מכון ויצמן למדע
WEIZMANN INSTITUTE OF SCIENCE

עבודת גמר (תזה) לתואר

דוקטור לפילוסופיה

מוגשת למועצה המדעית של

מכון ויצמן למדע
רחובות, ישראל

מאת

אמיר מיטשל

Thesis for the degree

Doctor of Philosophy

Submitted to the Scientific Council of the

Weizmann Institute of Science
Rehovot, Israel

By

Amir Mitchell

התניה סביבתית:

ניבוי שינויים סביבתיים על ידי מיקרו-אורגניזמים פרוקריוטים ואאוקריוטים

Adaptive Environmental Conditioning:

Prediction of Environmental Changes by Eukaryotic and Prokaryotic Micro-Organisms

מנחה:

פרופ' יצחק פלפל

ניסן ה', תשס"ט

Advisor:

Prof' Yitzhak Pilpel

April 1, 2009

Summary

Cells constantly sense and react to their surrounding environment in order to maintain their internal homeostasis. As conditions in the natural habitats of micro-organisms can dramatically fluctuate, evolution acts to preserve cells that are well adapted to a wide range of chemical, physical, and nutritional conditions. The adaptation of micro-organisms to the various stimuli they encounter has been studied extensively over the years, yet little attention has been given to clarify the temporal context in which these stimuli appear. In fact, some ecological niches expose organisms to stimuli in a predictable manner, thus offering the opportunity to prepare in advance for the next environmental change rather than merely respond to the change upon encounter.

In my thesis I propose that adaptation is not restricted to different types of environmental stimuli an organism encounters, but extends to act also on the temporal order in which these stimuli appear. I borrowed the conceptual framework of *Classical Pavlovian Conditioning* and ask whether genetic regulatory networks of Micro-organisms evolved to capture the temporal connections between subsequent stimuli in their habitats. I term a response strategy that captures the unidirectional temporal order of changes *Adaptive Environmental Conditioning*.

Here I present evidence indicating that Adaptive Conditioning exists in two model micro-organisms, *S. cerevisiae* and *E. coli*, under conditions found in their natural habitats, the switch from fermentation to respiration and the passage through the digestive tract, respectively. In both systems I studied the transcriptional response to conditions that resemble these habitats to reveal that the natural temporal order of the stimuli is embedded in the wiring of the regulatory network - early stimuli pre-induce genes that would be needed only later on while later stimuli only induce genes needed to cope with them. Yet the truly crucial question is whether such pre-exposure to the early stimulus enhances the fitness of the organism when it subsequently encounters the later stimulus. Such putative fitness enhancement might indicate that the observed response strategy was selected for during evolution. Indeed my experiments measuring fitness in a changing environment indicate that the proposed Adaptive Conditioning is likely an

adaptive trait that could be selected for during evolution. In particular and in accordance with the directionality of the transcriptional response, I observed a unidirectional fitness advantage as pre-exposure to the stimulus that typically appears early in the ecology improves the organism's fitness when challenged with a second stimulus. Finally, I also observed extinction of conditioned response in *E. coli* strains that were repeatedly exposed only to the first stimulus in a lab evolution experiment. This extinction was taken as evidence that Adaptive Conditioning can be selected against under an unsuitable environment.

The hypothesis of Adaptive Conditioning is further studied on two complementary levels. First I developed a mathematical model that predicts gain in fitness due to early preparation in various environments in order to better understand the key forces that select for this trait. The predictions of the model were experimentally validated using *E. coli*. This successful validation indicates that the model can also serve a predicative tool to identify ecologies potentially selecting for Adaptive Conditioning. Additionally, I tried, as a proof of concept, to evolve a lineage of *E. coli* cells in a laboratory evolution experiment to condition between previously unconnected stimuli. Although Adaptive Condoning failed to evolve in this experiment, insights from this system are now implemented to a new evolution experiment.

My work shows that Adaptive Conditioning is a selectable trait that was repeatedly selected for during evolution both in prokaryotes (*E. coli*) and eukaryotes (*S. cerevisiae*) and thus may be ubiquitous in biology.

Table of Contents

Summary **1**

1 Introduction **7**

2 Materials and Methods **14**

2.1 *E. coli* Experiments **14**

2.1.1 Strains and media 14

2.1.2 Population size and growth rate measurements 15

2.1.3 Monitoring operon activity with promtor-fused GFP library 15

2.1.4 Monitoring operon transcription using qRT-PCR 16

2.1.5 Measuring fitness advantage in alternating sugar environment 16

2.2 *S. cerevisiae* Experiments **17**

2.2.1 Strains and media 17

2.2.2 Cross-protection experiments 17

2.2.3 Sensitivity of deleted strains to heat and oxidative stress 19

2.2.4 Genome-wide expression experiments 19

2.2.5 Defining essential gene sets 20

2.2.6 Induction of the gene sets under different stresses 20

2.2.7 Identifying candidate genes that facilitate conditioning 20

3 Results **22**

3.1 A Mathematical Model of Adaptive Conditioning **22**

3.1.1 A toy model of the relative fitness of Adaptive Conditioning 23

3.1.2 A mathematical model of Adaptive Conditioning 25

3.1.3 Experimental validation of model's predictions 30

3.1.4 Model expansion – two-phase Adaptive Conditioning 32

3.2 Adaptive Conditioning in *E. coli* and the Intestinal Ecology 34

3.2.1 Intestinal ecology potentially selects for Adaptive Conditioning 34

3.2.2 Adaptive Conditioning in lactose and maltose pathways 37

3.2.3 Extinction of Adaptive Conditioning in evolved strains 39

3.2.4 Fitness advantage of Adaptive Conditioning 42

3.3 Adaptive Conditioning in *S. cerevisiae* and the Wine Ecology 44

3.3.1 Wine ecology potentially selects for Adaptive Conditioning 44

3.3.2 Adaptation to the order of stresses typical to wine production 45

3.3.3 Adaptive Conditioning in the transcriptional response to stress 47

3.4 Directed Evolution towards Adaptive Conditioning 54

4 Discussion 58

4.1 A Mathematical Model of Adaptive Conditioning 58

4.2 Adaptive Conditioning in *E. coli* and the Intestinal Ecology 60

4.3 Adaptive Conditioning in *S. cerevisiae* and the Wine Ecology 62

4.4 Directed Evolution towards Adaptive Conditioning 65

4.5 Closing Conclusions 66

5 Supplementary Results 69

5.1 A Mathematical Model of Adaptive Conditioning 69

5.1.1 Integrals of gain and cost functions 69

5.1.2 Two-phase conditioning model 70

5.1.3 Additional cost in the *E. coli* lactose system 73

5.2 Adaptive Conditioning in *E. coli* and the Intestinal Ecology 74

5.2.1 Response of operons to sequential addition of sugars 74

5.2.2 Adaptive Conditioning under different background carbon source .. 75

5.2.3 Control fitness experiments 77

5.3 Adaptive Conditioning in *S. cerevisiae* and the Wine Ecology 78

5.3.1 Standard deviations in cross protection experiments 78

5.3.2 Fitness contribution of conditioned genes under stresses 78

6 Literature 80

7 Independent Efforts and Collaborations 83

8 Acknowledgements 84

1 Introduction

Micro-organisms are constantly faced with environmental stresses and stimuli as external conditions in the habitats can rapidly and dramatically fluctuate. The adaptation of Micro-organisms to the various stimuli they encounter has been studied extensively over the years yet little attention has been given to clarify the cellular response to these challenges when they appear in the temporal context typical of the natural ecology. In fact, some ecological niches expose cells to stimuli in a predictable manner, i.e., a reoccurring set of stimuli that appears in a conserved order along the natural history of an organism. For example, during alcoholic fermentation *Saccharomyces cerevisiae* metabolizes glucose and produces ethanol until sugar depletion[1]. When the cells switch to respiration in order to utilize the ethanol they are exposed to oxidative stress originating from intercellular reactive oxygen species[2]. Therefore, in the context of alcoholic fermentation yeasts are exposed to two stimuli, ethanol and oxidative stress, in a sequential manner.

In my research I propose that adaptation is not restricted to different types of environmental stimuli an organism encounters, but extends to act also on the temporal order in which these stimuli appear. The capacity of higher organisms to capture the temporal connection between events in their environment serves as a classical paradigm in the study of learning. Here I borrow the conceptual framework of *Classical Pavlovian Conditioning*[3] and ask whether genetic regulatory networks of Micro-organisms adaptively evolved to capture the temporal connections between subsequent stimuli in their habitats. In a predictable environment, organisms that encode information about the temporal relationship between stimuli in their regulatory network have a significant advantage over organisms that react separately to each of the stimuli. I coin such a regulatory strategy *Adaptive Environmental Conditioning* (Figure 1).

