EXPERIMENTAL EVOLUTION OF FISSION YEAST GENOMES IN RESPONSE TO REPEATED ENVIRONMENTAL CHANGES

Laurent F. van Trigt

This dissertation is submitted for the

Degree Doctor of Philosophy

January 2013

University College London

London, UK

Declaration

I, Laurent van Trigt, confirm that the work presented in this thesis is my own and has not been submitted for a degree or any other qualification at any other university. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Laurent F. van Trigt

07/01/2013

Acknowledgements

First of all, a very big thank you to my supervisor, Jürg Bähler, who gave me the opportunity to carry out my PhD in his incredible laboratory. I thank him for always being ready to help, providing guidance, encouragement, and a contagious drive to learn and discover.

I would like to thank my second supervisor, Max Reuter, who has been involved in many meetings and discussions about the direction of the project.

My very grateful thanks are due to Daniel Jeffares, for his many useful suggestions with regards to experimental design, data generation and interpretation.

I would like to address my gratitude to Ville Mustonen who has been an essential collaborator, and gave a new momentum to the project at a time I could not see the forest for the trees.

I owe an enormous debt of gratitude to many past and present members of the laboratory. Samuel Marguerat, Martin Převorovský, Vera Pancaldi, Tannia Gracia, Luis López Maury, Babis Rallis, Sandra Codlin, Sophie Atkinson, Antonia Lock, Danny Bitton, Xavi Marsellach Castellví, Bruno Lages, Nazia Bibi, and Dora Sideri. All of them have provided a lot of very helpful practical guidance, and discussing my work with them has often been key to the development of my research.

I thank UCL for providing a PhD studentship.

Thank you to my parents for their unconditional love and support, and always letting me follow my passions.

Finally, this thesis could not even have started without the endless patience and support of my wife, Anna. She has made a lot of sacrifices which I will never be able to repay entirely, although I hope she takes this thesis as a first instalment.

Abstract

Evolutionary biology has mainly relied on comparative studies. Supported by the advances in genomics, several groups are now using microorganisms for experimental evolution studies to directly observe the evolution process and its genetic basis. To understand how cells adapt to an environmental stress, I developed experimental evolution assays during which fission yeast populations were repeatedly exposed to short heat-shocks, followed immediately by ideal growth conditions. Samples were frozen after each selection cycle for subsequent phenotypic and genotypic analyses. In total, I performed 10 independent such experiments for up to 150 selection cycles, including a control experiment with mock-treated samples.

In all experiments, except for the mock-treated control, cells acquired a markedly increased survival to heat and other stresses within only 10-20 selection cycles, while during later cycles there was a more subtle but continuous increase in stress resistance. Surprisingly, all populations, including the mock-treated control, also evolved a more rapid re-initiation of exponential growth following heat shock and a more efficient use of the growth medium, reflected by higher biomass in stationary phase. These rapid adaptations seem to be based on DNA variations as the new phenotypes are stable.

The cellular phenotypes and global regulation of gene expression of the evolved strains were determined using stress assays and DNA microarrays. Genomes of 96 strains at selected time points over the course of the evolution experiments were sequenced. We analyzed what genetic changes in regulatory and coding regions occurred in the different independent experiments, with the ultimate goal of uncovering the specific changes that are causing the striking alteration in phenotype. In collaboration with Dr. Ville Mustonen, we determined dynamic allele frequency trajectories that had swept through each population. Genetic variation among the parental cells seems to underlie the rapid initial adaptation of new phenotypes. This analysis led to the discovery of a common haplotype consisting of 9 mutations, which is likely associated with the more rapid recovery of growth after heat shock.

List of Abbreviations

bp	DNA base-pair
CFU	Colony forming units
IGV	Integrative Genomics Viewer (visualization tool for genomic datasets)
Max Slope Time	The time for a culture to reach maximum growth
HS	Heat shock
P	The parental population
E1, E2, E3	Evolved lines treated with a 10 min heat shock over 150 cycles
E0	Evolved line treated with 5 minute heat shocks
M0	Evolved line treated with 1 minute heat shocks
M1, M2	Replicated lines of E1, E2, and E3
S-P	A parental strain founded from a single genotype
S1, S2	Evolved lines starting from the clonal strain S-P
Mock	Line that was mock treated in parallel with evolving lines
chr	Chromosome
YES	Rich nutrient medium (yeast extract supplement)
SNP	Single nucleotide polymorphism
Indel	Short insertion or deletion
Fst	The fixation index as a measure of how populations differ genetically

Contents

CHAPTER 2

CHAPTER 5

CHAPTER 6

List of Figures

List of Tables

Chapter 1

Introduction to experimental evolution

Introduction to experimental evolution

This chapter gives an overview of some of the major experimental evolution approaches conducted over the past 25 years. I illustrate its power as a research approach but also present its challenges and limitations. Furthermore, I will introduce aspects of experimental design and show how recent advances in technology, which allow the detection of mutations on a genome-wide scale, are revolutionizing our understanding of evolutionary change.

Why experimental evolution?

The adaptation of living organisms to changes in the environment is critical to the survival of the species (Darwin 1859). There exist mainly two ways by which unicellular organisms adapt to novel conditions. First, there is a fast physiological response under the control of genetically hard-wired regulatory mechanisms (Lopez-Maury et al. 2008). Second, there is selection over many generations of genetic variants that confer fitness (Barrick et al. 2009). The difficulty in studying how phenotypes change is that they are influenced by many levels of biological organisation which will influence their reproductive success. Within the cell, there are interconnected regulatory steps that intervene between the genome and the phenotype (Lackner et al. 2007). Outside the cell, population dynamics processes and interactions between phenotypes and environmental conditions will act upon selection. Whether single cells or whole organisms, an important goal of science is to put together a global picture of all these interactive functional parts that influence the type of mutant that is being selected (Fig. 1.1). This will be particularly relevant in the field of infectious disease as both disease-causing pathogens and their hosts evolve. Pathogens have evolved

immune evasion and resistance to various drugs. In the field of cancer, accumulations of mutations in oncogenes and tumor suppressor genes that cause cells to outgrow their neighbours in a tissue are reminiscent of beneficial mutations that spread through a population of single cells (Boland, 2005). Therefore information gained by studies of evolution can help us understand the roots of genetic diseases and provide clues about how to treat them.

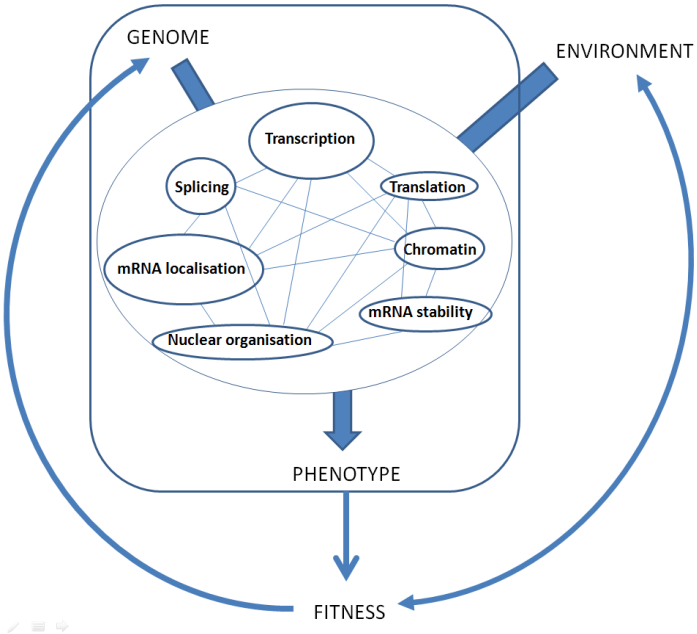


Figure 1.1. Summary of relationships between biological and ecological components. New mutations are introduced into the genome and can affect the phenotype depending on intracellular processes and environmental conditions. The resulting phenotype can in turn affect the fitness which will act upon population dynamics and change the frequency of the genotype.

To approach this general problem of integrating factors that act upon selection, more and more groups are running laboratory evolution experiments to study how genomes and phenotypic properties evolve over hundreds or thousands of generations (Elena and Lenski 2003; Kawecki et al. 2012). Evolutionary theories have often been tested with comparative approaches as in phylogeny and divergence between species or

populations. However, since past evolution occurred in unknown fluctuating conditions in which many neutral and deleterious mutations accumulated through drift or by hitchhiking with beneficial alleles, such approaches fall short of establishing cause and effect relationships between genotypes and phenotypes. Moreover, one is often limited to obtain sufficient samples to measure the rates of mutation for phylogenetic analysis and to address the repeatability of evolution.

In experimental evolution experiments, one typically propagates populations founded from a single genotype in a single nutrient medium in sterile flasks for hundreds or thousands of generations. Specific environmental conditions can be imposed by the experimenter and evolutionary processes can be studied in real time. By maintaining multiple replicate populations in each environment, the effects of a specific treatment and the identification of adaptive mutations can be systematically investigated in a manner that is often inaccessible with comparative approaches. Experimental evolution gives the power of reductionist biology that allows genetic analysis and physiological assays but within the context of long-term adaptations. The environment is artificial but what happens inside the flask through subsequent transfers should mimic natural selection and population dynamics and therefore provide an understanding of the broad patterns and principles of microbial evolution.

A long-term evolution experiment with *E. coli*

Consider a population placed in a particular environment. To what extent can we predict how it will respond to changing environments? Do we know if it will adapt? Can one predict which beneficial mutations will occur, how fast they will take place, and in what order they will likely appear throughout the genome? Or is adaptation unpredictable and controlled by chance? Understanding this bigger picture of biology would involve not only knowing how the organisms function but being able to address these questions. Sometimes, this seems straightforward as in the case of bacteria put under antibiotics selection. Here, commonly occurring mutations often arise in the targets of antibiotics. A point mutation in the *rpo*B gene can sufficiently alter the structure of rifampicin's target, the ß-subunit of the RNA polymerase, so that the antibiotic loses its inhibitory effect on bacterial transcription (Enright et al. 1998; Wang et al. 2001). But in many cases one sees many different alleles arising which all seem to contribute a small amount, and they don't necessarily occur in genes that are known to be related to the phenotype.

Let's consider a simple experimental set up like the one used in one of the most prominent evolution experiments to date. Richard Lenski and his team propagated 12 lines of *Escherichia coli* from the same ancestor over 40,000 generations in an experiment that took about 20 years to complete (Barrick et al. 2009). The basic experimental set up is a daily transfer diluting a fraction of each population 100-fold into minimal medium containing inorganic salts and glucose. Growing under these conditions at 37°C generated 6.6 divisions per day initially. It is worth noting that such environment is novel to these cells and adaptation in response to these particular

conditions is characterized by extreme fluctuations in the availability of the energy source. If one follows the cells through time, three different growth situations are seen that are standard to microbial growth (Fig. 1.2).

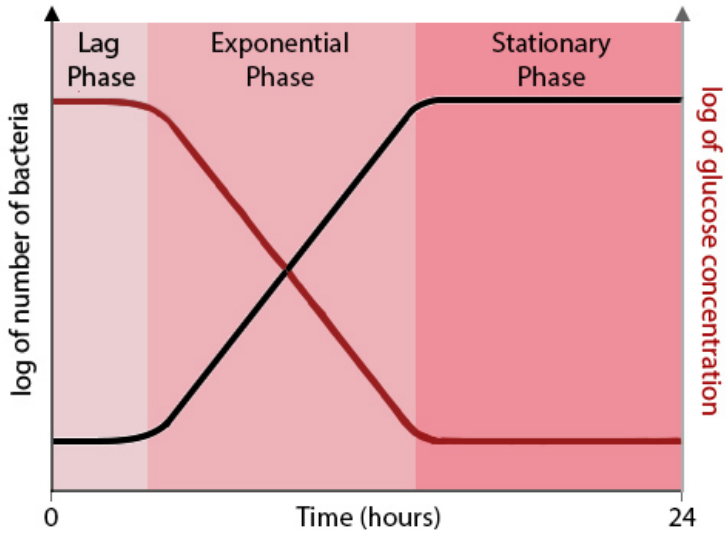


Figure 1.2. Theoretical growth curve of bacteria and substrate removal after dilution in fresh medium in which glucose is the only carbon source.

There is a lag phase in which cells prepare themselves to divide, an exponential growth phase in which doubling of the cells appears at a constant rate, and a stationary phase in which growth is inhibited following depletion of glucose (Tempest 1978).

Selection can act upon each one of these growth phases and differences in the lengths of each one of these phases can greatly affect which mutation will be selected for. For instance, a longer stationary phase would result in more cell mortality, and one would therefore select for mutants that survive the stationary phase. If the lag phase is long, the 'winning' mutants may be the ones with a shorter lag time. One could hypothesize that mutants with a shorter lag time and slower growth would be selected over mutants with a long lag phase but faster growth as the former could deplete the substrate by the time the fast growers come out of growth arrest. This illustrates how even a simple set

up makes it difficult to make predictions about adaptation and it is clear that small alterations in environmental conditions like temperature can affect the phenotypic outcome (Bennett and Lenski 1996).

One could try to limit selection to just the exponential phase by using a chemostat which is a well stirred tank of growth medium in which cells grow continually as fresh medium is added at a constant rate while old medium is simultaneously removed to keep a constant volume. However, it is hard to maintain a sterile system in a chemostat as well as maintaining a constant dilution. When cells adapt and increase their growth rate, one needs to change the time between dilutions or change the degree of dilutions while keeping the time intervals the same. Therefore running chemostats can be time consuming and impractical.

An important advantage of microbial evolution is that cells can be frozen in glycerol and revived later. By freezing populations at specific time intervals, not only is one able to capture a complete record of all mutational events in the genome of a single evolving lineage, but one can compare the fitness of different generations for example by competing the progeny against the ancestor. In Richard Lenski's long term experiment (Barrick et al. 2009) this was done by growing equal amounts of evolved and ancestral cells (distinguished by a neutral marker) in a common culture under conditions identical to the ones of the evolution experiment. By plating samples at different time points and comparing the number of colonies from each strain, his team noticed that selection had happened at the level of the lag phase and maximal growth rate. Different combinations of changes in these phases might produce similar fitness gains (Cooper and Lenski 2010). Since the stationary phase was relatively short and no cell death was expected, selection did not act upon survival during stationary phase.

Fitness was defined as the relative increase in net growth rates (encompassing all three growth phases) of evolved versus ancestral strain. They found that mean fitness increased over time with ~70% increase in the net growth rate after 20,000 generations (Barrick et al. 2009).

Genomic evolution

In a stroke of luck for the long-term evolution experiment with *E. coli*, new sequencing technologies became available which allowed for the rapid sequencing of whole genomes from the ancestral strain and bacteria sampled at generation 2000, 5000, 10,000, 15,000, 20,000 and 40,000 (Barrick et al. 2009). After 20,000 generations, in all 12 lines taken together, 45 targets of mutations were identified, comprising 29 single-nucleotide polymorphisms (SNPs), 15 indels (including transposable elements), and 1 inversion (Fig. 1.3).

These 45 events accumulated at a fairly constant rate, which is a signature for neutral evolution (Kimura 1983). However this would include synonymous mutations that have no effect on protein structure, whereas all 26 mutations that were found in coding regions were non-synonymous. Moreover, 9 of these mutations were individually introduced into the ancestor and the majority appears to be beneficial. A lot of events were intergenic which was surprising considering the vast majority of the *E. coli* genome is coding. Unfortunately, out of the 9 mutations tested, only 2 were intergenic SNPs and one of them did not confer fitness advantage. So it remains unclear if these intergenic mutations are mostly advantageous.

Among the 12 replicates, they observed a high degree of parallelism with 3 mutations occurring in the same gene in 11 populations. These mutations were never identical

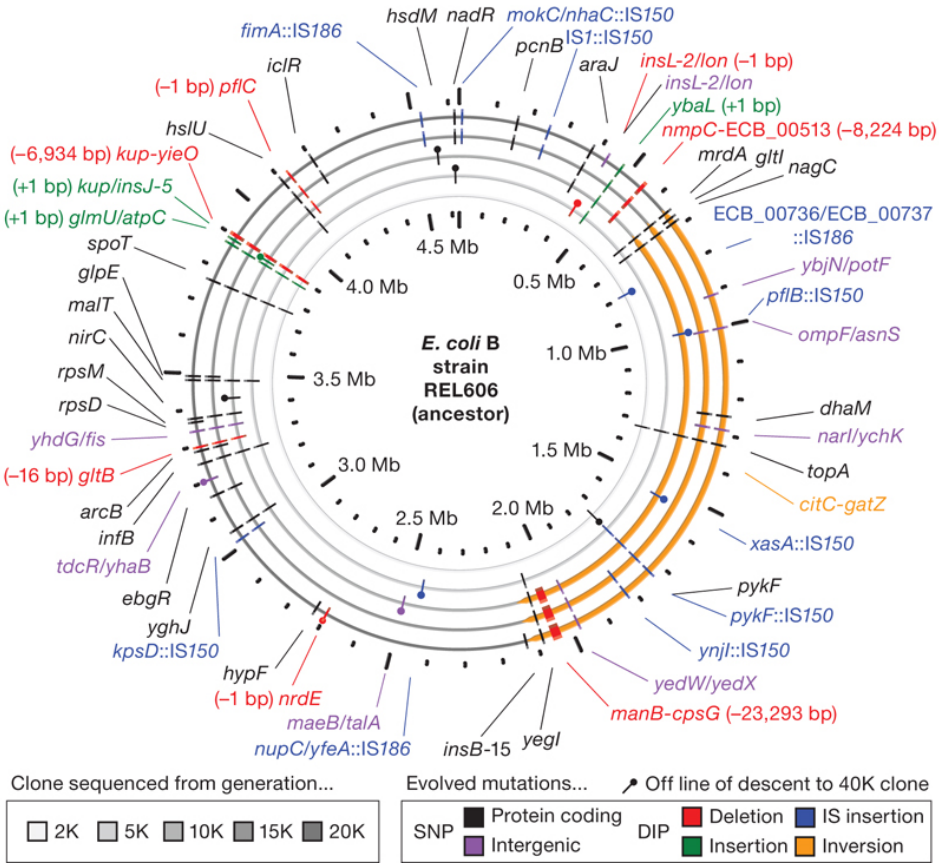


Figure 1.3. Mutations in evolved genomes between 2000 and 20,000 generations in an evolution experiment with *E. coli*. The circles correspond to different time points with the outermost one representing generation 20,000. The labels indicate about half of all mutational events are SNPs while indels, transposable elements and one inversion constitute the other half. This figure is taken from Barrick et al. (2009).

in terms of chromosome position, but they appeared in the same genes or intergenic regions.

What was striking in this experiment was that while mutations accumulated in an almost linear fashion, the net fitness improvements appear non-linear. Here, there is a rapid increase in fitness during the first 2000 generations, followed by a much slower increase the rest of the time. After generation 20,000, a SNP hit a mutation repair gene in several populations causing a mutator phenotype and the number of variants rose from 45 to over 600 events, a lot of which were synonymous with no impact on fitness.

Emergence of a novel trait

This long-term evolution experiment provided an opportunity to address how novel traits emerge (Blount et al. 2012). Glucose was not the only carbon source in the medium. Citrate had been added as a chelating agent to facilitate iron uptake. In the well-aerated conditions of the experiment, *E. coli* cannot digest citrate, but at around 31,000 generations one population found a way to successfully exploit it. How did this new trait emerge? First, they show that the mutations that had accumulated over 30,000 generations potentiated the emergence of this trait, possibly through epistatic interaction. By comparing sequencing reads to the ancestral genome, they identified chromosomal rearrangement joining fragments of an unrelated gene, rnk, to an existing gene that is used in citrate digestion in anaerobic conditions. As a result, transcription of citG and an adjacent gene citT (which encodes a citrate transporter) came under the control of the rnk promoter, allowing for their expression in oxic conditions. However, on its own, the expression of the rnk-citT module was too weak for bacteria to eat citrate, and thus they characterized a third 'refining' phase in which the rnk-citT copy number was increased through tandem duplications, increasing its expression. Cells containing duplications of the module were selected and formed a strain that could transition from glucose to citrate uptake. The ability to digest citrate had emerged from rewiring connections in the genome in a way that potentiated the redistribution of genetic elements which were then refined to be effective. Arguably, while the authors suggest this multi-step origin may be typical of many new functions, this study falls short of explaining the true source of new traits, as horizontal transfer redistributes functions that are already present in the genome.

How much of this could have been predicted beforehand? The fact that only 1 out of 12

populations evolved the ability to grow on citrate suggests adaptation can be unpredictable and controlled by chance. On the other hand, the high degree of parallelism in what genes were mutated during the first 20,000 generations indicates they were restricted to systems-level constraints under the given experimental conditions. Surprisingly, based on their known function, most of the 45 genes mutated in the first 20,000 generations appear unrelated to the phenotype. One exception is the mutation in DNA topoisomerase I which could affect expression levels across the genome. Such a genome wide change could be fine tuned to be beneficial and produce one of the big initial improvements in fitness. Also not surprising is the mutation in pyruvate kinase which when knocked out causes buildup of phosphoenolpyruvic acid (PEP) which is used in bacteria's phosphotransferase system to bring glucose into the cell. Interestingly, bacteria use mostly a different system in the gut where they need an external co-factor to convert glucose into gluconate, which is catabolised to pyruvate via the Entner-Doudoroff pathway (Peekhaus and Conway 1998).

Repeatability of adaptation

Recently, researchers addressed more rigorously if adaptation follows the same route for identical populations in the same environment, or if it proceeds along alternative means to achieve the same phenotypic outcome. Using a similar experimental procedure as in the long-term evolution experiment, researchers characterized the genetic response of 115 populations of *E. coli* to high temperature (42.2°C) over 2000 generations (Tenaillon et al. 2012). They noticed an increase in fitness in all lines with a mean of 1.42 when competed with the ancestor. Sequencing one genome from each population, they first identified a total of 1258 mutations, affecting more than 600

different sites with few mutations shared among replicates. However, when they analyzed the data further, a strong pattern of convergence emerged at the level of operons and functional complexes. The rich variety of individual mutations seemed to target two major pathways of high temperature adaptation, one centering around the RNA polymerase, another around the termination factor *rho*, and they tended to be mutually exclusive.

To address the same question in Eukaryotes, DNA microarray-based techniques were used to detect mutations in 24 cultures of budding yeast adapting over 200 generations in three different nutrient-limited chemostats (Gresham et al. 2008). Here they used global gene expression profiles to measure similarities among phenotypes. They found that evolved strains looked remarkably similar when adapted to sulfate limitation but a lot more diverse when adapted to glucose or phosphate limitation. Analysis of the genome of adapted clones revealed amplifications of sulfur transporter *SUL1* in all clones adapted to sulfate limitation. On the other hand, mutations in individuals adapted in the other environments exhibited diverse evolutionary responses. The high degree of parallelism in the sulfate-limited populations can be explained because *SUL1* amplifications provides a very large fitness increase, whereas in the other conditions no large-effect mutations were readily available resulting in the accumulation of various mutations conferring small selective advantages. This is consistent with studies on the evolution of drug and antibiotics resistance where typically only a small number of mutations in one or several genes are involved (Enright et al. 1998; Medeiros 1997; Costanzo et al. 2011).

These observations suggest that predictability of adaptation is largely dependent on environmental conditions. When only a few mutations are implicated in adaptation,

there is a high degree of parallelism even on the nucleotide level. When many more mutations are implicated and they appear in many different loci, these are confined to evolutionary trajectories at the level of functional complexes and pathways. Which one of these trajectories will be followed is thought to depend on which mutations happen first (Gresham et al. 2008).

Predictive behaviour within genetic networks

Cross-protection

Many microorganisms can increase their stress tolerance at the first sign of a problem. When pre-exposed to a mild dose of one stress, they induce a general stress response that makes them resistant to large, normally lethal doses of the same stress, as well as other types of stresses (cross-protection) (Jamieson 1998, Moradas-Ferreira et al. 2000, Chen et al. 2003, Berry and Gasch 2008). This cross-protection is thought to be an adaptation to fluctuating environments that are essentially unpredictable.

Adaptive anticipation

On the other hand, when variations are predictable, microorganisms have been found to adapt their genetic networks to the temporal correlations between perturbations. In *E. coli*, microarray experiments have revealed a remarkable overlap between the transcriptional responses to increased temperature and decreased oxygen levels: if exposed to a single change in oxygen or temperature level, the expression profile of the other factor will follow. Considering the covariation of these parameters upon transitions between the outside world and the mammalian gastrointestinal tract, this correlation is thought to be an evolutionary adaptation: *E. coli* have 'learnt' the

temporal correlation between an up-shift in temperature and down-shift in oxygen (Tagkopoulos et al. 2008). When these wild-type strains were placed in a new environment in which a temperature increase happened in the presence of high oxygen levels (as opposed to what they encounter in nature), they first grew poorly since the temperature increase caused repression of aerobic respiratory pathways in the presence of high oxygen levels. However, in less than one hundred generations, population growth-rate increased progressively by 50%. Examining the temperature and oxygen profiles of these evolved strains revealed that the original correlation was uncoupled. While no sequencing was performed for this study, the reproducible fitness advantage in the evolved strain and global changes in gene expression suggest that the coupling and de-coupling of regulatory networks are genetically hardwired.

When environmental parameters repeatedly occur in close succession, one study suggests that microorganisms' genetic networks can become hard-wired to 'foresee' what comes next in a sequence of stimuli by launching a response to the future stress before its onset (Mitchell et al. 2009). This adaptive anticipation is reminiscent of classical conditioning, first demonstrated by Ivan Pavlov in dogs in the 1890s. He would repeatedly ring a bell before giving them food, thereby training them to salivate in response to the ringing of the bell only. The dogs had learnt that the sound of the bells predicts the occurrence of food. Microorganisms can evolve a genetic conditioned response over many generations. For instance, during wine fermentation, a process done mainly by the budding yeast *Saccharomyces cerevisiae*, the environmental parameters always appear in the same order. As fermentation progresses, the yeast's sugar and acidity levels change, the environment heats up, alcohol levels rise, and this is followed by oxidative stress. This chronology of events appears to be written as an

asymmetrical anticipatory regulation in this yeast (Mitchell et al. 2009). When the cells 'feel' the heat, they begin activating genes for dealing with the stress conditions of the next stages, but not vice versa (when exposed to oxidative stress they do not activate the genes that protect them against heat). This indicates that the response is not generalized, but one that is precisely geared to a series of anticipated events. This may be a widespread means of evolutionary adaptation that enhances survival in many organisms – one that may also be related to circadian rhythms in higher organisms (Bell-Pedersen et al. 2005; Lopez-Maury et al. 2008).

Recently, using *S. cerevisiae* gene-deletion libraries, it was discovered that the tolerance to oxidative stress requires different genes when elicited by a mild dose of heat or osmotic stress (Berry et al. 2011). Thus, depending on context, tolerance to the same severe stress occurred through different mechanisms, which further supports the idea that distinct programmed stress response networks evolved in this yeast.

The rate of adaptation

Benefits of sex and genetic variation

Considering the changes humans are making in the environment, it is becoming increasingly important to understand how rapidly species are able to adapt to deal with the changes in their environment. 31,000 generations for a bacterial population to evolve the ability to use citrate may seem extremely slow, and this trait only evolved in one out of twelve replicate populations. However, the experimental set up in the bacterial evolution experiment has a few significant changes from what one would encounter in nature. First, all lines were started from a single clone. There is no genetic variation within the population and no diversity between them. This is a different

scenario from what one would encounter in the wild where there is standing genetic variation and evolution may act primarily on variants already present in the population without having to wait for a novel mutation to be introduced. Moreover, genetic variation present in the original population may be the result of selection through past environments and are more likely to be advantageous (Barrett and Schluter 2008). In the long-term *E. coli* experiment waiting time is needed for beneficial mutations to arise. On the other hand, this makes it a real evolution experiment as both the generation of allelic variation through *de novo* mutations and selection upon it occurs.

Secondly, *E. coli* is asexual and the strain used is not competent and does not allow horizontal gene transfer. Theoretically this lack of recombination and transmitting genetic material between cells greatly limits adaptation since one mutation can have a positive effect only when in the context of a second one. Moreover if a cell has a mutation that only works in combination with a second one found in a different cell and they have no way to combine the two, these individual mutations can quickly disappear from the population through drift or because another mutation that the cell acquires is a lethal one.

Finally, recombination enables the reassortment of beneficial alleles, giving rise to different lineages that will compete with each other for fixation. This competition within one population, known as clonal interference, can slow the spread of mutations and reduce the rate of fitness increase. Theoretically, sex will reduce clonal interference and increase the rate of adaptation (Muller 1932). Several studies have tested this hypothesis. One used the F plasmid to mediate recombination in *E. coli* (Cooper 2007). This seemed to have little effect in populations where mutations occurred rarely. However, in high mutation rate lines in which the DNA repair gene, mutS, was deleted,

beneficial alleles appeared to compete for fixation within hundreds of generations. In these lines, recombination significantly increased the rate of fitness improvement. When looking at the frequency of one beneficial allele at different time points up to 1000 generations, they found it reached fixation faster in the presence of recombination than in the absence of it.

That recombination leads to the fixation of beneficial alleles does not always seem to be true. One study followed heterogeneous populations of fruit flies through 600 generations (Burke et al. 2010). Each population had about 2000 flies and the researchers picked the flies that hatched earliest to be the parents of the next generation, thereby selecting for accelerated development. By the end of the experiment, the time to hatching in those flies had become 20 % shorter than in control populations that had not undergone selection on their development time. Not surprisingly, when looking for differences in allele frequencies between both pools at generation 600 they found many SNPs in development related genes. Although the population that was put under selective pressure showed less heterozygosity than the control group, there was no sign of a 'classic' sweep in which newly arising advantageous alleles reach fixation. This indicates selection had favoured a set of minor polymorphisms that were segregating in the initial gene pool. While in the case of asexual populations, as in the long-term evolution experiment with bacteria, one sees new mutants arise and sweep to fixation from 0 to 100 %, thereby replacing the ancestral variant, here selection seemed to favour a set of variants already present in the pool. They rise in frequency but do not sweep to fixation and sexual reproduction did not seem to eliminate clonal interference. Unfortunately there is no detail about how fast the phenotypes emerged, and while this may reflect a true picture of what is happening in the wild, compared to the thousands

of generations one can achieve in a microbial evolution experiment, 600 generations may simply not be enough time for a mutation to fix and produce genetic homogeneity. Arguably, whether or not an advantageous mutation gets fixed in a population may depend on its frequency in the original population. Being present in multiple copies increases the chances to fixation than when appearing as a single new mutation (Hermisson and Pennings 2005).

What is the effect of sex on the rate of adaptation? To address this question, one group ran evolution experiments with budding yeast (Goddard et al. 2005). Like with fission yeast, these cells grow vegetatively when supplied with sufficient nutrients, but undergo meiotic conjugation when starved, leading to the formation of spores. They grew populations of a wild-type recombining yeast strain and an engineered asexual mutant in a chemostat. Some were exposed to benign conditions, others to harsh conditions with elevated temperature and osmolarity. At several time points, sporulation was induced by replacing the medium with a potassium acetate solution. Spores were germinated by returning them to the normal growth medium. In total, cells underwent an estimated 300 vegetative generations spread over 5-10 cycles of growth and sporulation. No changes in fitness were observed in the benign environment when competing the offspring with the ancestor. However, there was a marked increase in relative fitness in the strains growing in the harsh environment where selective pressure was high, and this increase was significantly higher in the sexual strains which increased from 0 to 0.68, whereas the asexual lines increased from 0 to 0.59 (Fig. 1.4). There have been other studies demonstrating that sexual populations adapt faster than asexual populations (Rice and Chippindale 2001; Colegrave 2002), one providing direct support that sex creates a diverse array of genotypes (Becks and Agrawal 2012).

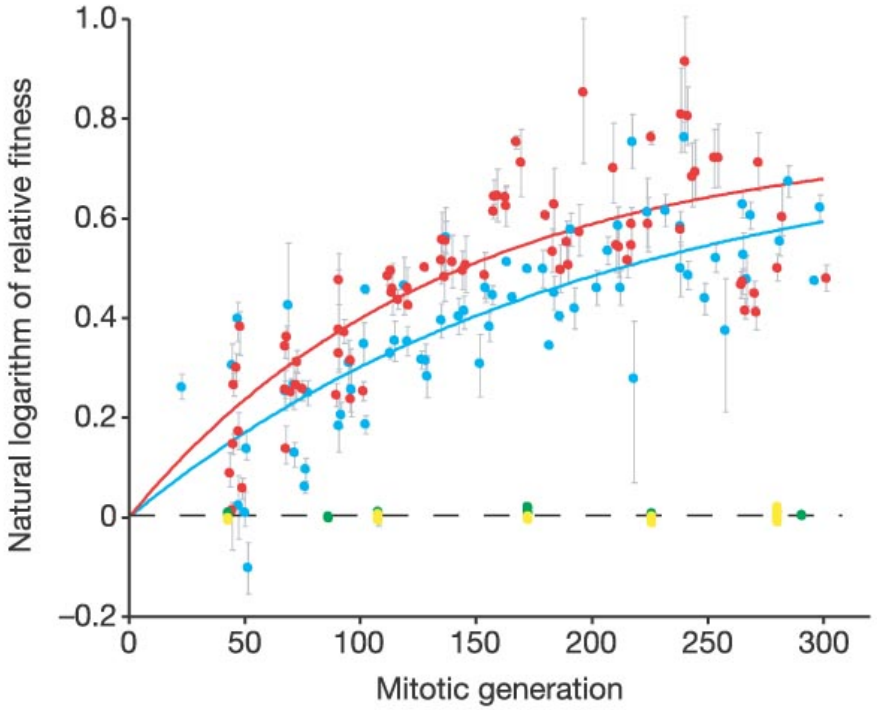


Figure 1.4. Sex increases the efficiency of adaptation in *S. cerevisiae*. Plotted changes in relative fitness for sexual strains in harsh and benign environments (red and green respectively), and asexual strains in the same regimes (blue and yellow). Fitness increases rapidly in the harsh environments and is markedly higher in the sexual strain. This figure is taken from Goddard et al. (2005).

Rapid adaptation

What is striking about the previous experiment (Goddard et al. 2005) is that the yeast populations were initiated from single clones and selection was therefore restricted to *de novo* mutations. Yet, even the fitness rate of the asexual mutant seems high, marked increases occurring within 100 generations. If we take a step back and consider the long-term evolution experiment with *E. coli* (Barrick et al. 2009), while the citrate-eating phenotype took 31,000 generations to evolve, by 400 generations all replicate lines had significantly improved fitness relative to their common ancestor (Bennett et al. 1990). When evolving the same strain at 42^{o}C, significant fitness improvements were

reported after 200 generations. Also, in the experiment mentioned above where *E. coli* grew in an oxygen-rich environment at 37 ^{o}C (Tagkopoulos et al. 2008), a rapid uncoupling of responses was observed within 100 generations. There are many other examples of rapid adaptation like the bacterial resistance to antibiotics or host defenses, or the origins of cancer (Matic et al. 1997; Cairns 1998; Andersson and Hughes 2009). When thinking in terms of beneficial point mutations outgrowing populations, such rates of adaptation may seem unrealistically high. Moreover, the rates of spontaneous mutations are estimated to be no more than 1/300 per genome per cell division (Drake et al. 1998). These frequencies are thought to be roughly constant in most DNA-based microbes regardless the size of the genome, which implies they have evolved under similar selective pressures. But microorganisms like bacteria and yeast have small genome sizes and each population in these experiments contains many millions of cells. Therefore the amount of genetic variation entering a population can become enormous within a few generations. Moreover, it has been proposed that mutation rates are elevated under stress (Bjedov et al. 2003; Matsuba et al. 2012) with a 30-fold increase in glucose-limited continuous cultures (Notley-McRobb et al. 2003). In addition to mutation rate that varies with the environment, cultures can contain mutator cells which owing to their elevated mutation rates are able to generate beneficial mutations more quickly. Competition experiments have shown that these mutators can out-compete cells with the wild-type mutation rate (Sniegowski et al. 1997; Le Chat et al. 2006; Thompson et al. 2006). Several studies revealed that these mutators only confer short-term benefits and are therefore transient in nature (Giraud et al. 2001; Herr et al. 2011), although many pathogenic strains have been found with high frequencies of mutators (Matic et al. 1997; Oliver 2000; Denamur et al. 2002).

