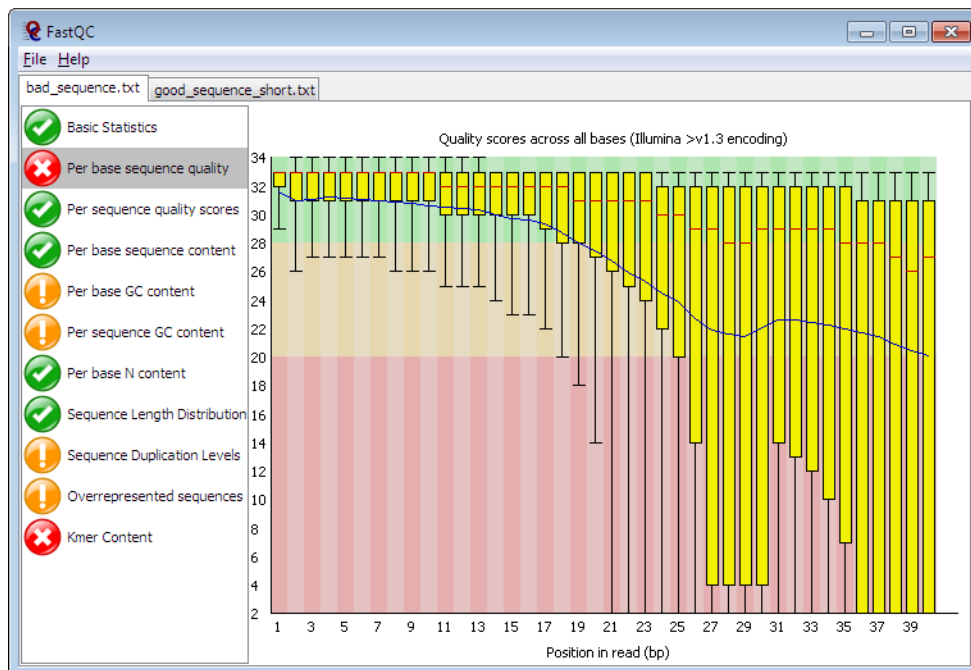

Image from https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

170

STEP 3: ALIGNMENT TO REFERENCE. The program used for aligning the sequence reads to the reference genome is BWA-MEM. The output is piped directly into several functions in SAMtools for sorting and indexing, which produces a sorted BAM (Binary sequence Alignment/Map) file. The sorted BAM file is processed by the 'IndelRealigner' function from the Genome Analysis ToolKit to improve alignment around insertions and deletions; this produces another BAM file. In SAM/BAM format, each read occupies one line; one example is rearranged into numbered lines below. The reads are ordered according to the reference coordinates of their primary alignment.

```
 1   ILLUMINA-D00365:412:HCMTVADXX:2:2101:4568:67853   163
 2     ref|NC_001133|   531   60   150M   =   565   184
 3   CATCATTATGCACGGCACTTGCCTCAGCGGTCTATACCCTGTGCCATTTACCCATAAC
 4   GCCCATCATTATCCACATTTTGATATCTATATCTCATTCGGCGGTCCCAAATATTGTA
 5   TAACTGCCCTTAATACATACGTTATACCACTTTT
 6   CCCFFFFFGHHHHJJJJIJJJJJIJJGIIIHHIIIJJIJIJHIJJFJJJJJIJJJJJJ
 7   JJHHFFFFEEEEDEEDDDDFEDBDDFDFEFFEEFDDEEEEDBBDDBDBDCDDDEEECC
 8   DDEDCDDDDDDCDEDDCDDCDCCDBDCCDDDDDD
 9   XA:Z:ref|NC_001135|,+828,150M,5; MC:Z:150M MD:Z:150
10   RG:Z:028A NM:i:0 MQ:i:60 AS:i:150 XS:i:125
```

The read information begins (lines 1-2 above) with the unique identifier from the FASTQ file followed by a bitwise FLAG value (**163**), the reference coordinate (**ref|NC_001133| 531** indicates Chromosome 1, base 531) and the mapping quality (**60**). Next is a CIGAR sequence (**150M**) that describes the alignment; in this case, the entire read of 150 bases was aligned (M denotes match/mismatch). The next part, **= 565 184**, indicates the coordinates of either the next read or the mate read (same chromosome, base 565) and the observed template length (i.e., insert length) determined from the mate's relative alignment (184). This information is followed by the sequence (lines 3-5), the base quality scores (lines 6-8), and a series of optional fields (lines 9-10). For interpretation of the bitwise FLAG value and the optional fields, see other SAM documentation.