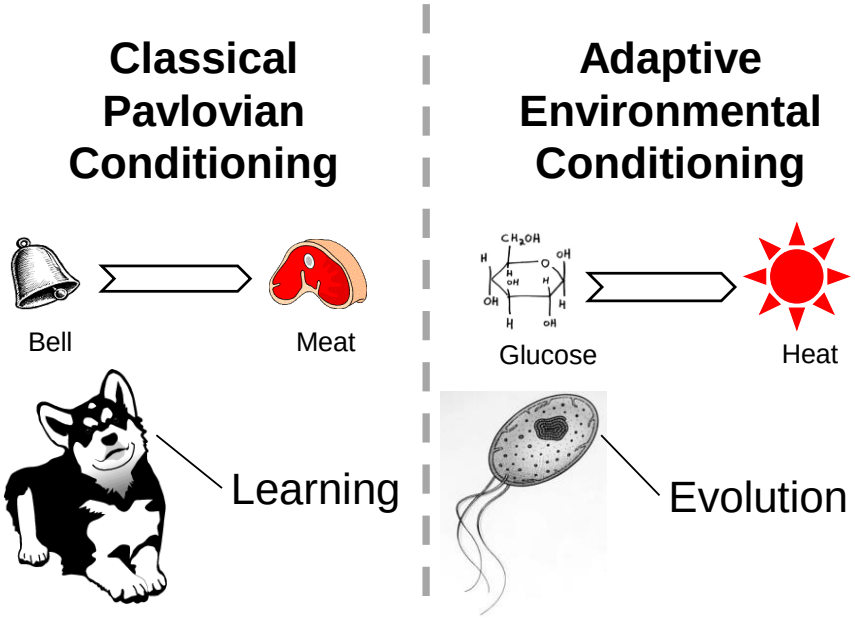


Figure 1. Analogous response strategies to sequential stimuli in the organism's environment. Pavlovian Conditioning is a cognitive capacity of higher multi-cellular organisms that is facilitated by a learning process. The proposed analogous Adaptive Environmental Conditioning is facilitated by natural selection during the process of evolution in Micro-organisms.

Over the years various response strategies were observed in Micro-organisms. The simplest response strategy frequently used by organisms is to monitor the environment and to directly respond to a stimulus upon encounter using a designated mechanism (*Direct Regulation*, Figure 2). The environmental stress response in yeast represents a complicated instance of the direct response strategy. The responses to the various stresses are partially overlapping, each of the stresses activates a designated response, yet there exist a common core response that is required for cope with all stresses (*Overlapping Response*, Figure 2)[4,5]. Recent studies have shown an alternative response strategy acting on the population level through creation of variation between individuals. Theoretical work indicates that when a population of Micro-organisms evolves under erratic environmental fluctuations, cells may not effectively monitor the environment, but may rather utilize stochasticity to randomly alternate between potential states ('Stochastic Switching', Figure 2). Stochastic Switching thus ensures that a portion of the population is prepared in advance to the unpredicted challenge[6,7,8].

Most recently, Tagkopoulos *et al.*[9] have discovered *Associative Regulation* (Figure 2), a cognitive-like capacity in which bacteria associate between environmental changes. Specifically, they studied the response of *E. coli* to temperature elevation which is shortly followed by a drop in oxygen availability during its entry to the digestive tract. Interestingly these two signals show a symmetrical associative regulation pattern - each signal affects the expression of genes needed to cope with both, although one of the stimuli shortly precedes the other. The authors successfully decoupled the two responses during a lab evolution experiment in which the two signals were presented out-of-phase from one another. The ability to decouple the two responses is an indication that the coupling seen in the wild-type is not a trivial overlapping response to the two stresses similar to the Environmental Stress Response.

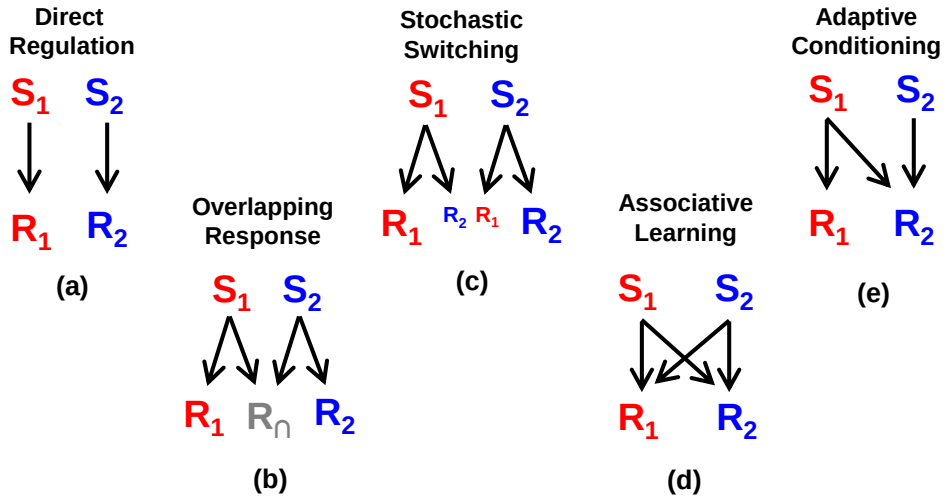


Figure 2. Five possible regulation strategies in response to environmental stimuli. (a) Under a Direct Regulation strategy each of the stimuli, S_1 and S_2, activates exclusively the responses R_1 and R_2 respectively. The Overlapping Response (b) represents a more complicated case of Direct Regulation; the two designated responses partly overlap. Under the Stochastic Switching strategy (c) cells randomly sample either R_1 or R_2 in response to either S_1 or S_2, thus a portion of the population is prepared for unexpected environmental changes. Under Associative Learning (d) each of the stimuli activates both responses. Under Adaptive Conditioning (e) the stimulus that typically appears first in the ecology activates both responses, while the stimulus that appears later induces only the second response.

In my thesis I show that biological systems that react to a unidirectional temporal order of environmental changes manifest a more elaborate predictive capacity than simple association. This capacity is reflected in a corresponding asymmetric response strategy - the first signal activates the responses to both first and second stimuli, yet since the second stimulus does not predict the appearance of the first it only activates its own response (*Adaptive Environmental Conditioning*, Figure 2).

I studied the hypothesis of Adaptive Conditioning on three fundamentally different levels: (1) I developed a mathematical model that predicts circumstances under which Adaptive Conditioning is evolutionarily beneficial, (2) I examined contemporary living organisms to identify examples of existing Adaptive Conditioning and (3) I conducted Pavlovian-like lab evolutionary experiments in order to select for lineages of cells that condition between previously unconnected stimuli. Figure 3 shows the outline of the basic paths of research.

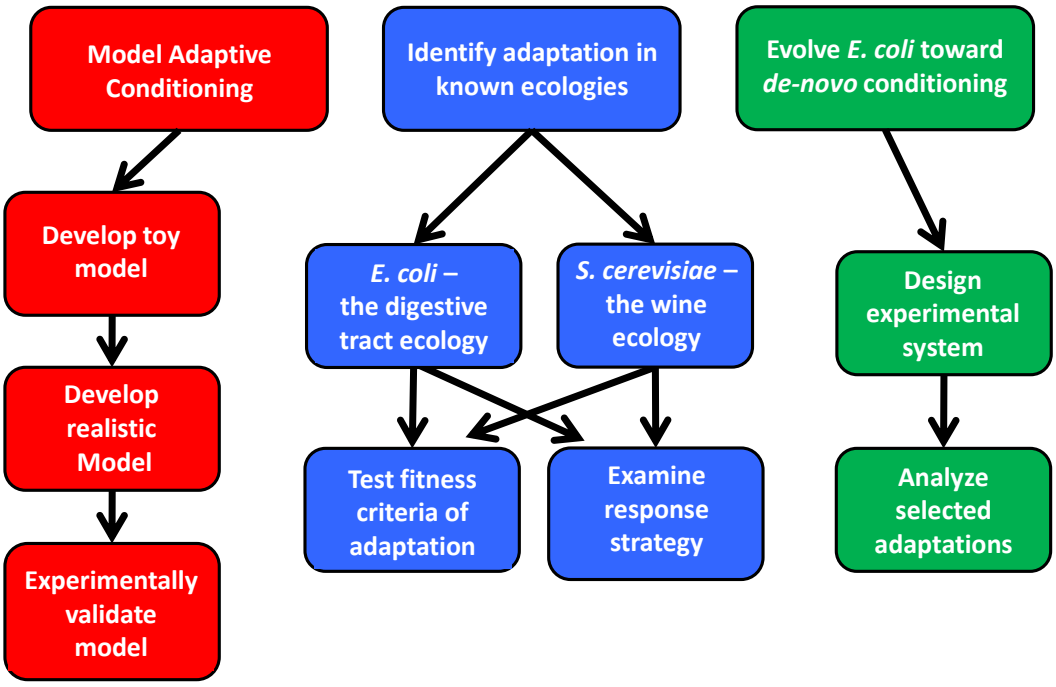


Figure 3. Outline of three fundamental levels for the study of the Environmental Adaptive Conditioning.