One difficulty in understanding how spontaneous mutations can sweep populations is that most mutations are thought to be lethal, deleterious, or represent a loss-of-function. How then can large fitness effects be gained? When it comes to loss-of-function, there are cases that can explain the adaptation of new traits. For example, in an attempt to study the emergence of multicellularity, an experiment was set up in *S. cerevisiae* (Ratcliff et al. 2012). Isogenic populations were grown in flasks of rich nutrient medium. After letting the flasks shake for a day, they let the yeast settle by gravity and transferred the cells that had settled first to a fresh new flask. Since clusters of cells sink more rapidly than single cells, selection favored clustering genotypes. In a matter of weeks, most of the populations were no longer growing as single cells but as snowflake-shaped clusters of hundreds of cells stuck together. When isolating individual cells and letting them grow, they would form new clusters, the newly budded daughter cells remaining stuck to their parents. Genome analysis has not been published yet, but one could hypothesize that the initial snowflake phenotype is caused by mutations that disrupt cytokinesis, preventing each daughter cell from cleaving the way they normally would when division is complete. Such mutations would represent a loss-of-function. Indeed, several knock outs that disrupt mitotic exit have been reported to form cell clusters (Oh et al. 2012).

Explanations for rapid adaptation

Hsp90 buffers genetic variation. Several mechanisms have been proposed to explain how certain variations can rapidly cause large fitness gains. It has been observed that pharmacological inhibition of the molecular chaperone protein Hsp90 leads to an increase in variation in a wide range of morphological phenotypes in *Arabidopsis thaliana*, and similar effects have been observed in *Drosophila* and zebrafish (Queitsch et al. 2002; Yeyati et al. 2007; Bergman and Siegal 2003; Jarosz and Lindquist 2010). How does this relate to increasing the rate of adaptation? Cells need to protect themselves from influences from new mutations in order to stabilize the functions that they require. Hsp90 has been found to stabilize such deleterious conformational changes in a large number of regulatory proteins thereby buffering genotypic variability. However Hsp90 function can be impaired through mutations or environmental stress. It can then exhibit the same effect as when treated with an inhibitory drug: unleash genetic variation. Mutations in Hsp90 could therefore result in large phenotypic consequences introducing variants in the population with large fitness advantages. Conversely, in some cases, it is functional Hsp90 that appears to potentiate the evolution of phenotypes. It was shown to enable rapid resistance to azoles in *Candida albicans* and *S. cerevisiae* (Cowen and Lindquist 2005). Since the initial discovery of Hsp90 as an evolutionary capacitor, other chaperones have been reported to buffer advantageous mutations (Tokuriki and Tawfik 2009) suggesting it is a widespread mechanism for rapid adaptation.

Unstable tandem repeats that adjust gene expression. Some explanations for rapid adaptation are related to the fact mutation rates vary within the genome. Many

promoters contain tandem repeats of the microsatellite type and these repeats are prone to errors during replication resulting in changes in the number of repeats at a much higher frequency than point mutations. By comparing *S. cerevisiae* strains with related yeast species, researchers have found that variations in repeat lengths are correlated with transcriptional activity. Because such repeat-containing regions are hotspots for mutations, they could act as "tuning knobs" to rapidly adjust gene expression to environmental change. To test this, an evolution experiment was carried out with a yeast strain that contained a selectable marker (*URA3*) driven by a repeat-containing promoter. After a few rounds of selection, increased growth was observed on medium lacking uracil, and significant changes were found in the repeat length of the promoter. After measuring expression levels in engineered tandem repeat constructs, they confirmed that in the evolution experiment the length of the repeat region had shifted to one that yielded maximal expression (Vinces et al. 2009). Consistent with the notion that repeat-containing promoters facilitate adaptation of gene expression when exposed to stressful conditions, there is increasing evidence that environmental stress can dramatically increase tandem repeat instability (Cooley et al. 2010).

Gene duplications facilitate genetic adaptation. Another common type of mutation is gene duplication and this too has been highlighted as a rapid mechanism to improve fitness. It has been estimated that in non-selectively growing populations of bacteria at least 10% contain a gene duplication somewhere in the genome at any given time, thereby providing a large reservoir of standing genetic variation (Andersson and Hughes 2009). Moreover, environmental insults can increase the rate of duplication formation (Hoffmann et al. 1978). Because increase in enzyme dosage occurs at a much higher

frequency through these gene duplications than through promoter mutations, these variants are likely the first to confer a quick solution to a selective problem. But gene duplications are unstable and are rapidly lost from a population if not under continuous selection (Kugelberg et al. 2006). Therefore, it is generally believed that as the population's growth is restricted, these variants divide further until they realize a more stable, long-term mutational solution.

Fission yeast as a model organism for experimental evolution

Why fission yeast?

Many characteristics of fission yeast (*Schizosaccharomyces pombe*) make it an appealing model organism for experimental evolution. Very large populations of yeast cells are easy to maintain, both technically and economically. They divide rapidly with a generation time of 2 hours in optimal conditions, have a relatively small genome with only 4,824 protein coding genes, can be frozen in glycerol and later revived, and like *S. cerevisiae* can reproduce sexually. The *S. pombe* genome was published in 2002 (Wood et al. 2002) and consists of 3 nuclear chromosomes: Chr 1 which is 5.7Mb in length, Chr 2 is 4.6Mb and Chr 3 is 3.5Mb. In addition, there is a 10Kb mitochondrial genome. *S. pombe* shares many features with cells of higher eukaryotes and is an excellent model for human physiology and cell biology. Many cellular processes such as cell division and cell cycle regulation, mRNA splicing, RNA-binding proteins (including those in the RNA interference machinery, chromatin- and centromere-binding proteins) and the stress response pathway are similar to those of mammals. It has proved useful in the research of human disease where it has helped identifying 23 genes implicated with cancer, all of which are involved in the maintenance of genomic stability (EBI, 2012).

As mentioned earlier, cancer can be seen as an example of cells increasing their fitness, and insights in adaptive evolution could be particularly relevant, for example by blocking known routes to adaptation. Surprisingly, no experimental evolution approaches have been published for *S. pombe* to date.

Exploiting yeast genetics

There have been many experimental evolution approaches in *S. cerevisiae*, and I have mentioned a few in the course of this introduction (Goddard et al. 2005; Berry et al. 2011; Ratcliff et al. 2012). However, most of these studies did not apply genome-wide DNA sequencing technologies the way it was done in the long-term evolution study in *E. coli*, matching genotypic changes to phenotypic outcomes. Whole genome sequencing has been applied however to study the genetic variation between wild and domestic isolates, matching phenotypic variation with genome-wide phylogenetic relationships (Liti et al. 2009). One more study in *S. cerevisiae* is worth noting to highlight the power of yeast genetics. One of the problems in identifying quantitative trait loci (QTL) is that it is difficult to establish with certainty which are the causative alleles for a given phenotype since many neutral mutations can exist within the linkage regions. To side-step this problem, two groups have developed a method involving the generation of very large segregating populations by forcing strains that exhibit and do not exhibit a phenotype through many rounds of random mating and sporulation (Ehrenreich et al. 2010; Parts et al. 2011). By growing a pool of these offsprings under selective conditions, one enriches for beneficial alleles that can be identified by sequencing total DNA and comparing the allele frequencies with the ones from a control pool that are grown under nonselective conditions. This three-step QTL

mapping strategy was applied in one of these studies to identify the loci responsible for 17 chemical resistance traits (Ehrenreich et al. 2010). What was striking about their results is that the level of genetic complexity causing the different resistances was highly variable, with some traits controlled by more than 20 loci, others by only one (Ehrenreich et al. 2010).

Aim and structure of this thesis

The overarching goal of this work is to understand how yeast cells adapt to an environmental stress. In Chapter 2, the Materials and Methods section is described, providing a detailed account of the procedures that were used in this study.

Chapter 3 describes how I designed, developed and tested an evolution assay to repeatedly expose cells to defined environmental changes.

In Chapter 4, different phenotypes are outlined that have emerged after many cycles of selection. To elucidate the molecular bases by which these phenotypes had evolved, genomic DNA from ancestral and evolved populations were sequenced, and mutational events that had accumulated over the course of the evolution experiment were identified. I examined in which parts of the genome they were and if they happened in coding genes or regulatory regions.

In Chapter 5, we sought to gain insights into the dynamics of adaptation. How did the accumulation of mutations and increase in allele frequencies correlate with phenotypic changes? Were the mutations shared amongst different replicates or did different beneficial mutations contribute to the same adaptive outcomes?

Finally, in Chapter 6, I provide a general discussion, overviewing the main findings, and putting this study into context.

Chapter 2

Materials and Methods

Materials and Methods

Generation of the parental population

To generate the starting population (termed 'P' for 'Parental') for the evolution experiment, we aimed at introducing some genetic variation to increase the scope for selection. To this end, we crossed three laboratory wild-type (WT) strains of different mating types. These strains were Leupold's 968 isolate which is the 'standard' homothallic h90 strain, Heim's SA21 isolate of mating type h+ (Heim 1990), and Leupold's 972 h- isolate which was sequenced to obtain the reference genome (Wood et al. 2002). In a first round, h90 cells were crossed with each other and h- cells were crossed with h+ cells on Malt Extract Agar (MEA). Forty-eight hours later, spores were scraped from the plates, pooled together (h90, h-, h+), digested overnight in H_2O with β-Glucuronidase (SIGMA) at a 1 to 100 dilution, and germinated over 48 hours on agar plates containing rich yeast extract supplement (YES). These cells were scraped from the plates and mixed together. They underwent a second round of meiosis on mating plates, this time also allowing crosses between h90 and h-/h+ partners. After a second round of digestion and germination, the resulting culture was grown in 200 ml YES medium for 18 hours, until $OD_{600nm} = 6.0$. Aliquots were made and stored in cryotubes at $-70^{o}C$.

Large fractions of these aliquots (500 µl) were used to generate starting populations for 8 out of the 10 evolving lines used in this study. Two lines (S1 and S2) originated from a common clone from a parental aliquot that was streaked out to single colonies.

Evolution assay

Starting with the Parental population, yeast lines were propagated in 500 ml sterile flasks of liquid YES medium in an incubator set at 25°C, and shaking cultures underwent cycles of selection in which they were treated with one heat shock/cycle. To ensure cells were in exponential growth at the time of the heat shock (HS), cultures were diluted to OD_{600nm} 0.3 in 100 ml YES medium in the morning and left shaking in an incubator (INFORS) at 25°C. Five hours later, ODs were approximately 0.8, and 5 ml of each growing culture was heat-shocked by being added to 45 ml YES medium that was pre-heated at 44°C in a different flask. This large pre-heated volume was used to produce an instant HS as opposed to letting the cultures gradually heat up. Cultures were left shaking for 10 min at 44°C in a water bath (INFORS or Grant). To stop the HS, 50 ml of fresh liquid YES at 25°C was added, and the flasks were placed in a second water bath set at 25°C for 10 min to allow rapid cooling. The flasks were then transferred to an incubator at 25°C where they were left shaking overnight. This cycle (Fig. 2.1) was repeated daily, except for most week-ends. During week-ends, flasks were left on a bench at room temperature, and cells were growing in large volumes of YES (300 ml) to prevent them from entering stationary phase.

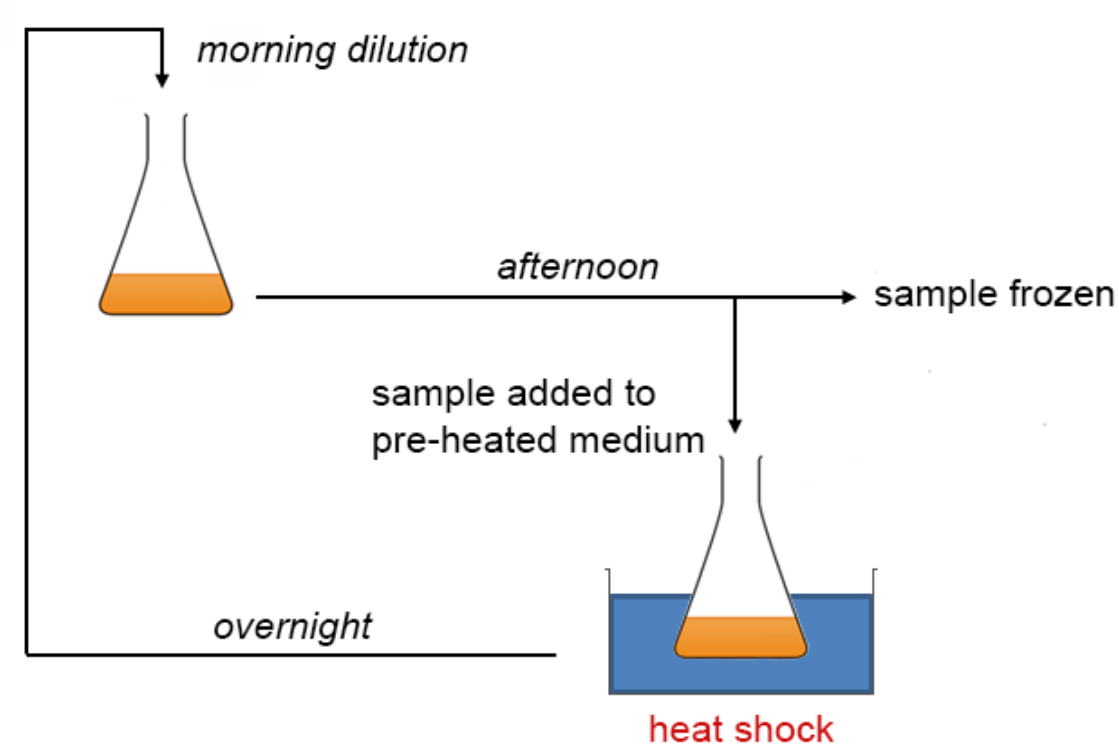


Figure 2.1. Flowchart of one cycle of selection in the main evolution experiment. In the morning, populations are diluted with new growth medium. Five hours later, they are treated with a 10 minute heat shock. A fraction is frozen at -80°C. Cells are left growing until the next morning.

This procedure was applied to 7 out of the 10 lines used in this study (E1, E2, E3, M1, M2, S1, and S2). For two other lines, M0 and E0, the HS only lasted for 1 and 5 min respectively. Finally, one mock treated line (also termed 'Mock') was added as a control. It was grown and diluted in parallel with the other lines in YES medium, but was never treated with the HS. So whilst other lines were heat-shocked at 44°C, 5 ml of Mock was added to 45 ml YES medium at 25°C and left shaking for 10 minutes in a water bath at 25°C.

20 ml of the remaining untreated cells were centrifuged down and resuspended in YES containing 50% glycerol (yellow freezing mix) and stored at – 70°C in a cryotube to capture a complete timeline of genomic changes over the course of the evolution

experiment. Experimental volumes were chosen to diminish the chance for a beneficial mutation to be lost. 5 ml fractions from the cultures in the dilution steps were estimated to contain not less than 80 million cells.

Determination of cell viability in response to HS treatment

High-resolution viability measurement (plate count assay)

Several methods were used for enumeration of viable cells after treatment with a heat shock. This assay is based on the accurate traditional method in which cells are diluted and plated onto YES agar plates, and viability is calculated based on the number of colonies arising (colony forming units; CFUs) on the nutrient agar. Cell lines were revived by scraping off approx. 200 µl from the frozen stocks and incubated on YES agar plates at 25°C over 3 days. Cells were transferred to liquid YES medium and incubated for 5 hours after which they were diluted to OD_{600nm} 0.01 and grown overnight at 25°C in YES medium to reach mid log phase the next morning. When concentrations reached an OD_{600nm} 0.5-0.6, cells were treated with a HS by adding 5 ml of the culture to 45 ml fresh YES medium pre-heated at 44°C and placed in a water bath shaker (INFORS or Grant) at 44°C for 10, 20, 40, or 60 min depending on the experiment. For the control plates, non heat-shocked cultures were diluted 200x, and 100 µl was plated on YES agar. This results in 100-300 CFUs/plate. These controls were presumed 100% viable. After the HS, aliquots of the heat-shocked cultures were diluted in eppendorf tubes and 100 µl was plated. We aimed to have roughly 100-300 CFUs/plate. For samples that showed low viability, the experiment was repeated, and dilutions were adjusted to obtain an adequate number of colonies. Plates were incubated at 25°C for 3-5 days. Colonies were counted manually with the help of a colony counter (Stuart) or

automatically using VisionWorks software (UVP). For this automated count, pictures were taken of each plate using the VisionWorks image capture function. The pictures were then processed one by one using the VisionWorks colony count function. The region of interest was defined using the Ellipse tool provided, the contrast was adjusted according to the preview window and the counting was started. Finally, errors in the outlined colonies were corrected using add, delete, and split colonies functions, and the results window was displayed to read the CFU count. For each sample, triplicate plates were made and the CFU counts were averaged. The percentage of viability was calculated by dividing the average CFUs appearing after HS by the number of average CFUs appearing on the control plate, ie. after no HS treatment. For biological repeats, the percentage of viability was averaged (Fig. 2.2). In the rest of this thesis, such data will only be displayed as in Fig. 2b or 2c.

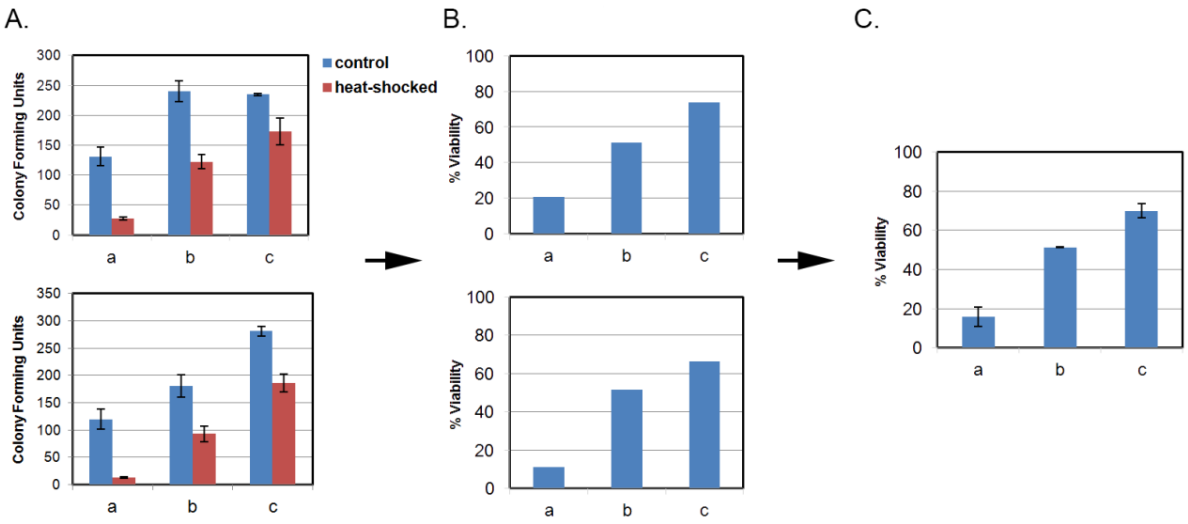


Figure 2.2. Steps describing how the average percentage of viability is obtained. (A) Two representative biological repeats are shown. Each bar represents the average CFU counts from 3 plates for samples a, b, and c. Blue bars are the average counts for control plates. Red bars represent the counts for heat-shocked cultures. (B) All values were normalized to the level of viability in the control plates (=100%). (C) % of viability averaged over the biological replicates. Error bars are the standard error of each average.

Low-resolution viability measurement (spotting assay)

This is a qualitative assay used to identify viability differences between populations. Frozen stocks were revived and grown in YES to mid-log phase as described before. Heat shock treatment was also done as described, except that after the HS, both control and heat-shocked populations were concentrated 10x through centrifugation (3 min at 2000 rpm) and then resuspended in 5 ml YES. Next, ten-fold serial dilutions of the indicated populations were spotted on YES agar plates using a replica plater (SIGMA). Plates were incubated for 3 days at 25^{o}C and scored visually. While quicker and using fewer resources than the plate count assay, this assay is less quantitatively rigorous and only useful to qualitatively identify large changes in stress resistance amongst populations.

Medium-resolution viability measurement (Biolector assay)

Frozen stocks were revived and grown in YES to mid-log phase as described above. Heat shock treatment was carried out in 48-well flowerplates (m2p-labs). Each sample was tested in duplicate or more where indicated. Plates were pre-heated in a plate shaker / incubator (Mikura) set at 44^{o}C. 750 µl of each sample was added to each well. 750 µl of fresh YES medium at 70^{o}C was added which immediately brought the samples to 44^{o}C, as previously determined. Plates were left incubating in the plate shaker for 40 min, were then sealed with a gas-permeable plate cover (m2p-labs) and transferred to a Biolector set at 25^{o}C. The biomass was monitored every 10 min over 72 hours. Data were transferred to an Excel spread sheet, measurements of experimental repeats were averaged and normalized by dividing each biomass value to the biomass

value measurement at t = 0. Scoring of the growth curves was done with a customized R script wrapper function based on the "grofit" package (Kahm et al. 2010) which extracts values for different features of the growth curve as seen in figure 2.3.

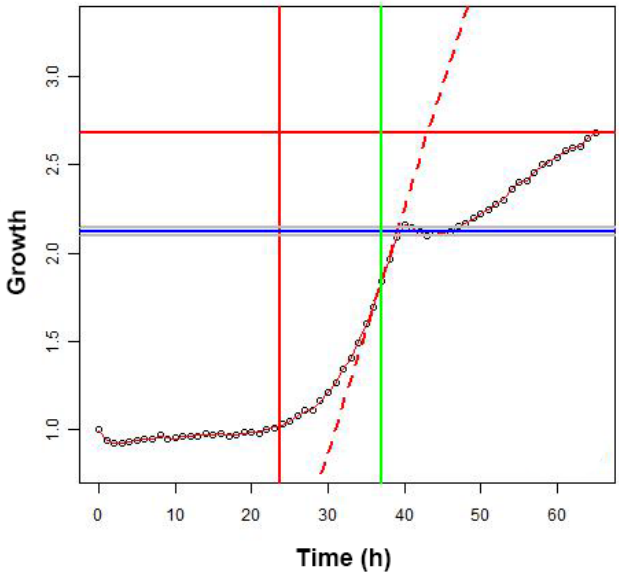


Figure 2.3. Plot for a Biolector growth curve after analysis with the R wrapper function. The colored lines show different features of the graph: maximal mass (highest point of the curve; horizontal red line), lag time (time when growth starts to be detectable; vertical red line), maximum slope (maximum growth rate; dashed line), integral (area under the curve), first, second and third best guess of relative culture density at start of stationary phase (horizontal blue and two grey lines), and finally the maximum slope time (time to maximum growth rate; green line).

Cell growth measurement

We looked at cell growth after exposure to HS as well as in the absence of HS. Cultures were grown in liquid YES medium to mid log phase as described before. HS treatment was done in flasks as in the evolution assay. Growth was measured with a Coulter Counter (Z1 Coulter Particle Counter) or a Biolector (m2p-labs). When using the Coulter Counter, cultures were growing in flasks in an incubator at 25^{o}C. At specific

time points 400 µl of culture was taken out of each flask and added to 1.6 ml formal saline fixative (0.9% saline, 3.7% formaldehyde). The next day 40 µl of culture and 20 ml of detergent (coulter isoton II diluent) were mixed into 20 ml cuvettes and the number of cells was counted using standard Coulter Counter settings. Duplicate measurements were made for each time point and the counts were averaged. When using the Biolector, 1.5 ml of heat-shocked or non-treated cells were transferred from flasks into 48-well flowerplates and transferred to a Biolector set at 25^{o}C. The biomass was monitored every 10 min, and the resulting data were transferred to an Excel spread sheet and analysed as described before.

Microarray protocol for 4x44k Agilent arrays

To study differences in transcriptional response to the 10 min HS between parental and evolved populations, custom made 4x44k Agilent Arrays were used, each containing 29957 probes covering all annotated coding genes, non-coding RNAs, as well as many introns and potential antisense transcripts (Table 2.1).

Probegroup name (as in eArray)	Total No of probes	No of unique genes	x Replicated on array	Description
controlgroup	18	18	230	spikes from holstege lab
Copy_of_CBP1_6_antisense_1253205239451	6	1	2	probes for CBP1 antisense
proteincoding-ext-annotated	5053	5053	1	protein coding from holstege lab
NonCodingRnaSpliced	804	402	1	ncRNA
top1-antisensecoding	4998	4998	1	antisense probes for protein coding
tags	8	8	2	tags
tRNAsplicedSeq	184	184	1	tRNA
tRNAantisenseSpliced	184	184	1	antisense probes for tRNA
top2-antisensecoding	3824	3824	1	antisense probes for protein coding
IntronAntisense	2016	2016	1	antisense probes for introns
noncodingRNA-external1	486	486	1	ncRNA from holstege lab
NonCodingRnaAntisense	802	401	1	antisense probes for ncRNA
IntronSeq	2016	2016	1	Intronic probes
proteincoding-anotated	9558	4779	2	protein coding

Table 2.1. Overview of the probe content for one 44k Agilent Array.

Total RNA extraction

In the microarray experiment, 100 ml of cells were grown to ODs of approximately 0.8 (as in the evolution assay), then centrifuged down (2 min at 2000 rpm) and snap frozen just before the HS (t=0), and 15 and 60 min after onset of the 10 min HS. Cells were stored at -70^{o}C. Total RNA was isolated as described previously (Lyne et al. 2003). Cells were thawed on ice and resuspended in 1 ml of pre-chilled DEPC water. After centrifugation, 750 µl of TES (10 mM Tris pH 7.5; 10 mM EDTA pH 8; 0.5% SDS) was added to the pellets and cells were resuspended with a pipette. 750 µl acidic phenol-chloroform (SIGMA) was immediately added, cells were vortexed and incubated in a 65^{o}C heat block. All mixtures were incubated in a heat block at 65^{o}C for 1 hour and vortexed for 10 s every 10 min. Next, the samples were placed on ice for 1 min, vortexed for 20 s, and centrifuged for 15 min at 14,000 rpm at 4^{o}C to separate the aqueous and organic phases. 700 µl of the aqueous phase was added to phase-lock tubes (Eppendorf/Qiagen) and extracted with 700 µl of acidic phenol-chloroform. In the same way, a third extraction was carried out with 700 µl of chloroform:isoamyl alcohol (24:1, SIGMA). 500 µl of the aqueous phase from the centrifuged samples was precipitated overnight at -20^{o}C in 3 volumes of 100% EtOH and 1/10th of 3M Na acetate (pH 5.2). Samples were pelleted (10 min, 14,000 rpm, RT), washed with ice-cold 70% EtOH, and air-dried to get rid of residual ethanol. The pellets were resuspended in 100 µl DEPC water. ODs were measured with a Nanodrop and 100 µg RNA of each sample was purified using the RNeasy mini spin columns (Qiagen) as described in the RNeasy Mini Handbook. Samples were eluted with 30 µl RNase-free water. Concentrations were determined with a Nanodrop and adjusted to 2 µg/µl.

cDNA labelling

Total RNA from evolved and control populations were converted to labelled first-strand cDNA by incorporating Alexa 647-aha-dUTP through reverse transcription. An RNA pool containing an equal fraction of each sample was used as a common reference, labeled by incorporating Alexa 555-aha-dUTP. For each experiment a dye-swap repeat was carried out. cDNA labelling was performed using the SuperScript Plus Direct cDNA Labeling System (Invitrogen) according to the manufacturer's instructions with the only modification that instead of 3 µl Alexa-dye, we used 1 µl Alexa-dye + 2 µl dNTP mix. Labelled sample and reference were mixed, purified together using the PureLink PCR Purification System (Invitrogen) and eluted with 46 µl water.

Microarray hybridizations and washes

11 µl of blocking agent and 55 µl of 2x GEx hybridization buffer (Agilent Technologies) was added, mixed, and denatured at 95^{o}C for 5 min. Agilent cover slips were placed in hybridization holders and 100 µl of each cDNA sample was placed into each chamber of the cover slip. Arrays were added onto the cover slips and the holders were tightly screwed. Hybridization happened for at least 17 hours at 65^{o}C. Washes were done in standard histochemistry slide holders. The microarray slides were washed in wash buffer 1 and 2 (Agilent Technologies) for 1 min each. Slides were slowly lifted from wash buffer 2 to prevent droplets and rinsed with acetonitrile for 10 s before a final 30 s wash in stabilization and drying solution.

Scanning and processing

Microarrays were scanned using a GenePix4000B (Axon Instruments) non-confocal laser scanner which acquires data at two wavelengths simultaneously providing real time access to a ratio image. The images were saved as TIFF files and processed with GenePix Pro 6.0. GenePix Pro 6.0 applies a thresholding algorithm to separate spots from background and places a grid over the spots, associating the integrated intensity in each spot with a probe name. We found that a composition pixel intensity (CPI) threshold of 200 was the best setting to filter out near background expression levels.

Standard normalization and data analysis

The data generated following the image processing in GenePix Pro 6.0 were further filtered and normalized with an R script developed by our group. This script applies cut-off criteria based on the standard deviation of pixel intensities in order to discard data from weak signals. Only spots that have >50% of pixels with a standard deviation (SD) >2 above median local background signal were kept for normalization. Also included were spots that have >95% of pixels with a SD >2 above local background in only one channel. The other spots were flagged "absent". The data were imported in GeneSpring (Agilent), and further normalized by applying the Lowess normalization to correct for intensity dependent effects. Finally, the 'normalize to specific samples' feature was applied. In this normalization, the signal ratios of a given sample are divided by the ones of any other sample. Since each sample was co-hybridized with the same reference pool, these divisions of 2 ratios cancelled out the reference signals. For instance, to plot the differences in mRNA abundance in evolved vs. parental populations, the signal ratios for each spot in evolved populations were divided by the

ratios of the corresponding spots in the parental background. As a result, any value >1 indicated the mRNA was enriched in the evolved population whereas a value <1 indicated a lower expression in comparison to the corresponding parental mRNA. As I will discuss in Chapter 4, a customized R script (created by Dr. Vera Pancaldi) was used to analyse these data. Lists were made of all mRNAs above and below an arbitrary cut off of 1.5. These lists were imported in Gene List Analyzer, a web-based program created by our group (http://www.bahlerlab.info/cgi-bin/GLA/GLA_input) to check for significant overlap with known functional lists or Gene Onthology (GO) terms. We also determined if existing lists are enriched in these experimental data using GeneSpring. The 'normalize to specific samples' feature was also used to create gene tree clusters using the Pearson correlation, by dividing the signal ratios for each mRNA at 15 and 60 min after onset of the HS by the ratio of the same mRNA at time point 0.

Genomic DNA extraction protocol

For all populations sequenced in this study, genomic DNA was extracted using a protocol based on that described in the Qiagen Genomic DNA Handbook 08/2001 (p. 37, 44-46). Cells from each line were revived by scraping off approx. 200 µl from the frozen stocks and incubating them on YES agar plates at 25^{o}C over 3 days. As many cells as possible were collected from the plates, transferred to rich YES medium and grown overnight at 25^{o}C to reach stationary phase the next morning. Cells were centrifuged at 3000 rpm for 5 min at 4^{o}C, supernatants were discarded and pellets were frozen at -80^{o}C. On the day of the extraction, pellets were thawed and resuspended in 10 ml lysis buffer (50 mM Citrate/ Phosphate pH 5.6 (35.5 g/l Na2HPO4.2H2O, 42.38 g/l citric acid), 40 mM EDTA pH 8.0, 1.2M Sorbitol). Glucanex (Lysing Enzymes from

Trichoderma harzianum, SIGMA) was added to 20 mg/ml and samples were incubated at 32°C for 2 hours. After centrifuging down the spheroblasts (4000 rpm, 4°C, 10 min), they were resuspended in 5 ml of Qiagen buffer G2 containing 10 µl of DNAse-free RNAse A (Wood et al. 2002) and 100 µl of Proteinase K stock solution (20 mg/ml) (Qiagen), then incubated at 50°C for 30 min. Cellular debris was pelleted (4000 rpm, 4°C, 10 min) and discarded. QIAGEN Genomic-tip 100/G were equilibrated with 4 ml of Buffer QBT. Supernatants were vortexed for 10 s and applied to the equilibrated QIAGEN Genomic-tips and allowed to enter the resin by gravity flow. Genomic-tips were washed with 15 ml of Buffer QC, and genomic DNA was eluted with 5 ml of Buffer QF. DNA was precipitated by adding 3.5 ml isopropanol. Samples were then pelleted (10,000 rpm, 15 min, 4°C), washed with ice-cold 70% EtOH, air-dried and resuspended in 0.5 ml TE (10 mM Tris pH 7.5; 10 mM EDTA pH 8). DNA was dissolved by shaking overnight at 32°C and quantified by Nanodrop. The absence of RNA was tested by running an aliquot on a 1% gel.

Library preparation

For the data of the 96 samples described in chapter 5, sequencing was done at the Sanger Institute using a single Illumina HiSeq lane (containing 96 multiplexed samples), thanks to the collaboration of Dr. Ville Mustonen.

With the help of Dr. Sendu Bala, the pipeline for calling genetic variation was also done at the Sanger Institute, and Dr. Ville Mustonen carried out data analysis where stated. For the sequencing data of the 4 populations described in chapter 4, libraries were prepared in our laboratory and the sequencing was carried out at the UCL Cancer

Institute. To generate these genomic DNA libraries for paired end Illumina sequencing we used adapter sequences and NEB and Illumina reagents as follows.

DNA shearing was carried out on a CovarisS2. For each sample, 3µg of genomic DNA was dispensed into a lo-bind 1.5 ml tube and the volume was adjusted to 120 µl using TE buffer. Each sample was transferred to a glass Covaris micro-tube and the genomic DNA was sheared at 4^{o}C using settings aimed at producing 200 bp fragments (duty cycle 10%, intensity 5, cycle/burst 200, 6 cycles of 60 s/sample). Sheared DNA was transferred into a fresh lo-bind 1.5 ml tube and concentrated using AmPure XP Beads. AmPure XP Beads were first equilibrated to room temperature for at least 30 min. The beads were vortexed to achieve a homogenous mix of beads. 180 µl of beads was added to the sheared DNA, gently vortexed and incubated at room temperature for 5 min. Beads were pelleted by placing the DNA/Bead tubes in the magnetic particle concentrator (MPC) for 5 min. Supernatants were removed and pellets were washed with 500 µl of 70% ethanol for 1 min and dried in a 37^{o}C incubator. To elute DNA, beads were resuspended in 30 µl DNase/RNase free water, mixed by pipetting, incubated for 2 min, and pelleted in the MPC for 1 min. The supernatants containing the sheared DNA were removed into a 0.2 ml PCR tube. 1 µl was ran on an Agilent Bioanalyser DNA 1000 chip according to standard procedure to test if genomic DNA was sheared successfully with fragment sizes peaking at approximately 170 bp (Fig. 2.4).