In paired-end sequencing data, each read has a mate. The first part of the read information for mate of the example above is shown below.

```
 1   ILLUMINA-D00365:412:HCMTVADXX:2:2101:4568:67853   83
 2     ref|NC_001133|   565   60   150M   =   531   -184
```

The template length of the mate is given the opposite sign (**-184**). Because each read is 150 bases, a template length of less than 300 indicates that some bases of the template were read twice. This is accounted for by the next processing step.

STEP 4: SOFT-CLIP OVERLAPPING BASES. For paired-end data, the realigned BAM file is passed to the 'clipOverlap' function from BamUtil; this function identifies bases that were sequenced in both reads of a pair and 'soft-clips' one read to exclude the overlapping bases from downstream analysis. A new BAM file is created in which the CIGAR sequence for the read has changed to indicate which bases were clipped. In the example below, the 150 aligned bases are now 34 aligned bases followed by 116 soft-clipped bases. The BAM file still contains the sequence and quality data for these bases, but they will not be used to generate the pileup file in the next step.

Before clipping:

```
1   ILLUMINA-D00365:412:HCMTVADXX:2:2101:4568:67853   163
2     ref|NC_001133|   531   60   150M   =   565   184
```

After clipping:

```
1   ILLUMINA-D00365:412:HCMTVADXX:2:2101:4568:67853   163
2     ref|NC_001133|   531   60   34M116S   =   565   184
```

STEP 5: GENERATE PILEUP FILE. The final BAM file is processed by the SAMtools 'mpileup' function to create a pileup file. This file compiles the bases mapped to each location in the reference. Four lines are excerpted from a pileup file below.

```
ref|NC_001133| 534 C 50
  ,,,.......................,,.,,,,..............$,,,.,....^K,
  FJCIDDCEEFFFFHJJJJJJDIJDJJBJGGJJIJIGJIJBJHCGHFFCCC
ref|NC_001133| 535 A 49
  ,,,.......................,,.,,,,...............,,,.,....,
  FJCGCDEECFFEEHIJJJGIFHJDJIBJIIIJJJGEJJGIHCHEFFCCC
ref|NC_001133| 536 T 49
  ,,,.......................,,.,,,,...............,,,.,....,
  FJCIDDDDEFFEEHIIHJIGDIIDJIBJIGIJIJIIJJIGDAHEDFFFC
ref|NC_001133| 537 T 49
  ,cCc.....CC...CC.C.C,..,..cc..CCCCC......,cc.c....c
  FJDFEEEEEFFEEFIJHJJHDJJDIJFJJJIJJJJIJJJIHGCHIFFFFC
```

Each line begins with the reference sequence coordinate (in this case, Chromosome 1, bases 534 through 537) followed by the reference base and the number of reads mapped to that location. The string of symbols indicates whether the mapped read is a match to the forward (.) or reverse (,) strand, or a mismatch to the forward (ACTGN) or reverse (actgn) strand. The symbol ˆ marks the beginning of a read, the character following it indicates the read's mapping quality, and the symbol $ marks the end of a read. The string of letters are the corresponding base qualities.