Adaptive Conditioning is expected to provide a selective advantage in a limited set of environmental and cellular conditions. Intuitively, highly predictable habitats, in which an early stimulus is a reliable predictor for a subsequent stimulus, are ideal for this type of adaptation. On the other hand, an irregular environment may generate many wasteful "false alarms" events in which the organism will futilely prepare for the second stimulus. I started my study by developing a mathematical model in order to theoretically map the environmental and cell internal parameters under which Adaptive Conditioning confers a selective advantage. This mapping clarifies the key forces that select for the proposed adaptation and the relationship between them. Moreover, the mathematical model's quantitative predications are verified experimentally using *E. coli*. Thus I conclude that the model can be used as a predictive tool to identify natural ecologies that are likely to exhibit Adaptive Conditioning.

I next turned to study Adaptive Conditioning in contemporary living organisms. I focused on the two model micro-organisms, *S. cerevisiae* and in *E. coli* under conditions found in their natural habitats, the switch from fermentation to respiration and the passage through the digestive tract, respectively. In each of the systems I uncover an asymmetric response strategy that fits the temporal order of stimuli in the natural habitats. I continue to test three criteria to determine whether the observed cross regulation pattern forms an adaptive trait that was likely selected for during evolution:

1. **Asymmetric fitness advantage**: Pre-exposure to the naturally preceding stimulus, S_1 enhances the fitness under the subsequent stimulus, S_2. However, exposure to the same two stimuli in reverse order does not enhance fitness.

2. **Preparation is stimulus specific**: The conditioned response is specific to S_1 and not to other unrelated stimuli that the organism did not encounter prior to S_2 during its evolution.

3. **Preparation is costly** Pre-induction of genes needed to cope with S_2 is costly and not beneficial during S_1. Such early induction is preserved due to a future benefit that is expected to exceed the cost, upon encounter with S_2.

During its life cycle *E. coli* alternates between two principal habitats, intestines of mammals and water, sediment, and soil[10]. Focusing on the intestinal ecology reveals a predictable environment that can potentially select for lineages of conditioned bacteria. Specifically in my research, I focused on regulation of sugar metabolism pathways as during passage along the digestive tract, exposure to lactose precedes exposure to another sugar, maltose[11]. My experiments revealed an asymmetrical regulation strategy compatible with Adaptive Conditioning - the sugar that is encountered early in the digestive tract, lactose, also activates in advance genes that would be required only further down the tract for metabolism of maltose. Additional experiments revealed that this conditioned strategy forms an adaptive trait according to the three criteria outlined above: Pre-exposure to lactose improves cells growth on maltose, yet reversing the order of sugars does not improve fitness. Additionally, conditioning proves to be lactose specific, pre-exposure to other sugars does not improve growth on maltose. Finally I observed extinction of the conditioned response in *E. coli* strains that were repeatedly exposed only to the first stimulus in a lab evolution experiment. This extinction indicates that induction of the conditioned response is indeed maladaptive in the presence of first stimulus.

Wine making is an ancient human trait, dating back thousands of years ago. Since yeast cells that survive this stressful process are not killed but rather propagate and may even be intentionally reused for future wine fermentations[12], this process forms an evolutionary cycle that potentially selects for lineages of conditioned yeasts. Specifically, I focused on stresses that are typical to the switch from fermentive to oxidative metabolism to reveal an example of existing Adaptive Conditioning. My experiments revealed the asymmetric fitness advantage expected under the hypothesis, i.e., cell survival is higher when consecutive stresses are introduced in their natural order, rather than in the reverse order. Additionally, I observed that this advantage is restricted to stresses encountered during wine production in line with the requirement of high specificity. Finally, I examined the genome-wide transcriptional response under relevant

stresses and revealed the underlying molecular mechanism – genes required to cope with the second stress are already pre-induced upon encounter the first stress.

Adaptive conditioning is proposed to be the result of explicit selective circumstances – a consistent environment of recurring stimuli. In a third path of research I conducted directed evolution experiments in order to select for lineages of cells that condition between previously unconnected stimuli. In these experiments *E. coli* cultures were diluted daily into a new growth medium and were sequentially exposed to two stimuli, e.g., addition of maltose followed by exposure to heat shock. Unfortunately, even after three months of serial transfers a conditioned response was not observed between the stimuli pair. However, other, more trivial, adaptations were fixated, e.g., increased resistance to heat shock. Conclusions and insights from this experiment series were implemented in the design of a new set of experiments that will be conducted by another student in the lab.

2 Materials and Methods

2.1 E. coli Experiments

2.1.1 Strains and media

E. coli MG1655 (*E. coli* genetic stock centre) was used for model validation experiments and for the fitness experiments.

E. coli W2244[13] was used as a *lacZ* mutant strain (*lacZ39*Δ) for model validation experiments. W2244 has a deletion of a third of the *lacZ* gene, the mutant shows no β-galactosidase activity yet has an active permease[13].

An *E. coli* GFP reporter library[14] was used to monitor operon expression in response to treatment with various carbon sources.

The evolved *E. coli* strains were kindly provided by Erez Dekel and are described Oxman et al.[15]

All experiments were done in M9 defined medium consisting of M9 salts (1 mM MgSO4, 0.1 mM CaCl2, 0.05% casamino acids, and 5ng/ml Thiamine) supplemented with the appropriate carbon source.

Three basic media types were used:

1. M9
2. M9-Gly [0.1% glycerol (Baker)]
3. M9-Glu [0.1% glucose (Sigma) + 20 mM cAMP (Sigma)], cAMP was added to avoid glucose repression[16].

To test the effect of various treatments the media were supplemented with the following:

- Lactose 10mM (Fluka). This concentration was chosen as it allows maximal growth on this carbon source.

- Maltose 5mM (Sigma). This concentration was chosen as it allows maximal growth on this carbon source.
- Galactose 10mM (Sigma). This concentration was chosen to be equal to the concentration of lactose.
- Sucrose 10mM (Sigma). This concentration was chosen to be equal to the concentration of lactose.
- IPTG 0.15mM (Sigma). This concentration was chosen to achieve saturation of the lactose operon induction[16].
- Kanamycin 25 μg/ml (Sigma) was used in the GFP expression experiment to maintain selection of the plasmid.

2.1.2 Population size and growth rate measurements

Population size and growth rate measurements were preformed similarly to Dekel and Alon[17]. Overnight cultures were diluted into fresh media and were grown for 1 hour at 37^0C. The cultures were then transferred to a 96-well plate to perform the experiments. Population size and growth rate were monitored using a multi-well spectrophotometer at 595 nm (GENios and INFINITE200; TECAN). Alternatively treated cultures were plated in a "checkerboard pattern" to account for different growth rates on different locations on the plate[17]. The effect of a given treatment was calculated by dividing the OD at a given treated well to the average OD of its four alternatively treated neighbors. This yielded 48 ratios which greatly increase the resolution of measurement and the ability to detect reliably even small differences in population sizes.

2.1.3 Monitoring operon activity with promtor-fused GFP library

Overnight cultures, each carrying a unique plasmid with a specific promoter fused to GFP, were diluted into fresh M9-Glu media and were grown for 1 hour at 37^0C. The cultures were then transferred to a 96-well plate and treated with sugars. Expression and cell growth (at 30^0C) were monitored simultaneously using INFINITE200 (TECAN) a multi-well reader (fluorescence at 495/520 nm, OD at 595 nm). OD and GFP signals were smoothed using local regression with the MATLAB "smooth" function (rlowess

option). The normalized GFP level and promoter activity were calculated similarly to Kaplan et al[16]. Briefly, the GFP signal was calculated after subtraction of the media fluorescence and cell auto fluorescence. GFP per cell was calculated by dividing the GFP signal by the OD measured. Promoter activity ((dGFP/dt)/OD) was taken as the average promoter activity measured in a time window of 1 hour of exponential growth.

The following operons were examined in my research: MalEFG, MalK-lamB-malM, MalPQ, MalS , MalZ, LacZYA. For simplicity I refer to the operons by the name of the first gene only. The maltose operon MalT was not included in my analysis as it shows no responsiveness to maltose (results not shown).

2.1.4 Monitoring operon transcription using qRT-PCR

Overnight cultures were diluted into either M9-Gly or M9-Glu media alone or supplemented with either lactose or maltose, and grown until reaching the logarithmic phase. RNA was extracted using RNeasy Mini kit (QIAGEN) and used as a template for the reaction. qRT-PCR was done using the LightCycler 480 system (Roche) according to the manufacture instructions.