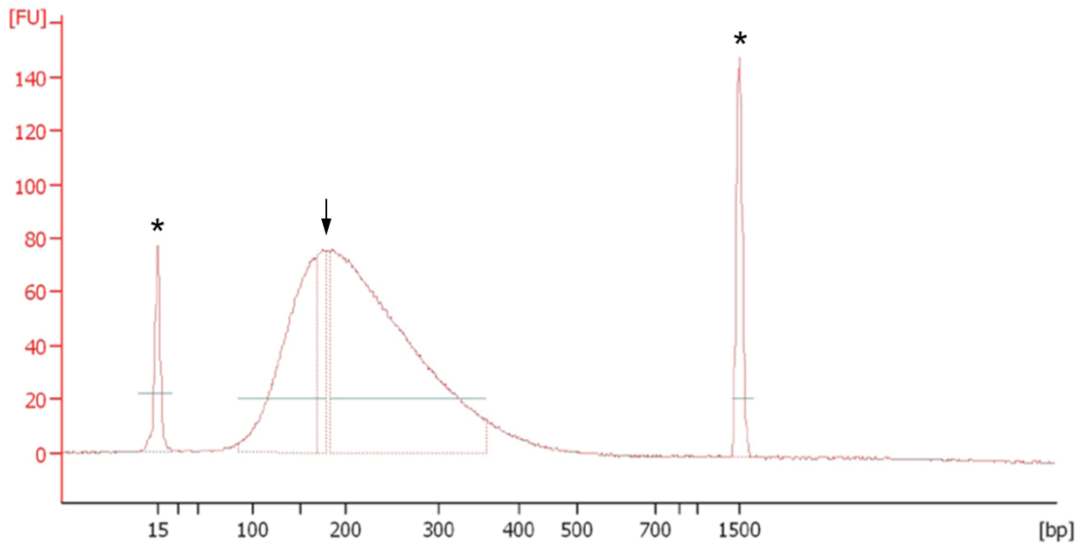


Figure 2.4. Agilent DNA 1000 Bioanalyser trace of successfully sheared genomic DNA showing a peak at approximately 170 bp (arrow). Peak values mark the height of fluorescence units (FU). The asterisks (*) indicate upper and lower marker peaks.

To repair the ends of the fragmented DNA we used the NEBNext DNA Sample Prep Reagent Set 1 which is Illumina compatible. Fragmented DNA, end repair enzymes and buffer were mixed according to the manufacturer's instructions and incubated in a thermal cycler for 30 min at 20^{o}C without a heated lid. Samples were purified using the AmPure XP beads as described above, with the only difference that the beads were resuspended in 32 µl DNase/RNase free water instead of 30 µl. The repaired DNA was transferred to a fresh 0.2 ml PCR tube.

Next, the 3' ends were adenylated using components from the NEB Reagent Set 1 in the following proportions: 32 µl DNA sample + 5 µl 10X NEB Buffer 2 + 10 µl 1 mM dATP mix + 3 µl Klenow exo-. Samples were incubated in a thermal cycler for 30 min at 37^{o}C and purified using AmPure XP Beads as before, except that 90 µl of beads was added to the repaired DNA instead of 180 µl, and the adenylated DNA was resuspended in 14 µl DNase/RNase free water before being removed into fresh 0.2 ml PCR tubes.

To ligate adaptors to dA-Tailed DNA the following mixture was made using components from the NEB Reagent Set 1 and pre-annealed PE adapters from the Multiplexing Sample Preparation Oligonucleotide Kit (Illumina): 14 µl DNA sample + 25 µl 2X DNA Ligase Buffer + 6 µl Pre-annealed PE Adapters (15µM) + 5 µl DNA Ligase. Samples were incubated in a thermal cycler for 30 min at 20°C and purified again with AmPure XP beads, this time adding 90 µl of beads to the adaptor ligated fragments and resuspending them in 30 µl DNase/RNase free water. 1 µl was run on an Agilent Bioanalyser High Sensitivity DNA chip to check that the library peak is approximately 80bp larger (the length of the adapter sequences) than before.

The library was enriched using the Multiplexing Sample Preparation Kit (Illumina) and the Phusion HF PCR Master Mix (Finnzymes). The following mixture was prepared in a 200 µl thin wall PCR tube using the appropriate PCR Index Primer for each separate sample: 3 µl DNA sample + Ultra Pure Water to increase DNA volume to 22 µl + 1 µl PCR primer InPE2.0 + 1 µl PCR primer InPE1.0 + 1 µl PCR Primer Index<#> + 25 µl Phusion DNA Polymerase. This mixture was amplified using the following PCR process: 30 s at 98°C / 18 cycles of: 10 s at 98°C - 30 s at 65°C - 30 s at 72°C / 5 min at 72°C. This product was purified following the instructions in the QIAquick PCR Purification Kit using QIAquick MinElute columns, eluting the DNA in 30 µl of QIAGEN buffer EB. Finally, the samples were purified using the AmPure XP beads as described above, adding 90 µl of beads to the adaptor ligated fragments and eluting the amplified DNA in 30 µl DNase/RNase free water. 1 µl of each sample was run on an Agilent Bioanalyser High Sensitivity DNA chip to check if the library peak is approximately 40bp larger than before (the length of the PCR extended adapter sequences). 10 µl of each sample was sequenced at the UCL Cancer Institute. It should

be noted that our group has since then optimized this protocol and replaced this last purification step with a gel purification procedure as described in the Multiplexing Sample Preparation Guide. This is to ensure a narrow range of fragment sizes.

Pipeline for calling genetic variations

To call single-nucleotide polymorphisms (SNPs) and indels from one population in the sequence data described in chapter 4, a pipeline was used to map the paired end Illumina reads with the reference genome using the programs Stampy and BWA, and process the output files using the Genome Analysis Toolkit (GATK). SNPs and indels were called using the GATK Unified Genotyper. A quality score was applied to reduce false positives. This pipeline was developed by Dr. Daniel Jeffares from our group. The pipeline for the 96 samples described in chapter 5 was also based on this but run at the Sanger Institute in collaboration with Dr. Ville Mustonen and Dr. Sendu Bala.

First a directory was made using the command mkdir and a Perl script was used to set up the files in the directory.

```
cp /mypath/Setup001.pl .
```

In this directory, links were made to the raw sequence files.

```
mypath/myreads1.fastq . ln -sf
```

The mapping and variant finding script was run.

```
./MapVariantCall005.pl --path . --ref Spombe_genome.fasta --strain name --fastq1 name1.fastq --fastq2
name2.fastq
```

The following results appeared in the specified directory:

File	What's in it
strain.snp.raw.vcf	Raw (unfiltered) SNP calls
strain.snp.filtered.vcf	Filtered SNP calls
strain.indel.raw.vcf	Raw (unfiltered) indel calls
strain.indel.filtered.vcf	Filtered indel calls
strain.realign.bam	Alignment of all reads to reference (binary)
strain.realign.bam.stats	Information on this alignment (plain text)

To verify if the filtered SNP and indel calls were real, each one of them was visually inspected using a visualization tool called Integrative Genomics Viewer (IGV, Broad Institute). IGV was installed as part of the Setup001.pl script (See http://www.broadinstitute.org/igv/UserGuide). This program was started and the alignments were loaded.

```
java -Xms10g -Xmx40g -jar ./igv.jar &
```

Next, the genome was loaded together with the alignment of mapped reads and genome annotations. One could now see all SNPs and indels in the reads aligned with the reference genome, as well as the annotation and a score for the mapping quality. Two genome viewers, Artemis and the web based program Annmaps, were used to identify in what region of the genome a SNP or indel was located (coding part of a gene, intron, UTR, intergenic, non-coding RNA or antisense transcript). In the case a SNP occurred in a coding region, Artemis was also used to determine if the substitution was synonymous or non-synonymous.

Identifying differences in allele frequencies

To compare allele frequencies for SNPs between populations and to identify significant differences, we used Popoolation2, a program specifically designed for this purpose (Kofler et al. 2011). This required the BAM files to be converted into synchronized files.

```
perl <popoolation2-path>/mpileup2sync.pl --fastq-type sanger --min-qual 20 --input p1_p2.mpileup --output p1_p2.sync
```

This synchronized file format contains the allele frequencies for every population at every base in the reference genome. Next, Fst values were calculated using a sliding window approach. We chose a coverage of 2-2000 to include a very large amount of data, including low quality regions with only 2 reads.

Perl <popoolation2-path>/fst-sliding.pl --input p1_p2.sync --output p1_p2.mcount1.mincov2.maxcov2000.win10kb-1kb.fst --suppress-noninformative --min-count 1 --min-coverage 2 --max-coverage 2000 --min-covered-fraction 0 --window-size 10000 --step-size 1000 --pool-size 500 >& log2 &

This was converted so that it could be loaded into IGV.

Perl /usr/local/popoolation2-read-only/export/pwc2igv.pl --input p1_p2.mcount1.mincov2.maxcov2000.win10kb-1kb.fst –output p1_p2.mcount1.mincov2.maxcov2000.win10kb-1kb.fst.igv &

To identify significant differences in Fst values between two populations, a Fisher test was applied.

perl <popoolation2-path>/fisher-test.pl –input p1_p2.sync –output p1_p2_coverage2-2000.sync.fet --min-count 1 --min-coverage 2 --max-coverage 2000 --suppress-noninformative &

This too was converted into IGV format.

perl <popoolation2-path>/export/pwc2igv.pl --input p1_p2_coverage2-2000.sync.fet --output p1_p2_coverage2-2000.sync.fet.igv &

Finally, these data were visually verified in IGV by loading them together with the genome and alignment of the reads as before (Fig. 2.5).

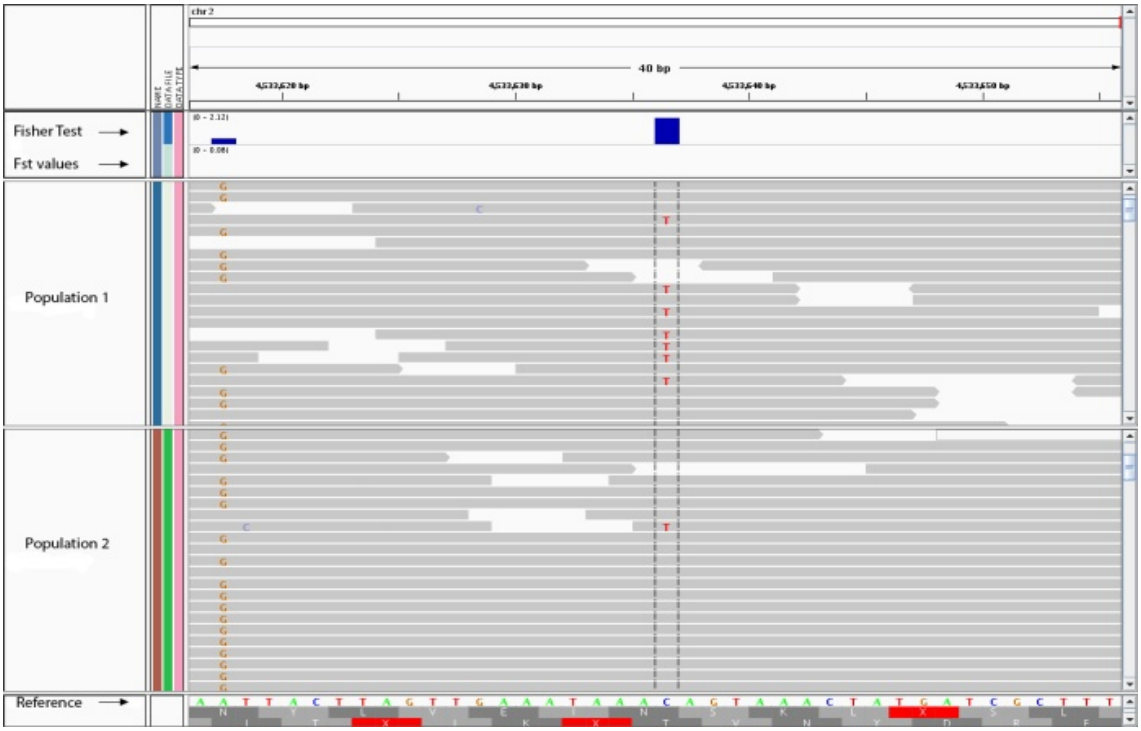


Figure 2.5. Example of a SNP that had a significant difference in allele frequency in two populations. The reference C has been replaced by a T at a higher frequency in population 1 than in population 2 as indicated by the Fisher test.

It is important to stress that this method only identifies SNPs and short indels (shorter than 20 nucleotides). It is not designed to find large deletions, insertions, transposon insertions or inversions.

Determining allele frequency trajectories

For the 96 samples described in Chapter 5, rather than comparing the difference in allele frequency between two time points, a method was applied to estimate allele counts per time point and per locus (Illingworth et al. 2012). These were modelled to result from a binomial distribution where the number of draws is defined by sequencing depth. As such, allele frequency estimations in regions with large sequencing coverage are more reliable and are given small distributions widths, whereas large estimation errors are assigned for SNPs in locations with shallow sequencing depth. Next, each set of allele counts forming a trajectory is subjected to a maximum likelihood inference given the observation and a model. In a neutral model, each trajectory is explained with a single underlying allele frequency. For a model with selection, the fitness advantage (or disadvantage) of the variant is inferred based on the allele frequency increases (or decreases) over time. Each trajectory is then given a likelihood score, ΔS, under each model associated with it. This score will be positive if the model under selection explains the observation better than the neutral model. If ΔS is close to zero, both selection and neutral models explain the observation equally well. In summary, if a trajectory is a horizontal straight line it is considered neutral. If the shape of a trajectory shows gradual increases or decreases over time, it will fit a selection model and give a positive score difference. In our analysis, trajectories with $\Delta S > 5$ were considered to be under selection, whereas those <5 were considered neutral.

Chapter 3

Designing an experimental evolution approach in fission yeast

Designing an experimental evolution approach in fission yeast

This chapter will describe how the experimental evolution approach for fission yeast was developed. This approach was inspired by previous studies around the yeast stress response as well as experiments aimed at characterizing the parental population. First, I therefore provide an overview of the stress response in fission yeast.

Introduction

My initial motivation for starting an experimental evolution approach in fission yeast was the study described in Chapter 1 on a predictive behaviour in *E. coli* (Tagkopoulos et al. 2008). They inferred that wild-type *E. coli* have 'learnt' the correlation between an up-shift in temperature and down-shift in oxygen and showed this correlation can be uncoupled by growing the bacteria in an inverted environment. I was interested in how regulatory networks evolve and wanted to create an environment for *S. pombe* that could rewire stress response networks as was shown here in *E. coli*. The ability to uncouple genetic regulatory networks that had adaptively evolved to capture temporal connections between stimuli was later confirmed in *S. cerevisiae* (Mitchell et al. 2009). One of the problems in setting up an experimental evolution approach in *S. pombe* is that little is known about their natural ecology. Therefore one cannot use information gained from their natural habitat in the same way as that for *E. coli* and *S. cerevisiae*, by creating environments in which parameters vary in an opposite pattern to the ecologically native structure. However, fission yeast has long proved useful as a model system for the analysis of cellular stress responses. Many features of the yeast stress response have been characterized (Wilkinson et al. 1996; Chen et al. 2003; Lopez-

Maury et al. 2008), and extensive laboratory and computational resources have become available for its study.

Stress responses in fission yeast

In fission yeast, a conditioned response which predicts the presence of a specific stimulus as the ones described above in *E. coli* and *S. cerevisiae* has not been shown experimentally. However, pre-exposure to moderate environmental stresses triggers a general stress response which results in a transient resistance to large, normally lethal doses of the same stress, as well as an increased resistance to other types of stress (see Chapter 1). The general stress response is mediated mainly by Sty1 which is a kinase that is phosphorylated (and thus activated) in response to a wide range of environmental signals (Berlanga et al. 2010, Chen et al. 2003, Shieh et al. 1998).

The Sty1 protein is a mitogen activated protein kinase (MAPK) which belongs to one of the best characterised eukaryotic stress-signalling systems. Components of this MAPK cascade are extensively evolutionary conserved from yeast to man, both in the degree of homology of key stress response regulators as well as in function (Toone and Jones 1998).

Eukaryotic cells respond to stress sequentially. First, there is a pre-transcriptional phase in which a cellular sensor recognizes a stress and relays its presence to the Sty1 protein kinase. Whilst still in the cytoplasm, Sty1 together with a bound target, the Srk1 protein kinase, rapidly modifies already synthesized proteins to repair damage (Smith et al. 2002). This has been experimentally shown in *S. cerevisiae* where the Sty1 kinase homologue, Hog1, controls mostly osmotic regulation. Salt stress transiently dissociates proteins from DNA, which presumably would prevent stress-induced

transcription. To combat this problem, Hog1 phosphorylates a plasma membrane ionic transporter, the Na+/H+ antiporter, extruding Na+ from the cells thereby permitting transcriptional regulatory proteins to re-associate with their genomic targets (Proft and Struhl 2004).

Next, the short term repair mechanism permits the subsequent long-term transcriptional response. Once translocated to the nucleus, Sty1 dissociates from Srk1 and starts activating transcription factors which will induce the expression of stress-related genes that encode defence proteins such as chaperones that stabilize proteins, as well as enzymes involved in the protection of cells against further stress. Atf1 is a major transcription factor directly activated through phosphorylation by Sty1, and it induces more than half of the stress-related genes. The *atf1* gene encodes a factor that is related to the bZip domain of the mammalian ATF-2, notably a transcription factor also regulated by the MAPK signalling pathway in mammals. Atf1 also targets *pyp2* which encodes a tyrosine-specific phosphatase that inactivates Sty1, thereby creating a negative feedback loop to regulate the extent of signalling (Wilkinson et al. 1996, Toone and Jones 1998, Chen et al. 2003, Lawrence et al. 2007). In parallel, the dissociated Srk1 kinase negatively regulates cell cycle progression by inhibiting Cdc25, a universally conserved protein phosphatase that promotes mitotic entry. In addition to blocking cell cycle entry through Cdc25, entire growth-related gene expression programmes are antagonistically controlled to stress-related gene expression programmes, including those involved in nucleotide and ribosome biogenesis (Lopez-Aviles et al. 2008, Lopez-Maury et al. 2008, Hartmuth and Petersen 2009). Therefore the induction of the general stress response involves a shutting down of growth related genes and leads to transient growth arrest.

Disruption of *sty1* and *atf1* results in a plethora of overlapping phenotypes, including increased sensitivity to osmotic and heat stress and inability to undergo sexual conjugation. *sty1*$^{-}$ cells also show increased sensitivity to heavy metal treatment and oxidative stress and rapidly lose viability when the stationary phase is reached (Toone and Jones 1998). One study shows that *sty1*$^{-}$ cells show little decrease in cell survival in comparison to wild-type (WT) when temperatures are raised from 30^{o}C to 48^{o}C in the first 10 min (Degols et al. 1996). However, after 40 min, all *sty1*$^{-}$ cells have died while the WT still shows about 80% survival. This indicates that cells do not require the Sty1-mediated stress response to deal with short-term stress, but they need it to survive long (severe) stress conditions where new protein synthesis is needed.

Whilst it is unknown whether fission yeast would adapt in a laboratory evolution experiment, we have seen that populations grown in an environment different from the organism's evolutionary past, adaptation is very likely to occur, and its rate increases depending on standing genetic variation, strength of selective pressures, and the presence of sexual reproduction (Barrett and Schluter 2008, Goddard et al. 2005, Yu and Etheridge 2008). Growing cells in rich nutrient medium in flasks is a new and artificial environment compared to that what they would encounter in the wild where nutrients are not always readily available and where cells grow mostly in colonies. However, there is the potential problem in using laboratory strains of wild-type yeast that we cannot know for sure how 'wild-type' the laboratory strain has remained since 1950, when originally isolated by Leupold, although it is preserved as stocks at -70ºC in laboratory freezers.

Standing variation in the starting population

In order to potentially increase the likelihood of adaptation in a laboratory experiment, rather than starting from a clonal population, we added genetic variation in the starting parental progenitor thereby increasing the scope for selection. To this end, we crossed 3 laboratory wild-type strains: Leupold's h- and h90 isolates, and Heim's h+ strain (see Chapter 2). Though at the time we did not know how much genetic variation existed between these 3 strains, in a project ran by Dr. Daniel Jeffares the genomes of Leupold's h- and h90 (also termed JB22 and JB50 in our laboratory) have since been sequenced. Comparing their genomes revealed little genetic variation (between 50 and 100 single nucleotide polymorphisms [SNPs]; personal communication from Dr. Daniel Jeffares). Moreover, about 50 SNPs were evident in the sequence of our own Leupold's h- strain after mapping it to the reference genome. This finding is surprising considering this strain was used to make the reference genome (Wood et al. 2002). Heim's h+ strain (also termed JB32) has not been sequenced, but since it was derived from Leupold's h- in the first place it is likely very closely related (Heim 1990). Still, there is some variation present in the population, although some of these SNPs will reflect sequencing errors.

Adding selection pressure

Whereas many experimental evolution approaches impose constant environmental conditions, such as in nutrient-limited chemostats, here we chose a heat shock as a source of selective pressure. Based on previous studies, we hypothesized that this would increase the rate of adaptation (Bennett et al. 1990, Goddard et al. 2005, Yu et al. 2008). Such heat shock should exert strong selective pressure, but also allow surviving

cells to recover rapidly so that if presented once daily there would be several cell divisions between 2 consecutive heat shocks. This is critical because it is during these divisions that most *de novo* mutations are likely to occur. Heat shock treatments, as opposed to other types of stress, also require minimal handling of the cells and allow for good control over dosage and time of exposure as flasks can be quickly cooled down.

Rewiring the stress response

Another reason for adding a heat shock in the evolution assay was related to our initial aim to rewire the stress response. We hypothesized that by repeatedly exposing cells to a short moderate stress, without ever following it by a severe stress, we would create an environment in which the anticipatory stress response of the cells no longer conferred a competitive advantage. On the contrary, it would be "crying wolf" when one does not need it, and mutations that result in a decreased stress response would have a competitive advantage because they will be released more quickly from the long growth arrest that comes with the general stress response. Experimentally, such adaptation should be evident in increased cell death when evolved populations are exposed to a second severe stress. In addition, it could be seen as shortening in the growth delay. Similarly to the *E. coli* experiment, in which cells had uncoupled their high temperature/low oxygen genetic profiles when brought to a high temperature/high oxygen environment (Tagkopoulos et al. 2008), we postulated that the yeast cells could uncouple the pathways involved in the short-term pre-transcriptional repair response from the long-term response involved in cross-protection. Mutants that have lost the predictive behaviour could therefore be selected and outcompete the ancestral cells.

Choosing a heat shock

To choose a suitable heat shock for the evolution assay, I used the parental population to measure cell survival and cell cycle arrest after treatment with 10 minute heat shocks ranging between 42°C and 48°C. For the heat shock (HS) to be instant, rather than bringing flasks of cells into a pre-warmed water bath and letting them warm up gradually, 5 ml of cells were added to 45 ml YES medium that was pre-warmed in a water bath at the desired temperature. To terminate the HS, medium at 25°C was added and flasks were transferred to a second water bath at 25°C (see Chapter 2). A 10 min HS treatment at 42°C caused no cell death (Fig. 3.1) and showed no detectable growth arrest (data not shown). Incubating cells at 47°C for 10 min resulted in ~70% cell death (Fig. 3.1). Interestingly, increasing the temperature by just 1 more degree (i.e., 10 min HS at 48°C) killed more than 99% of the cells (Fig. 3.1). Importantly, exposure to 44°C for 10 min caused minimal cell death (~20%; Fig. 3.1), which is ideal and can be expressed as a selection coefficient s = 0.2. This fits a strong selection model that dominates random drift (Higgs and Attwood 2005).

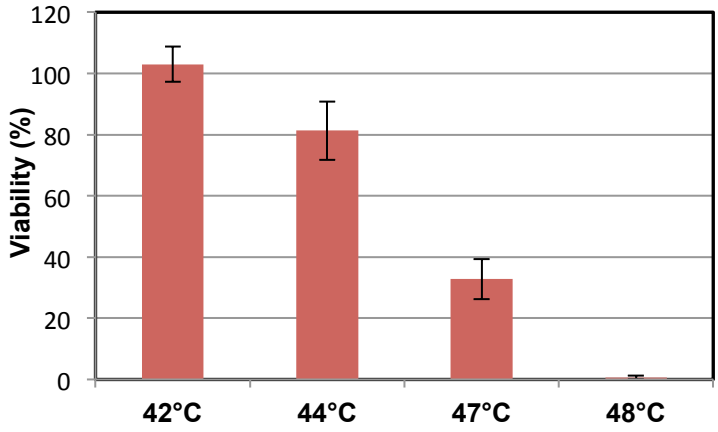


Figure 3.1. Survival of the Parental population to a 10 minute HS of different temperatures. The averages of two biological repeats are shown. The error bars indicate the standard error. A 44°C HS kills ~20% of cells.

Cell growth was measured using a Coulter Counter, following treatment with a 10 min HS at 44°C. This approach revealed a growth delay of ~4 hours (Fig. 3.2). Monitoring growth after HS treatment was later repeated with the Biolector which showed a similar duration for the growth arrest (see Chapter 4).

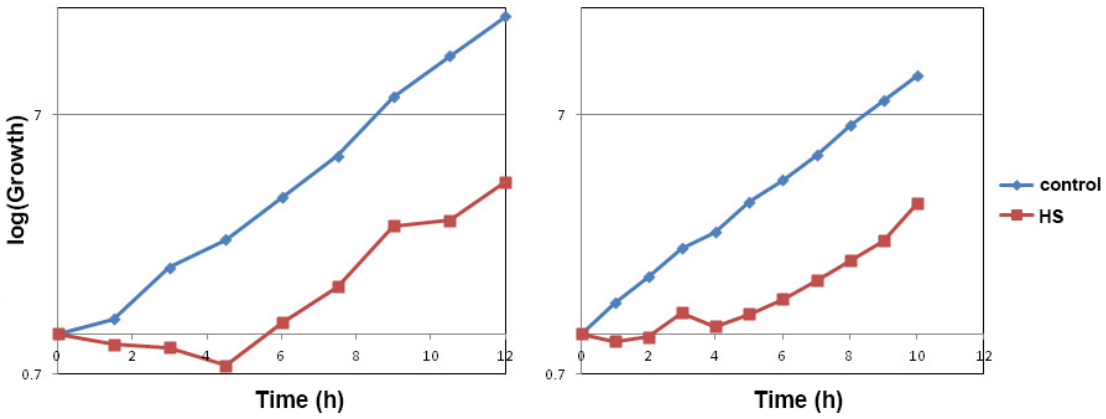


Figure 3.2. Curves of growth rates of cells over 12 hours. The control group (blue) is a population of cells that was left growing at 25°C. The HS group (red) received a heat stress of 44°C for 10 min at time = 0 and was then brought back to 25°C. Two independent biological repeats are shown: in the left graph cell numbers were measured every 1.5 hours; in the right one every hour.

I tested whether pre-exposure to this stress offers resistance against a normally lethal stress. Cells were first heat-shocked for 10 min at 44°C at t = 0. A 10 min HS of 48°C was then delivered at times 30min, 2h, 4h, 6h and 9h. Cells were collected at these time points for measurement of cell survival using the plate count assay (see Chapter 2). We found a transient stress protection that is maximal after ~4 hours of the first stress (Fig. 3.3). Interestingly, this coincides with the length of cell cycle arrest described above.

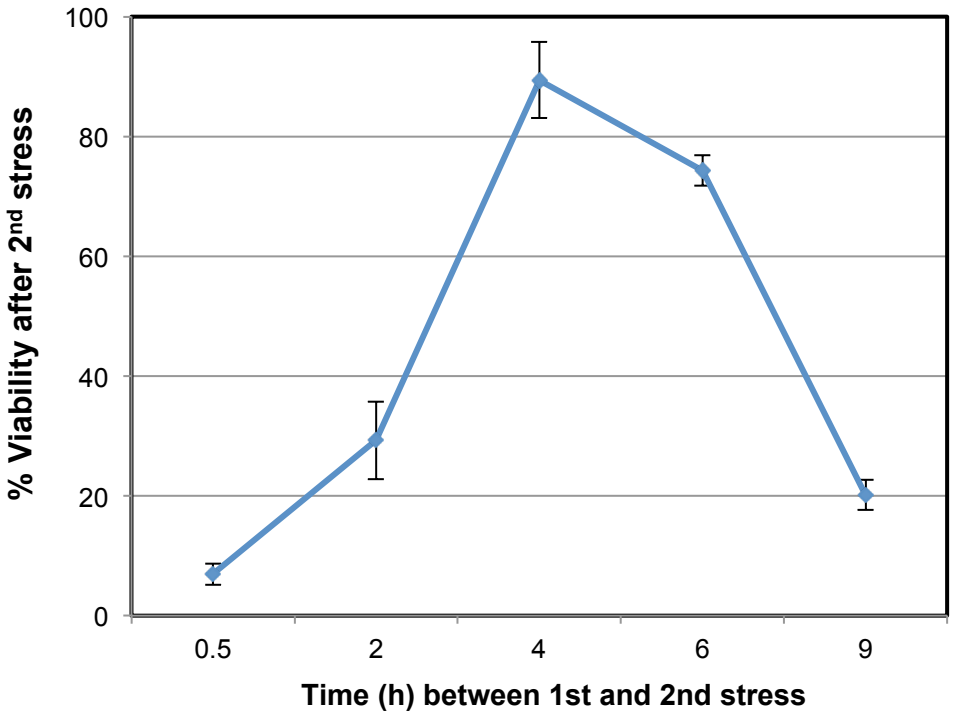


Figure 3.3. Pre-treatment with a 44°C HS provides protection against lethality from 48°C. Cells were exposed to 48°C at several time points after pre-exposure to the milder 44°C HS. The resistance against 48°C is maximal at ~4 hours, with 90% cell survival. Error bars indicate the standard error over two independent biological repeats.

The long growth arrest in addition to the conferred resistance against severe stress is indicative of the general stress response mediated by Sty1. Do the cells need activation of stress response networks to deal with the 44°C HS itself or does it mainly act as an alarm to elicit an anticipatory response for future protection of cells? To test this question, we exposed *sty1*$^{-}$, *atf1*$^{-}$, and *sty1*$^{-}$/*atf1*$^{-}$ deletion strains to 44°C for 10 min. Each of these strains showed the same level of survival as the parental population (~80%) (Fig. 3.4), indicating that the general stress response mediated by Sty1 is not required for surviving the 44°C HS. Although the experiment would need to be repeated including a WT control heat-shocked in parallel with the deletion strains.

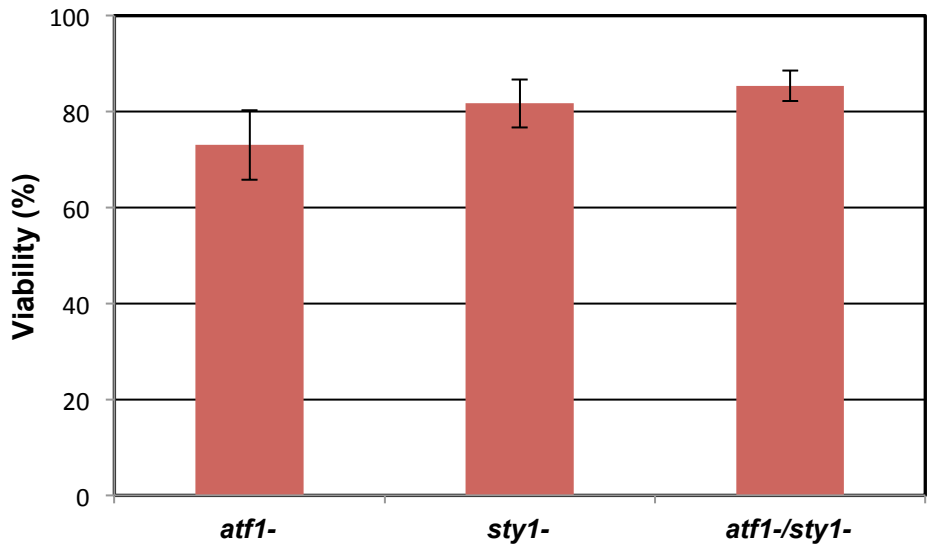


Figure 3.4. Survival of *sty1*$^-$, *atf1*$^-$, and *sty1*$^-$/*atf1*$^-$ deletion strains to a 10 min HS of 44°C. Error bars indicate the standard error based on the % viability in three biological repeats.

Starting a laboratory evolution experiment

Taken together, these data suggested that the 10 min HS at 44°C creates a selective pressure that is suitable for a laboratory evolution experiment. We designed an evolution assay in which cells grow in rich nutrient medium and are exposed to this HS at every cycle of selection (see Chapter 2). Three parallel, independent lines derived from the same parental progenitor progressed through such cycles of selection. They were termed 'E' for 'Evolved' (E1, E2, and E3). One mock treated line was run in parallel exposed to the same manipulations except that it was never treated with the 44°C HS. In total, these 4 lines underwent 150 cycles of selection. Between every round, samples were frozen and preserved at -80°C (see Fig. 2.1 in Chapter 2).

Conclusion

We have set up an experimental evolution approach in which replicated cultures of yeast progress through rounds of selection by repeated treatment with a 44^{o}C heat shock, immediately followed by ideal growth conditions. This HS causes ~20% cell death and a growth arrest of ~4 hours. It causes cells to launch a stress response that protects them for several hours against lethality from a 48^{o}C HS. This finding is in line with previous studies showing that a moderate stress can act as an alarm causing cells to launch a general stress response to brace themselves against expected further stress. In the evolution experiment, no further stress is occurring, and cells that launch the general stress response and coincident growth arrest could be at a disadvantage. Selection could therefore favour mutants that are diminishing their stress response and "learn" that the 44^{o}C HS acts as a "false alarm". However, selection could also favour cells that are HS tolerant or grow more efficiently in YES medium. Given the strong selective pressure, the different levels selection can act upon, the large population sizes, and standing variation present in the starting population, we expected that these cultures would gradually adapt and exhibit altered phenotypes after multiple rounds of selection.

Chapter 4

Global characterization of evolved populations

Global characterization of evolved populations

This chapter explores the nature of adaptation of the lines that were propagated through 150 cycles of selection. I will present various phenotypic analyses for heat shock evolved and mock treated populations. Microarray gene expression profiling was performed to identify changes at the gene expression level between these populations. I also present genomic sequencing data for two evolved cultures sampled at one time point and compare their allele frequencies with those from parental and mock treated populations.

Introduction

When we look at the evolution assay (Chapter 2, Fig. 2.1) and plot cell concentrations over the time of one round of selection, we can examine more closely the key steps where selection can act upon (Fig. 4.1). Selection could mainly occur at the level of the heat shock (HS) (phase '4' in Fig. 4.1) in which case cells that are HS tolerant would outgrow the population. Or we could be selecting for fast growers (phase 3 and 6) that exploit the nutrient medium more efficiently. As postulated before, mutations that result in a decreased stress response can also have a competitive advantage because they will recover faster to initiate exponential growth after HS which should reflect in a shorter lag phase (phase 5). There could also be improvements in recovery of growth following stationary phase (phase 2). Less likely is that selection happens at the level of the stationary phase itself, because *S. pombe* cultures show no loss of viability for 4 days in this phase (Matsuo et al. 2007). The cultures in our evolution assay are never in stationary phase for this length of time. Important to mention, and not shown in the diagram, is that during large breaks in the evolution experiments or in the case of

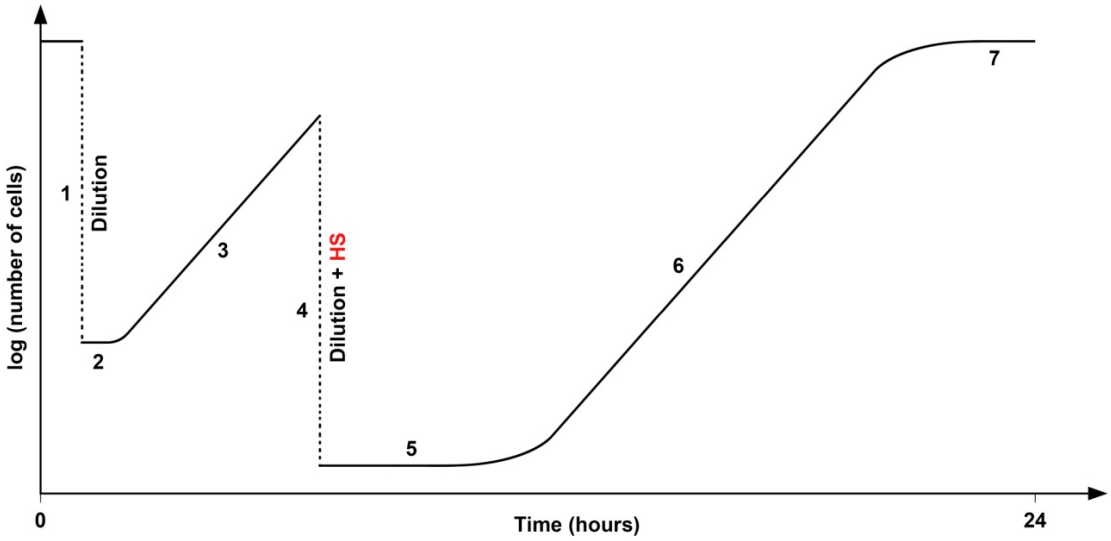


Figure 4.1. Schematic representation of how cell concentrations change over one cycle of selection. At time 0, cells are at OD ~1.0 (approximately 2 x 10^7 cells/ml), which marks the beginning of stationary phase. Cells are diluted to OD 0.3 (1). This dilution is followed by a short lag time (2) before resuming exponential growth (3). Five hours later, these exponentially growing cells are heat-shocked during which ~20% of cells die (4). With this HS, they also undergo a 20-fold dilution because 5 ml of cells is added to 45 ml pre-heated medium and 50 ml fresh nutrient medium is added to end the HS. This is followed by a growth arrest of ~4 hours (5). Cells then resume exponential growth (6), and eventually re-enter stationary phase (7).

a bacterial contamination, the last frozen sample was revived to continue the experiment. Such problems occurred 11 times over the course of the experiment. While little is known about the effects of freezing and thawing in yeast, it is likely that this caused some selection. In one study, populations of *E. coli* were propagated for 150 freeze-thaw-growth cycles. These populations showed improved survival during freezing and thawing as well as a more rapid entering into exponential growth after thawing (Sleight and Lenski 2007). On average, the fitness increased by 89% relative to the ancestral population. While they have started to uncover the genetic basis for this adaptation, the physiological mechanism for this fitness improvement remains unknown (Sleight et al. 2008). However, in our assay, we never repeatedly froze and thawed the

same cells and selection should therefore be minimal.