172

FIND MUTATIONS IN CLONES. The analysis pipeline finds mutations in clones using the VarScan 'somatic' function. VarScan was written to find sequence variants (SNPs and indels) between normal and tumor samples, and it classifies variants as either germline (both samples differ from the reference genome) or somatic (only the tumor differs). When identifying mutations in evolved strains, the evolved clone is considered the tumor sample and the ancestor is the normal sample. The first 15 columns of three lines are excerpted from a VarScan .snp file below. `normal/tumor_reads1` and `reads2` are the reads with the reference (`ref`) and variant (`var`) base, respectively; `normal/tumor_gt` is the consensus genotype.

```
chrom          position  ref  var  normal_reads1  normal_reads2
ref|NC_001133| 537       T    C    29             20
ref|NC_001133| 610       G    A    18             13
ref|NC_001142| 107919    C    A    14             0


normal_var_freq  normal_gt  tumor_reads1  tumor_reads2  tumor_var_freq
40.82%           Y          36            26            41.94%
41.94%           R          26            32            55.17%
0%               C          0             14            100%


tumor_gt  somatic_status  variant_p_value            somatic_p_value
Y         Germline        3.7752715862120977E-17     0.5304477608132048
R         Germline        1.4812831017457745E-17     0.16669057568084028
A         Somatic         1.0                        2.4927336813189342E-8
```

FIND MUTATIONS IN POPULATIONS. The analysis pipeline finds mutations in populations using the VarScan 'pileup2snp' and 'pileup2indel' functions, which compare a single pileup file to a reference genome. This function outputs similar information as the VarScan 'somatic' function but cannot distinguish pre-existing mutations from evolved mutations (i.e., germline versus somatic). These files are used in conjunction with the ancestor-versus-evolved clone VarScan files generated by the 'somatic' function to identify evolved mutations present at high frequency in either a back-crossed pool or an evolving population. Thirteen columns of three lines are excerpted from a VarScan .snp file below.

```
Chrom          Position  Ref  Cons  Reads1  Reads2  VarFreq  Strands1
ref|NC_001133| 537       T    Y     77      42      35.29%   2
ref|NC_001133| 610       G    R     40      28      41.18%   2
ref|NC_001142| 107919    C    A     0       39      100%     0


Strands2  Qual1  Qual2  Pvalue                       ...  VarAllele
2         38     38     2.750496826499323E-15        ...  C
2         36     36     1.1003828151779401E-10       ...  A
2         0      37     3.6741722220670473E-23       ...  A
```

CLASSIFY MUTATIONS. The final step of the automated pipeline uses the "mutantanalysis.py" script written by John Koschwanez. This script matches the sequence variants identified by VarScan to annotated features on the reference genome. It determines whether the mutation affects coding or non-coding DNA and, if in a coding region, the effect of the mutation on the amino acid sequence. It outputs this information in an HTML file with links to the Saccharomyces Genome Database and in two text files that list the information by gene and by mutation. It also compiles the reads from the BAM file that overlap the mutated region and enables viewing them in the HTML file.

HAL5	**Protein mutation**	**Base 1543 (Chromosome 10, Position 107919)**
ORF	DNA fasta	Read alignment
S000003701	DNA alignment	033B : forward mutation snp from C to A at 100 percent of 14 reads.
SGD	Protein fasta	119_NaCl_HU segregates A at 100 percent of 39 reads.
Fungal alignment	Protein alignment	

```
AYGVV ref_sequence_HAL5_ORF
AYGVV 028A_HAL5_ORF
AYWVV 033B_HAL5_ORF
```

Tabbed output by gene:

```
gene_name   sgdid       chr_num    position   sample_name   snp_indel
HAL5        S000003701  10         107918     033B          snp
WHI2        S000005569  15         412304     033B          snp
IRC23       S000005570  15         412304     033B          snp


ref         read        fraction   tot_reads  mutation_type
C           A           100        14         nonsynonymous
A           G           100        10         nonsynonymous
A           G           100        10         promoter
```

Tabbed output by mutation:

```
chr_num   position   snp_indel   seg_percent   genes       sample_name
10        107918     snp         100           HAL5        033B
15        412304     snp         77            WHI2,IRC23   033B
```

DETERMINE LIKELIHOOD OF CAUSALITY. This is where the human analysis takes over. The user must evaluate the likelihood of the identified mutations to cause the phenotype of interest. This analysis could take into consideration the effect of the mutation on the amino acid sequence, the previously-described function of the gene product, and any phenotypes associated with mutations in the locus. In the case of bulk segregant analysis, it is important to consider whether there is linkage to other loci under selection, e.g., a locus used to generate the pool of spores.