2.1.5 Measuring fitness advantage in alternating sugar environment

Overnight cultures were diluted either into M9-Gly+S_1 (e.g., lactose) (treated) or M9-Gly (untreated). After 3 hours the size of the populations was determined (OD). The OD values were used in order to dilute (1:100) an equal amount of cells into a new growth medium containing low levels of S_2 (e.g., maltose) as a sole carbon source. In order to account for a potential metabolic effect a residual S_1 that remains after dilution of the treatment culture, an identical trace amount of $S_1$was added to the control growth medium. The populations' size was measured after 2 hours of growth. The population size ratio (treated/untreated) was used as an indication of fitness advantage originating from pre-exposure to S_1. In order to account for the possibility that the growth advantage observed due to early exposure to lactose originates from the metabolic value of this sugar as a carbon source rather than its role as a conditioning signal, I preformed a control experiment, See Additional Results section.

2.2 S. cerevisiae Experiments

2.2.1 Strains and media

Strain BY4741 (*MATa*; *his3Δ1*; *leu2Δ0*; *met15Δ0*; *ura3Δ0*) was used in all yeast experiments.

All deletion strains used in this study are based on Strain BY4741 and were obtained from the Saccharomyces Genome Deletion Project[18].

Experiments were carried out in liquid YPD medium (2% yeast extract, 1% peptone, 1% dextrose) at 30°C. Osmotic stress was applied by addition of KCl (Baker), oxidative stress was applied by addition of hydrogen peroxide (Frutarom) and heat shock was applied by transferring the cell cultures into a preheated water bath. Parallel application of heat shock to multiple strains was done using a multiwall plate in a PCR machine. For specific concentrations see

Table **1**.

2.2.2 Cross-protection experiments

An overnight culture was diluted into fresh YPD medium and grown to a cell concentration of $2*10^6$ cells/ml. Cells were diluted 1:2 into the first stress (S_1). Transfer into the second stress (S_2) was done by centrifuging the culture, removing the medium containing the stress and adding new medium containing the second stress. Stress levels and exposure periods were calibrated to achieve a mild effect (50% survival) for S_1 and a severe effect (0.5% survival) for S_2 (see example of calibration experiment in Figure 4). A full list of stress levels used appears in

Table **1**. This experimental design is crucial since it allows to measure fold protection values which span two orders of magnitude.

Table 1 – Stresses levels applied for measuring cross protection phenotypes

Stress type	Stress level	Exposure Time
High osmolarity: KCl (mild)	1M	30'
low pH (mild)	2	30'
High pH (mild)	8.2	45'
Copper: CuSO4 (mild)	4 mM	45'
Acetic acid (mild)	60 mM	45'
Heat shock (mild)	42°C	45'
Heat shock (severe)	49°C	45'
Ethanol (mild)	4%	30'
Ethanol (severe)	16%	35'
Oxidative stress: H2O2 (mild)	1mM	30'
Oxidative stress: H2O2 (severe)	30mM	20'

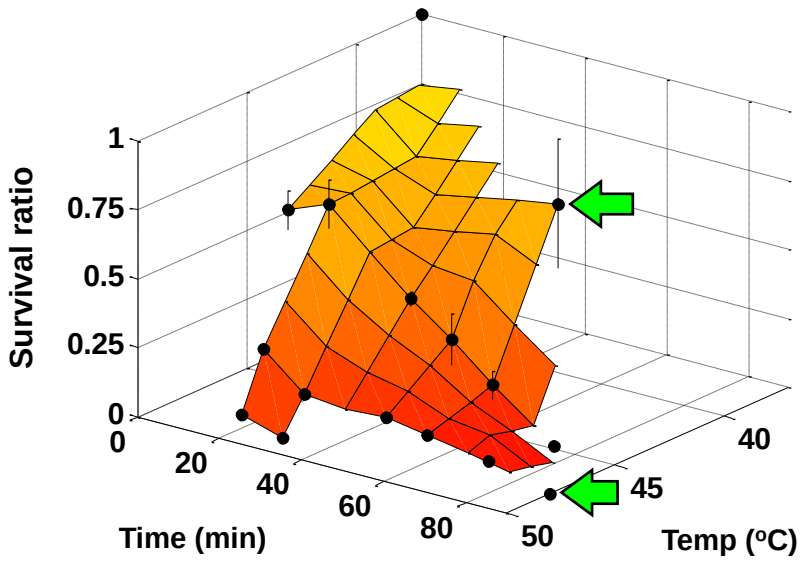


Figure 4. Calibration of heat shock levels. Heat was applied while varying the both the temperature and exposure time. Black circles mark the mean survival ratio measured for at least three independent repeats. Bars denote the standard deviations. Green arrows mark the stress levels found to result in mild and severe effects.

Samples were taken from cultures at the end of each treatment and plated on YPD plates. The plates were incubated for 48 hours at 30°C before counting the colonies formed. The fold protection (FP) entailed by pre-exposing cells to a mild level of S_1 before a severe level of S_2 was calculated after measuring three independent survival ratios (Sur):

$$FP_{s1\to s2} = \frac{Sur_{s1\to s2}}{Sur_{s1} \cdot Sur_{s2}}$$

2.2.3 Sensitivity of deleted strains to heat and oxidative stress

Cultures of both wild type and deleted strains were grown overnight at 30°C in a 96-well plate. Cultures were then diluted (1:20) into fresh YPD medium and grown for 3 hours. Cultures were then diluted into either mild heat shock (37°C for 30 minutes) or mild oxidative stress (H_2O_2 1mM for 30 minutes). The stresses were calibrated to result in a similar effect on the survival of the wild type strain (90%). As a control, untreated isogenic cultures were diluted into rich medium. All cultures were then diluted (1:20) into fresh YPD medium and grown for 6 hours. Population size was monitored using a multi-well spectrophotometer at 595 nm (INFINITE200; TECAN). Survival ratio was calculated by dividing the OD of stress treated cultures by the OD of untreated cultures. Each strain was measured in 4 independent experiments. A t-test was used to determine whether the difference in the mean survival ratio of a deletion strain and the wild type is statistically significant in a given stress.

2.2.4 Genome-wide expression experiments

An overnight culture was diluted into fresh YPD medium and grown to cell concentration of $2.5*10^7$ cells/ml. Cells were diluted 3:4 into a fresh medium with the S_1 treatments (KCl 0.8M, heat shock 40°C or YPD as control) and grown for 45 minutes. Cells were then diluted 1:10 into the second treatments, S_2 (H_2O_2 0.66mM or YPD as control). First stress (S_1) levels were calibrated to have an identical exposure period and relatively minor effects on survival across all treatments.

Aliquots (5ml) were removed after 0, 30, and 45 minutes into S_1 while in S_2 aliquots (50ml) were removed after 15, 30, and 45 minutes. Samples were frozen in liquid nitrogen, and RNA was extracted using MasterPure™ (EPICENTER Biotechnologies). The quality of the RNA was assessed using the BIOANALYZER 2100 platform (AGILENT); the samples were then processed and hybridized to Affymetrix yeast 2.0 microarrays using the Affymetrix GeneChip system according to manufacturer's instructions.

2.2.5 Defining essential gene sets

Essential gene sets for each stress were defined according to the gene annotation in the Proteome BioKnowledge Library (a manually curated, proteome-wide compilation of scientific literature). For example, the set of oxidative essential genes was obtained by finding genes annotated as conferring either sensitivity to oxidative stress (when mutated) or resistance to oxidative stress (when over-expressed). In order to discard genes which are essential for more than one of the tested stresses I defined sets of non overlapping unique essential genes. For example, a set of unique oxidative essential genes comprises of genes essential under oxidative stress but not essential under osmotic stress or heat shock.

2.2.6 Induction of the gene sets under different stresses

I tested whether the above sets of essential genes are over-expressed in response to a given stress using a simple binomial test. Towards this aim I used the time point within the stress that showed a peak in the genome-wide averaged response. Genes were then categorized as induced or repressed according to their fold change at this peak time point. Finally, the binomial test was used to check whether the number of essential genes that were induced at a given condition is significantly high, given the genome wide a-priori probabilities.

2.2.7 Identifying candidate genes that facilitate conditioning

I define an ideal expression profile describing the expected dynamics of a gene that can facilitate the observed cross-protection phenotype (heat shock protects against

oxidative stress while osmotic stress does not protect against oxidative stress). Thus the profile is defined accordingly:

1. Level of induction in response to heat shock- 0.2 in log2 scale or higher.
2. Level of induction in response to oxidative stress- 0.2 in log2 scale or higher.
3. Induction under osmotic stress is significantly lower than in heat shock- 0.4 in log 2 scale (this high threshold was used to completely avoid genes induced under both osmotic and heat shock to similar levels).
4. Induction is maintained higher in the transfer from heat shock to oxidative stress relative to the expression level seen when heat shock-treated cells are transferred to YPD.

A control profile is similarly defined by using osmotic stress instead of heat shock in the above requirements. The results are qualitatively similar under alternative cut-off values (results not shown).