How many divisions have occurred over 150 rounds? At 25^{o}C, it takes ~3 hours for populations in rich medium to double. I estimate 1.5 divisions to occur during the exponential growth preceding the HS (phase 3) and 5 divisions in the exponential phase following the HS (phase 6). So the number of generations/day at the beginning of the experiment is 6.5. During most week-ends, flasks were left on a bench at room temperature which greatly decreased growth rates as opposed to growing them under constant shaking in a 25^{o}C incubator. A detailed log was kept of the times of dilution and OD values. Based on the estimations above and the information from this log, I inferred the total number of generations for E1, E2, and E3 to be approximately 1200. The mock population was not heat-shocked and therefore did not have the growth arrest following the HS. I estimate this population to have undergone a total of 1500 divisions.

Considering the OD of the populations at the time of the HS was ~0.8, which is the equivalent of 1.6 x 10^{7} cells/ml, and 5 ml was transferred, the population bottleneck is 8 x 10^{7} cells. This number should be large enough to preserve genetic variation, minimize the chance of losing beneficial alleles, and therefore allow evolutionary adaptability in the face of the changing environment.

Before the populations underwent a total of 150 cycles of selection, I did several experiments aimed at identifying phenotypic changes at multiple time points, in particular at cycle 80 and 120. The term 'cycle' is used to describe a full 24 h round of selection. In experimental evolution approaches in *E. coli,* phenotypes are typically measured through direct competition experiments with the ancestral population. Replicated lines in these experiments often originate from two ancestors that are

genetically identical with the exception that one has a mutation in the araA gene that renders it unable to metabolize the sugar L-arabinose. When these ara- cells are grown together with ara+ cells on tetrazolium arabinose plates, they form red and white colonies respectively. After growing equal amounts in a flask of ara- or ara+ evolved cells with ancestral cells of the opposite marker state and plating samples at the end of the competition experiment, the relative fitness is calculated as the ratio of the number of evolved to ancestral cells.

Since we do not have such marker strains, we cannot do this assay. However, since we hypothesized that selection will favour variants at the level of the HS or during growth phases, I first tested if there were changes in viability after treatment with the 10 min HS. To detect evolved changes in growth dynamics, our laboratory is equipped with Coulter Counter (Z1 Coulter Particle Counter) and a Biolector (m2p-labs) which can accurately monitor cell growth (see Chapter 2). Since one growth phase happens in the absence of stress (phases 2 and 3), and one happens following a HS (phases 5, 6, and 7), I monitored growth in both of these regimes.

Stress resistant phenotypes in evolved lines

Heat shock tolerant phenotype

Viability to the 10 min HS was measured for the evolved lines (E1, E2, E3) as well as the mock population at cycle 120. We termed these evolved samples E1-120, E2-120, and E3-120, while Mock-120 is the sample that was mock treated over these 120 cycles of selection. P is the ancestral population that was kept frozen at -80^{o}C. The results showed no significant differences between populations (Fig. 4.2). It is however possible that there are small differences, too subtle to be detected with this assay.

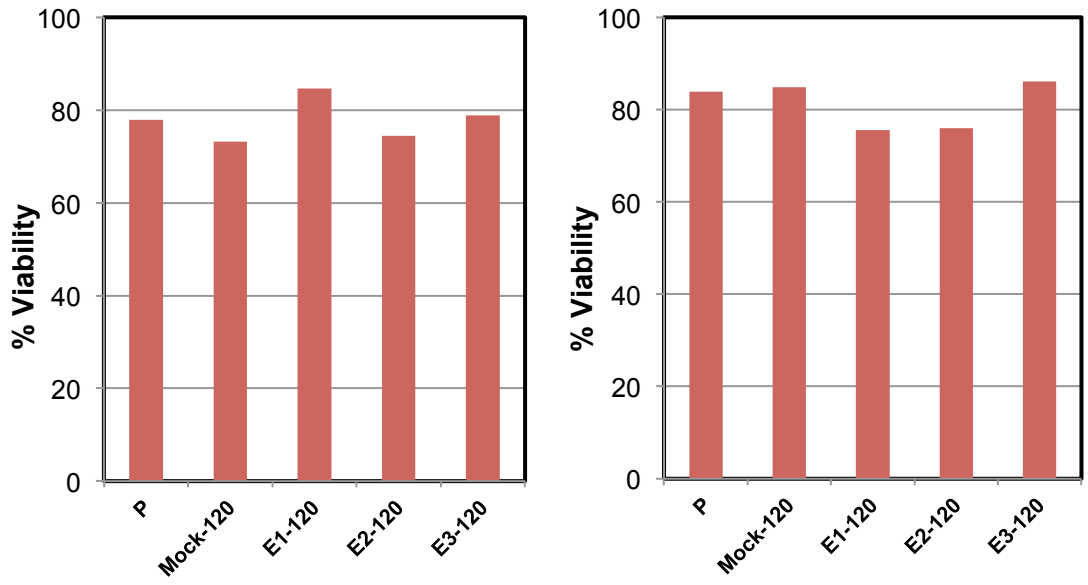


Figure 4.2. Survival of parental, mock and heat shock evolved populations at cycle 120 to a 10 min HS of 44°C. No increases in survival are detected in evolved populations compared to the parental. Two independent biological repeats are shown.

Focussing on E1, I then exposed this population to longer time intervals at 44°C. This revealed that E1 had adapted a strong HS tolerance to longer heat treatments, with the clearest difference after a 40 min exposure (Fig. 4.3).

While we had previously hypothesized that cells could treat the HS as a 'false alarm', because it was never followed by further stress and therefore lose their anticipatory stress response, these data indicate that we have instead selected for mutants that do the opposite, i.e. they are better prepared against stress that follows the 10 min HS. Because the largest difference in cell viability between the evolved and parental population was seen after 40 min, this long stress became our standard assay to measure adapted phenotypes.

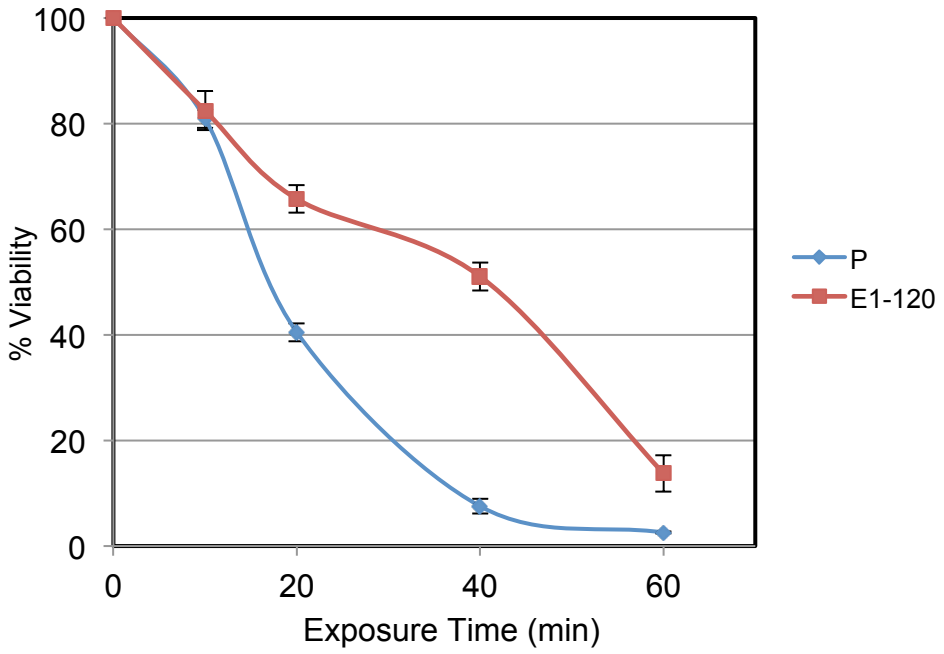


Figure 4.3. Cell viability for E1-120 and P after different times of exposure to 44°C. Cells were grown to mid-log phase in YES medium at 25°C and heat-shocked as described. Cell viability was measured using the plate count assay by plating dilutions of the cultures after 10, 20, 40, and 60 min of HS treatment. Error bars are the standard error of the average % viability in two biological repeats.

Applying this assay on all evolved lines, including Mock at 80 and 120 cycles (Fig. 4.4) revealed strong improvements in long HS tolerance for E1, E2, and E3. The effect seems weaker in E3 than in E1 and E2, which could mean that the different populations have not evolved identically at the molecular level. Less than 10 % of the parental population survived the 40-minute HS. The mock population, however, shows significantly higher survival compared to P (an average of 10% when sampled at cycle 120, while P had 5 % survival), but much less than E1, E2 and E3. Since the Mock population has never been heat-shocked, how can we explain this increase in survival to the heat shock? One possibility is that the conditions the mock line evolved to were in some ways similar to a stress condition. While this population was kept growing in rich YES medium, it also spent time in stationary phase and was subjected to passages of

freezing and thawing during large breaks in the evolution experiment (see Chapter 3). Both these conditions may exert stress on the cells. Given the Sty1 protein kinase signalling pathway that is activated in response to a wide range of environmental signals (Chen et al. 2003, Berlanga et al. 2010), we can hypothesize that evolving resistance to these different stresses may also confer resistance to HS. Another possibility is that some of the variants that have accumulated in Mock are useful to survive the HS just by chance.

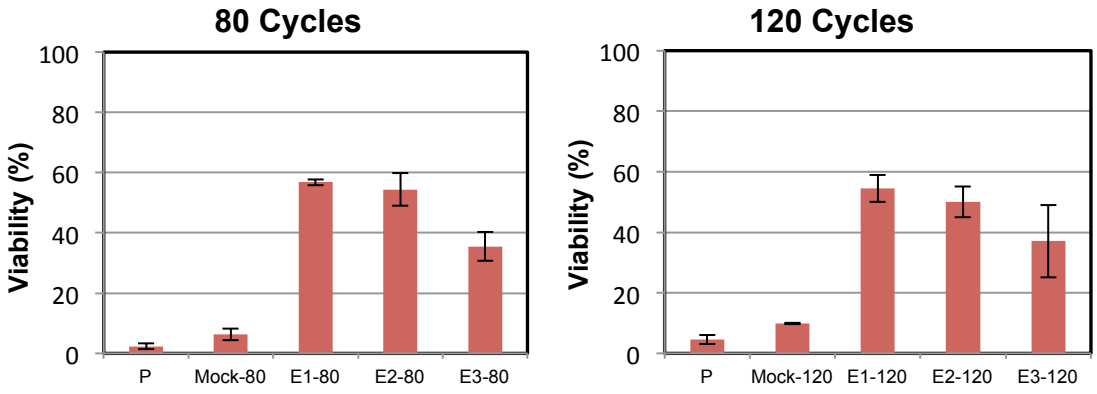


Figure 4.4. Cell viability for parental and evolved lines at 80 and 120 cycles of selection after treatment with a 40 min HS at 44°C. Viability was measured using the plate count assay (Error bars are based on the % viability in 3 biological repeats for 80 cycles and 2 for 120 cycles). Strong HS tolerance is seen in all evolved lines.

At this point, we became interested in knowing how fast this heat resistance phenotype arose in the different, independent populations. Because the plate assay to measure cell viability after HS treatment is time consuming, we reverted to the qualitative spotting assay (Chapter 2). While this is a low resolution way for looking at cell survival, if the phenotype is robust enough it would be reflected in spot densities. This approach was performed for P, Mock-120, E1-120, E2-120, and E3-120 (Fig. 4.5). The results correspond to those obtained via the plate count assay: large survival improvements in the lines evolved under the HS regime and a small improvement in Mock.

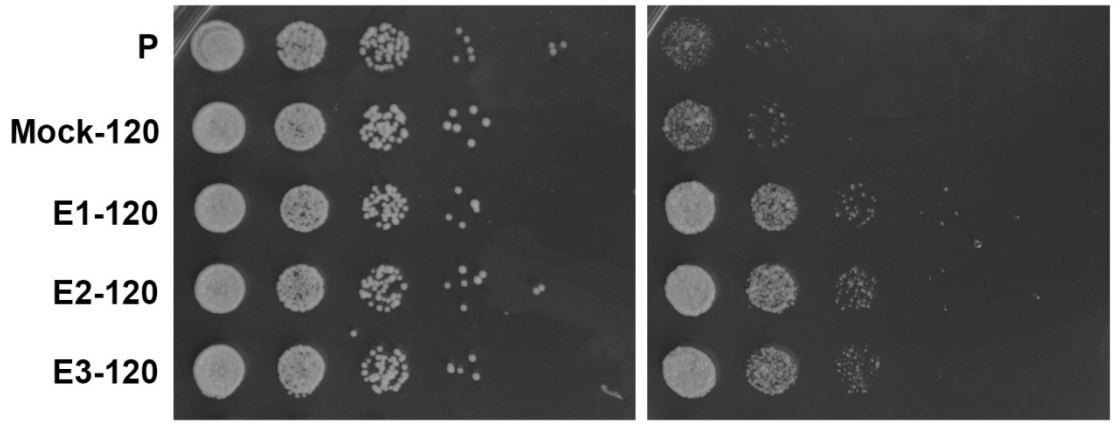


Figure 4.5. Cell viability using the spotting assay after treating parental and evolved populations at cycle 120 with the 40 min HS. The left panel is a control with non heat-shocked cultures to show the same concentration of cells was plated for each population. The right panel shows more cell survival in the evolved compared to Parental and Mock.

Focusing on the evolved line E1, this assay was done for samples taken every 10^{th} cycle over the first 120 cycles (Fig. 4.6A). These results suggest that most of the adaptation happened early, with large gains in survival between 10 and 20 cycles. When repeating this approach at higher resolution, on each cycle between 10 and 19 cycles of selection, a large adaptive step became visible between 10 and 13 cycles (Fig. 4.6B).

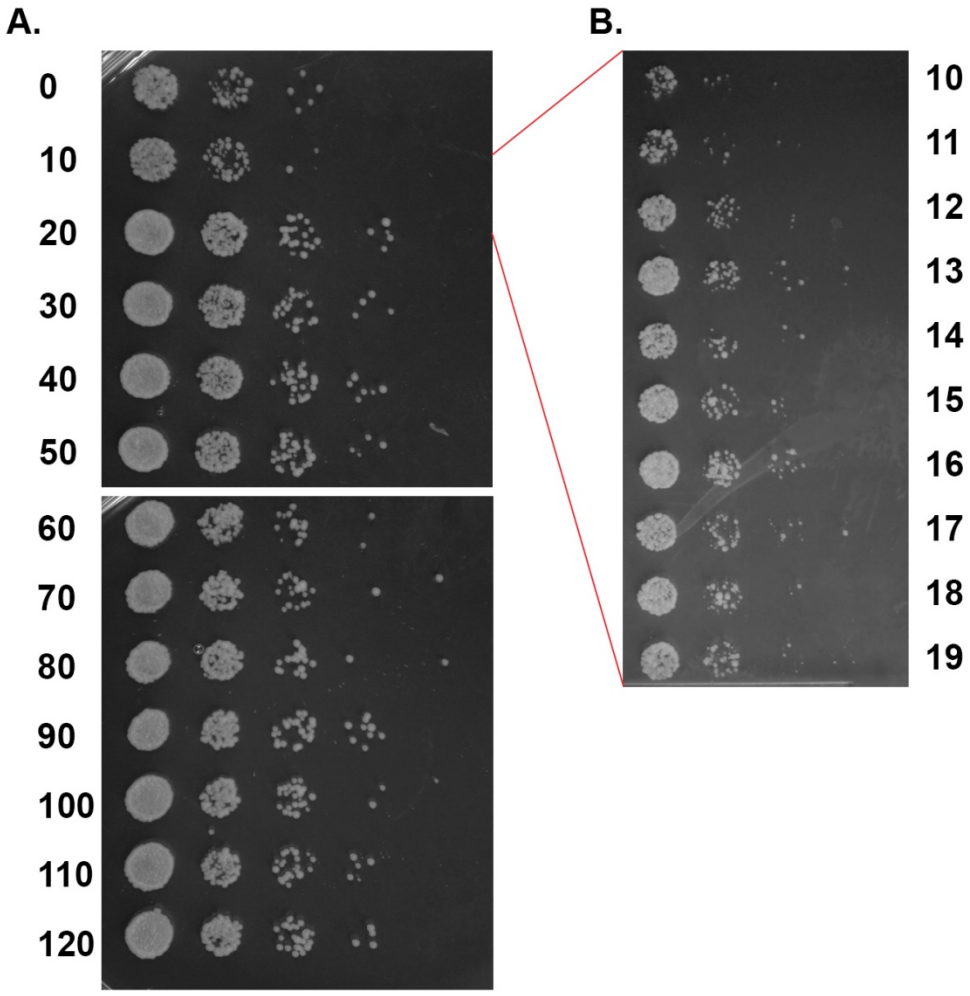


Figure 4.6. Viability of E1 populations heat-shocked for 40 min using the spotting assay. (A) E1 sampled every 10th cycle revealed that survival improvement occurred early within 20 cycles of selection. (B) Sampling every cycle from 10-19 indicates that the evolved phenotype emerged between 10 and 13 cycles. Control plates of untreated cells showed that equal concentrations were plated (not shown).

We performed this spotting assay on the same cycles (10-13) for E2 and E3, and found similar early adaptive steps (Fig. 4.7). The fact adaptation happened so early and simultaneously in all three lines may seem surprising at a first glance. However, these lines are initially replicates from the same starting population (P population) and could possibly evolve the same trajectories. Furthermore there was standing genetic variation present in this initial population, and selection can act upon pre-existing genetic variants

which is thought to increase the rate of adaptation (Hermisson and Pennings 2005, Barrett and Schluter 2008). We later repeated the evolution experiment up to 35 cycles to refine this conclusion (see Chapter 5).

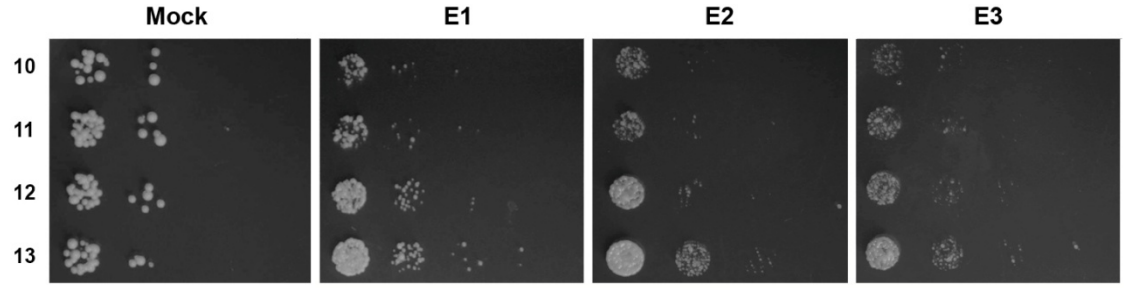


Figure 4.7. Cell viability for samples from 10-13 cycles of selection across all evolved lines. All HS evolved but not the mock treated line showed a rapid increase in survival to a 40 min HS between samples taken at cycles 10, 11, 12, and 13.

For the E1 cell line, I also used the plate count assay in order to see the magnitude of the adaptive step and examine if there were other steps not visible with the low resolution spotting assay (Fig. 4.8). This approach confirmed the rapid adaptation within 20 cycles but also showed cells continuing to adapt until 40 cycles after which an adaptive plateau seems to be reached. Possible another adaptive step happened between cycles 60 and 80; the viabilities measured at 80, 120, and 150 cycles are all higher than at cycle 60 but did not increase, or even slightly decrease, after ~80 cycles. Given these results, we decided to stop continuing the cycles of selection and focus on the molecular basis of the observed adaptation.

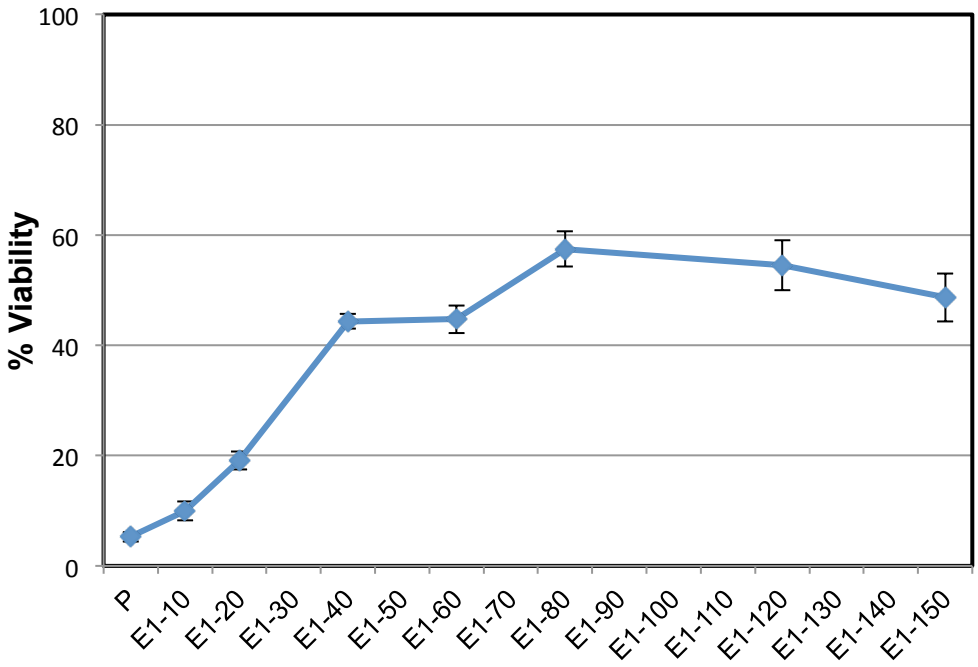


Figure 4.8. Cell viability to the 40 min HS as measured with the plate count assay at different cycles of selection for E1. A rapid, gradual adaptation is seen in the first 40 cycles with some more happening after cycle 60. Each data point is the average of 2 biological repeats. Error bars are the standard error of each average.

Oxidative stress phenotype

Do these populations also exhibit resistance to different types of stress? To test this idea, mid-exponentially growing cells from evolved, mock and parental cultures were spotted on YES plates with H_2O_2 to test for oxidative stress, and KCl to test for osmotic stress (Fig. 4.9). Resistance to high dosage of H_2O_2 was observed in the evolved populations and, much less pronounced, the mock population at cycle 80, but not to high levels of KCl. This finding suggests a coupling of the adaptation to heat stress specifically with the response to oxidative stress. Adaptation may not have changed the general stress response but rather specific pathways.

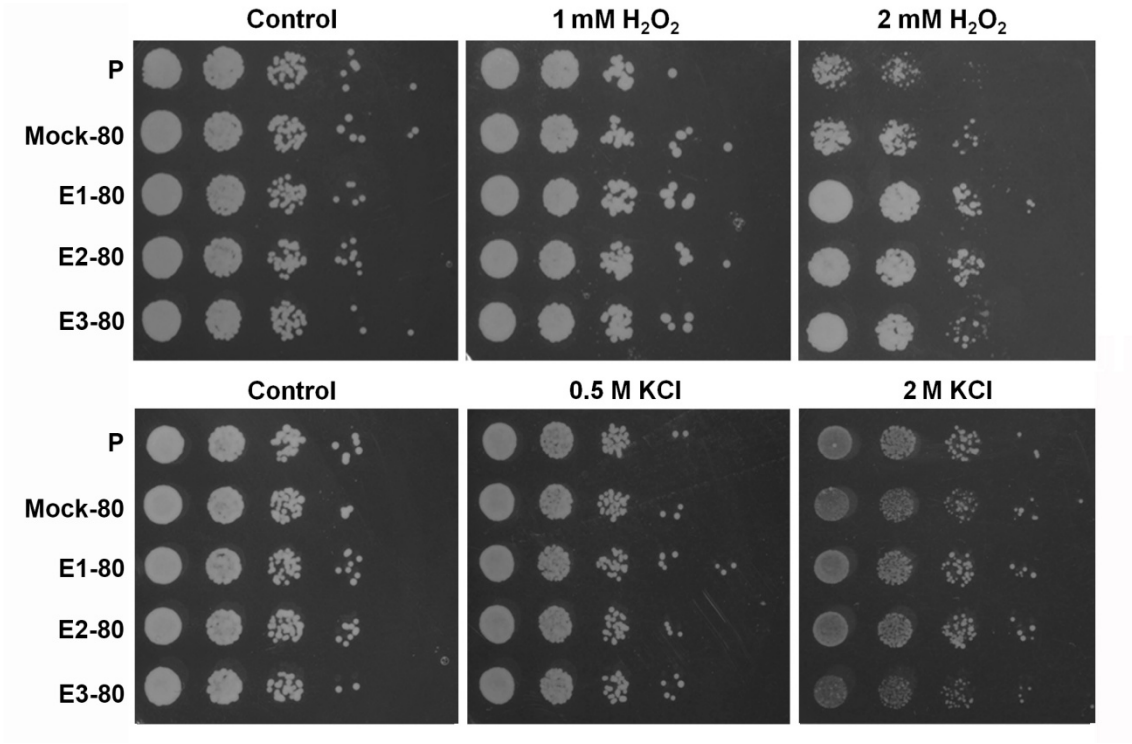


Figure 4.9. Heat shock evolved populations at cycle 80 have adapted an increased H_2O_2 resistance. Cells were grown in flasks to mid-log phase and then spotted on YES plates with or without different concentrations of KCl and H_2O_2. Note the elevated tolerance to H_2O_2 but not KCl in the evolved populations.

The experiment was repeated for the evolved populations at cycle 120. Due to instability issues with H_2O_2, t-butyl hydroperoxide (tBH) was used to induce oxidative stress. The effect was not as strong as with H_2O_2 for populations at cycle 80 (Fig. 4.10). The mock treated population seems to respond very similar to E3-120. Again, no improvement in resistance was observed in response to KCl.

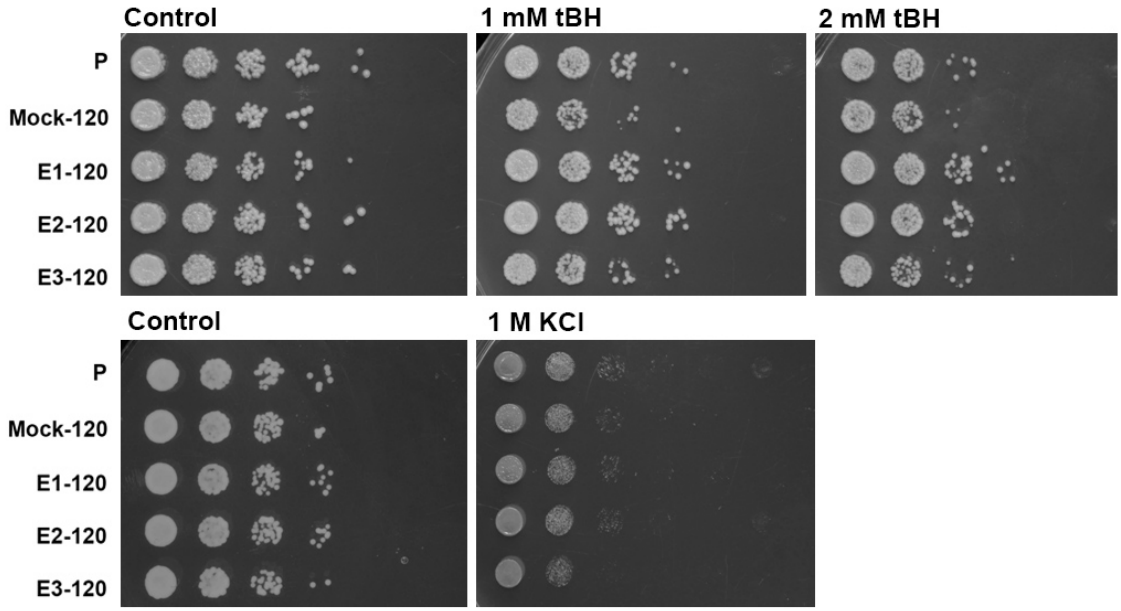


Figure 4.10. Heat shock evolved populations at cycle 120 have adapted an increased H_2O_2 resistance. Cells were grown in flasks to mid-log phase and then spotted on YES plates with or without different concentrations of KCl and H_2O_2. Note the elevated tolerance to H_2O_2 but not KCl in the evolved populations.

I have tried to test many different stresses on these evolved populations using a Singer RoToR robot for high throughput screening. However, I could not proceed with this assay because many colonies appeared sticky and the replica plating pads that pin the colonies on plates would lift entire colonies rather than a few cells. Interestingly the "sticky" colonies were exclusively the evolved populations and not parental or mock populations. While this is not a refined assay to phenotype cell adhesion, it raises the possibility that an adhesion phenotype evolved alongside the stress resistance. Cell-to-surface adhesion has been observed in many transcription factor mutants (Dodgson et al. 2009). It has also been found in populations in which a histone deacetylase has been deleted, which causes transcriptional activation of adhesion genes that are located in the subtelomeres (Shwab et al. 2007). However these mutants have been associated with heat sensitivity and not heat resistance (Kimata et al. 2008). One way forward to carry

out the high throughput stress assay would be to add soluble galactose into the plates, as galactose has been shown to inhibit cell-to-surface adhesion (Dodgson et al. 2009).

Growth phenotypes in mock and evolved lines

Growth dynamics in the absence of heat shock

Are there phenotypic differences between the evolved populations and the parental progenitor as measured by growth curves? I measured the growth of parental, mock and evolved populations at cycle 120 in the absence of heat shock using the Coulter Counter and Biolector assays (see Chapter 2). Figure 4.11A shows the results from one Coulter Counter assay with measurements done over 11 hours. One final measurement was done after 24 hours to reveal differences in cell density during stationary phase (Fig. 4.11B). These results suggest a slightly slower growth for the parental population in comparison to all the others. They also show a lower value for the stationary phase.

To gain further insight into the growth parameters, I repeated the experiment when our laboratory acquired a Biolector. In Figure 4.12, data are separately plotted for two biological replicates. However, each curve was calculated by averaging 3 replicate growth trajectories as each sample was tested over 3 wells in the same plate. We ran an R script to calculate different graph parameters (see Chapter 2). Scores for plateau 2 and 3 (the 2nd and 3rd best estimate for stationary phase) were excluded as they reflected the same trait as plateau 1. Calculations of the duration of lag phase are not listed, because the computation factored in experimental artefacts thereby resulting in unreliable estimates. Here we do not expect to see a lag phase since these exponentially growing cells were not heat-shocked. Notice that the curves are not fully linear. It is possible that handling the cultures while setting up the assay caused some stress on the

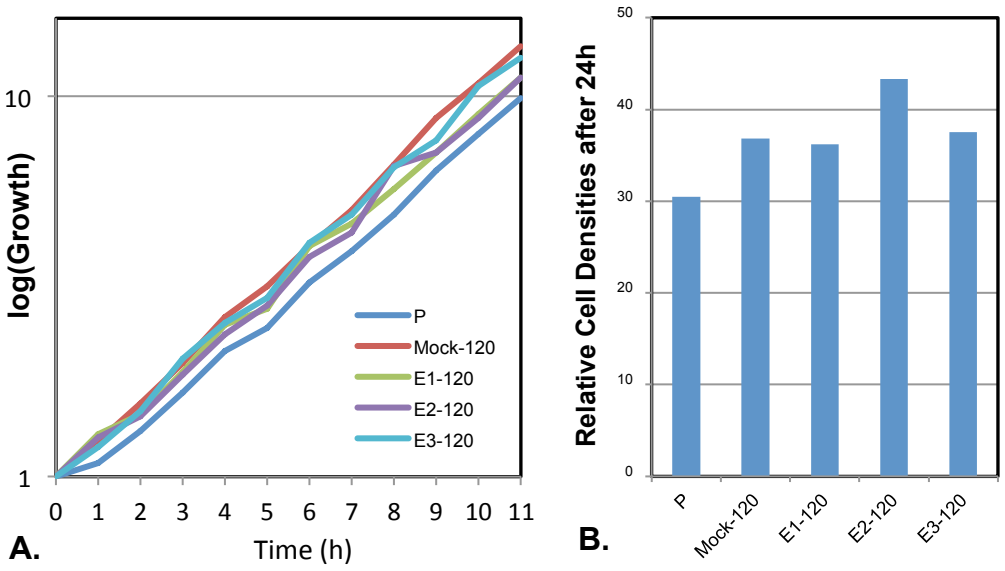
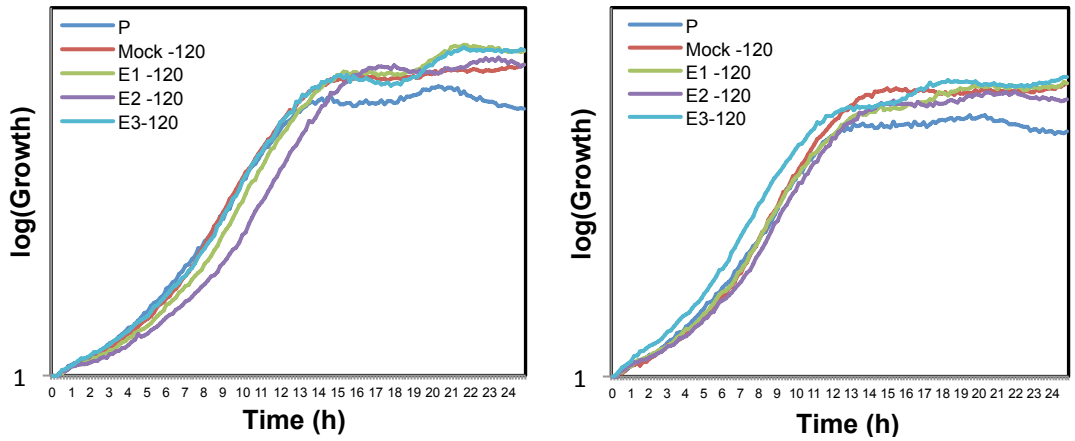


Figure 4.11. Growth curves in the absence of HS obtained with the Coulter Counter assay. (A) Samples from P, Mock and HS evolved populations at cycle 120, taken every hour over 11 hours. (B) Measurement at 24 hours after the 1st time point. All cell densities are normalized to the first value time point. The parental culture shows slower growth rates and reaches a lower concentration at stationary phase, indicative of a less efficient use of the growth medium compared to the evolved populations.

cells resulting in slower growth at the start. Moreover, with the room temperature being lower than 25ºC, there is some cooling down of the medium before it re-gains 25ºC when shaking in the Biolector. This might explain the non-linear curves during log phase.