MEASURE COPY NUMBER VARIATION. The script "copynumber.py" finds differences in copy number between two pileup files using the VarScan 'copynumber' function. Three lines are excerpted from a resulting .copynumber file below.

```
chrom           chr_start   chr_stop   num_positions
ref|NC_001133|  9372        9471       100
ref|NC_001133|  9472        9571       100
ref|NC_001133|  9572        9671       100


normal_depth    tumor_depth log2_ratio gc_content
28.7            22.1        -0.376     39.0
27.2            21.3        -0.352     37.0
20.4            24.0        0.232      32.0
```

VISUALIZE COPY NUMBER VARIATION. The copy number data can be visualized using the script "plot_copynumber.py" for one sample or the script "batch_plot_copynumber.py" for multiple samples. Each plot displays read depth normalized to the entire genome versus chromosome position. The resulting plots for all chromosomes of a sample are saved in a single PDF. The script generates one file for each sample plus one file for each comparison (e.g., ancestor versus evolved) listed in the input text file; for the latter, the read depth of the evolved sample is divided by that of the ancestor.

Read depth of a single sample:
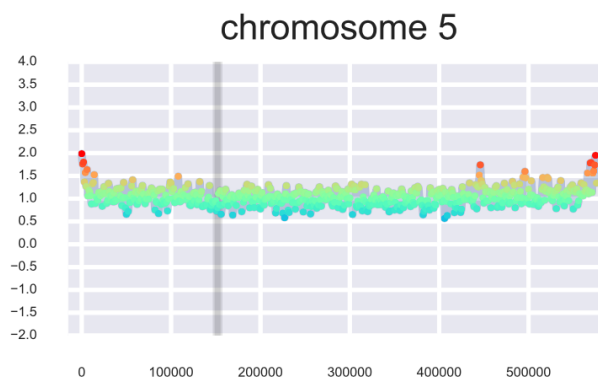
Read depth relative to ancestor:

175
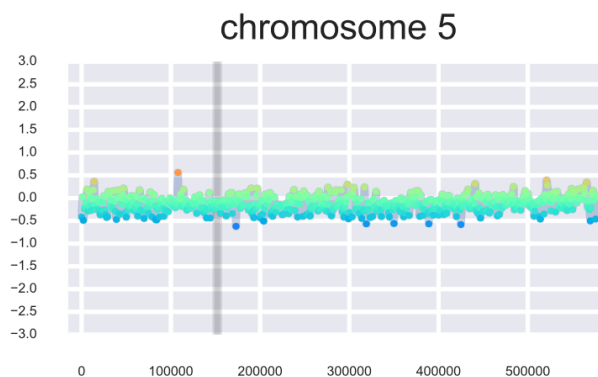
EVALUATE ALIGNMENT. The picard-tools function 'Collect Alignment Summary Metrics' yields information about the alignment of reads in a SAM/BAM file, such as the numbers of aligned reads and aligned bases and the mean read length. These metrics can be collected for each sample by the script "picard_align_report.py" and then collated for many samples by the script "compile_picard_report.py." Snippets of the individual sample report and the compiled report are below.

Picard alignment summary metrics:

```
CATEGORY   TOTAL_READS   PF_READS   PCT_PF_READS   PF_READS_ALIGNED   ...
PAIR       2631456       2631456    1              2626143            ...
```

Compiled metrics:

```
sample        total_reads   aligned_reads   percent_reads_aligned   ...
033B          2516654       2411302         0.958138                ...
119_NaCl_HU   6532322       6229601         0.953658                ...
028A          2631456       2626143         0.997981                ...
```