3 Results

3.1 A Mathematical Model of Adaptive Conditioning

Adaptive Conditioning confers a selective advantage under specific environmental conditions. Intuitively, highly predictable habitats, in which an early stimulus is a reliable predictor for a subsequent stimulus, are ideal for this type of adaptation. Additionally, in the responding cellular system it is crucial that the cost of the early preparation will be smaller than the benefit gained from encountering the second stimulus in a prepared mode. In order to better understand the economic considerations of cells when conditioning between stimuli in their environment, a simple mathematical model and experimental test system were developed. For simplicity I focus on the minimal environmental scheme that can potentially select for conditioning. The environment includes two stimuli, S_1 and S_2, which require two independent cellular responses R_1 and R_2, respectively. The model calculates the relative fitness advantage of cells employing a conditioned response over cells employing a direct response strategy in a given environment. The fitness advantage is given in an experimentally measurable quantity – population size ratio after a single encounter with the stimuli pair.

The mathematical model is developed here in a few stages. I start by presenting a simple toy equation mapping the key environmental and cell internal parameters that affect the fitness of Adaptive Conditioning over the Direct Response strategy. This toy equation is then further developed to a more realistic mathematical model that includes biologically meaningful and measureable parameters. The accuracy of the model's predictions is experimentally validated using *E. coli*. After validation the model is further developed to consider a more complicated manifestation of the Adaptive Conditioning – a two-phase conditioning strategy in which the cells only partially commit to the R_2 response upon encounter with the predicative stimulus.

3.1.1 A toy model of the relative fitness of Adaptive Conditioning

Adaptive Conditioning confers a selective advantage under specific environmental conditions. The predictability of the habitat is a key environmental parameter that affects the relative fitness a conditioned organism will have over un-prepared cells. Let us consider an environment that includes both coupled and uncoupled appearances of the two stimuli, S_1 and S_2, and denote p as the probability that S_2 will occur given S_1 after a Δt time period. Two key parameters of the biological system affect the relative fitness of the conditioned organism, the energetic cost of preparation (denoted C), and the benefit gained from encountering the second stimulus in a prepared mode (denoted B). The gain and cost parameters are, in turn, dependent on the typical time constants of the habitat. The cost is expected to be proportional to duration of "untimely" response, i.e., the period in which the conditioned organism induces the R_2 response although S_2 is not present. The gain is expected to be proportional to the preparation period before actual encounter with S_2, e.g., a short period might not be sufficient for full preparation.

Equation 1 incorporates the three relevant key parameters to calculate the fitness difference ($\Delta F_{c\text{-}d}$) between an organism that uses the Adaptive Conditioning strategy and a naïve organism that adheres to the Direct Regulation strategy:

$$\Delta F_{c-d} = \underbrace{p \cdot \left(B(\Delta t) - C(\Delta t)\right)}_{\text{i}} - \underbrace{\left(1-p\right) \cdot \left(C(t_{S1})\right)}_{\text{ii}} \tag{1}$$

Thus the fitness difference is calculated as the difference between two terms, term (i) represents the net contribution to fitness difference due to cases in which S_1 was followed by S_2. Such cases will contribute to the fitness advantage of the conditioned organism (given that the benefit is larger than the cost). In contrast, term (ii) represents uncoupled events, i.e., cases that may be regarded as "false alarm", in which S_1 occurred, yet S_2 did not follow it; such cases will diminish the advantage of the conditioned organism. The two terms are weighted by the respective probabilities p and the

complementary probability (1-*p*). Note that appearance of S_2 alone does not affect ΔF as both under Direct Regulation and the conditioned strategy encounter with S_2 alone will result in identical induction of R_2. The dependence of fitness on the interplay between the three key factors, the cost, gain, and predictability are shown in Figure 5.

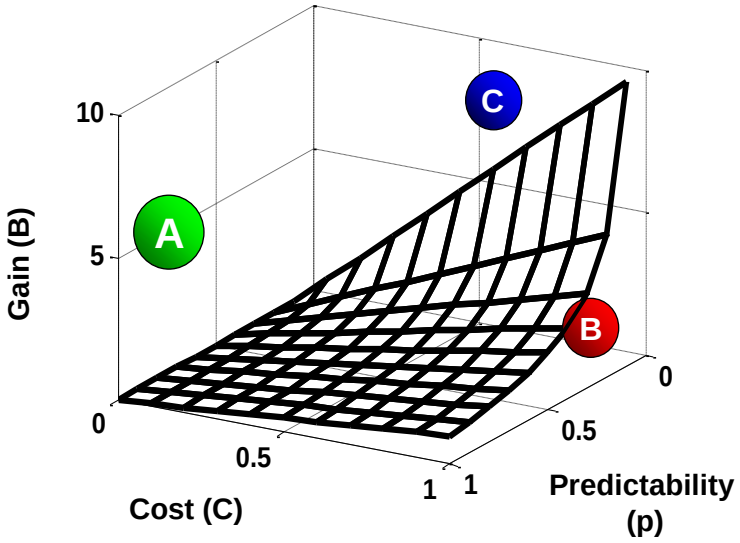


Figure 5. Relationship between three key parameters that affect the relative fitness advantage of an organism that applies Adaptive Conditioning. Each point in the 3D space marks a system that is characterized by different values of the gain, cost, and predictability parameters. The subspace above the curved plane represents combinations of values that favour Adaptive Conditioning, whereas points below it correspond to combinations that favour Direct Regulation. For example, point A favours conditioning since the environment is relatively predictable and the gain from early preparation to S_2 is much larger than the cost. Point B marks a parameter combination that is not suitable for Adaptive Conditioning since the environment is relatively unpredictable while preparation is costly relative to the potential benefit. Point C marks a case in which the low predictability is compensated by the low cost and high gain from preparation.

The results of the simple toy model reflect a non-linear dependency between the relative fitness of conditioning and the basic environmental and biological parameters. Furthermore, the results are in agreement with the basic intuition regarding parameter combinations that select for Adaptive Conditioning over the Direct Response strategy – highly predictable habitats and biological systems that are characterised by a high gain to cost ratio.

3.1.2 A mathematical model of Adaptive Conditioning

The toy model displayed above reflects the basic parameters that affect the relative fitness of conditioning and the relationship between them. However, in order to derive quantitative predictions from the above phenomenological equation I further developed the model to include biologically meaningful and realistic parameters, such as measurable gain, cost and time scales.

The mathematical model for Adaptive Conditioning follows the growth dynamics of two organisms in the given environment while taking into account the alternative response strategies. Environments are characterized by the stimuli, the typical time constants of the stimuli presence and the predictability of their coupling. The organisms are characterized by their alternative response strategies and the values of gain and cost parameters. I restricted the model to a minimal set of biological assumptions: (1) R_2 requires a protein system to be transcribed and translated. (2) The energetic cost of the R_2 protein system is proportional to the R_2 production rate, and (3) the benefit gained upon exposure to S_2 is assumed to be linearly proportional to the level of the R_2 protein system at a given time point. Note however that this model can be accommodated to include other, more complex, assumptions regarding the biological system and the environment.

In this section I will describe the equations of the mathematical model in a few stages. First I present the organism's response function - a function describing the time dependent changes in the R_2 response level in a given environment. The gain and cost effects, manifested as changes in the basal growth rate, are derived from the response function. By integrating over the gain and cost functions I calculate the accumulative effect of a response strategy in the given environment. Finally, the accumulative effects are plugged back into the toy equation to yield an experimentally measurable fitness parameter. The fitness difference ($\Delta F_{c\text{-}d}$) is represented as the predicted population size ratio between alternatively responding cells after encounter with the considered environment.

In order to model the time dependent dynamics of the R_2 protein system I relied on the function describing an exponential approach to a steady state protein level:

(2) $$Y(t) = Y_{st}(1 - e^{-\alpha t})$$

Where t is the time from induction, Y_{st} is the steady state level of the protein and α is the dilution/degradation rate. For simplicity, active degradation of the protein is neglected and hence α equals the growth rate[19]. This basic function is used to develop the two alternative response functions $r_d(t)$, $r_c(t)$, which denote the relative response level (normalized to Y_{st}) under the Direct Regulation and Adaptive Conditioning, respectively (Figure 6, red and blue graphs)

(3) $$r_d(t) = 1 - e^{-\alpha(t-\Delta t)}$$

(4) $$r_c(t) = 1 - e^{-\alpha t}$$

With a limitation for $r_d(t)$, $\Delta t < t$.