Examining these results shows only marginal differences. In support of the data from the Coulter Counter assay, the stationary phase is reached at a lower biomass in the parental population relative to the mock treated and those evolved under the heat shock regime (see the values under 'plateau.1'). This comes with longer exponential growth phases and could reflect a more efficient use of the YES medium, possibly metabolizing other carbon sources that are present. This increase is the main component contributing to the increased values for 'integral' which is a measurement for the overall fitness in

this regime. We do not see slower growth rates for the parental population as suggested with the Coulter Counter assay.



name	integral	max.slope	plateau.1	max.slope.time
P	42.16	0.15	2.09	9.17
Mock-120	44.51	0.17	2.34	9.33
E1-120	44.06	0.16	2.36	9.50
E2-120	43.33	0.15	2.30	9.00
E3-120	45.82	0.16	2.39	8.17

name	integral	max.slope	plateau.1	max.slope.time
P	43.87	0.18	2.31	10.00
Mock-120	45.57	0.17	2.43	9.83
E1-120	45.74	0.18	2.63	11.17
E2-120	43.93	0.18	2.48	12.67
E3-120	46.07	0.17	2.40	10.00

Figure 4.12. Biolector results for growth in the absence of HS in two replicate assays. Cultures of parental, mock and evolved populations at cycle 120 were grown in YES to mid-log phase and transferred to Biolector flowerplates. Two independent biological repeats of the same experiment are shown. Biomass values were measure every 10 min over 24 hours. Stationary phase is consistently at a higher cell mass in evolved and mock populations relative to the parental, while the maximum growth rate is similar for all populations.

Growth dynamics after heat shock treatment

Figure 4.13 shows the data for growth obtained with a Coulter Counter after treatment with the 10 min HS at 44^{o}C. Again the biomass values after 24 h, which falls in the stationary phase, are lower for the parental population, as is the case in the absence of HS. Possibly there is a faster recovery to exponential growth in the evolved populations.

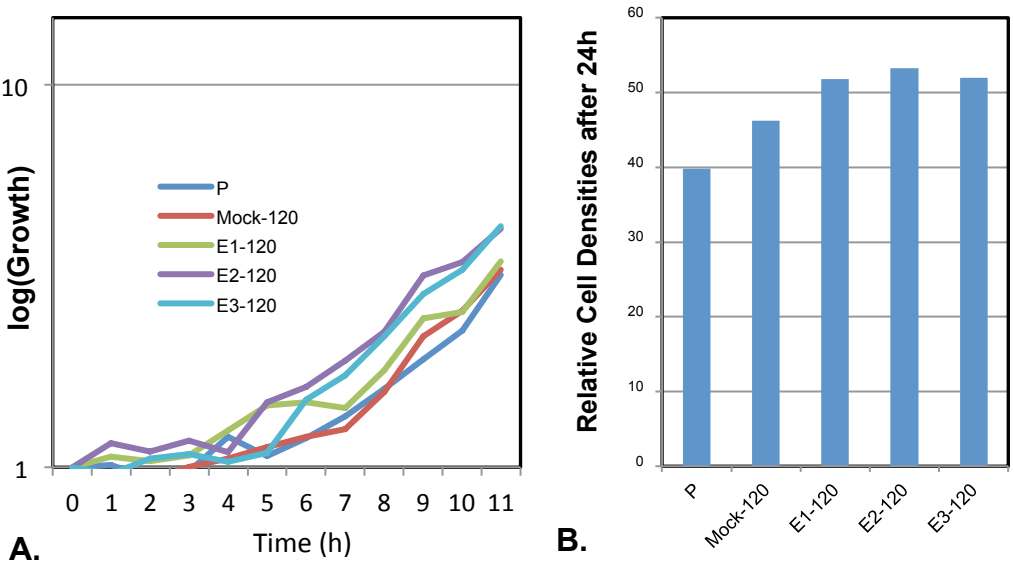


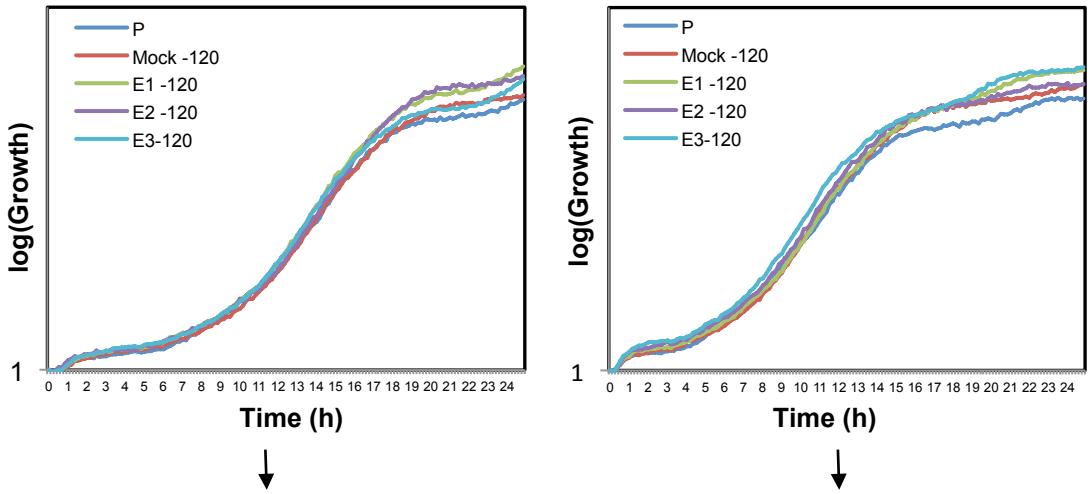
Figure 4.13. Growth curves after HS obtained with the Coulter Counter assay. A, samples taken every hour over 11 hours. B, measurement at 24 hours after the 1st time point. All cell densities are normalized to the first value time point.

Repeating the assay using a Biolector, calculations for the stationary phase show marginal though consistent increases in mock and evolved populations compared to the parental population (Fig. 4.14). There are possibly increases in growth rate (max.slope) and a faster transition into exponential growth, as the duration of time before reaching maximum growth is longer in P than in all the other populations (with the exception of one value for E2-120).

What is striking is that the increased cell densities for stationary phase (and thereby the longer exponential growth phases) in the HS evolved lines are also seen in Mock-120. As stated before, this could mean that other carbon sources in the medium are being metabolized and the medium can thus be used more efficiently. Both mock and HS evolved lines seem to have adopted this trait.

Regarding the faster recovery after HS, the differences between populations are as small as ~2%. It is surprising to see this trait also for the mock population since it could not

have evolved faster recovery of their growth capacity specifically following HS treatment. As stated, it could be that they evolved faster recovery after stationary phase and freeze-thaw treatments and that these conditions share similarities with recovery from HS. Alternatively, this trait could be connected to the increased efficiency in medium usage. Taken together, these data revealed fitness improvements for both mock and HS evolved lines, in particular at the level of the stationary phase but possibly also in the recovery following HS treatment.



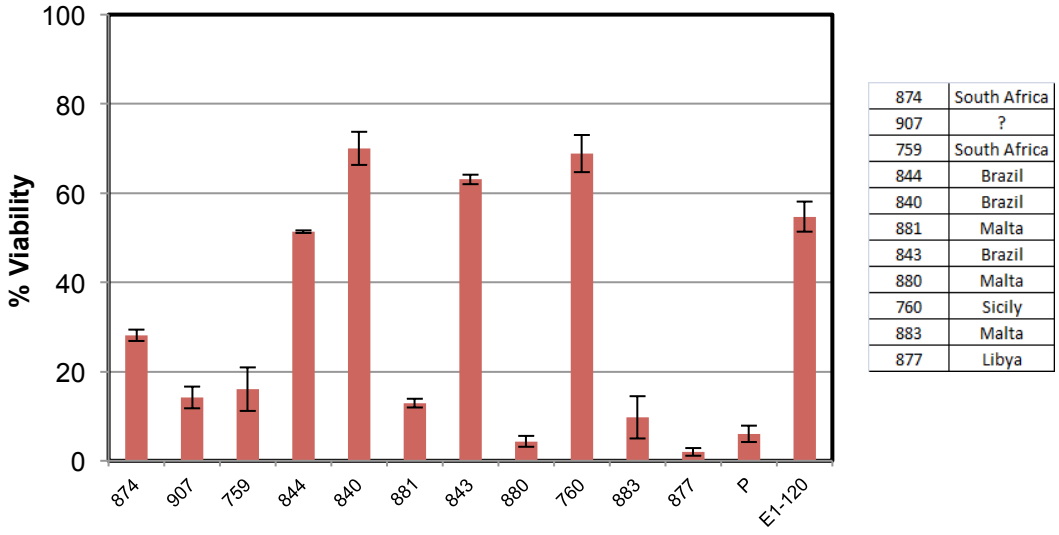
name	integral	max.slope	plateau.1	max.slope.time
P	36.53	0.13	2.16	15.00
Mock-120	36.82	0.14	2.28	14.67
E1-120	38.06	0.15	2.31	14.67
E2-120	37.96	0.16	2.36	16.00
E3-120	37.33	0.15	2.20	13.50

name	integral	max.slope	plateau.1	max.slope.time
P	39.72	0.13	2.07	11.50
Mock-120	40.64	0.14	2.27	10.67
E1-120	41.31	0.13	2.47	11.17
E2-120	41.37	0.14	2.39	11.00
E3-120	42.35	0.15	2.48	11.00

Figure 4.14. Two replicate assays for growth dynamics following the 10 min HS. Parental, mock and evolved populations at cycle 120 were grown to mid-log phase, heat-shocked for 10 min at 44°C in flasks as described, and transferred to Biolector flowerplates. Two independent biological repeats of the same experiment are shown. Though subtle, the calculations for integral show overall fitness improvements for HS evolved populations and also for Mock.

Effect of the 40 min heat shock on wild S. pombe isolates

Most of the studies using *S. pombe* were using derivatives of a single strain isolated in 1950 (Leupold's 968 homothallic strain). In a recent work, more than 81 geographically diverse *S. pombe* isolates were studied (Brown et al. 2011). We obtained this collection, and the survival to the 40 min HS was tested in 11 of these isolates. These strains were chosen because at the time we also had obtained their genomic sequences. While these sequences are still being analyzed, we know they are much more divergent than the laboratory strains that were used to generate the parental population (Dr. Daniel Jeffares; personal communication). I observed a lot of variation in viability among these strains, ranging from 2 to 70% (Fig. 4.15).



874	South Africa
907	?
759	South Africa
844	Brazil
840	Brazil
881	Malta
843	Brazil
880	Malta
760	Sicily
883	Malta
877	Libya

Figure 4.15. Cell viability to the 40 min HS in 11 wild yeast isolates as measured with the plate count assay. Much variation is seen amongst the different strains. Parental and E1-120 population results are also plotted for comparison. Error bars are based on averages of two biological repeats. The table on the right gives the country of origin for each strain.

Unfortunately, little is known about the natural environments in which these strains have evolved, apart from the substrate they have been isolated from (grape must, wine, cane sugar...) and their country of origin. These results suggest that yeast is able to

evolve a strong heat resistance also under natural conditions. They also raise the question whether the adaptation to heat is the emergence of a novel trait or rather the activation, or enhancement, of an existing one.

Homogeneity of populations

We wanted to address whether the resistant phenotype is the result of genomic changes that have swept to fixation to 100%, replacing the ancestral variants, or whether the populations contain a mixture of genomic variants. To this end, 5 clonal populations were obtained from 5 randomly picked colonies derived from single cells, from E1-80 exposed to the long stress experiment (Fig. 4.16A). The results show that there are differences in survival ranging from 50 to 78%, but none of the clones show the weak parental phenotype. Repeating the assay on clonal cultures from E1-120, the conclusion was similar (Fig. 4.16B). Here, I also added an experimental repeat for one clone to test whether the variation was not due to technical noise.

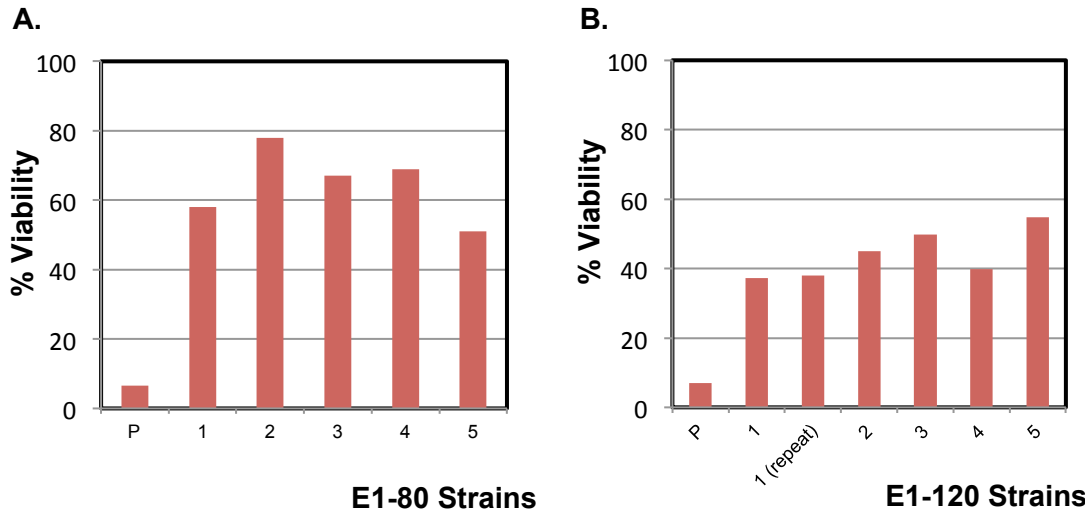


Figure 4.16. Viability of strains derived from single clones from E1-80 (A) and E1-120 (B). All strains show strong resistance to the 40 min HS compared to the parental population P, though there is some variation suggesting slightly different genotypes (n=1).

Though many more clones will be required to represent these populations, the present data suggest that the populations mainly consist of variants showing strong resistance to the 40 min HS.

In addition, the fraction of h90 positive cells was identified through iodine staining. By streaking out the populations to single cells and replica plating them to malt extract, only h90 positive cells will undergo sporulation. These colonies will turn dark brown when held over a petri dish containing iodine crystals. For this, 400-500 colonies were plated for each population over 3 plates. Counting the iodine positive colonies revealed that the parental population contained 12% of h90 positive cells, while E2-120 had 9%, and Mock-120, E1-120, and E3-120 all had 0%. This result shows that a fraction of the parental population has been outgrown during the experiment. One could proceed identifying the fraction of h- and h+ positive cells by PCR. However, since at the start of the experiment the 3 strains of different mating types were crossed, the same beneficial alleles can exist in cells of different mating types. Therefore these experiments reveal little about the homogeneity of the population. Without giving away the results of Chapter 5, this will be further elucidated with the analysis of the sequencing data.

Genome-wide transcriptional profiling of evolved and parental populations

We questioned whether the differences in physiological phenotypes are reflected in changes in gene expression. I used Agilent DNA microarrays to analyse the pattern of gene expression at time point t = 0 (just before the 10 min HS), and at 15 and 60 min after onset of the 10 min HS (see Chapter 2). First this analysis was carried out for P, E1-80, E2-80, and E3-80. When processing the data and clustering the genes based on similar expression patterns for these parental and evolved populations, the clusters appeared very similar at a global level, indicating no large differences in expression profiles between parental and evolved populations (Fig. 4.17A).

This assay was repeated for parental and evolved populations at cycle 120. We did not include the 15 min time point for this repeat, because the microarray analysis at 80 cycles of selection showed most differences at 60 minutes after the onset of the HS. We did however include Mock-120, but it was hybridized on a new batch of arrays. Unfortunately there was a quality problem with this microarray, and many genes appeared missing (depicted in grey in the cluster analysis). However, clustering led to the same outcome, indicating no large changes at a global level (Fig. 4.17B).

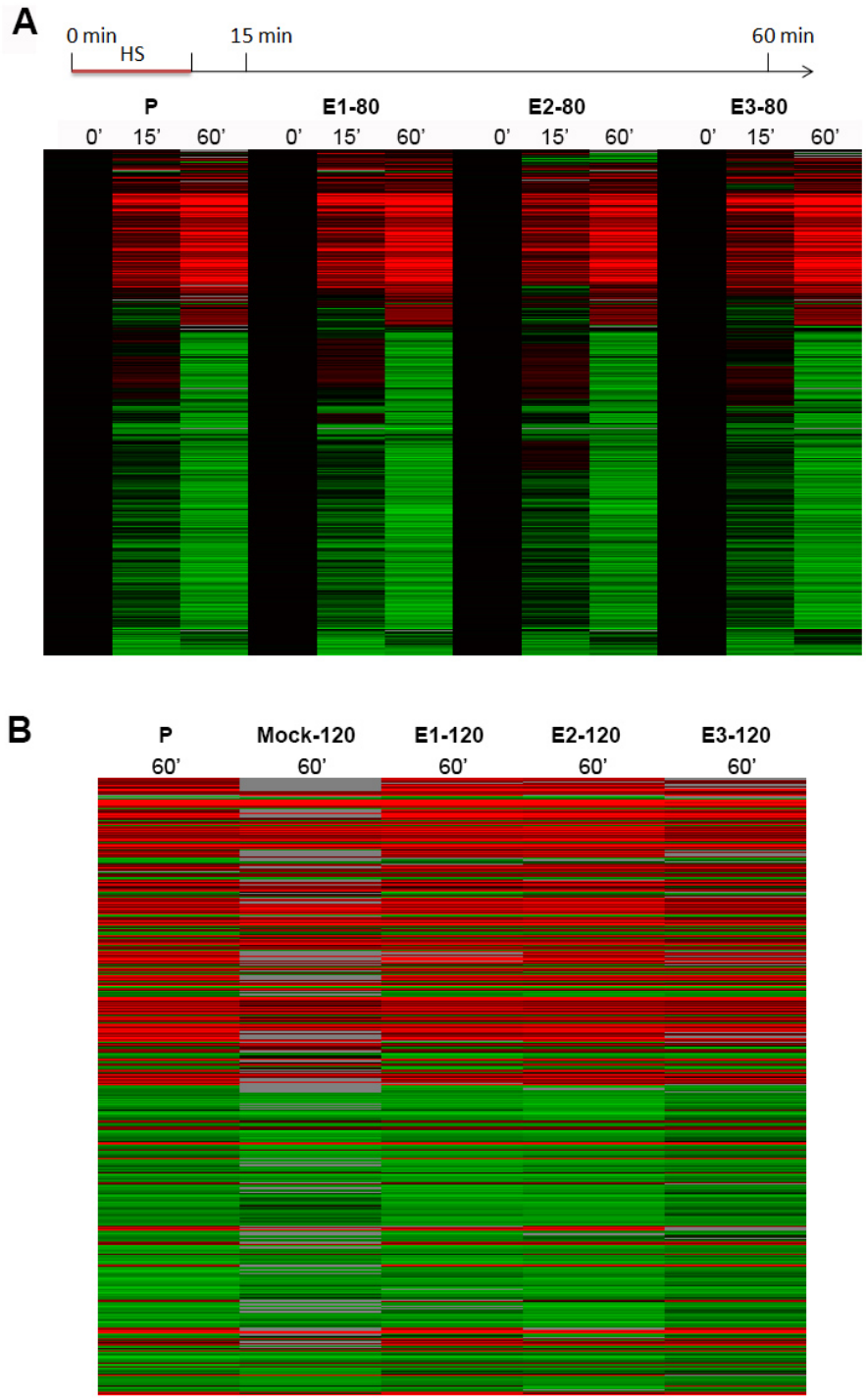


Figure 4.17. Changes in gene expression in response to 10 min HS (44°C) in parental and evolved populations. (A) Gene expression data for parental and evolved populations at cycle 80, presented as the ratio of each gene's expression value at 15 and 60 min versus t = 0. The scheme on top indicates the timeline for the timepoints. (B) Expression data for Parental, Mock and Evolved at cycle 120, presented as the ratio at 60 min versus 0 min. Data were clustered according to the Spearman correlation.

This finding is somewhat surprising considering the robust phenotype like the tolerance of the evolved populations for the 40 min HS. However, here cells were not treated with a 40 min HS, because it would have killed most parental cells needed for RNA extraction. Nevertheless, we assumed that genes required for resistance to long heat shocks would start being expressed within the first 10 minutes and therefore reflect in gene expression data. Thus, these results suggest that the heat resistance phenotype is not brought about by extensive reprogramming of gene expression, although altered regulation of a few specific genes might still be important.

By normalizing the data to the first time point it is possible that gene expression differences at that time point are not spotted. We were interested in this because it is known that constitutively expressed genes, despite representing a permanent cost, can provide immediate benefit when the protein is needed at once (Lopez-Maury et al. 2008; Geisel 2011). Given the nature of the HS, we hypothesized that cells could have adapted a constitutive expression of genes involved in long term HS resistance, genes that would normally be expressed "on demand". In order to test this, I divided the signal values at t = 0, 15, and 60 min for evolved populations (cycle 80) by the corresponding ones for the parental population. Therefore each value represents the times a gene is up- or down-regulated relative to this ancestral population at each one of these time points. I used an over-representation analysis (created by Dr. Vera Pancaldi) that partitioned these values among GO (Gene Ontology) slims which are cut-down versions of GO ontology lists. This method does not depend on a cut-off but calculates the median expression for each GO slim and compares it to the median of all other signals taken together. The output shows which GO slim lists are highlighted on the microarrays relative to the parental population. Table 4.1 displays GO slim terms that

were significantly different as determined by a Wilcoxon test after Bonferroni correction ($p < 0.05$). At first, the results seemed promising. It appears that certain sets of genes are already up- or down-regulated before the onset of the HS and that these sets show the typical stress signature with 'response to stress' being up-regulated and growth-related genes mostly down-regulated. This suggests that cells have evolved a way to 'foresee' the 10 min HS as they already show a low level of response before its onset. Similar sets are seen at 15 and 60 min, again suggesting that the stress response is more pronounced in these populations which could explain the resistance to 40 min HS. This gene set enrichment method was applied for the samples at cycle 120 (lists not shown). Contrary to the previous results, 'response to stress' was only listed at 60 min and many growth related terms were highlighted both in the up- and down-regulated categories.

	0 min	15 min	60 min
E1-80	response to stress cytokinesis	response to stress protein catabolic process protein modification by small protein conjugation	response to stress protein catabolic process meiosis protein modification by small protein conjugation vacuole organization
E2-80	x	response to stress	cell adhesion cellular amino acid metabolic process
E3-80	translation	x	response to stress protein catabolic process meiosis

	0 min	15 min	60 min
E1-80	ribosome biogenesis	translation ribosome biogenesis lipid metabolic process cellular amino acid metabolic proc.	translation ribosome biogenesis lipid metabolic process cellular amino acid metabolic proc. tRNA metabolic process urea metabolic process
E2-80	translation ribosome biogenesis cellular amino acid metabolic proc.	translation lipid metabolic process	translation
E3-80	lipid metabolic process	x	translation ribosome biogenesis cellular amino acid metabolic proc. vitamin metabolic process

Table 4.1. GO slim lists that are possibly enriched in evolved versus parental populations at 0, 15 and 60 min. The stress response list is only seen in lists that are significantly up-regulated relative to parental (pink table). Lists that are significantly downegulated contain only growth related profiles (blue).

Furthermore, when I examined the plots of these GO slims in GeneSpring, it turned out that their differences in expression are so marginal that they could be technical or biological noise (Fig. 4.18). For instance looking at the plots for "translation" revealed the values 1.024, 0.93 and 0,94 for E1-120, E2-120, and E3-120, respectively. This is <7% different from the overall average (= 1) which in terms of microarray analysis is minimal. I observed the same effects for all other GO slims. With such small differences, it is difficult to be conclusive. One could validate these results by doing an RT-PCR, although the present results indicate we would not expect big changes.

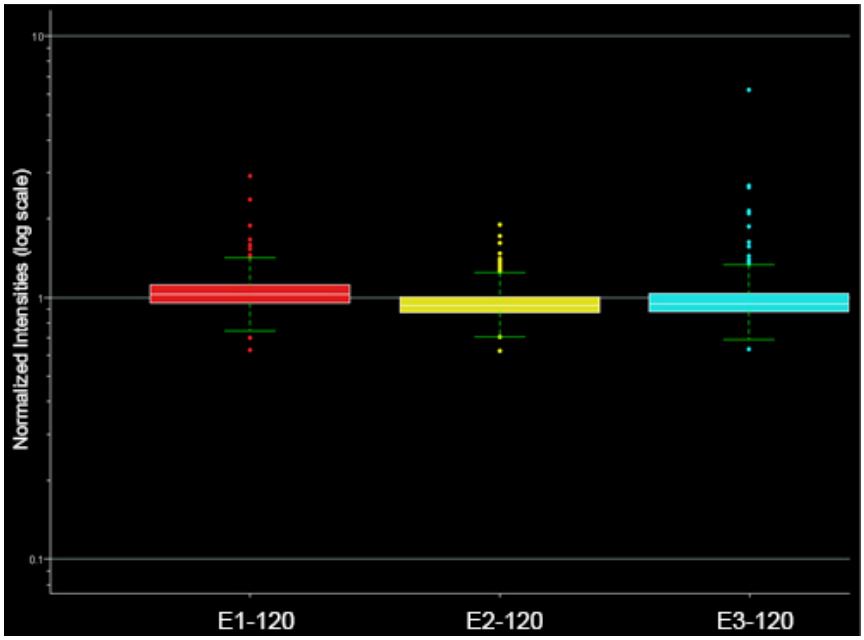


Figure 4.18. Box plot display in Genespring for the ratio values of the genes in the onthology list 'translation'. While this list was highlighted as significantly up-regulated in E1-120 and down-regulated in E2-120 and E3-120 relative to the parental population, this plot shows that the differences are marginal.

However, when this analysis was repeated on a biological replicate for E1-80 at 60 min, we obtained similar results (Table. 4.2). In this repeat, the labelling dyes had also been swapped to control for dye biases.

E1-80 / 60 min	repeat
response to stress	response to stress
protein modification by small protein conjugation	protein modification by small protein conjugation
protein catabolic process	protein catabolic process
meiosis	signaling
vacuole organization	regulation of mitotic cell cycle

E1-80 / 60 min	repeat
translation	translation
ribosome biogenesis	ribosome biogenesis
cellular amino acid metabolic process	cellular amino acid metabolic process
tRNA metabolic process	tRNA metabolic process
urea metabolic process	urea metabolic process
lipid metabolic process	

Table 4.2. GO slim analysis between two biological repeats. Of the 11 lists that are enriched in E1-80 according to the analysis, 8 are also seen in the dye-swap repeat.

I adopted a different way for identifying enrichment of specific functional categories. Gene expression data were presented as before by taking the ratio of each value for evolved versus parental population at 0 and 60 min. Lists were made above and below an arbitrary cut off of 1.5 in at least 1 out of the 6 conditions and imported into Gene List Analyzer developed in our group. Instead of using expression values, this method compares gene names and applies a Fisher Exact test to evaluate whether a known functionally defined group of genes is represented more than expected by chance within each gene list. The results were difficult to interpret. In both up- and down-regulated lists, Gene List Analyzer found significant overlap with mostly stress-related gene lists (such as stress module, Sty1-activated, CESR up- all, and intron-less genes). We would expect to see this in only the up-regulated list. Also in this up-regulated list there is overlap with Ribosome occupancy and Ribosomal density genelists. This is surprising considering an elevated stress response is known to shut down growth-related gene expression programmes (Lopez-Maury et al. 2008). When plotting these normalized gene expression data in Genespring, one interesting observation was made. In Fig. 4.19

each gene is coloured according to its ratio value at 60 min for E1-120. Blue means it is down-regulated versus the parental and red shows it is up-regulated. The same colours are then shown for these genes in the other 5 groups. We find that there are similarities in all 3 evolved populations. More importantly, many of the genes that are more up- or down-regulated at 60 min versus the parental population already show this expression state (though at a lower level) at t = 0, before the onset of the HS. While these genes do not fit known categories, these profiles suggest coherent changes in transcription levels that the evolved populations may have adapted in the face of the heat shock regime.

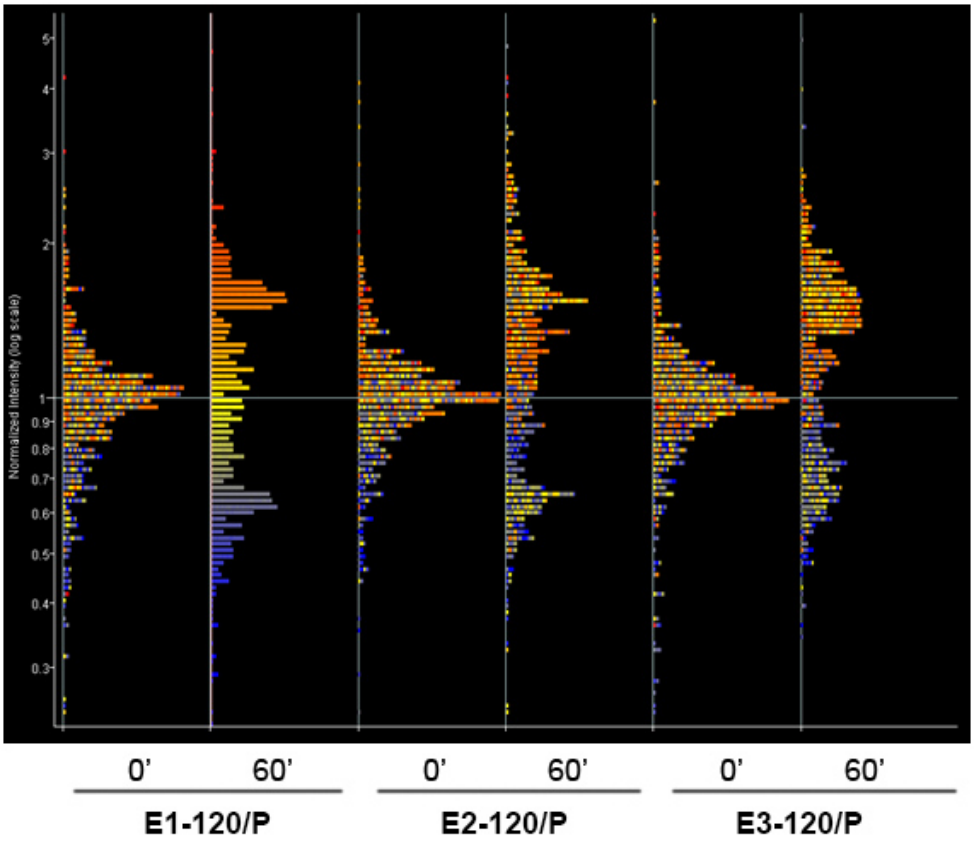


Figure 4.19. Gene expression distribution for ratio values (evolved vs. parental) that are higher or lower than 1.5 in at least 1 of 6 conditions. These data suggest that there are several specific genes that have evolved a different expression state before the onset of the heat shock.

Whole genome sequence analysis of evolved and parental populations

Identifying SNPs and indels

Libraries were prepared for Parental, Mock-120, E1-120, and E2-120 and sequenced with Multiplexed paired end sequencing technology (Illumina) at the UCL Cancer Institute. Because there was only space for 4 samples, we excluded E3-120 as it showed the weakest phenotype. Dr. Daniel Jeffares applied a pipeline which mapped the reads against the reference genome, refined them, and called SNPs and short indels (see Chapter 2). These variants were filtered automatically based on criteria such as low coverage, low mapping score, and strand bias. Table 4.3 lists the numbers of SNPs and indels for every population.

	SNPs	Indels
P	301	207
Mock-120	311	222
E1-120	403	250
E2-120	307	217

Table 4.3. Number of SNP and indel calls after filtering for variant quality score.

The average sequencing depth (also termed 'coverage') was 30X which is relatively low. Importantly, while DNA sequencing is often done using clonal populations, here genomic DNA was isolated from mixed colony populations consisting of millions of cells. When one allele is present in 1/3rd of the population, we expect it to be seen in approximately 10/30 reads. Most likely we are losing rare SNPs and indels, especially in regions with low coverage. Since there was standing genetic variation in the parental population, focussing on SNPs and indels that appear in the evolved populations but are completely absent from the parental population may be a too conservative approach,

because it excludes the possibility of adaptation due to standing genetic variation. For this reason, we used the program *Popoolation2* which identifies significant differences in allele frequencies rather than finding alleles that are present in one population and absent in another. Popoolation2 was used to calculate the Fst values for each pairwise comparison between Mock-120, E1-120, E2-120 and P (see Chapter 2). I focused on SNPs with allele frequency differences that were highly significant (Fisher exact test; $p < 0.1$). In the Integrated Genome Viewer (IGV), the validity of these data was tested by viewing each allele frequency difference directly in the aligned reads. Many were discarded because they were located in low quality reads, most of which were in telomeric regions which are known to be difficult to map due to repetitive sequences.

Table 4.4 lists the output after running Popoolation2 for Mock-120 and P. These are the SNPs that were selected in the absence of HS, indicating the growth in rich nutrient medium was far from neutral. Many of these were shared between E1 and E2, which is not surprising considering they were propagated in the same regime except they were also treated with the heat shock. Out of interest for the heat shock resistant phenotype, for E1-120 and E2-120 any SNP they had in common with Mock-120 was subtracted. This should highlight SNPs that are possibly linked to the long stress phenotype. However, for E1-120 no such SNP could be identified that was unique to this population and not shared with Mock-120. For E2-120, 10 such SNPs were found (Table 4.5).

One limitation of Popoolation2 is that it only applies to SNPs. Therefore, for the indels, I reverted to the variant calls from the pipeline. The mapped reads for all 4 populations were loaded in IGV. I looked at each one of the indels for E1-120 and E2-120 and, using strict criteria such as quality score, read depth and strand bias, I determined

manually whether there were differences in their frequencies compared to Mock and Parental. This analysis resulted in 12 indels for E1 and 12 for E2. For the non-annotated intergenic indels in E1-120, the nearest genes were listed (Table 4.5). Three of these were parallel mutations found in both populations. For Mock-120 this has not been done. Such an approach is not ideal, and we plan in the future to customize Popoolation2 to include indels.