C.2   Executing the pipeline

The user runs three scripts from the command line, corresponding to parts I-III of the analysis described in Section C.1. The user must be logged into the Harvard Odyssey Research Computing cluster, access to the cluster requires registration and must be arranged through Research Computing. The user's main folder on the cluster must contain the appropriate Bash profile, which loads the modules required for executing the pipeline software programs. An example of this Bash profile is available on the lab server in `/murraylab/Sequencing/Scripts/Instructions`. The modules required include: Anaconda (for Python), cutadapt, FastQC, BWA, SAMtools, bamUtil, bamtools, VarScan, ClustalW, and Picard-tools. Also, the environmental variables GATK, VARSCAN, and PICARD should be set to the respective paths for these programs. For

the versions of these programs used in this dissertation's analysis, see Section 4.6.

The complete file tree for running the pipeline is shown in the image below. The programs simply titled "script1.py," "script2.py," and "script3.py" are executed by the user in that order; the "segtools" program and files in the "array_scripts" and "mutantanalysis" directories are called by these programs. The programs "copynumber.py," "plot_copynumber," "batch_plot_copynumber," "picard_align_report.py," and "compile_picard_report.py" are for analyses described in part IV in Section C.1.
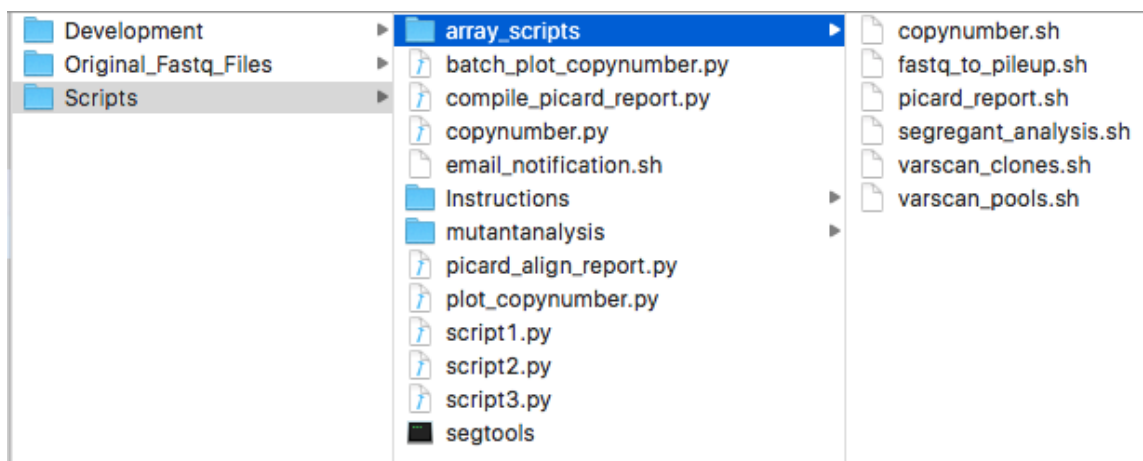
All three scripts take as input (1) a directory where the FASTQ, pileup or BAM files are located, (2) an output directory, and (3) a CSV file containing names of the samples to be analyzed, with ancestors, clones and pools listed in separate columns. Each script first checks whether the files it will create already exist and whether the input files it needs are in the directory provided. It then submits this information to the cluster as a separate job for each sample to be run in parallel, using the relevant array script to execute functions either from "segtools" or from "mutantanalysis.py." The required inputs and the format for command-line execution for each script are detailed below.

I. Required input

FASTQ FILES. The original FASTQ files can be placed in the lab server's "Sequencing" directory. No analysis files should be written to this directory. I recommend storing the files in a directory tree that enables other lab members to easily discern the relevant experiment, such as:

```
/murraylab/Sequencing/Original_Fastq_Files/Nichole/2015-01-15/
```

CSV FILE. Execution of the analysis pipeline requires a CSV file containing sample names that uniquely identify the corresponding FASTQ files. In the examples below, there are samples named "006A" and "006B;" an entry of "006" will find the FASTQ files for both samples and result in an error. These sample names should be arranged in two or more columns according to the comparisons to be made and the headings must follow one of the four formats below.