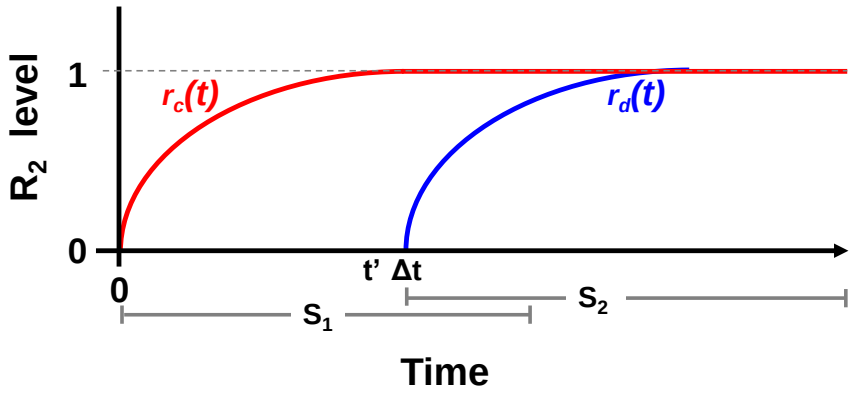


Figure 6. Response functions under different regulation strategies. The functions $r_d(t)$ and $r_c(t)$ describe the relative R_2 level in a given environment under direct and conditioned response strategies, respectively.

The benefit gained under each regulation strategy is modeled through a temporary increase in the basal grow rate. I assume that the gain is linearly proportional to the response level at a given time point, yet it is important to note that a benefit is gained

only when S_2 is present. As the gain depends on the biological system under study, a scaling parameter is required.

(5) $$b_i(t) = \kappa \cdot r_i(t)$$

Where κ is a system specific scale parameter and *i* marks the type of the response strategy (Direct or Conditioning). An example of the derived gain graphs is shown in Figure 7.

The cost under each regulation strategy is modeled through a temporary decrease in the basal grow rate. I assume that the cost is linearly proportional to the relative rate of R_2 production (β). This cost reflects constant rate processes such as transcription and translation which were recently shown to account for the observed cost during protein expression[17,20]. As the R_2 production rate, β, is s binary parameter (β=0 or β=1) in both regulation strategies, upon induction the cost is a time independent constant (scaled by a system factor η):

(6a) $$c(t) = \eta$$

In specific systems the cost might show a more complex and nonlinear dependency on activation period (for example, if accumulation of R_2 harms the cell). Thus the cost function might be further developed to include an additional cost factor η^* and the cost function becomes:

(6a) $$c(t) = \eta + \eta^*(t)$$

In order to calculate the accumulative effect of all changes on the basal growth rate I integrate over the gain and cost functions during the relevant time intervals (see Supplementary Results for integration details). The fitness of each regulation strategy in a given environment is given by the sum of the gain and cost integrals (see example in Figure 7). Note that as this fitness is calculated independently for each strategy. This fitness is in fact relative to a "no-response" strategy, in which the basal growth rate is maintained (no benefit is gained and a cost does not exist).

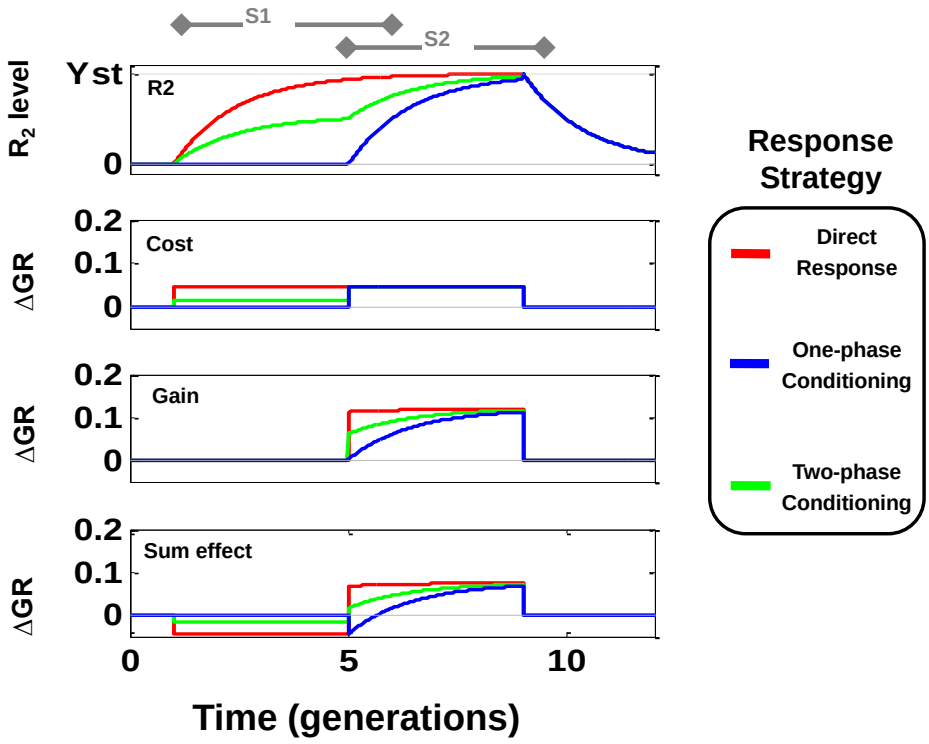


Figure 7. Calculation of fitness difference according to the detailed mathematical model. The blue, red and green graphs mark direct, conditioned, and two-phase conditioning regulation strategies, respectively. Details on the two-phase conditioning strategy appear later in this section. The durations of S_1 and S_2 are marked on top. Upon induction the R_2 response is assumed to follow an exponential approach to a steady state level. The cost and the gain are manifested by temporal changes in the growth rate (GR) relative to the basal growth rate. The cost is constant and exists throughout the response period. The benefit is linearly proportional to the R_2 level when S_2 is present. The final effect on the growth rate is given by subtracting the cost from the gain. By integrating over all changes in growth rate one can calculate the predicted population size under each strategy after an encounter with a pair of stimuli

After acquiring biologically informative cost and gain expressions the parameters from the toy model describing the fitness difference can be further developed (equation 1). The gain (B in equation 1) is essentially the difference between the integrals of the gain functions (equation 5):

(7) $$B(\Delta t) = \int_{\Delta t}^{\infty} \kappa \cdot r_c dt - \int_{\Delta t}^{\infty} \kappa \cdot r_d dt = \kappa \cdot \int_{\Delta t}^{\infty} r_c - r_d dt$$

The cost (C in equation 1) is essentially the difference between the integrals of the cost functions (equation 6) during the relevant time intervals. In coupled cases (S_2 was preceded by S_1) the period of surplus response is Δt, thus the cost in coupled cases (C_{cp}) is:

(8a) $$C_{cp} = \int_0^{\infty} \eta dt - \int_{\Delta t}^{\infty} \eta dt = \eta \cdot \Delta t = C(\Delta t)$$

while in uncoupled cases (only S_1 appeared) the period of surplus response is whole duration of S_1 (t_{s1}) since under direct response there is no response at all. Thus the cost in uncoupled cases (C_{un}) is:

(8b) $$C_{un} = \int_0^{t_{S!}} \eta dt - \int_0^{0} \eta dt = \eta \cdot t_{S1} = C(t_{S1})$$

To conclude, the toy equation describing the fitness difference between a conditioned and a direct response strategies can be now be rewritten:

(1a) $$\Delta F_{c-d} = p \cdot \left(\kappa \cdot \int_{\Delta t}^{\infty} r_c - r_d dt - \eta \Delta t \right) - (1-p) \cdot (\eta t_{S1})$$

Note that the ΔF expression denotes the difference in reproduction capability using the new cost and gain expressions which in turn reflect changes in the basal growth rate of the organism. Assignment of the solved integrals of the gain and cost functions in the above equation results in following expression (see also Supplementary Results section):

(1b) $$\Delta F_{c-d} = p \cdot \left(\kappa \frac{(1-e^{-\delta \Delta t})}{\delta} - \eta \Delta t \right) - (1-p) \cdot \eta t_{S1}$$

this expression can be easily solved and is dependent only on the environmental parameters (the time constants and the predictability parameter) and on the biological scaling parameters.

The new developed fitness expression (1b) is extremely useful as it can be used to calculate an experimentally measurable fitness parameter – population size ratio using a simple transformation:

$$N_c / N_d = \exp(\delta \cdot \Delta F_{c-d}) \quad (9)$$

Where δ is the basal growth rate. A population ratio above one indicates that the tested environment favors conditioning. For example, a ratio of 1.05 indicates that after a single cycle in the environment (weighted by the probability factor *p*), a conditioned population will be larger by 5% than a population with a Direct Regulation strategy.

3.1.3 Experimental validation of model's predictions

I devised an experimental system to test the model's predictions under a specific environmental setup – addition of lactose, a superior carbon source, to *E. coli* cells growing on glycerol, an inferior carbon source (Figure 8A). I reasoned that cells pre-inducing the lactose operon will be able to better utilize lactose resources upon encounter than naïve cells.

The values of the gain and the production cost (κ and η, respectively) have been previously measured for the lactose operon[17,20] and validated by myself. However, an additional cost parameter (η^*) is also known to exists in this system. This cost originates from excessive transport of lactose and consequent loss of membrane potential in cells already expressing the lactose permease when encountering the sugar[21]. I estimated the value of the additional cost using a *lacZ* mutant strain that can transfer lactose but cannot metabolize it (Supplementary Figure 2). Using the gain and cost measurements it is possible to test the relative fitness of conditioning for different delay times between pre-induction of the operon and lactose exposure. Note that the experiments allow testing the accuracy of the model's predictions without the use of any free parameters.