Mock-120					
SNPs					
# CHROM	POS	Annotation	REF	ALT	Region
chr1	**999959**	**eng1**	**C**	**T**	**coding**
chr1	**999960**	**eng1**	**C**	**T**	**coding**
chr1	**3416193**	**near rp13202**	**C**	**T**	**Intergenic**
chr1	**1906124**	**5'UTR of ppk1**	**C**	**T**	**5'UTR**
chr1	**1816034**	**between sgt1 and gaa1**	**C**	**T**	**Intergenic**
chr1	**3479252**	**tps1**	**C**	**T**	**5' UTR**
chr1	3846335	pyk1	C	T	coding
chr1	1960611	hsr1	G	A	coding
chr1	3504996	gna1	C	A	coding
chr1	**3204845**	**SPAC1486.06**	**G**	**A**	**coding**
chr1	1787158	SPAC17A5.16	T	A	coding
chr1	8666	SPAC212.09c	G	A	coding
chr1	**3028375**	**SPAC31G5.20c**	**A**	**G**	**coding**
chr1	4794116	SPAC22F8.05	G	T	coding
chr1	3726309	SPAC15E1.05c	A	C	coding
chr1	5357556	SPAC4D7.04c	A	C	intronic
chr1	4134706	bg12	A	C	coding
chr1	**4733863**	**near SPAC458.03**	**G**	**C**	**miscRNA**
chr1	1486244	SPAC9.11	G	T	coding
chr1	2643596	hcs1	C	A	3'UTR
chr1	1960704	hsr1	A	G	coding
chr1	**2559681**	**SPAC4A8.07c**	**A**	**G**	**coding**
chr2	**3357187**	**near 5'UTR of alp1**	**G**	**A**	**intergenic**
chr2	3938648	SPRRNA.36	G	A	rRNA
chr2	**3194729**	**SPBC776.10c**	**G**	**A**	**coding**
chr2	**154212**	**mal1**	**C**	**T**	**coding**
chr2	**4533636**	**near SPBCPT2R1.08c**	**C**	**T**	**non-coding**
chr2	**1352172**	**SPBC27B12.12c**	**G**	**A**	**coding**
chr2	**653857**	**SPBC947.10**	**G**	**A**	**coding**
chr2	**1895892**	**SPBP16F5.03c**	**G**	**A**	**coding**
chr2	154062	mal1	T	A	coding
chr2	154176	mal1	A	T	coding
chr2	**2394638**	**tlg1**	**C**	**T**	**3'UTR**
chr2	689032	SPBPJ4664.02	C	T	coding
chr2	154150	mal1	A	G	coding
chr2	2437688	map4	A	G	coding
chr2	1518794	SPBC83.05	C	G	coding
chr2	270986	SPBC800.14c	C	A	5'UTR
chr2	**694186**	**SPBPJ4664.02**	**T**	**C**	**coding**
chr2	154065	mal1	G	A	coding
chr2	695427	SPBPJ4664.02	T	C	coding
chr2	**4336867**	**near SPBC16C6.05**	**G**	**A**	**intergenic**
chr2	**696261**	**SPBPJ4664.02**	**C**	**G**	**coding**
chr2	1442298	SPBP22H7.05c	G	A	coding
chr2	**1280612**	**SPBC3D6.06c**	**C**	**T**	**coding**
chr2	2595300	SPBC2G5.08	C	A	coding
chr2	693595	SPBPJ4664.02	C	T	coding
chr3	**2128341**	**SPCC126.07c**	**C**	**T**	**coding**
chr3	**679772**	**orb4**	**T**	**C**	**5' UTR**
chr3	**1797171**	**ist1**	**C**	**A**	**coding**
chr3	**1739643**	**wtf16**	**T**	**A**	**coding**

Table 4.4. Lists of SNPs for Mock-120 with significantly different allele frequencies compared to P. In bold font are the SNPs shared with E1-120 and E2-120.

E1-120

SNPs						
?						
Indels						
# CHROM	POS	Annotation	REF	ALT	Region	syn/non-syn
chr1	1765306	ulp2	C	CA	Intronic	N/A
chr1	650099	close to pis1	C	CAA	Intergenic	N/A
chr1	965438	close to hem13	C	CAAAA	Intergenic	N/A
chr1	1595077	close to srp102	T	TT	Intergenic	N/A
chr1	4431856	between sfp1 and dus3	G	GA	Intergenic	N/A
chr1	4648159	SPNCRNA.1008	TG	T	Intergenic	N/A
chr1	5173430	SPNCRNA.1055	G	GT	Intergenic	N/A
chr1	5271465	SPNCRNA.1065	TA	T	Intergenic	N/A
chr2	484534	ctf18	A	AT	Intronic	N/A
chr2	1555919	SPNCRNA.1443	A	ATTTATT	Intergenic	N/A
chr2	4517047	close to unknown	T	TTACTCATC	Intergenic	N/A
mito	8378	cox1	A	ACTA	Coding	non-syn

E2-120

SNPs						
# CHROM	POS	Annotation	REF	ALT	Region	syn/non-syn
chr1	4979869	rad1	G	T	3' UTR	N/A
chr1	5456432	SPAC1039.04	G	A	3' UTR	N/A
chr1	2617220	SPAC7D4.12c	G	T	Coding	non-syn
chr1	4614504	SPAC1093.03	G	C	Coding	non-syn
chr2	1707007	gua1	G	C	Coding	non-syn
chr2	694838	SPBPJ4664.02	T	C	Coding	non-syn
chr2	808485	SPBC530.07c	G	T	Coding	non-syn
chr2	246572	Pep12	G	T	Coding	non-syn
chr3	1820329	Nup60	G	T	Coding	non-syn
chr3	45214	nic1	A	G	3'UTR	N/A
Indels						
# CHROM	POS	Annotation	REF	ALT	Region	syn/non-syn
chr1	427504	tht1	C	CT	5' UTR	N/A
chr1	1507275	SPAC5D6.04	C	CAA	5' UTR	N/A
chr1	3301538	SPNCRNA.900	C	CT	Intergenic	N/A
chr1	3609896	SPAC17G6.11c	T	TA	5' UTR	N/A
chr1	3707192	Ubr11	A	AT	Coding	non-syn
chr1	4442621	SPNCRNA.988	AAT	A	Intergenic	N/A
chr1	4648159	SPNCRNA.1008	TG	T	Intergenic	N/A
chr2	889920	SPNCRNA.1395	TC	T	Intergenic	N/A
chr2	2579458	SPBC2G5.02c	C	CAA	5' UTR	N/A
chr2	484534	ctf18	A	AT	Intronic	N/A
chr3	822701	SPCC1393.11	G	GT	3' UTR	N/A
mito	8378	cox1	A	ACTA	Coding	non-syn

Table 4.5. Lists of SNPs and indels for E1-120 and E2-120 that show significant allele frequency changes compared to P. The SNPs that are shared with Mock-120 have been excluded.

Analysis of SNPs and indels

Of the combined 21 indels for E1-120 and E2-120, 12 were intergenic, 2 intronic, 5 were 3' or 5'UTRs, and only 2 were in coding regions. This is not surprising since coding indels are always non-synonymous (unless they are 3 bases long), changing the amino-acid sequence of the translated protein. However, 70% SNPs in E2-120 were coding and non-synonymous. The others were 3'UTR SNPs.

Not shown in these lists is that out of the 31 combined events, 25 were already seen in the Parental population at a very low frequency. This is indicative of standing genetic variation that has changed frequencies during the evolution experiments. Moreover, many of the parallel mutations appear in the same loci (ie. chromosome position), which also indicates they were already present in the parental progenitor. However, examining the function of these mutations, it is difficult to establish their cause and effect relationships with the observed phenotypes. Worth noting is *rad1* which is involved in the repair of damaged DNA (Tomkinson et al. 1993). Some targets are growth related such as *gua1* which is involved in purine biosynthesis (Karaer et al. 2006), and *nic1* that is needed for urease biosynthesis (Eitinger et al. 2000). In Mock-120 one target is *mal1* which is required for maltose utilisation (Goldenthal et al. 1987). Whereas in most cases, the alleles rise in frequency, viewing the reads in IGV revealed that *mal1* has several SNPs at low frequency in the parental population which were lost in the mock control. So the mock population has the WT version of *mal1*, whereas the parental also has mutated versions of this gene. Also worth noting in the mock population is *pyk1* (pyruvate kinase) which is known to control carbon flux in yeast (Pearce et al. 2001). Its homolog in *E. coli* was also mutated in the long-term evolution experiment (see Chapter 1).

Are these targets differentially expressed compared to the parental population? I looked back at the microarray data and examined the expression levels for each of the targets from Table 4.5 that are represented on the array. Like in the over-representation analysis, the expression values for t = 0 and 60 min for evolved populations were divided by their matching value for P. Consequently, a relative expression of 1 shows no difference (Fig. 4.20). Two genes appear down-regulated: SPBC530.07c which encodes a TENA/THI family protein, and hem13 which codes for coproporphyrinogen oxidase. However they are also repressed in Mock-120. Other changes are small though many appear slightly down-regulated in comparison to P.

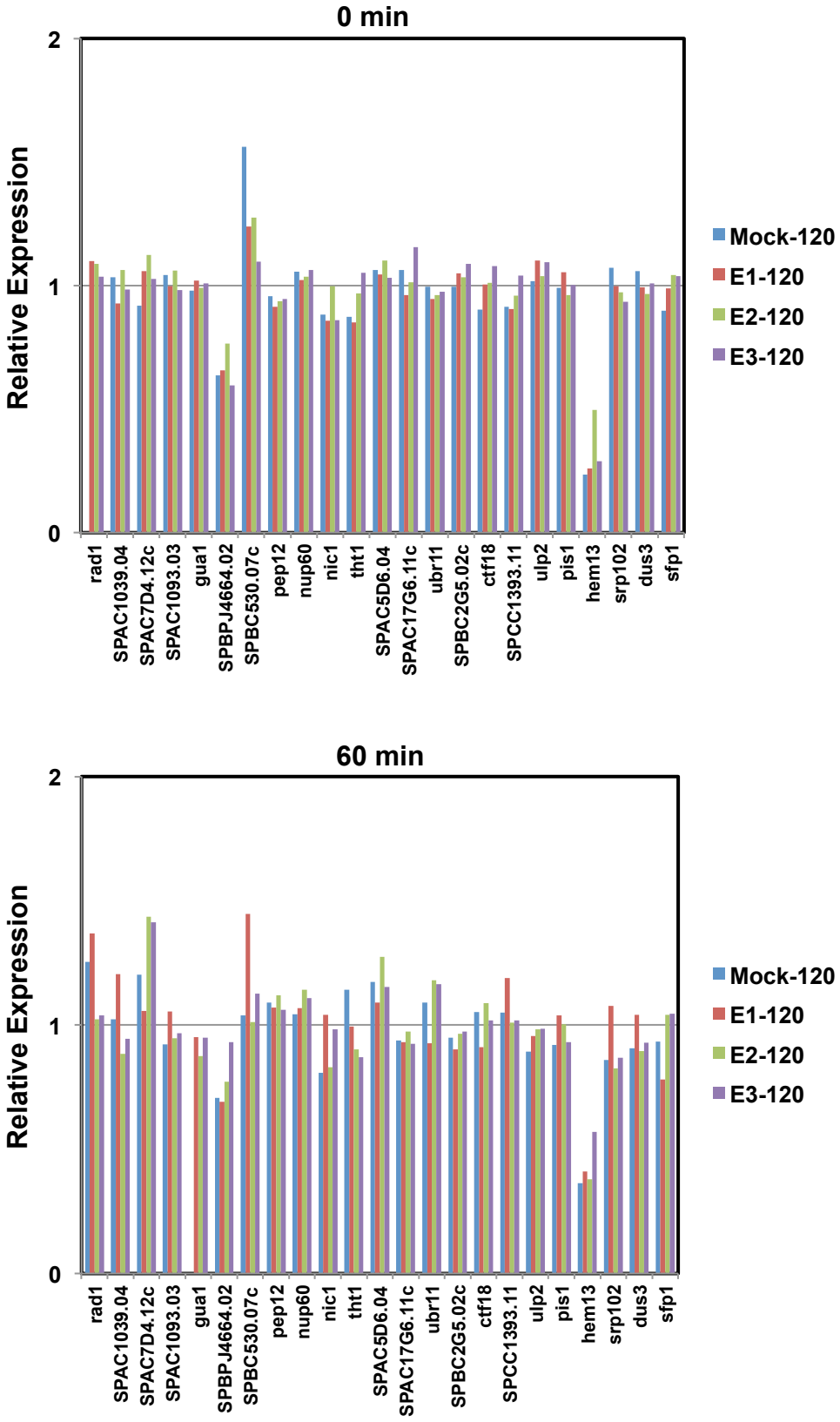


Figure 4.20. Signal ratios (Mock and Evolved vs. Parental) for the genes that were mutated in E1-120 and E2-120.

Identifying duplications

We have so far only identified SNPs and indels and currently do not have a pipeline to discover structural variants like large deletions, duplications, insertions, and inversions. However, as we have seen in Chapter 1, some of these structural variants can play an important role in evolution. For instance, duplications can provide a reservoir of standing variation as selection can favour cells that have an increased copy number of a particular region (Kugelberg et al. 2006; Andersson and Hughes 2009).

Given a sufficient number of reads, it is possible to use read depth to infer copy number variants (Zhu et al. 2012). Dr. Daniel Jeffares made a Perl script that calculated the read coverage for each population using a sliding window of 100 nt. Presence of duplications should show 2-fold changes in read coverage. None were detected, which suggests that there are no obvious large duplications, though there could still be smaller ones. Figure 4.21 shows these plots for E1-120. The ones for E2-120 and Mock-120 look very similar (not shown).

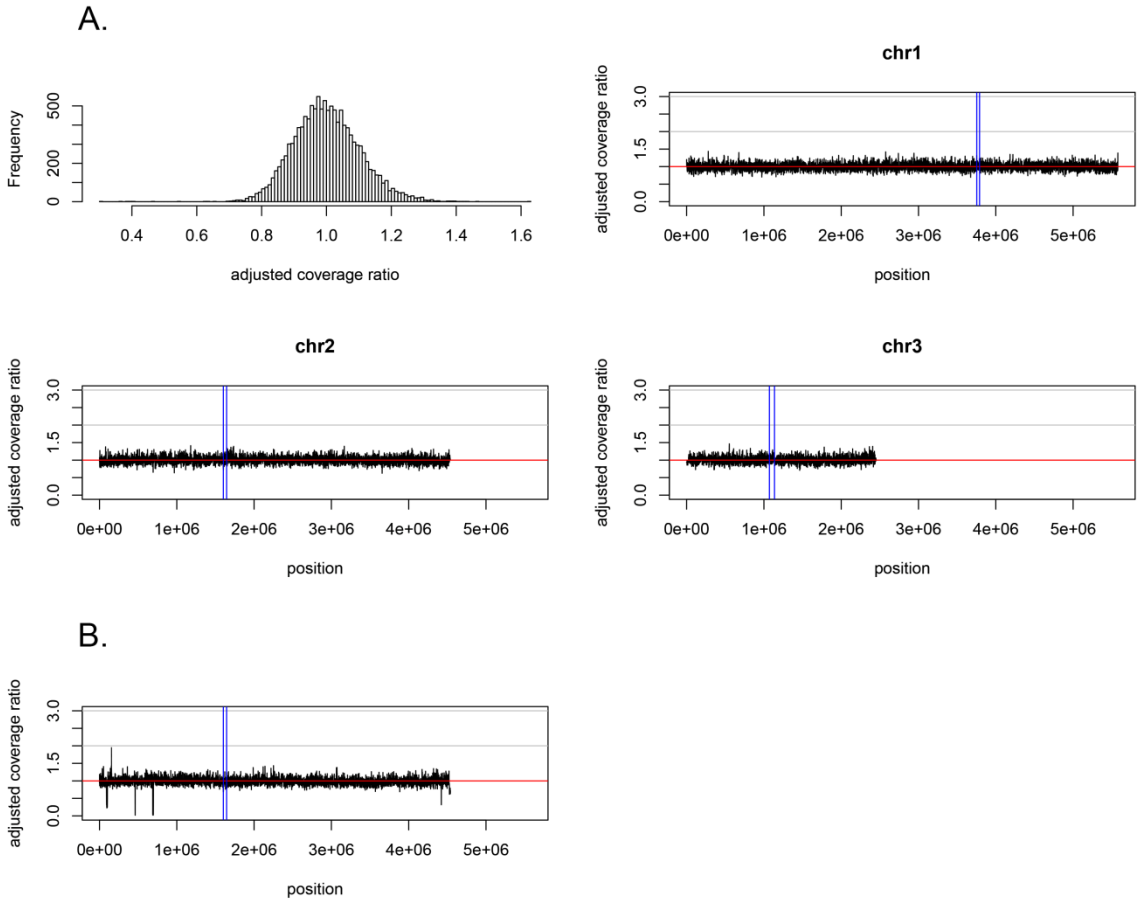


Figure 4.21. Adjusted coverage ratios for E1-120. (A) Analysis indicating there are no large duplications in E1-120. The blue, vertical lines mark the centromeres, the red, horizontal line is the median, and the grey lines are 2x and 3x the median. In the top left panel the distribution is plotted for the relative sequencing depth. (B) Example of a plot in which indels (and one duplication) are seen (courtesy of Dr. Daniel Jeffares).

Conclusion

We have shown that the evolved populations have adapted tolerance against extended stress such as a 40 min heat shock, and that most this adaptation happened rapidly, within 40 cycles of selection. This may seem surprising since the populations are never exposed to longer stress in the evolution assay, and so acquiring resistance to the 40 min HS does not improve fitness. It is contrary to our initial hypothesis that populations that are repeatedly exposed to the short heat shock would evolve a decreased response to longer stress. We did not measure improved resistance to the 10 min HS which they received during the evolution experiment, although small improvements are difficult to detect due to the experimental error that comes with the plate count assay.

Examining the growth dynamics, both mock and evolved populations showed longer exponential growth phases which reflect a more efficient use of the nutrient medium.

Globally, no large changes could be found in gene expression when comparing evolved and parental populations. In an overrepresentation analysis of GO slim terms, a stress-related gene category appeared enriched before the onset of the heat shock, although further analysis showed only marginal changes in gene expression.

Whole genome sequencing analysis of parental, mock, and two evolved populations revealed a combined 31 mutations that show significant allele frequency differences in the evolved compared to the parental progenitor, and that were not present in the mock control. Out of these 31 events, 10 were 3'UTR or coding SNPs, and 21 were indels, located mainly in intergenic regions. The majority of these SNPs exist in the parental population at a low frequency which is indicative of standing genetic variation.

Chapter 5

Repeatability of evolution and mutation dynamics

Repeatability of evolution and mutation dynamics

To address the repeatability of adapted phenotypes, the previously described evolution experiment was repeated over 35 cycles of selection in 6 populations. In order to gain insights into the dynamics of genomic adaptation, all evolution lines were phenotyped and sequenced at many different time points over the course of the experiments. Genome sequencing and analysis was done in collaboration with Dr. Ville Mustonen at the Sanger Institute.

Introduction

In Chapter 4, I described a rapid adaptive heat shock tolerance in 3 replicated evolution lines. These results raised several questions. Firstly, considering that the selection pressures seemed relatively low, we were wondering whether there could have been a technical cause behind the rapid adaptation, independent from the cycles of selection. For instance, due to two contaminations in the first 15 cycles, on two occasions, frozen populations from an earlier cycle were revived to continue the experiment. These procedures could have exerted strong selective pressure, thereby increasing the rate of adaptation. Secondly, if standing genetic variation had played a crucial role, rather than *de novo* mutations, we wondered whether this could be tested by running the experiment starting from clonal populations. In these clonal lineages, adaptation should proceed more slowly as evolution depends on the random generation of *de novo* variants. Thirdly, coming back to the initial hypothesis that repeated exposure to a mild stress may cause populations to lose their long-term stress response, the fact that adaptation showed the opposite effect could mean that the10 min HS was too strong and that selection favoured HS resistant cells rather than cells that recover more rapidly

from the lag phase following the HS. Even though we could not detect resistance to 10 min HS in the evolved population, small improvements would be difficult to detect considering the experimental error that comes with the plate count assay.

To address these three questions, we repeated the evolution experiment with several lines. Two of them were exact biological replicates of the previous experiment, starting from the same parental background, and heat-shocked for 10 min at 44^{o}C. In this evolution experiment, samples were only revived once from frozen stocks (at cycle 20) due to a technical issue with the water bath. To test for standing variation as a cause for rapid adaptation, two lines were added starting from an isogenic, clonal population. Finally, in order to test whether the HS was too strong to make cells lose their anticipatory response, two lineages were evolved from the parental population but in order to induce milder stress levels, they were treated with shorter durations of the HS. Because previous results showed that the most dramatic phenotypic changes happened within 30 cycles, these new experimental lines were stopped after passing through 35 cycles of selection.

With these many additional lines came the need for a high-throughput phenotyping method. The plate count assay is time and resource consuming and not suited for high-throughput analysis. Recently, a high-throughput method for quantitatively measuring chronological life span (CLS) was described by inoculating aged cultures in rich medium in individual wells of a Bioscreen 100-well plate (Powers et al. 2006, Murakami et al. 2008). Absorbance readings were taken every 30 min, and viability was quantified by a shift in the growth curve compared to the growth curve at the initial age-point. Using a Biolector (m2p-labs) which monitors light scattering as a measure of biomass, we established a similar approach to quantify cell viability after 40 min HS

treatment (see Chapter 2). Less cell death in a culture would be reflected in a shorter duration of the lag phase, because it would take surviving cells fewer divisions to reach the mid-log phase compared to a culture with more cell death. However, as we will see, a shorter duration of the lag phase may also indicate a more rapid recovery of exponential growth. This assay was applied on samples from many time points across all lines, providing insight into the rate of phenotypic adaptation.

Finally, we wanted to examine how the accumulation of mutations and changes in allele frequencies correlate with these phenotypic changes, and how many of these events are shared among different replicates. Based on the Biolector phenotype data, 96 samples were selected and sequenced using a single Illumina HiSeq lane. This was done in collaboration with Dr. Ville Mustonen who also analysed and modelled allele frequency trajectories as previously described (Illingworth et al. 2012). Rather than using Popoolation2 as in Chapter 4 to pinpoint causative alleles, this approach was specifically designed to quantify allele frequency changes over multiple time points. The SNP calling pipeline was run by Dr. Daniel Jeffares with the help of Dr. Sendu Bala at the Sanger Institute (see Chapter 2).

Repeating the evolution experiment

The evolution experiment was repeated with 6 lines. One of them, E0, was run in parallel with the last 30 cycles of Mock, E1, E2, and E3 which underwent 150 rounds of selection in total. The E0 lineage was started from the same parental population and treated with a 5 min HS at 44^{o}C (while E1, E2, and E3 were undergoing a 10 min HS, E0 was taken out of the water bath after 5 min).

After this experiment, 5 new lineages were started. A frozen aliquot of the parental population was revived and divided into 3 populations which evolved in separate flasks over 35 cycles. Two were exact biological repeats of E1, E2, and E3, heat-shocked for 10 min at every cycle. These lines are termed M1 and M2 ('M' for 'Mixed', as they started with the original mixed parental population). The third line ('M0') was treated with only a 1 min HS at 44^{o}C, thus reducing the selection pressure for heat resistance.

In one experiment, growth arrests were measured for the parental population after treatment with 1, 5, and 10 min HS (Fig. 5.1). Lag phases or slower entry into exponential growth are clearly observed, even when treated with a 1 min HS. Cell viability was also measured after treatment with these different heat shocks (n = 1). No cell death was detected after 1 minute HS, while 5 min HS resulted in 3% cell death, although this is likely within experimental error. The data for growth arrest and cell viability after 10 min HS reproduced previous results (see Fig. 3.1 and 3.2 in Chapter 3). Although repeats are needed to be conclusive, these data suggest that there is little, if any, effect on cell mortality, although there is still a growth delay following these shorter heat shocks. As such, rather than selecting HS resistant cells, selection could favour variants that recover more rapidly from the lag phase.

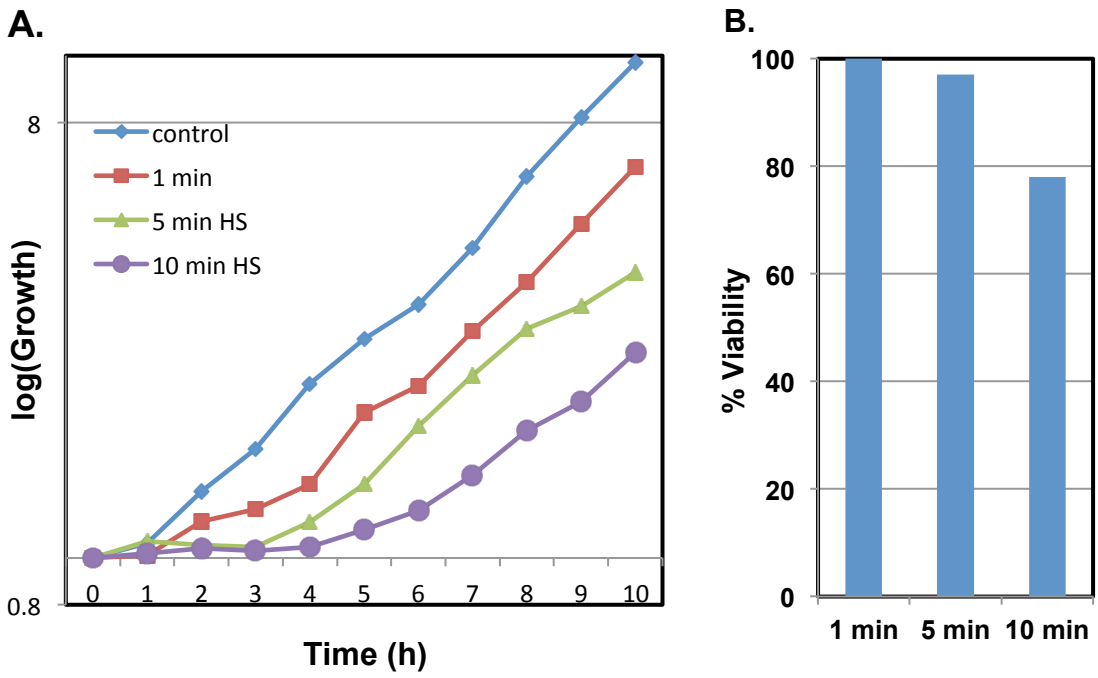


Figure 5.1. 1 and 5 min HS treatments have little or no effect on cell viability but still elicit a delay in recovery of exponential growth. (A) 10 min HS elicits a growth arrest of ~4 h as previously shown, while 1 and 5 min HS show ~1 and ~3 hours, respectively. (B) Viability measurements with the plate count assay reveal that heat-shocking cells for only 1 or 5 minutes causes little cell death (n = 1).

For the single colony experiment, two lineages were started from a clonal population. They are termed S1 and S2 ('S' for 'Single'). For these, the parental population was first streaked to colonies derived from single cells. One colony was picked at random and transferred to liquid YES medium to form an isogenic haploid strain (termed S-P). The viability to 40 min HS was tested for this strain to ensure it exhibits the weak resistance of the parental population (Fig. 5.2). This population was divided to form 2 replicate lines, which then each underwent the standard evolution assay over 35 cycles of selection.

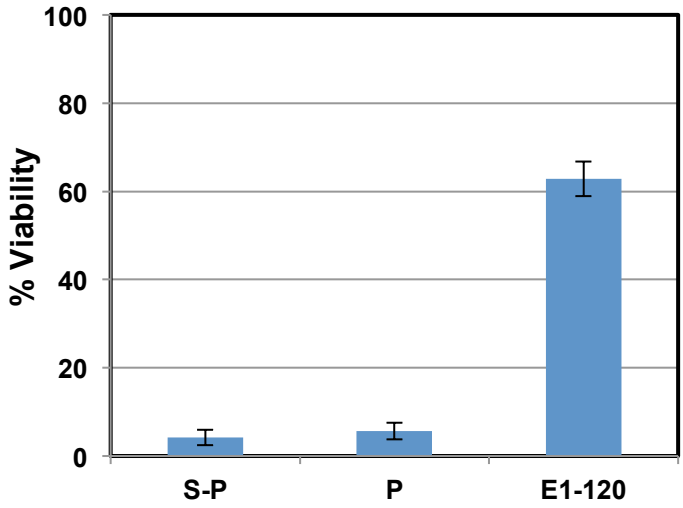


Figure 5.2. Viability to 40 min HS in the isogenic population S-P, which was the starting population for S1 and S2. P and E1-120 were added as controls and showed viability of a similar range as in previous experiments. The error bars are the standard error of the average of two biological repeats.

The Biolector assay as a high-throughput phenotyping method

An R script was used to score the Biolector growth curves (see Chapter 2). Since the estimates of the lag phases were often biased due to experimental noise, we settled on 'max slope time' as a robust measure for cell viability (see Chapter 2). This is the time it takes for the maximum growth rate to be reached. Like the lag phase, it is inversely correlated with cell viability. In one experiment, the viability from 8 samples was measured with both the plate count and Biolector assay. Plotting the data from both sources revealed a linear correlation (Fig. 5.3), although we will see below that these changes may not reflect heat shock survival.

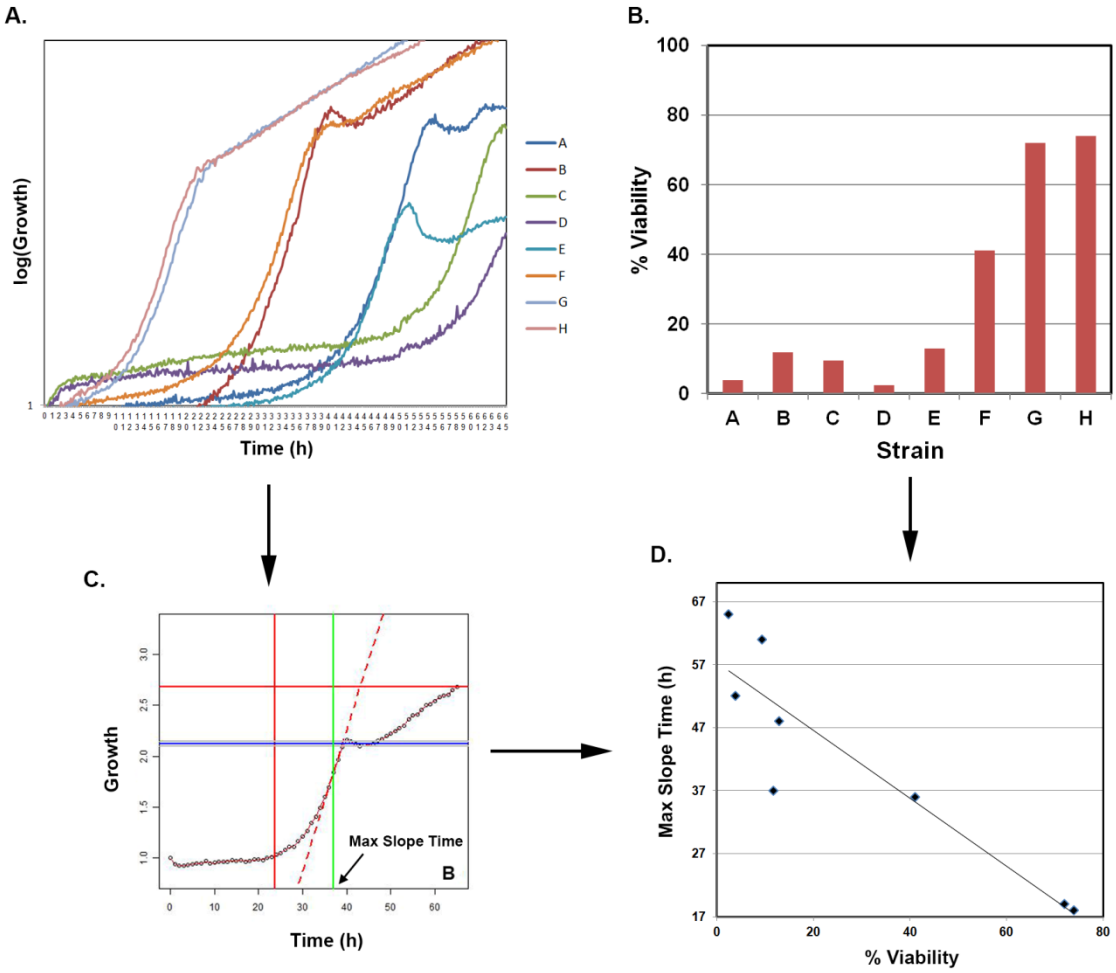


Figure 5.3. Validation of the Biolector assay as a measurement for cell survival. (A) Plotted Biolector measurements of relative changes in cell mass for 8 strains (A-H) after treatment with 40 min HS. (B) Viability to 40 min HS in the same strains, measured with the plate count assay. (C) Scoring of the curves using R, here only plotted for strain 'B'. The vertical green line indicates the 'Max Slope Time'. (D) Viability is plotted against 'Max Slope Time' and shows a linear, inverse correlation between both data sets (Pearson = -0.91).

Figure 5.4 provides an overview of all 10 evolution lines indicating which samples were run through the Biolector assay for high-throughput phenotyping. Since previous results indicated that most adaptation happened early, for the lines that ran over 150 cycles, a dense sampling was used for the first 30 cycles, followed by a shallower sampling for cycles 30 - 150. For all other lineages, sampling was done every 5th cycle.

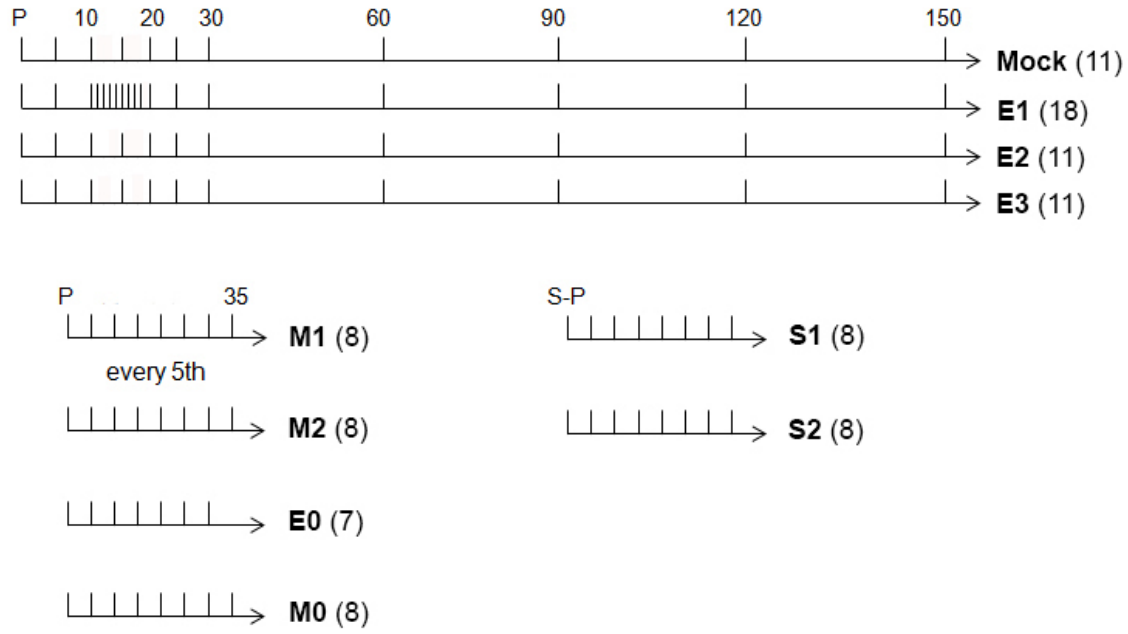


Figure 5.4. Overview of all 10 evolution lines. The vertical marks indicate which samples were phenotyped using the Biolector assay. Indicated in braquets are the number of samples for each line.

Results of the Biolector assay

For each of these samples, the Biolector assay was run, growth curves were plotted, and the maximum slope time was calculated. Each curve was the average of at least 2 experimental repeats, as each sample was tested over 2 or 3 wells in the same plate. Biological repeats were done for most lines (except for M0 and E0), although often on fewer cycles (for example, every 10^{th} instead of every 5^{th}). These repeats were not averaged due to the qualitative nature of the assay, although they each gave very similar outcomes. Figure 5.5A illustrates how such growth curves look like after running the Biolector assay. In Figure 5.5B, I have plotted the corresponding maximum slope times for each curve. The results for all populations are shown in Figures 5.6 and 5.7, where only the plots for maximum slope times are displayed. Looking at these data, in support of previous results, most adaptation seems to occur in the first 30 cycles for E1 (Fig. 5.6, top left).