For single clone analysis: For multi-clone analysis:

```
Ancestor  Clone          Ancestor  Clone1  Clone2  Clone3  ...
006A      113            028A      033A    034A    035A    ...
006A      114            006A      113     114             ...
006B      115            006B      115     116     117     ...
```

For segregant analysis: For segregant analysis with multiple pools:

```
Ancestor  Clone  Pool     Ancestor  Clone  Pool1        Pool2       ...
028A      033A   066      028A      033A   119_Initial  119_NaCl ...
028A      034A   067      028A      034A   152_Initial  152_NaCl ...
028A      035A   068      028A      035A   121_Initial  121_NaCl ...
```

Files could contain just two columns (Ancestor and Clone) or many columns for clones or pools. The analysis to be run – segregant analysis with clone (sawc) or comparing multiple clones (cmc) – is determined from the column headings; e.g., a "Pool" column indicates that segregant analysis should be run. In multi-clone analysis, each clone in a row is compared to the ancestor, and the results for all clones in the row are output into a single set of files. For segregant analysis with multiple pools, each ancestor-clone-pool trio will have its own output.

OUTPUT DIRECTORY. The output directory should be given as a full path to the user's folder on the lab server; i.e., :

```
/murraylab/Users/nmcollin/Sequencing/2015_Seq_Data/
```

SCRIPT1.PY. The first script finds FASTQ files for the samples listed in the CSV file and creates pileup files. All output files are placed into a new folder with the sample name in the given output directory. The script determines whether FASTQ files are present as single or paired-end reads and uses either the 'singlef2p' or the 'pairedf2p' function from the "segtools" script to generate pileup files.

```
input:    python /full/path/to/script1.py
          -i /full/path/to/fastq_files/
          -d /full/path/to/output_directory/
          -f /full/path/to/csv_file.csv

output:   /output_directory/alignment_files/sample/sample.pileup
```

An output pileup file is created for each individual sample listed in the CSV file. The output folder will also contain other files created in pileup process, including the FastQC HTML report, the cutadapt summary file, and the BAM file required for script 3.

SCRIPT2.PY. The second script finds pileup files for the samples listed in the CSV file and creates .snp and .indel files for each ancestor-clone pair and for each pool. All output files are placed into the relevant subdirectory of a new directory "varscan_files." The VarScan files are created by calling the 'varscan_clone' or 'varscan_pool' function from "segtools." These are submitted as two separate array jobs to the cluster: one for ancestor-clone pairs and one for pools. If -d is omitted, the directory is assumed to be the same one specified by -i.

```
input:    python /full/path/to/script2.py
          -i /full/path/to/pileup_files/
          -f /full/path/to/csv_file.csv

          [-d /full/path/to/output_directory/] (optional)

output:   /output_directory/varscan_files/anc_vs_clone/...
             ...anc_vs_clone.snp
             ...anc_vs_clone.indel
          /output_directory/varscan_files/pool/pool.snp
          /output_directory/varscan_files/pool/pool.indel
```

179

SCRIPT3.PY. The third script finds BAM files created by script 1 and .snp and .indel files created by script 2 and submits these to the "mutantanalysis.py" script to obtain the comparison output for each ancestor-clone-pool or ancestor-clone (-clone...) group. The user can provide separate paths to the BAM and VarScan files, or a single path that encompasses both. If -v and/or -d are omitted, the directories are assumed to be the same one specified by -b. The optional parameters -m and -s are integers between 0-100 that indicate the percent of reads needed to call a mutation (m) or segregant (s). Defaults are m=90 and s=70. Recommended usage for diploid clones is m=35 or lower. The output for each comparison is placed into the relevant subdirectory of a new directory "filename-p-m%-s%" where filename = name of CSV file, p = pipeline (sawc or cmc), m% = mutation percent, and s% = segregation percent.

```
input:   python /full/path/to/script3.py
         -b /full/path/to/bam_files/
         -f /full/path/to/csv_file.csv

         [-v /full/path/to/varscan_files] (optional)
         [-d /full/path/to/output_directory/] (optional)
         [-m mutation percent cutoff] (optional; default:  90)
         [-s segregation percent cutoff] (optional; default:  70)

output:  /output_directory/csv_file-sawc-m90-s70/anc_clone_pool/...
            ...compare.html
            ...tabbed_output_by_gene.txt
            ...tabbed_output_by_mutation.txt
```

Other output files present in this folder are used by the HTML file for viewing aligned reads and FASTA and protein sequence alignments.