Figure 8B shows a comparison between the predicted and observed growth advantage of a pre-induced population over a control naive population, as a function of the time interval, Δt. As the figure shows, a good fit is observed between the predictions

and the experimental results. It is important to note that the pre-induction exercised here is an artificial mimicry of a more natural conditioning situation, yet the ability to accurately predict the fitness difference in this simple system indicates that the model can be used to explore natural, more complex situations.

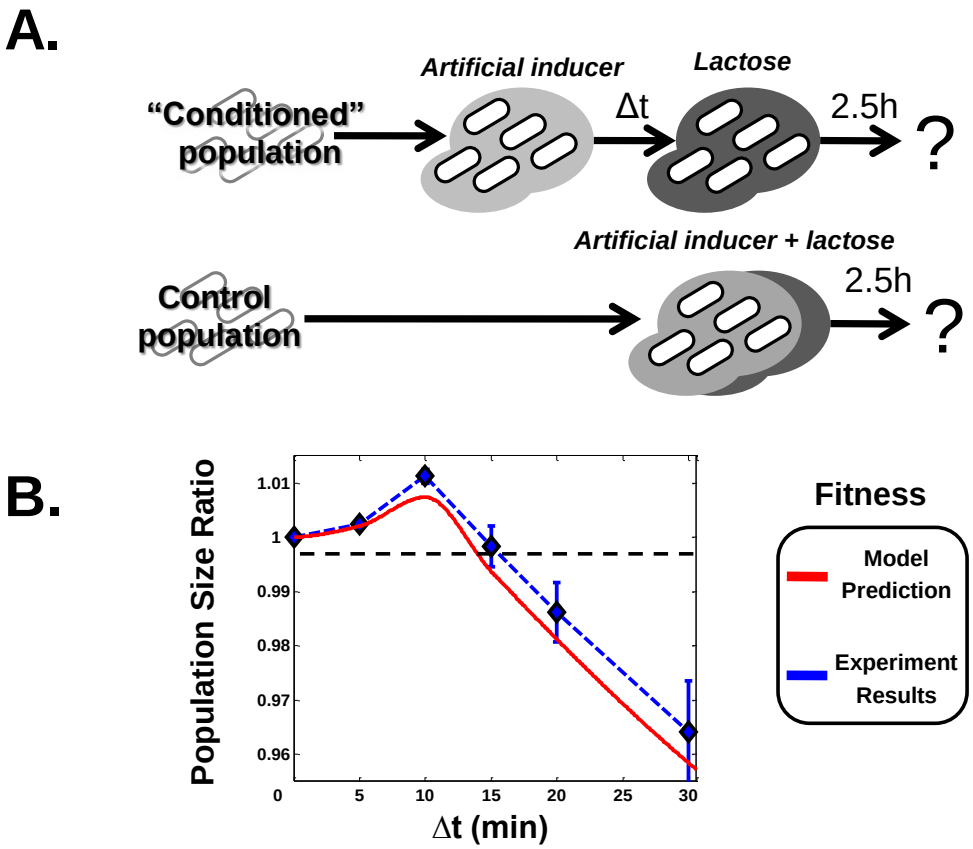


Figure 8. Validation of mathematical model's predictions. (A) Experimental setup to measure the fitness of inducing the lactose operon prior to lactose exposure. Two identical populations of *E. coli* are exposed to identical stimuli at a different order. The lactose operon of one culture is pre-induced by an artificial inducer. The size of the two populations is measured two and a half hours after lactose addition. (B) Mathematically predicted and experimentally observed fitness advantage of pre-induction of the lactose operon. The y-axis depicts the relative fitness of conditioning, measured as the ratio between the population sizes of two alternatively treated cultures. Model predictions were calculated by substituting the known cost and gain values of the lactose system in Equation (1b): $\Delta F = 0.17 \cdot (1 - e^{-0.924\Delta t}) \cdot \frac{1}{0.924} - 0.04 \cdot \Delta t - \eta_2(\Delta t)$. Calculation of the expected population ratio from ΔF was done using Equation (9). The red and blue graphs depict the ratio between the two populations ($OD_{conditioned}/OD_{unconditioned}$) as predicted by the model and experimentally observed, respectively. Bars denote the standard errors of three repetitions.

3.1.4 Model expansion – two-phase Adaptive Conditioning

The strategy of Adaptive Conditioning described above represent a full commitment of an organism to the R_2 response already upon encounter with the preceding stimulus S_1. However, under a more adjustable strategy S_1 might activate R_2 to either a full or a *partial* level. Such partial activation would reflect only a limited commitment of the cell to the R_2 in response to S_1, while only S_2 would fully trigger the R_2 response. I term this strategy *two-phase conditioning* to indicate the two steady state levels of R_2 – first, an intermediate level when S_1 is encountered, and then to a full level, when S_2 appears. An example of the response dynamics under two-phase conditioning is shown by the green graphs in Figure 7. Under some environments the two-phase response strategy may be optimal. Intuitively, this strategy can be viewed as a risk management strategy, allowing an organism to be moderately prepared for only a part of the cost.

I further developed the mathematical model to include the two-phase response strategy. Basically, this expansion is achieved through changes in the production rate parameter β. Previously induction of the R_2 response followed a binary pattern under the direct and one-phase conditioning strategies. Upon response, the R_2 relative production rate parameter β is 1, otherwise β is 0. The difference between the Direct Response and one-phase conditioning was manifested by the different assignment to β upon encounter with S_1. Under two-phase conditioning R_2 induction level is stimulus dependent. Upon encounter with S_1, R_2 is induced only to a partial level ($0<\beta<1$) while upon encounter with S_2, R_2 is fully induced ($\beta=1$). The relative fitness of two-phase conditioning can be analytically calculated, given the intermediate level β. A full account on the extended model is detailed, for easier readability of the main text, in the Supplementary Results section.

After deriving the fitness functions of the two-phase response strategy I can use the mathematical model as a predictive tool to identify the most beneficial response strategy in a given habitat. Specifically, given the typical gain and cost values of the biological system and the predictability level of the environment, the model can analytically find the

optimal fractional level of R_2 induction in response to the preceding stimulus S_1. Figure 9 shows the predicted optimal R_2 activity in response to S_1 as a function of the gain to cost ratio and the typical delay time. The figure illustrates the value combinations that select for a two-phase conditioned strategy. As the gain to cost ratio decreases at a given delay, the optimal strategy shifts toward the two-phase conditioning. Likewise, at a given gain to cost ratio, a negative correlation between the optimal intermediate expression level and the delay time is observed, e.g., long delay periods are optimal for two-phase conditioning or even only Direct Regulation.

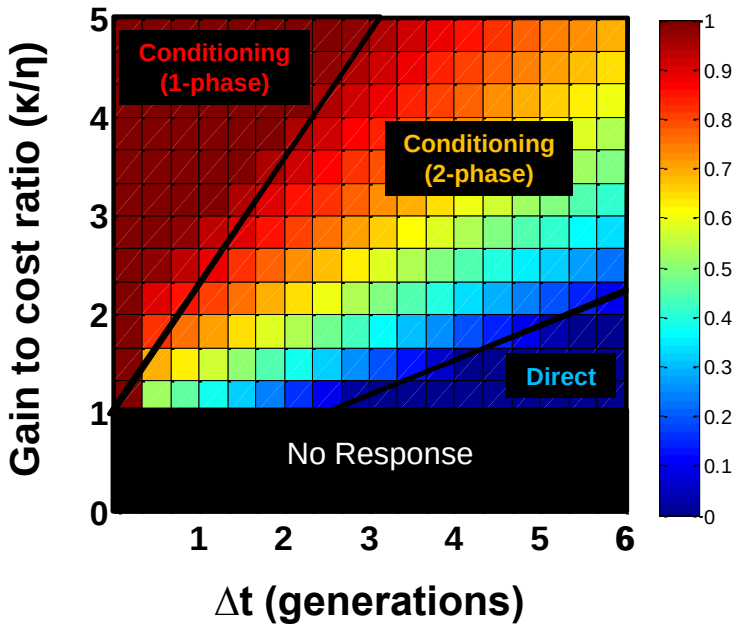


Figure 9. Model-predicted regulation strategies under different value combinations. The model was used to calculate the optimal intermediate level of induction after encounter with S_1 as a function of the gain to cost ratio and the delay time. The color code marks the predicted relative intermediate level. In cases where the cost exceeds the gain not responding is the optimal strategy.