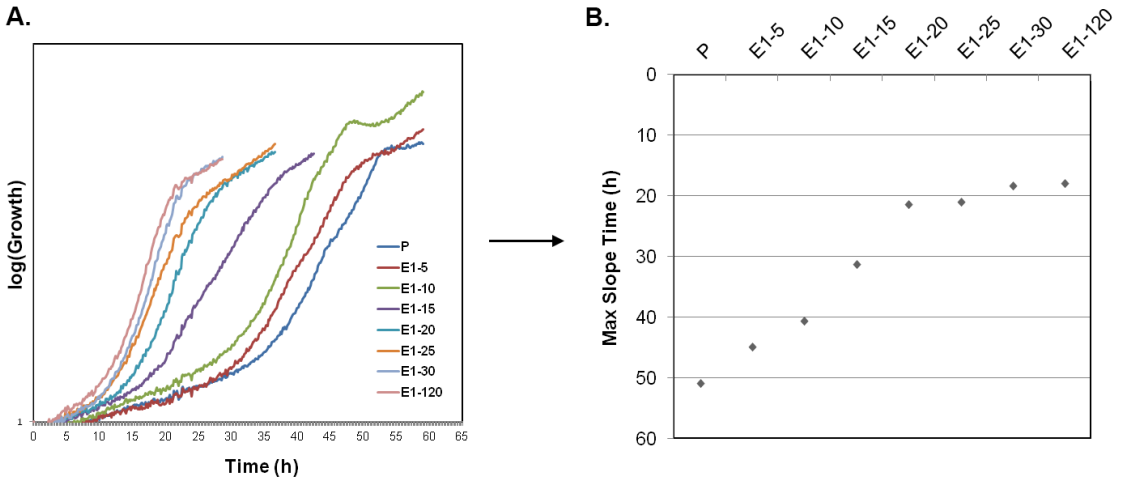


Figure 5.5. Example of growth curves and their corresponding maximum slope times. (A) Growth curves for E1 samples after treatment with 40 min HS in the Biolector assay. (B) Times to the maximum growth rate for each sample.

Such rapid and continuous improvements with the number of selection cycles are seen in each experiment. Surprisingly, the same trend in changes of Max Slope Time were observed in all lineages, including M0 and E0 which received shorter durations of the stress over the evolution experiment.

Even more surprising is that the plot for Mock was very similar to the ones for E1, E2, and E3. Previously, cell viability measurements using the plate count assay revealed some increased resistance in this population but to a much lesser extent than the heat shock evolved populations. The results here suggest that Mock behaves similarly to E1, E2, and E3. There can be several explanations for this discrepancy. First, there are differences in the assays themselves. In the plate count assay, the HS is presented in flasks in a water bath, and cells are plated on agar plates. For the Biolector assay, the HS happens in a plate by adding hot medium to each well. These plates are then transferred to the Biolector, and the cell mass is measured every 10 min. Possibly, cells from the mock population respond differently to these two assays. Secondly, it is difficult to distinguish whether a decreased lag phase reflects a higher number of viable

cells, or whether it reflects a growth related phenotype. Indeed, if cells would recover exponential growth more rapidly following the stress, this too would reflect in a shorter duration for the maximum slope time. This observation made us redefine the phenotypes obtained with the Biolector protocol. Despite the correlation previously discussed in Fig. 5.3, rather than being a measure for cell viability, Max Slope Time could represent how fast cells resume exponential growth. We will therefore attribute the term "lag related" to describe the phenotypes which show a shortened Max Slope Time relative to the parental population.

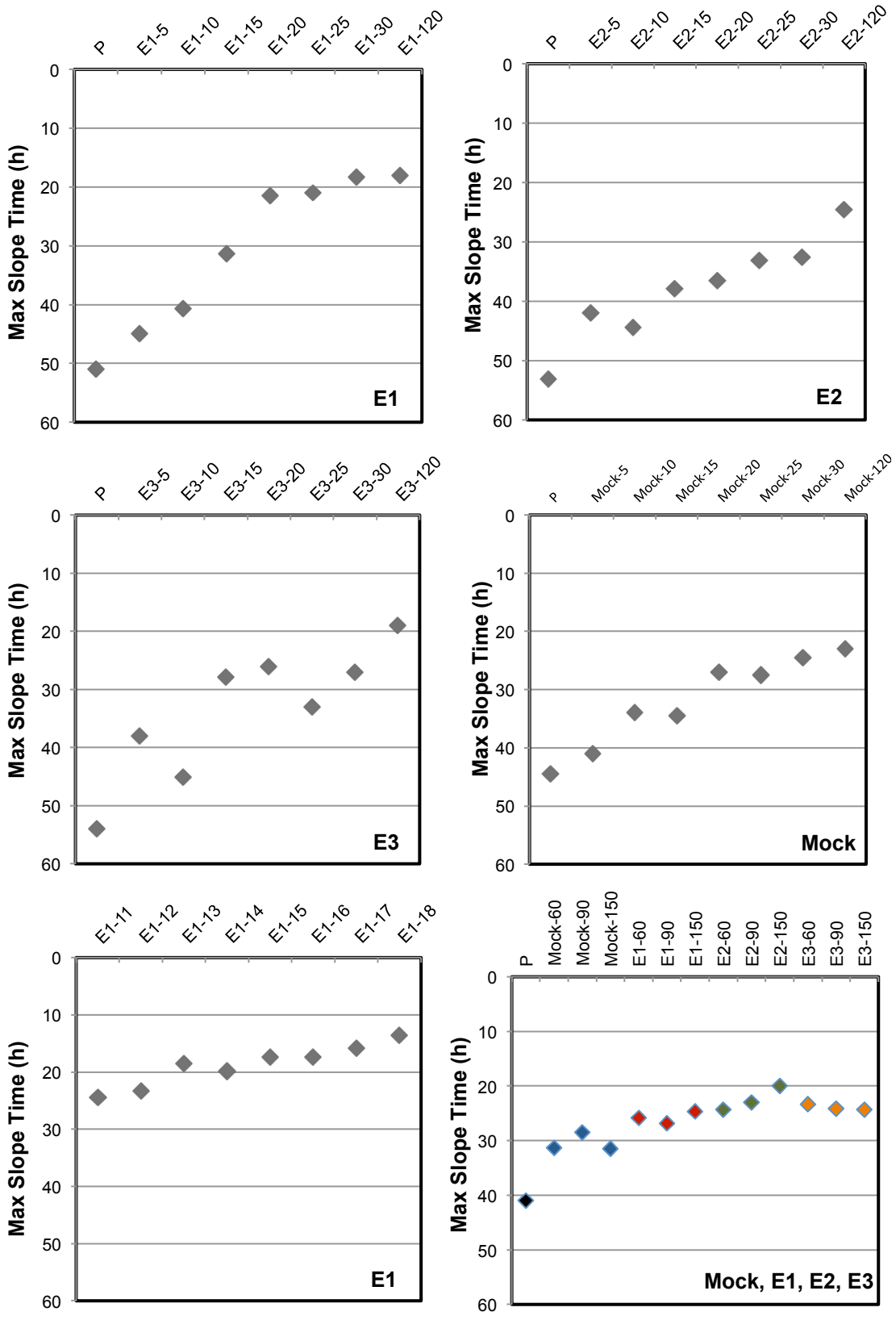


Figure 5.6. Times to maximum growth rate for Mock, E1, E2 and E3. The data in the bottom right plot were acquired in an idependent, parallel experiment.

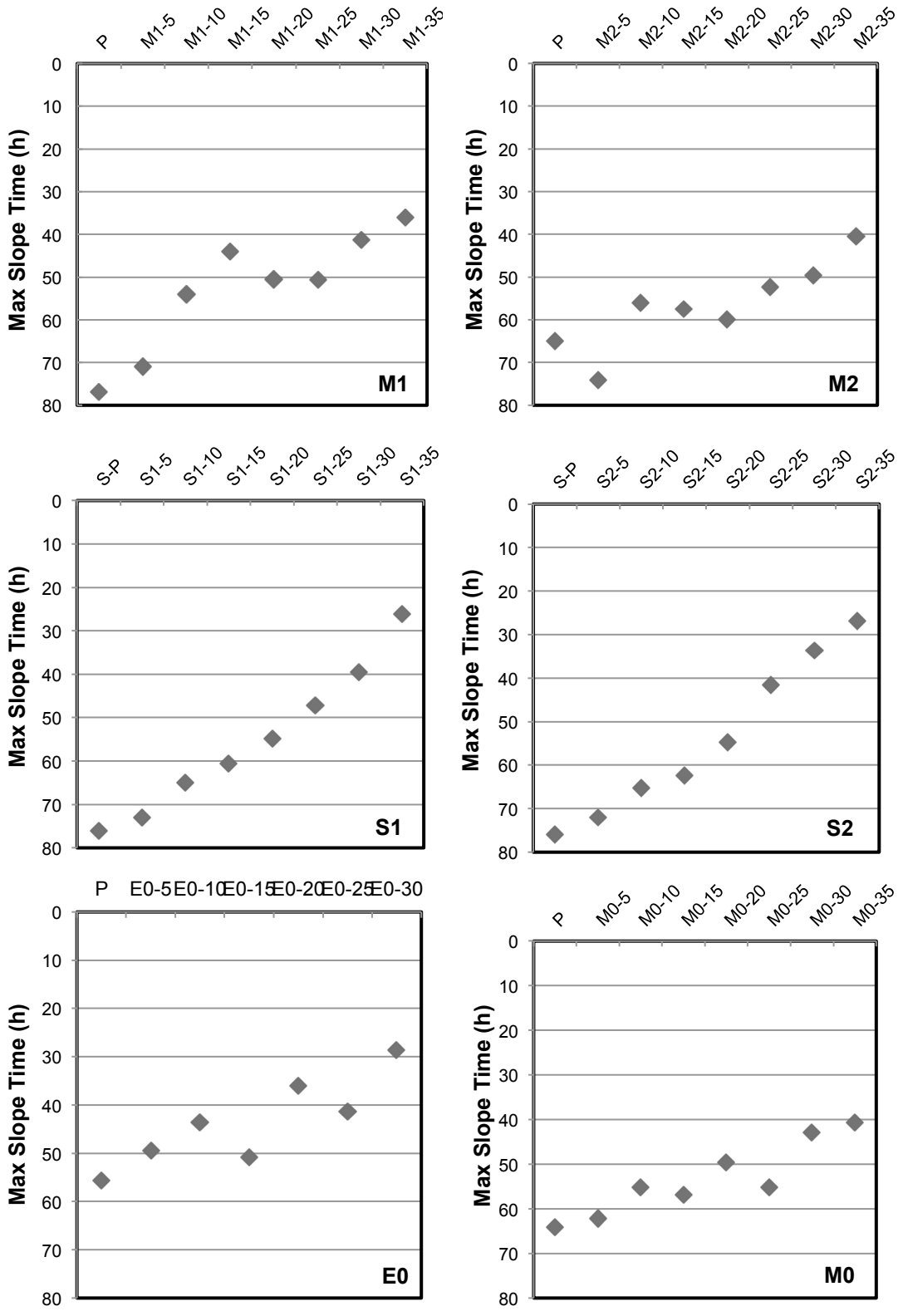


Figure 5.7. Calculations of the time to maximum growth rate for M1, M2, S1, S2, E0, M0.

Analysis of allele frequency trajectories

Genomic DNA was extracted from 96 samples for deep sequencing using a single Illumina HiSeq lane (see Chapter 2). These 96 samples were identical to the ones that were phenotyped (see the overview in Fig. 5.4), except for the cycles between 10 and 20 for Mock, E1, E2, and E3. For the latter, we sequenced every other cycle, i.e. E1-10, E1-12, E1-14, E1-16, etc... Note that the parental line P, as well as S-P, only needed to be sequenced once. Dr. Daniel Jeffares ran a pipeline for SNP calling. Indel calls have not yet been identified.

The analysis is still in progress and what is presented here has been carried out in close collaboration with Dr. Ville Mustonen who applied a method that he and his team have designed to observe allele frequency changes (Illingworth et al. 2012). Here, allele frequency trajectories were modelled for each SNP according to the mathematical framework described for the unlinked driver model in their publication (see Chapter 2). This is different from Popoolation2 which calculates allele frequency differences in a pairwise comparison, only looking at two time points at a time. In the analysis here, allele counts per time point and per locus are calculated and using Mathematica, trajectories are modelled for each variant. Since allele frequency estimations are more reliable for SNPs located in regions with large sequencing depth, this modelling takes into account sequence coverage, i.e. it assigns large estimation errors for SNPs in locations with shallow sequencing depth, and smaller distribution widths for SNPs in high coverage locations.

Next a score ΔS is calculated for each variant. This score is the difference between the likelihood a variant's trajectory is under selection and the likelihood it fits a neutral

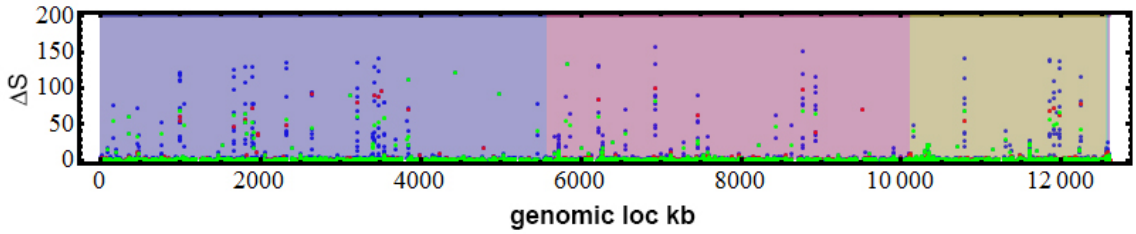


Figure 5.8. Genome-wide score differences estimating the likelihood a variant is under selection. The blue dots represent variants from the nine heat shock evolved lineages. The red dots represent variants from the mock population. Green represents the mean score calculated over all but the mock experiment. The background colours represent the 3 chromosomes.

model (see Chapter 2). In Figure 5.8, ΔS is plotted for every loci for all 10 experiments taken together.

What is striking from these plots is that many variants seem to occur in the same loci for several independent experiments. Analysing the intersection lists of these SNPs whose trajectories fit a selection model led to the identification of 9 events that are shared amongst all 10 evolution experiments (Table 5.1). Of these 9 events, 7 had been previously identified in the Popoolation2 analysis as being shared between mock and evolved populations (see Chapter 4).

Looking at the known functions of these genes, some could be involved in biochemical and physiological pathways by which these lines might have improved their transition to cell growth and division. *Eng1* encodes a glucanase that is involved in septation (Martín-Cuadrado et al. 2003). SPAC1486.06 belongs to the family of glycosyltransferases (Pombase database). *Orb4* is essential during cytokinesis and polarized growth (Cortés et al. 2005). *Alp1* mutants result in defects in microtubule function (Radcliffe et al. 1999). SPCTRNAGLN.05 is a tRNA, and SPBC27B12.12c encodes a magnesium ion transporter (Pombase database). Finally, *tps1* is required for

SNPs					
# CHROM	POS	Annotation	REF	ALT	Region
chr1	**999959**	**eng1**	**C**	**T**	**coding**
chr1	**999960**	**eng1**	**C**	**T**	**coding**
chr1	1664907	near 5'UTR of SPAC343.21	T	A	intergenic
chr1	**3204845**	**SPAC1486.06**	**G**	**A**	**coding**
chr1	**3479252**	**tps1**	**C**	**T**	**5'UTR**
chr2	**1352172**	**SPBC27B12.12c**	**G**	**A**	**coding**
chr2	**3194729**	**SPBC776.10c**	**G**	**A**	**coding**
chr2	**3357187**	**near 5'UTR of alp1**	**G**	**A**	**intergenic**
chr3	**679772**	**orb4**	**T**	**C**	**coding**
chr3	1862803	near SPCTRNAGLN.05	A	G	intergenic

Table 5.1. List of 9 SNPs that show likely selection trajectories in all 10 experiments. One is a double event in *eng1*. In bold font are the SNPs that had previously been identified as shared between Mock-120, E1-120, and E2-120 (see Table 4.4 in Chapter 4).

the synthesis of trehalose and is induced by the stress response (Soto et al. 1999). The function of the other targets is unknown.

Examining the plots of the allele frequency trajectories resulted in some interesting observations. In Figure 5.9 an example is given of what the raw trajectories look like before applying the model prediction.

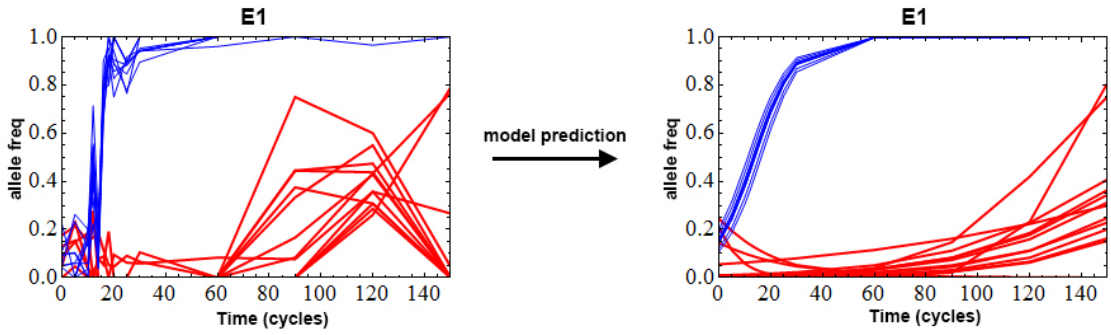


Figure 5.9. Allele frequency trajectories before and after applying the model prediction (Illingworth et al. 2012). The allele frequency trajectories in blue are from 10 SNPs that are shared between all lines. The trajectories in red are seen in E1 but not in Mock.

Because previous results revealed phenotypic differences between the mock and the heat shock evolved lines (see Chapter 4), in addition to the 9 common SNPs, here we also plot those trajectories that differ from Mock (Fig. 5.10). These are not necessarily shared amongst all the other experiments. They were obtained by determining every likely selection trajectory for each heat shock evolved line, and removing the ones that were found in a list of all possible trajectories for Mock.

A first observation is that the 9 common events spread as one haplotype block in each line (Fig. 5.10). Furthermore, overall, the timings of their appearance seem to coincide visually with the phenotype dynamics in the Biolector readouts, though this correlation still needs to be tested quantitatively. This suggests that they may be the causative alleles for the lag related phenotype. However, since no meiotic recombination occurred over the course of the evolution experiment, only 1 of these SNPs may be causative for the observed phenotype whilst the others are 'hitchhikers'. In fact, since so far only SNPs have been considered, these 9 events may be linked with a structural variant we have yet to identify.

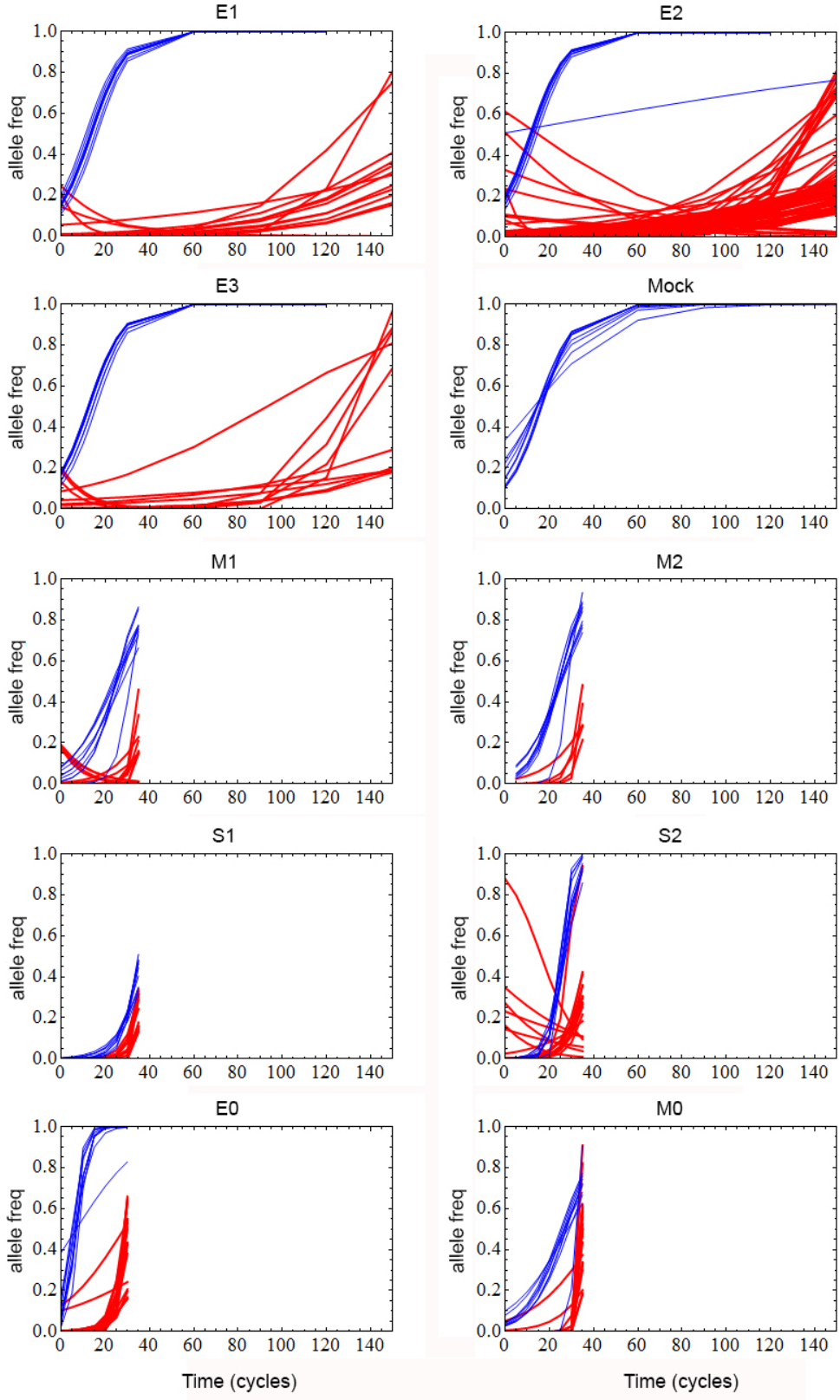


Figure 5.10. Predicted allele frequency trajectories for all 10 evolution lines. The selection trajectories in blue are shared amongst all lines. The trajectories traced in red are absent from Mock.

Several observations are indicative of standing genetic variation. First, not only do the 9 common SNPs share the same genomic regions, but their genomic position and molecular substitutions are identical. It seems very unlikely that the same SNPs would have affected 10 populations independently as *de novo* mutations. Second, standing variation could explain the rapid rate of adaptation, as copies of the haplotype beneficial under the selection regime may already exist in the parental population. Indeed, looking at the allele frequency values at the beginning of each experiment shows that many of these SNPs are present in a substantial fraction of the population. For instance, the maximum estimation of the allele frequency for the standing variation haplotype is ~0.2. However, this explanation is not sufficient for the single colony experiment, which was added to distinguish effects of standing variation from *de novo* mutation. Here, the expectation is that the starting strains contain no standing variation. S1 and S2 both share the haplotype as well, though it arises later. This is unlikely reflecting the occurrence of *de novo* mutations, as all 9 SNPs would have to start in one single cell and this would take much longer to spread through the population. A possible explanation is that a small cross-contamination has occurred early in the experiment. Selection still acts on standing variation, but on a background containing a very small proportion of the beneficial haplotype (note that the allele frequency in S1 and S2 starts at 0). We cannot exclude this possibility, because the single colony experiment was run in parallel with the lines M1 and M2 that started with the parental population. However, if beneficial mutations took longer to spread through populations S1 and S2, this does not correlate with the Biolector readouts where changes in the maximum slope times happen rapidly (Fig. 5.7).

Another observation is that the same haplotype sweeps rapidly through each population, reaching fixation in the lines which evolved over 150 cycles, and also in S2 and E0 (allele frequency = 1). Selection was strong and there is no sign of clonal interference. Given the common haplotype in all 10 lineages, including Mock, how can one explain the specific resistance to long stress described in Chapter 4 for the heat shock evolved populations? To address this question, we examined the predicted trajectories for each variant that is absent in Mock (traced in red in Fig. 5.10). For M1, M2, and also E0 and M0, additional events can be seen that seem independent from the common haplotype, as they arise later (in E0, they start after the haplotype has been fixed). However, since the Biolector readouts may not reflect resistance to 40 min HS, we lack high-resolution viability counts for these lines. If they do exhibit HS tolerance, it will be interesting to test whether the trajectories of these SNPs correlate with the phenotype dynamics. When examining the red trajectories for E1, E2, and E3, they sweep much later through the populations, after cycle 60. We have previously shown that the HS tolerant phenotype appeared early, with large improvements occurring between 10 and 20 cycles (see Fig. 4.5 and 4.6 in Chapter 4). Therefore, the trajectories here are likely unrelated to the HS resistance phenotype, and indicate the emergence of another trait. The precise targets of these SNPs that are not shared with Mock have yet to be analysed.

The haplotype is likely a segregant

Since the parental background was obtained by crossing 3 laboratory wild-type strains (h-, h+, and h90), it consists of a very large number of segregants, and may also include 'pure' h-, h+, and h90 cells. Since two of these laboratory strains, h- and h90, had been sequenced, I examined if any of the 9 events that constitute the common haplotype was

present in these strains. This was done by looking directly at their mapped reads using IGV (see Chapter 2). Out of the 9 mutations, 4 were seen in h90 (Fig. 5.11). These were the ones targeting *orb4*, SPAC1486.06, SPAC343.21 and SPCTRNAGLN.05. The other 5 were not seen in either h- or h90. Possibly they therefore derive from Heim's h+ strain which is the 3rd strain that was used, and has not been sequenced.

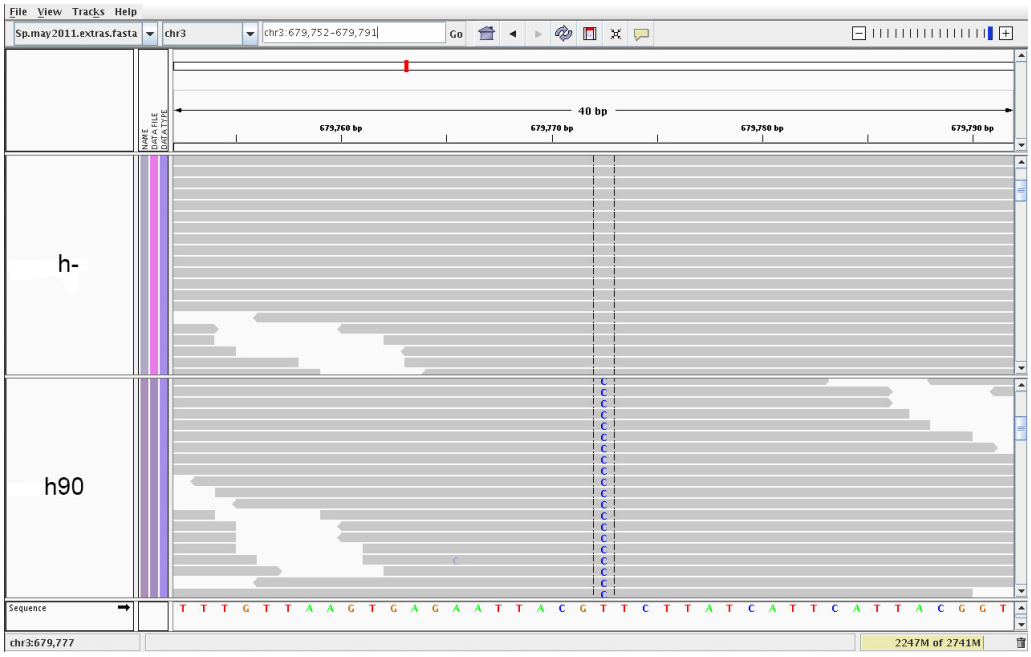


Figure 5.11. Sequenced reads for h- and h90 visualized in IGV. The h90 strain carries the T-to-C substitution at position 679772 on chromosome 3 (targeting *orb4*) that is seen in the haplotype.

To test this hypothesis, DNA fragments containing the SNPs were PCR amplified using h+ genomic DNA. Parental genomic DNA was used as a control. These samples were shipped for traditional Sanger sequencing at UCL. Unfortunately, the sequencing run did not work well and only generated sequence files for 2/5 SNPs with multiple errors. However, both of these show the variant alleles from the haplotype. One of these is shown in Figure 5.12. As such, it is likely that the haplotype consisting of 9 events is a segregant derived from recombining h90 and h+ SNPs. This finding suggests that the

lag-related phenotype is a complex trait, and that rather than a single causative allele, several or all of the SNPs from the haplotype block may be required.

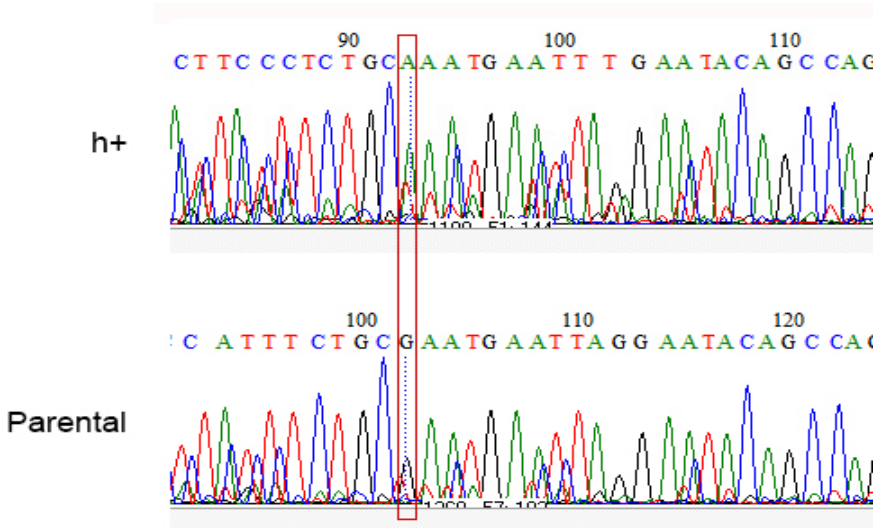


Figure 5.12. The G-to-A substitution in position 3357187 of chromosome 2 is present in the h+ strain. Sequencing chromatograms for DNA from parental and h+ strains. Contoured in red is the position of the tested allele.

Measuring the viability of each strain to the 40 min HS, no differences were detected between h-, h+, h90, and P (data not shown). When looking at exponential growth of each strain, the growth curve for P is closest to h+, though, once again, the stationary phase is reached at a lower cell mass (Fig. 5.13). These results suggest that the novel combination of different SNPs in a segregant haplotype is required for the tested phenotype and has been selected among all segregants in the parental population.

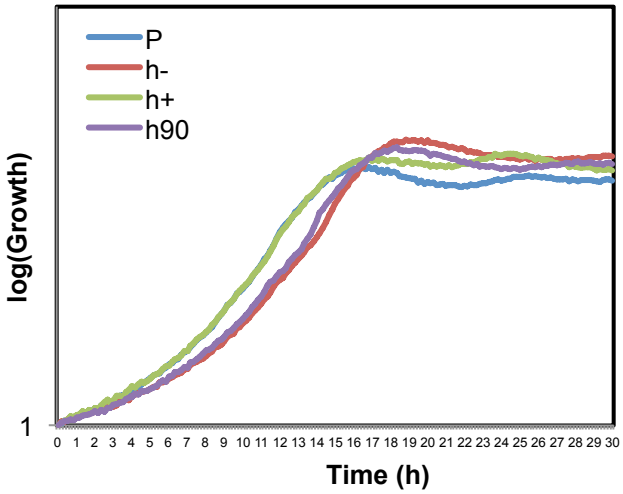


Figure 5.13. Biolector results for growth in the absence of HS, for P, h-, h+ and h90 strains. The cross (P) shows a slightly lower overall fitness in terms of stationary phase cell mass.

Conclusion

We have repeated the evolution experiment for 6 additional lines, including 2 which were exposed to shorter durations of the stress, and 2 which started with a clonal population. A Biolector high-throughput phenotyping method was developed and revealed that all evolution lines have evolved a more rapid recovery of growth following a 40 min HS that was presented in plates. Because the mock line also exhibits this adaptation, this phenotype is likely independent of the heat shock tolerance phenotype described in Chapter 4, which was much more pronounced for the lines experiencing repeated cycles of HS. This evolution appears rapidly with gradual increases over 35 cycles of selection. Analysis of the allele frequency trajectories reveals a shared haplotype of 9 mutations that swept to fixation (or near-fixation) in all 10 populations. Overall, these trajectories seem to correlate with the emergence of the lag related phenotypes, indicating that the common haplotype consists of potentially causative alleles. Furthermore, this haplotype appears to combine SNPs from both h+

and h90 strains, two of the 3 strains that were used to generate the parental population consisting of segregants among the three strains. The single colony experiment remains inconclusive as the haplotype is also found in the populations which were presumed isogenic, raising the possibility that a cross-contamination has occurred between different lines handled in parallel. Finally, to elucidate the heat shock tolerance phenotype, we looked at trajectories that are not shared with Mock. None of these trajectories seem to correlate with the emergence of the specific HS phenotype in E1, E2, and E3. Indel calls and identification of structural rearrangements remain to be done, and will hopefully elucidate the molecular basis for the evolved tolerance against extended heat.

Chapter 6

General discussion

General discussion

In this chapter, I will discuss further the main results from previous chapters and put them into a wider context. I will first examine the new, evolved phenotypes that have been discovered, followed by a discussion of the main findings of our genome sequencing analysis. Promising avenues for future work will also be highlighted.

Evolved populations show fitness improvements in new regimes

We performed experiments to investigate the evolutionary adaptation of fission yeast cells to stress. The basic setup of the experiment consisted of daily serial transfers to new growth media and a stress. Populations grew exponentially upon transfer to new nutrients and were then treated with a 10 min HS, at which point they entered a transient growth arrest before they resumed cell growth and division. In a first round of experiments, 3 populations underwent 150 cycles of growth and heat shock treatment. One mock treated line was added to distinguish between adaptation through stress and other selective pressures from the experimental regime. These 150 cycles encompassed more than 1200 generations. Later, the experiment was repeated on a further 6 lines for over 35 cycles. With the exception of 2 lines, which started as clonal populations, all these lineages were founded by the same parental progenitor cells which contained a small degree of genetic variation as it was obtained by inter-crossing 3 laboratory wild-type strains of mating types h-, h+ and h90.

To examine new phenotypes, we first looked for fitness improvements in assays that mimic the selection regime. To assess differences in resistance to the 10 min HS, viability was measured for both parental and evolved cells. No increased resistance to 10 min HS could be detected, although the plate count assay may not be sensitive

enough to pick up any small improvements. Next, growth rates were measured for each cell line, both during exponential phase, and following heat shock treatment. Evolved lines showed small improvements in overall fitness in both these experiments, with the clearest effect seen for stationary phase which reached a higher cells mass than in the parental cells. The mock treated line showed the same growth-related traits as the ones that were heat-shocked, indicating these improvements are independent from any adaptation to stress. Rather, they reflect a more efficient use of the nutrient medium, possibly by uptake and metabolism of other carbon sources present in the rich YES medium. One way to test this idea would be to grow parental and evolved cells in minimal media containing only glucose and see if the improvements still exist. If the effect is no longer observed, it would show that the evolved lines have adopted the ability to use other sugars as an energy source.

Surprisingly, in all 10 cell lines, robust phenotypes were found in conditions that differed from the evolutionary regime. First, a heat shock of 40 min kills ~95 % of the parental cells, but only 40-60% of cells from the lines that have been heat shock evolved over 120 cycles. The mock cells at cycle 120 also exhibited some increased resistance but far below the levels observed in the heat shocked lines (~90% killing). A similar effect was seen when treating these cell lines with a severe dosage of H_2O_2 (see Fig. 4.9 in Chapter 4).

Based on studies which report that cells can adapt predictive behaviours, part of the reason for the experimental setup was to cause uncoupling of the short term alarm response from the long term transcriptional response that is induced in the presence of a mild HS, and that then protects the cells against enduring and future stresses. The observed phenotype, however, was the opposite from what we expected. Despite the

fact these populations were repeatedly exposed to short heat shocks, followed by ideal growth conditions, they have adapted increased resistance to long stress. Because these differences are not seen in Mock indicates that it is the result of adaptation to the heat shock regime. Microarray analysis revealed very similar gene expression profiles between parental and evolved cells following HS, although there was a potential set of stress-related genes that showed small differential expression before the onset of the HS. Possibly, cells have adapted an advanced preparation to the heat shock treatment, which could explain the increased survival during 40 min treatments.