EMAIL_NOTIFICATION.SH. When a script executes successfully, it returns a job ID number. The user will be prompted with the option to have an email notification sent once the executed program is completed. To use this option, the user must copy the "email_notification.sh" script from the common directory to their own directory, open the script and change the email address in line 8 to their own. Then, type the following into the command line, with the job ID in the brackets:

```
input:   sbatch --dependency=afterok:[JOB ID #]
            /full/path/to/email_notification.sh
```

To check the status of an executed job, type the following into the command line, substituting the relevant information in the brackets:

```
input:   squeue -u [username] -j[JOB ID #]
```

COPYNUMBER.PY. This script is similar to "script2.py." This script finds pileup files for the samples listed in the csv file and creates .copynumber files for each ancestor-clone pair. All output files are placed into the relevant subdirectory of a new (or existing) directory "varscan_files." The VarScan files are created by the 'varscan_copynumber' function from "segtools."

```
input:    python /full/path/to/copynumber.py
          -i /full/path/to/pileup_files/
          -f /full/path/to/csv_file.csv
          [-d /full/path/to/output_directory/] (optional)

output:   /output_directory/varscan_files/anc_vs_clone/...
               ...anc_vs_clone.copynumber
```

To visualize the results of the copynumber analysis, run "plot_copynumber.py" for one sample or run "batch_plot_copynumber.py" for multiple samples using a CSV file as input. These files create PDF files with plots of copy number for each chromosome.

PLOT_COPYNUMBER.PY. This script takes a single .copynumber file as input and generates three PDF files: one for the ancestor, one for the clone, and one for the clone normalized by the ancestor.

```
input:    python /full/path/to/plot_copynumber.py
          -f /full/path/to/anc_vs_clone.copynumber
          -d /full/path/to/output_directory/

output:   /output_directory/anc_copynumber.pdf
          /output_directory/clone_copynumber.pdf
          /output_directory/anc_vs_clone_copynumber.pdf
```

BATCH_PLOT_COPYNUMBER.PY. This script takes a CSV file as input and generates one PDF file for each individual sample and one for each ancestor–clone comparison, in which the clone's coverage is normalized by the ancestor's coverage. The format of the CSV file should match one of the formats described in part I.

```
input:    python /full/path/to/batch_plot_copynumber.py
          -i /full/path/to/copynumber_files/
          -f /full/path/to/csv_file.csv

          [-o /full/path/to/output_directory/] (optional)
          [-w window size for smoothing] (optional; default:  1000)

output:   /output_directory/anc_copynumber.pdf
          /output_directory/clone_copynumber.pdf
          /output_directory/anc_vs_clone_copynumber.pdf
```

PICARD_ALIGN_REPORT.PY. This script executes the 'Collect Alignment Summary Metrics' function from Picard tools on the BAM file of each sample listed in the CSV file.

```
input:    python /full/path/to/picard_align_report.py
          -d /full/path/to/bam_files/
          -f /full/path/to/csv_file.csv

output:   /full/path/to/bam_files/alignment_files/sample/sample_asm.txt
```

This script could be modified to execute any of the functions in the Picard tools package on sample BAM files by replacing the function name in the "picard_report.sh" array script.

COMPILE_PICARD_REPORT.PY. This script takes the alignment summary files generated by the function above and compiles this information into a single text file. The output file includes computations of coverage levels using a yeast genome size of 12,157,105 bases.

```
input:    python /full/path/to/compile_picard_report.py
          -d /full/path/to/bam_files/
          -f /full/path/to/csv_file.csv

output:   /full/path/to/bam_files/csv_file_alignment_data.txt
```