Another strong phenotype was observed in all 10 evolution lineages when exposing them to 40 min HS in well plates. While this assay was initially thought to act as a high-throughput phenotyping method to assess viability of cells to 40 min HS, it may reflect a separate characteristic: recovery of growth following severe stress. In these experiments, all 10 lines (including Mock) showed striking improvement in the speed of their physiological recovery from the 40 min HS. In the future, it will be important to disentangle the two confounding factors in this Biolector assay. One being cell mortality during the 40 min HS, which contributes to delayed cell mass increases at a population level, while the other is the speed of recovery from a growth arrest. Note that the latter phenotype will confer selective advantage both for cells undergoing the HS regime and for cells undergoing the mock regime, as both regimes involve repeated cycles of re-growing cells from non-growing states. It is possible that the 40 min HS in the well plates is weaker than the one presented in flasks and that the Biolector readouts mostly reflect a lag related phenotype. This could be tested by using the plate count assay to measure cell viability from samples taken directly from the well plates.

These fitness gains are remarkable because, evidently, the conditions in these assays do

not bear directly on the reproductive success of the cells when they are being propagated through the cycles of selection. This is reminiscent of generalist genotypes, i.e. organisms evolving the ability to consume a wide variety of substrates (Barrett et al. 2005). These are known to be favoured in fluctuating environments, as opposed to specialist genotypes which are favoured in constant environments and adapt one specific phenotype (Cooper and Lenski 2000). In one study, *E. coli* populations were evolved for 2000 generations in several regimes that differed in the type of carbon source. Interestingly, the lines that had adapted to glucose as a sole energy source lost fitness when tested in a lactose-only environment. However, all lactose-evolved populations displayed some increased fitness when grown in glucose compared to the ancestor (Cooper and Lenski 2010). It would be interesting to test whether the Biolector readouts for cell recovery after treatment with 40 min HS correlate with the small gains in growth we measured in the experimental regime. For these, so far we have only tested 1 time point, at cycle 120. Many more repeats would be needed for this experiment to apply statistics.

In retrospect, the new phenotypes in the heat shock evolved cells may not be surprising. While we postulated that cells would decrease their tolerance to long stress and increase their speed of recovery following the short HS in the evolution assay, they seem to have evolved a phenotype in which they do recover more rapidly yet have also have become more resistant to stress. This finding is interesting given the induction of stress-related gene expression programmes involves the shutting down of growth-related gene expression programmes, causing transient cell cycle arrest (Lopez-Maury et al. 2008). Despite the fact this still holds true as seen from the microarray expression profiles, these populations seem to have found a way to partially side-step this trade-off,

allowing them to recover more rapidly while still being prepared for severe stress.

Adaptation from standing variation

To date, not many evolution experiments exist for which whole-genome sequencing data are available. Most of these studies are from asexual systems, initiated with clonal populations, studying the adaptation based on new mutations. Sexual systems can be used to study adaptation from standing genetic variation, though there are little direct experimental data describing this process. One study, which was designed as a QTL mapping tool, described a strategy capable of informing us about mutational dynamics in *S. cerevisiae* populations with standing variation (see Introduction; Parts et al. 2011). In this strategy, strains of different heat tolerance were crossed to generate large segregants pools. These pools were subjected to heat stress, enriching for causative alleles required for stress resistance. The pools were sequenced at several time points and allele frequency differences were monitored (Fig. 6.1).

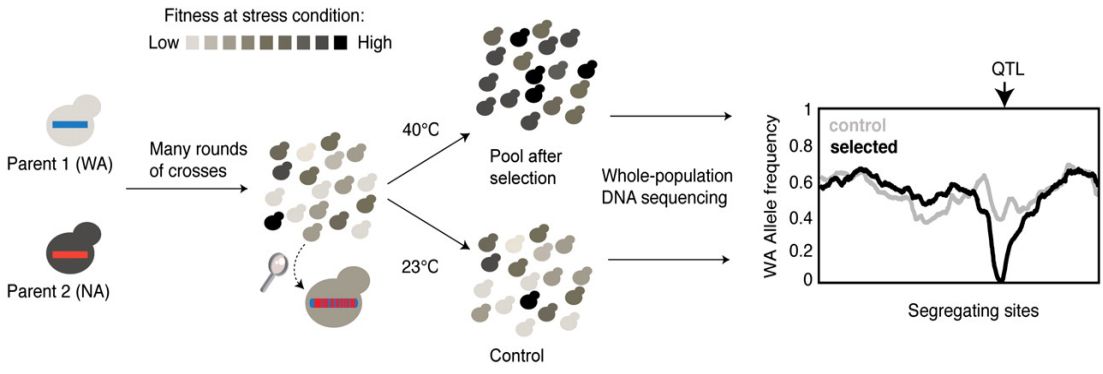


Figure 6.1. Schematic of a three-step QTL mapping strategy. Isogenic parental strains with diverse phenotypic effects undergo multiple rounds of sexual recombination, generating a pool of millions of segregants. This pool is then forced through rounds of selection, and the populations are monitored for changes in allele frequency. This figure is taken from Parts et al. (2011).

No complete fixation was found, though the experiment encompassed only 200 selection generations, likely not giving enough time for standing variants or *de novo* mutations to fix. While it was not our main goal to study mutational dynamics of standing variation, our experimental approach bears similarities with the one described by Parts et al. (2011). By crossing 3 laboratory strains, many different haplotypes were generated, and these were subjected to cycles of selection in the absence of further meiotic recombination. Populations were sequenced at various time points and allele frequency trajectories were determined. Theoretically, such an experimental setup allows for both the spreading of new mutations, or sweeps from the standing variants. And this is reflected in our findings. Already in the initial sequencing analysis in Chapter 4, where we used Popoolation2 to detect differences in allele frequencies: of the 31 combined events (see Table 4.5), 25 already existed in the Parental population at a very low frequency. The selective sweep dynamics described in Chapter 5 provide strong genetic evidence for a haplotype block consisting of 9 SNPs that has swept through all 10 populations. These data are indicative of a strong selection on a small fraction of each population. This is somehow surprising since we could not find dramatic fitness increases for the evolved populations under the experimental regime, although increased efficiency of using the growth medium (higher biomass in stationary phase) and reduced delay of re-entering cell growth and division (shortened lag phase) will certainly be advantageous across several cycles. Because this group of SNPs also sweeps through the mock treated population, we conclude that its selection was independent from the HS treatments. This also indicates that the rich nutrient medium is highly selective, possibly reflecting that it is rarely, if ever, encountered by cells in the 'wild'. Furthermore, while correlations will need to be quantified, the selective

sweep dynamics of this haplotype seem consistent with the occurrence of growth recovery gains from the Biolector readouts. As such, the haplotype is likely responsible for the observed changes in growth recoveries. Analysis of the 9 SNPs that constitute the haplotype shows that most have functions related to cell growth and division, most notably e*ng1,* which is involved in septation, and *orb4,* which is required for cytokinesis and polarized growth. One exception is *tps1,* which is required for the synthesis of trehalose. Moreover, we have shown that this haplotype is likely a recombinant of two laboratory strains, indicating that the association of at least some of the SNPs in the haplotype block is required for the phenotype. However, since in our protocol, h90 and h+ strains were only crossed once, this may not have been sufficient to reduce linkage between desirable and undesirable or neutral loci. Some of these SNPs may still be hitchhikers. It is also possible that the new combination of different SNPs is important for the phenotype by creating particular genetic interactions.

Our results are in line with some classical views on adaptation via standing variation. First, there are good reasons why populations adapt mainly from standing genetic variation rather than waiting for *de novo* mutations to arise. Beneficial mutations can be immediately available, and be present in the population at a relatively high frequency (Innan and Kim 2004). It is also believed that as soon as the sweep of one allele occurs, standing genetic variation is wiped out (Burke 2012). We find that in our allele frequency trajectories, most of the SNPs belonging to the common haplotype already exist at a certain frequency before the experimental evolution (see Fig. 5.10 in Chapter 5), though for most of the trajectories that arise later, this is not the case. While this needs to be fully analyzed, the trajectories which are not shared with Mock could be selective sweeps from *de novo* mutations. The best illustration for this is seen in E0

where trajectories arise after the haplotype is fixed. Consistent with the model of periodic selection (Atwood et al. 1951), there is a period during which the population 'waits' for the next beneficial mutations to arise. This is less clear for S1 where the additional trajectories move so close to the ones from the common haplotype that they could have been linked to it from the start.

These sweep patterns do not fit the clonal interference model which is thought to slow adaptation (Gerrish and Lenski 1998; Hermisson and Pennings 2005; de Visser and Rozen 2006). None of the 10 populations displays the coexistence of multiple ancestral haplotypes that interfere with the selective sweep of a single haplotype.

While we knew we had some variation in the parental population, arguably the 3 laboratory wild-type strains are almost genetically identical. This is interesting in its own right as it shows how few genetic variations, and new combinations thereof, in laboratory strains presumed to be identical can have a large impact on physiological responses.

Future perspectives

While there are good indications that the common haplotype that sweeped the different cell lines is responsible for the lag related phenotype, it is still likely that some alleles are neutral hitchhikers given the absence of meiotic recombination. Like in the long-term evolution experiment with *E. coli*, one could construct 9 isogenic strains with the derived alleles. Each of these genetic variants can then be tested for phenotypic effects. However, such strategy would fall short of distinguishing if one allele is beneficial in combination with others from the haplotype block. This would need to be tested by recreating different combinations of the 9 SNPs in yeast cells and check for resulting phenotypic effects.

S. pombe offers the same high meiotic recombination rate as *S. cerevisiae*, and therefore one could apply published innovative approaches to map trait loci (Fig. 6.1) (Ehrenreich et al. 2010; Parts et al. 2011). By subjecting our evolved populations to many rounds of crosses with the parental background, we would generate a pool of several million highly recombined haplotypes, thereby unlinking adaptive mutations from other genomic variation. Applying selective pressure by growing the pools in the evolution regime will enrich for fit individuals with beneficial alleles. If we then sequence the pools at various time points, and analyse for allele frequency differences as described in Chapter 5, new sweeps of standing variation should be detectable. With this strategy we should be able to dissect which SNPs in the haplotype block are responsible for the changes in growth recovery after HS.

Our current data still beg the question of which are the causative alleles for the HS tolerant phenotype. Since the Mock line did not display this trait to the same extent, this phenotype is the result of selection caused by the 10 min HS. In Chapter 4 we have listed candidates of alleles which we detected in E1 and E2 but not in Mock. Since 7 out of the 9 SNPs from the haplotype were also detected in our Popoolation2 analysis, this validates the approach we took in Chapter 4. However, none of the trajectories for E1, E2, and E3, other than the common haplotype, emerge before cycle 60 (see Fig. 5.10 in Chapter 5), which is much later than the appearance of the HS tolerant phenotype. Such trajectories indicate the emergence of another selected trait. Therefore, the candidate alleles we identified to be unique to E1 and E2 at cycle 120, could be associated to another adaptation.

Since so far we have only considered SNPs, it will be interesting to apply the analysis for indels. This should be straightforward since both the pipeline for variant calling and

the method to determine selection trajectories are identical for indels. This may reveal new trajectories, some that could coincide with the emergence of the HS resistance. Likewise, using the strategy described above to filter out alleles that are unrelated to growth recovery, one could subject HS evolved populations to many rounds of crosses with the parental or mock cells. Since the evolved cells are resistant to 40 min HS, growing these segregants pools to rounds of selection in which they are exposed to 40 min HS should strongly enrich for those variants that confer HS resistance.

The question has been raised whether epigenetics might be the underlying cause for the improvements in cell survival during HS. In fission yeast, DNA methylation has not been observed (Becker et al. 2012), but histone modifications are present and functional (Sinha et al. 2006; Shimada and Murakami 2010). While it is not clear for how many generations the transmission of a phenotype that is under epigenetic control can be observed, most report epigenetic effects for up to 3 generations (Curley et al. 2011). In our survival assay, this number is greatly exceeded as frozen samples are first revived on agar plates until colony formation, and then transferred to flasks where they grow exponentially overnight. Still, the phenotype persists. Moreover, through the remodelling of chromatin, one would expect to find global changes in gene expression, which we could not detect in the microarray analysis. As such, we strongly favour that the adaptation happened at the DNA level rather than through epigenetic influences.

Finally, it is unfortunate that the single colony experiments remain inconclusive. The safest way to repeat these experiments is not to run them in parallel with other lines. To date, a direct comparison of adaptation from new mutations versus adaptation from standing variation has not been tested in a single experimental system. Certainly, this project has shown that this is a realistic future goal, and it has laid down a useful

framework for targeted follow-up studies and for zooming into the molecular causes of the observed newly acquired phenotypes.

References

Andersson, D. I. and Hughes, D. 2009. Gene amplification and adaptive evolution in bacteria. *Annu Rev Genet*, 43, 167–95.

Atwood, K. C., Schneider, L. K. Ryan, F. J. 1951. Periodic Selection in *Escherichia coli*. *Proc Natl Acad Sci USA*, 37, 146–155.

Barrett, R. D. H., MacLean, R. C., Bell, G. 2005. Experimental evolution of *Pseudomonas fluorescens* in simple and complex environments. *Am Nat*, 166, 470–480.

Barrett, R. D. H. and Schluter, D. 2008. Adaptation from standing genetic variation. *Trends Ecol Evol*, 23, 38–44.

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

Becker, M., Müller, S., Nellen, W., Jurkowski, T. P., Jeltsch, A., Ehrenhofer-Murray A. E. 2012. Pmt1, a Dnmt2 homolog in *Schizosaccharomyces pombe*, mediates tRNA methylation in response to nutrient signaling. *Nucleic Acids Res*, 40, 11648-11658.

Becks, L. and Agrawal, A. F. 2012. The evolution of sex is favoured during adaptation to new environments. *PLoS Biol*, 10, e1001317.

Bell-Pedersen, D., Cassone, V. M., Earnest, D. J., Golden, S. S., Hardin, P. E., Thomas, T. L., Zoran, M. J. 2005. Circadian rhythms from multiple oscillators: lessons from diverse organisms. *Nat Rev Genet,* 6, 544–556.

Bennett, A. F., Dao, K. M., Lenski, R. E. 1990. Rapid evolution in response to high temperature selection. *Nature*, 346, 79-81.

Bennett, A. F. and Lenski, R. E. 1996. Evolutionary Adaptation to Temperature. V. Adaptive Mechanisms and Correlated Responses in Experimental Lines of *Escherichia coli*. *Evolution*, 50, 493–503.

Bergman, A. and Siegal, M. L. 2003. Evolutionary capacitance as a general feature of complex gene networks. *Nature*, 424, 549–552.

Berlanga, J. J., Rivero, D., Martín, R., Herrero, S., Moreno, S., de Haro C. 2010. Role of mitogen-activated protein kinase Sty1 in regulation of eukaryotic initiation factor 2alpha kinases in response to environmental stress in *Schizosaccharomyces pombe*. *Eukaryot Cell*, 9, 194–207.

Berry, D. B. and Gasch, A. P. 2008. Stress-activated genomic expression changes serve a preparative role for impending stress in yeast. *Mol Biol Cell*, 19, 4580–4587.

Berry D. B., Guan Q., Hose J., Haroon S., Gebbia M., Heisler L. E., Nislow C., Giaever G., Gasch A. P. 2011. Multiple means to the same end: the genetic basis of acquired stress resistance in yeast. *PLoS Genet*, 7, e1002353.

Bjedov I., Tenaillon O., Gérard B., Souza V., Denamur E., Radman M., Taddei F., Matic I. 2003. Stress-Induced Mutagenesis in Bacteria. *Science*, 300, 1404–1409.

Blount Z. D., Barrick J. E., Davidson C. J., Lenski R. E. 2012. Genomic analysis of a key innovation in an experimental *Escherichia coli* population. *Nature*, 489, 513–518.

Boland, C.R. and Goel, A. 2005. Somatic evolution of cancer cells. *Semin Cancer Biol*, 15, 436–450.

Brown, W. R., Liti G., Rosa C. et al. 2011. A Geographically Diverse Collection of *Schizosaccharomyces pombe* Isolates Shows Limited Phenotypic Variation but Extensive Karyotypic Diversity. *G3*, 1, 615–626.

Burke M. K., Dunham J. P., Shahrestani P., Thornton K. R., Rose M. R., Long A.D. 2010. Genome-wide analysis of a long-term evolution experiment with *Drosophila*. *Nature*, 467, 587–590.

Burke, M. K. 2012. How does adaptation sweep through the genome? Insights from long-term selection experiments. *Proc Roy Soc Lond B,* 279, 5029-5038.

Cairns, J. 1998. Mutation and cancer: the antecedents to our studies of adaptive mutation. *Genetics*, 148, 1433–1440.

Le Chat, L., Fons, M, Taddei, F. 2006. *Escherichia coli* mutators: selection criteria and migration effect. *Microbiology*, 152, 67–73.

Chen D., Toone W. M., Mata J., Lyne R., Burns G., Kivinen K., Brazma A., Jones N., Bähler J. 2003. Global transcriptional responses of fission yeast to environmental stress. *Mol Biol Cell*, 14, 214–229.

Colegrave, N. 2002. Sex releases the speed limit on evolution. *Nature*, 420, 664–666.

Cooley M. B., Carychao D., Nguyen K., Whitehand L., Mandrell R**.** 2010. Effects of environmental stress on stability of tandem repeats in *Escherichia coli*. *Applied Environ Microbiol*, 76, 3398–3400.

Cooper, T. F. 2007. Recombination speeds adaptation by reducing competition between beneficial mutations in populations of *Escherichia coli*. *PLoS Biol*, 5, e225.

Cooper, T. F., and Lenski, R. E. 2010. Experimental evolution with *Escherichia coli* in diverse resource environments. I. Fluctuating environments promote divergence of replicate populations. *BMC Evol Biol*, 10, 11.

Cooper, V. S. and Lenski, R. E. 2000. The population genetics of ecological specialization in evolving *Escherichia coli* populations. *Nature*, 407, 736–739.

Cortés J. C., Carnero E., Ishiguro J., Sánchez Y., Durán A., Ribas J. C. 2005. The novel fission yeast (1,3)beta-D-glucan synthase catalytic subunit Bgs4p is essential during both cytokinesis and polarized growth. *J Cell Sci*, 118, 157–174.

Costanzo, M. S., Brown, K. M. and Hartl, D. L. 2011. Fitness trade-offs in the evolution of dihydrofolate reductase and drug resistance in *Plasmodium falciparum*. *PLoS ONE*, 6, e19636.

Cowen, L. E. and Lindquist, S. 2005. Hsp90 Potentiates the Rapid Evolution of New Traits: Drug Resistance in Diverse Fungi. *Science*, 309, 2185–2189.

Curley, J. P., Mashoodh, R., Champagne, F. A. 2011. Epigenetics and the origins of paternal effects. *Hormones and Behavior*, 59, 306–314.

Darwin, C. 1859. On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life. *England: John Murray*.

Degols, G., Shiozaki, K. and Russell, P. 1996. Activation and regulation of the Spc1 stress-activated protein kinase in *Schizosaccharomyces pombe*. *Mol Cell Biol*, 16, 2870–2877.

Denamur E., Bonacorsi S., Giraud A., Duriez P., Hilali F., Amorin C., Bingen E., Andremont A., Picard B., Taddei F., Matic I. 2002. High Frequency of Mutator Strains among Human Uropathogenic *Escherichia coli* Isolates. *J Bacteriol*, 184, 605–609.

Dodgson J., Avula H., Hoe K. L., Kim D. U., Park H. O., Hayles J., Armstrong J. 2009. Functional genomics of adhesion, invasion, and mycelial formation in *Schizosaccharomyces pombe*. *Eukaryot Cell*, 8, 1298–1306.

Drake J. W., Charlesworth B., Charlesworth D., Crow J. F. 1998. Rates of spontaneous mutation. *Genetics*, 148, 1667–1686.

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

Eitinger T., Degen O., Bohnke U., Muller M. 2000. Nic1p, a relative of bacterial transition metal permeases in *Schizosaccharomyces pombe*, provides nickel ion for urease biosynthesis. *J Biol Chem*, 275, 18029–18033.

Elena, S. F. and Lenski, R. E. 2003. Evolution experiments with microorganisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet*, 4, 457–469.

Enright M., Zawadski P., Pickerill P., Dowson C. G. 1998. Molecular evolution of rifampicin resistance in *Streptococcus pneumoniae*. *Microb Drug Resist*, 4, 65–70.

European Bioinformatics Institute, EBI. 2012. "Eukaryotes Genomes – S*chizosaccharomyces pombe*". [Online] <http://www.ebi.ac.uk/2can/genomes/eukaryotes/Schizosaccharomyces_pombe.html>.

Geisel, N. 2011. Constitutive versus responsive gene expression strategies for growth in changing environments. *PLoS ONE*, 6, e27033.

Gerrish, P. J. and Lenski, R. E. 1998. The fate of competing beneficial mutations in an asexual population. *Genetica*, 102-103, 127–144.

Giraud A., Matic I., Tenaillon O., Clara A., Radman M., Fons M., Taddei F. 2001. Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut. *Science,* 291, 2606–2608.

Goddard, M. R., Godfray, H. C., Burt, A. 2005. Sex increases the efficacy of natural selection in experimental yeast populations. *Nature*, 434, 636–640.

Goldenthal, M. J., Vanoni, M., Buchferer, B., Marmur, J. 1987. Regulation of MAL gene expression in yeast: gene dosage effects. *Mol Gen Genet*, 209, 508–517.

Gresham D., Desai M. M., Tucker C. M., Jenq H. T., Pai D. A., Ward A., DeSevo C. G., Botstein D., Dunham M. J. 2008. The repertoire and dynamics of evolutionary adaptations to controlled nutrient-limited environments in yeast. *PLoS Genet*, 4, e1000303.

Hartmuth, S. and Petersen, J. 2009. Fission yeast Tor1 functions as part of TORC1 to control mitotic entry through the stress MAPK pathway following nutrient stress. *J Cell Sci*, 122, 1737–1746.

Heim, L. 1990. Construction of an h+S strain of *Schizosaccharomyces pombe*. *Curr Genet*, 17, 13–19.

Hermisson, J. and Pennings, P. S. 2005. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*, 169, 2335–2352.

Herr A. J., Ogawa M., Lawrence N. A., Williams L. N., Eggington J. M., Singh M., Smith R. A., Preston B. D. 2011. Mutator suppression and escape from replication error-induced extinction in yeast. *PLoS Genet*, 7, e1002282.

Higgs, P. G. and Attwood, T. K. 2005. Bioinformatics and Molecular Evolution, *Blackwell Publishing.*

Hoffmann, G. R., Morgan, R. W., Harvey, R. C. 1978. Effects of chemical and physical mutagens on the frequency of a large genetic duplication in *Salmonella typhimurium*. I. Induction of duplications. *Mutation Res*, 52, 73–80.

Illingworth C. J., Parts L., Schiffels S., Liti G., Mustonen V. 2012. Quantifying selection acting on a complex trait using allele frequency time series data. *Mol Biol Evol* 29, 1187–1197.

Innan, H. and Kim, Y. 2004. Pattern of polymorphism after strong artificial selection in a domestication event. *Proc Natl Acad Sci U S A*, 101, 10667–10672.

Jamieson, D.J. 1998. Oxidative stress responses of the yeast *Saccharomyces cerevisiae*. *Yeast*, 14, 1511–1527.

Jarosz, D. F. and Lindquist, S. 2010. Hsp90 and Environmental Stress Transform the Adaptive Value of Natural Genetic Variation. *Science*, 330, 1820–1824.

Kahm, M., Hasenbrink, G., Lichtenberg-Frat, H., Ludwig, J., Kschischo, M. 2010. grofit: Fitting Biological Growth Curves with R. *J Stat Software*, 33, 1–21.

Karaer, S., Sarikaya, A. T., Arda, N., Temizkan, G. 2006. The 3’ terminal sequence of the inosine monophosphate dehydrogenase gene encodes an active domain in the yeast *Schizosaccharomyces pombe*. *Genet Mol Biol*, 29, 551–557.

Kawecki T. J., Lenski R. E., Ebert D., Hollis B., Olivieri I., Whitlock M. C. 2012. Experimental evolution. *Trends Ecol Evol*, 27, 547–560.

Kimata Y., Matsuyama A., Nagao K., Furuya K., Obuse C., Yoshida M., Yanagida M. 2008. Diminishing HDACs by drugs or mutations promotes normal or abnormal sister chromatid separation by affecting APC/C and adherin. *J Cell Sci*, 121, 1107–1118.

Kimura, M. 1983. The Neutral Theory of Molecular Evolution. *Cambridge University Press.*

Kugelberg E., Kofoid E., Reams A. B., Andersson D. I., Roth J. R. 2006. Multiple pathways of selected gene amplification during adaptive mutation. *Proc Natl Acad Sci U S A*, 103, 17319–17324.

Lackner D. H., Beilharz T. H., Marguerat S., Mata J., Watt S., Schubert F., Preiss T., Bähler J. 2007. A network of multiple regulatory layers shapes gene expression in fission yeast. *Mol Cell*, 26, 145–155.

Lawrence, C.L., Maekawa, H., Worthington, J. L., Reiter, W., Wilkinson, C. R., Jones, N. 2007. Regulation of *Schizosaccharomyces pombe* Atf1 protein levels by Sty1-mediated phosphorylation and heterodimerization with Pcr1. *J Biol Chem*, 282, 5160–5170.

Liti G., Carter D. M., Moses A. M., Warringer J., Parts L., James S. A., Davey R. P., Roberts I. N., Burt A., Koufopanou V., Tsai I. J., Bergman C. M., Bensasson D., O'Kelly M. J., van Oudenaarden A., Barton D. B., Bailes E., Nguyen A. N., Jones M., Quail M. A., Goodhead I., Sims S., Smith F., Blomberg A., Durbin R., Louis E. J. 2009. Population genomics of domestic and wild yeasts. *Nature*, 458, 337–341.

Lopez-Aviles, S., Lambea, E., Moldon, A., Grande, M., Fajardo, A., Rodriguez-Gabriel, M. A., Hidalgo, E., Aligue, R. 2008. Activation of Srk1 by the mitogen-activated

protein kinase Sty1/Spc1 precedes its dissociation from the kinase and signals its degradation. *Mol Biol Cell*, 19, 1670–1679.

Lopez-Maury, L., Marguerat, S., Bahler, J. 2008. Tuning gene expression to changing environments: from rapid responses to evolutionary adaptation. *Nat Rev Genet*, 9, 583–593.

Lyne, R., Burns, G., Mata, J., Penkett, C. J., Rustici, G., Chen, D., Langford, C., Vetrie, D., Bahler, J. 2003. Whole-genome microarrays of fission yeast: characteristics, accuracy, reproducibility, and processing of array data. *BMC Genomics*, 4, 27.

Martín-Cuadrado A. B., Dueñas E., Sipiczki M., Vázquez de Aldana C. R., del Rey F. 2003. The endo-beta-1,3-glucanase eng1p is required for dissolution of the primary septum during cell separation in *Schizosaccharomyces pombe*. *J Cell Sci*, 116, 1689–1698.

Matic I., Radman M., Taddei F., Picard B., Doit C., Bingen E., Denamur E., Elion J. 1997. Highly Variable Mutation Rates in Commensal and Pathogenic *Escherichia coli*. *Science*, 277,1833–1834.

Matsuba C., Ostrow D. G., Salomon M. P., Tolani A., Baer C. F. 2012. Temperature, stress and spontaneous mutation in *Caenorhabditis briggsae* and *Caenorhabditis elegans*. *Biol Lett*, 9, 8–12.

Matsuo Y., Fisher E., Patton-Vogt J., Marcus S. 2007. Functional characterization of the fission yeast phosphatidylserine synthase gene, pps1, reveals novel cellular functions for phosphatidylserine. *Eukaryot Cell*, 6, 2092–2101.

Medeiros, A. A. 1997. Evolution and dissemination of beta-lactamases accelerated by generations of beta-lactam antibiotics. *Clin Infect Dis*, 24, 19–45.

Mitchell, A. and Pilpel, Y. 2009. Adaptive prediction of environmental changes by microorganisms. *Nature*, 460, 220–224.

Moradas-Ferreira, P. and Costa, V. 2000. Adaptive response of the yeast *Saccharomyces cerevisiae* to reactive oxygen species: defences, damage and death. *Redox Rep*, 5, 277–285.

Muller, H. J. 1932. Some genetic aspects of sex. *Am Nat*, 66, 118–138.

Murakami, C.J., Burtner, C. R., Kennedy, B. K., Kaeberlein, M. 2008. A method for high-throughput quantitative analysis of yeast chronological life span. *J Gerontol A Biol Sci Med Sci*, 63, 113–121.

Notley-McRobb, L., Seeto, S., Ferenci, T. 2003. The influence of cellular physiology on the initiation of mutational pathways in *Escherichia coli* populations. *Proc Roy Soc Lond B*, 270, 843–848.

Oh Y., Chang K. J., Orlean P., Wloka C., Deshaies R., Bi E. 2012. Mitotic exit kinase Dbf2 directly phosphorylates chitin synthase Chs2 to regulate cytokinesis in budding yeast. *Mol Biol Cell*, 23, 2445–2456.

Oliver A., Cantón R., Campo P., Baquero F., Blázquez J. 2000. High Frequency of Hypermutable *Pseudomonas aeruginosa* in Cystic Fibrosis Lung Infection. *Science*, 288, 1251–1253.

Parts L., Cubillos F. A., Warringer J., Jain K., Salinas F., Bumpstead S. J., Molin M., Zia A., Simpson J. T., Quail M. A., Moses A., Louis E. J., Durbin R., Liti G. 2011. Revealing the genetic structure of a trait by sequencing a population under selection. *Genome Res*, 21, 1131–1138.

Pearce A. K., Crimmins K., Groussac E., Hewlins M. J., Dickinson J. R., Francois J., Booth I. R., Brown A. J. 2001. Pyruvate kinase (Pyk1) levels influence both the rate and direction of carbon flux in yeast under fermentative conditions. *Microbiology*, 147, 391–401.

Peekhaus, N. and Conway, T. 1998. What's for Dinner ?: Entner-Doudoroff Metabolism in *Escherichia coli*. *J Bacteriol*, 180, 3495–3502.

Powers , R. W., Kaeberlein, M., Caldwell, S. D., Kennedy, B. K., Fields, S. 2006. Extension of chronological life span in yeast by decreased TOR pathway signaling. *Genes Dev*, 20, 174–184.

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

Queitsch, C., Sangster, T. A ., Lindquist, S. 2002. Hsp90 as a capacitor of phenotypic variation. *Nature*, 417, 618–624.

Radcliffe P. A., Hirata D., Vardy L., Toda T. 1999. Functional dissection and hierarchy of tubulin-folding cofactor homologues in fission yeast. *Mol Biol Cell*, 10, 2987–3001.

Ratcliff W. C., Denison R. F., Borrello M., Travisano M. 2012. Experimental evolution of multicellularity. *Proc Natl Acad Sci U S A*, 109, 1595–1600.

Rice, W. R. and Chippindale, A. K. 2001. Sexual Recombination and the Power of Natural Selection. *Science*, 294, 555–559.

Shieh, J. C., Martin, H., Millar, J. B. 1998. Evidence for a novel MAPKKK-independent pathway controlling the stress activated Sty1/Spc1 MAP kinase in fission yeast. *J Cell Sci*, 111, 2799–2807.

Shimada, A. and Murakami, Y. 2010. Dynamic regulation of heterochromatin function via phosphorylation of HP1-family proteins. *Epigenetics*, 5, 30–33.

Shwab E. K., Bok J. W., Tribus M., Galehr J., Graessle S., Keller N. P. 2007. Histone deacetylase activity regulates chemical diversity in *Aspergillus*. *Eukaryot Cell*, 6, 1656–1664.

Sinha, I., Wirén, M., Ekwall, K. 2006. Genome-wide patterns of histone modifications in fission yeast. *Chromosome Res,* 14, 95–105.

Sleight, S. C. and Lenski, R. E. 2007. Evolutionary adaptation to freeze-thaw-growth cycles in *Escherichia coli*. *Physiol Biochem Zool*, 80, 370–385.

Sleight S. C., Orlic C., Schneider D., Lenski R. E. 2008. Genetic basis of evolutionary adaptation by *Escherichia coli* to stressful cycles of freezing, thawing and growth. *Genetics*, 180, 431–443.

Smith, D.A., Toone, W. M., Chen, D., Bahler, J., Jones, N., Morgan, B. A., Quinn, J. 2002. The Srk1 protein kinase is a target for the Sty1 stress-activated MAPK in fission yeast. *J Biol Chem*, 277, 33411–33421.

Sniegowski, P. D., Gerrish, P. J., Lenski, R. E. 1997. Evolution of high mutation rates in experimental populations of *Escherichia coli*. *Nature*, 606, 703–705.

Soto T., Fernandez J., Vicente-Soler J., Cansado J., Gacto M. 1999. Accumulation of trehalose by overexpression of tps1, coding for trehalose-6-phosphate synthase, causes increased resistance to multiple stresses in the fission yeast *Schizosaccharomyces pombe*. *Appl Environ Microbiol*, 65, 2020–2024.

Tagkopoulos, I., Liu, Y. C., Tavazoie, S. 2008. Predictive Behavior Within Microbial Genetic Networks. *Science*, 320,1313–1317.

Tempest, D. W. 1978. Dynamics of Microbial Growth. *John Wiley & Sons Ltd.*

Tenaillon O., Rodríguez-Verdugo A., Gaut R. L., McDonald P., Bennett A. F., Long A. D., Gaut B. S. 2012. The Molecular Diversity of Adaptive Convergence. *Science*, 335, 457–461.

Thompson, D. A., Desai, M. M., Murray, A. W. 2006. Ploidy controls the success of mutators and nature of mutations during budding yeast evolution. *Curr Biol,* 16, 1581–1590.

Tokuriki, N. and Tawfik, D. S. 2009. Chaperonin overexpression promotes genetic variation and enzyme evolution. *Nature*, 459, 668–673.

Tomkinson, A. E., Bardwell, A. J., Bardwell, L., Tappe, N. J., Friedberg, E. C. 1993. Yeast DNA repair and recombination proteins Rad1 and Rad1O constitute a single-stranded-DNA endonuclease. *Nature*, 362, 860–862.

Toone, W.M. and Jones, N. 1998. Stress-activated signalling pathways in yeast. *Genes Cells*, 3, 485–498.

Vinces M. D., Legendre M., Caldara M., Hagihara M., Verstrepen K. J. 2009. Unstable Tandem Repeats in Promoters Confer Transcriptional Evolvability. *Science*, 324, 1213–1216.

De Visser, J. A., Rozen, D. E. 2006. Clonal interference and the periodic selection of new beneficial mutations in *Escherichia coli*. *Genetics*, 172, 2093–2100.

Wang G., Wilson T. J., Jiang Q., Taylor D. E. 2001. Spontaneous mutations that confer antibiotic resistance in *Helicobacter pylori*. *Antimicrob Agents Chemother*, 45, 727–733.

Wilkinson M. G., Samuels M., Takeda T., Toone W. M., Shieh J. C., Toda T., Millar J. B., Jones N. 1996. The Atf1 transcription factor is a target for the Sty1 stress-activated MAP kinase pathway in fission yeast. *Genes Dev*, 10, 2289–2301.

Wood, V., Gwilliam, R., Rajandream, M. A. 2002. The genome sequence of *Schizo-saccharomyces pombe*. *Nature*, 415, 871–880.

Yeyati P. L., Bancewicz R. M., Maule J., van Heyningen V. 2007. Hsp90 selectively modulates phenotype in vertebrate development. *PLoS Genet*, 3, e43.

Yu, F. and Etheridge, A. 2008. Evolutionary Biology from Concept to Application. *Springer.*

Zhu M., Need A. C., Han Y., Ge D., Maia J. M., Zhu Q., Heinzen E. L., Cirulli E. T., Pelak K., He M., Ruzzo E. K., Gumbs C., Singh A., Feng S., Shianna K. V., Goldstein D. B. 2012. Using ERDS to infer copy-number variants in high-coverage genomes. *Am J Hum Genet*, 91, 408–421.