3.2 Adaptive Conditioning in E. coli and the Intestinal Ecology

During its life cycle *E. coli* mainly alternates between two habitats, intestines of mammals and water, sediment, and soil[10]. Focusing on the intestinal ecology reveals a predictable environment that can potentially select for Adaptive Conditioning. Specifically, during passage along the digestive tract, exposure to lactose precedes exposure to another sugar, maltose[11]. I thus expect that this environment can select for Adaptive Conditioning in sugar metabolism – bacteria that link between the presence of lactose and future exposure to maltose may be better utilize maltose resources upon encounter. Note that as explained in the Introduction section, recent work has already shown that this ecology selects for a *symmetrical* association between stimuli[9]. Genes required for coping with either elevated temperature or decrease in oxygen availability are induced in response to each of the stimuli separately.

The study of existing conditioning in *E. coli* is presented here in a few stages. I start by briefly describing the main habitats of *E. coli* as a predictable environment. Specifically, I focus on regulation of maltose and lactose operons to uncover molecular evidence for a two-phase conditioned response strategy that fits the order of stimuli in the intestinal tract. Moreover, I examined whether the observed conditioned response provides a fitness advantage to the organism and is not merely a result of a neutral cross talk between the metabolic pathways. In line with the three criteria outlined the Introduction section, I observed a unidirectional fitness advantage when cells growing on maltose were pre-exposed to lactose. Specificity of the fitness advantage was also observed, pre-exposure to other sugars did not improve cells growth on maltose. Finally, I observed extinction of the conditioned response in *E. coli* strains that evolved in an environment that repeatedly exposed them to lactose (S_1) without the subsequent maltose (S_2), indicating that conditioned response is maladaptive in the presence of S_1.

3.2.1 Intestinal ecology potentially selects for Adaptive Conditioning

During its life cycle *E. coli* alternates between two principal habitats, intestines of mammals and water, sediment, and soil[10]. The population of *E. coli* cycling through the digestive track can be roughly divided into three categories according to their

colonization in this system. The colonization of the newborn intestine occurs within the first few days after birth. From that stage on it is difficult for new strains to become implanted there (yet this process does regularly occur). The crucial point is that the majority of strains found in fecal samples are transient *E. coli* strains that persist in the colon for shorts periods of a few days to a few weeks[22]. Thus, although colonization is a lifelong event from the perspective of the host, the majority of individual *E. coli* cells cycle through this habitat within much shorter periods. A third population of *E. coli* is expected to constantly cycle through the digestive track without any colonization and thus to exit within a few hours after entering it. To conclude, individual cells do experience repeated exposure to the alternating habitats and may eventually condition between some of the typical stimuli in those habitats.

Figure 10A shows the evolutionary cycle composed from the two alternating environments and the typical stimuli. The external environment is rich in oxygen and is usually colder than the temperature controlled anaerobic environment of the intestines. Additionally, nutrients availability is known to dramatically change along the intestinal tract as the host's body breaks down large molecules or absorbs some products. Specifically, during passage along the digestive system, exposure to lactose precedes exposure to another sugar, maltose[11].

The mathematical model was used in order to predict which type of regulation strategy is most beneficial for the induction of the lactose and maltose metabolic pathways. Calculations were based on the known parameter values of the digestive tract (the values of the typical delay and the basal growth rate used are taken from Savageau[11]). Unfortunately, the values of the gain and cost parameters are not known for this system and direct measurement within the natural habitat is extremely difficult. Thus I define a wide range of possible gain to cost ratios. As *E. coli* respond to maltose it is safe to assume that a gain is greater than the cost yielding a lower boundary of one to the ratio. As an upper boundary I used the gain (κ) to cost (η) ratio of the lactose system which reflects the energy yield in a carbon source superior to maltose (the additional cost, η^*, found in the lactose system is ignored as it reflects unique damage phenomenon).

As Figure 10C shows, under this range of possible values the model predicts that a two-phase conditioning strategy will be most beneficial. Thus upon encounter with lactose cells are expected to fully induce the lactose operon and additionally induce the maltose operon to an intermediate level. As this is an asymmetric regulation strategy, upon exposure to maltose, cells are expected to induce only the maltose operons.

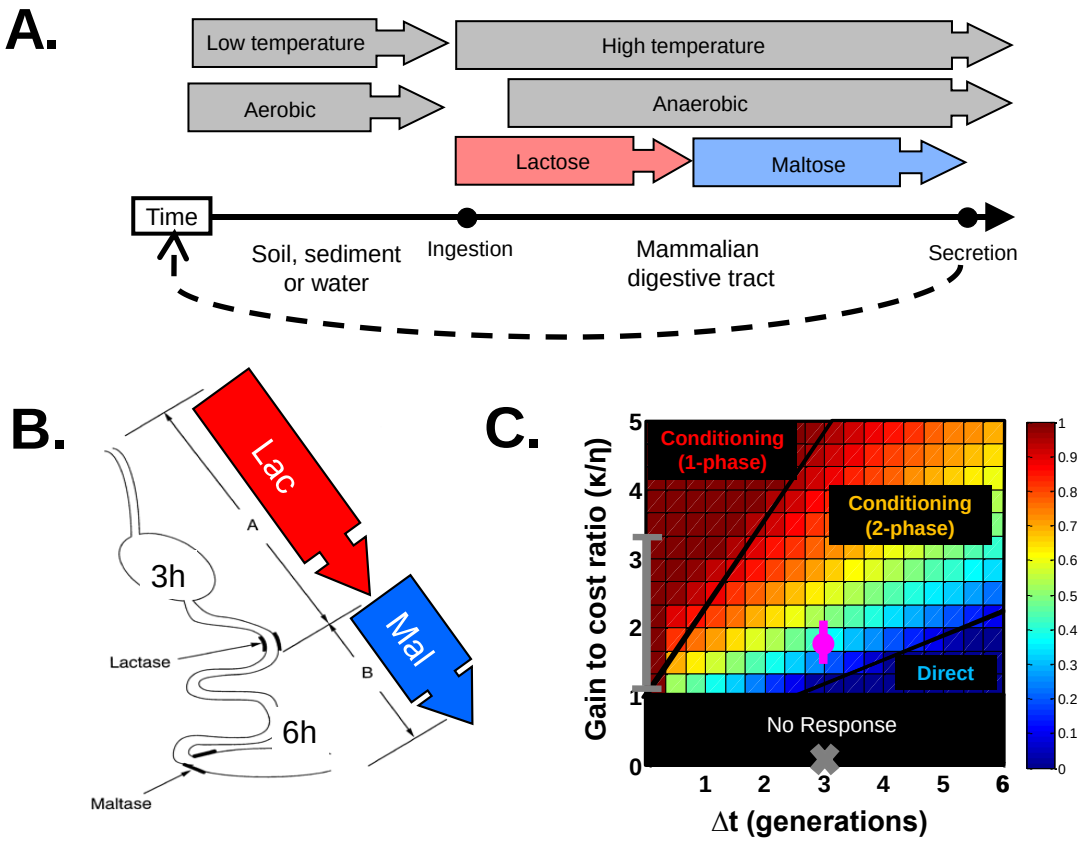


Figure 10. Principle habitats of *E. coli* and possible conditioning between to characteristic stimuli. (A) Two habitats of *E. coli* acting as an evolutionary cycle. (B) Focus on two stimuli in the digestive track, lactose and maltose, and their typical time constants (figure adopted with changes from Savageau[11]). (C) Model-predicted regulation strategies under different combinations of parameter values (similarly to Figure 9). The color code marks the predicted relative intermediate level. The typical delay time and feasible range of gain to cost ratios are marked in grey. The magenta bar depicts the mean intermediate induction levels (0.37±0.2) observed for the five maltose operons in response to lactose in the wild type *E. coli*.

3.2.2 Adaptive Conditioning in lactose and maltose pathways

The mathematical model indicates that the most advantageous strategy in regard to sugar metabolism is a two-phase conditioned response strategy that captures the temporal order of sugar availability in the digestive tract. I reasoned that if the response strategy is implemented through wiring of the transcriptional network, evidence of conditioning can be observed by monitoring the activity of the relevant promoters. In order to systematically test the promoter activity of the operons in the presence of maltose and lactose, I used the *E. coli* promtor-fused GFP library constructed and previously by Zaslaver[14]. The experiments revealed the asymmetrical regulation pattern predicted by the model (Figure 11). Maltose operons that are induced by maltose are also induced, yet to a lower level, by lactose, thus manifesting the predicted two-phase conditioning strategy. Interestingly, basic responsiveness of some of the maltose operons to lactose was observed in a different *E. coli* strain as well[23]. In contrast, and as expected by the hypothesis, the lactose operon remains unaffected by the presence of maltose (Figure 11). The intermediate response level of the five maltose operons to lactose is similar, showing a surplus activity of 0.37±0.2 above the basal promoter activity. This level of partial activity can be plugged back into the model (Figure 10C) in order to estimate the gain to cost ratio in the system. Given an expected three hours time interval and the observed partial activation level, I predict that the ratio is around 1.7. Supplementary Figure 3 demonstrates the two-phase induction dynamics of the maltose operons in response to sequential addition of sugars (lactose followed by maltose).

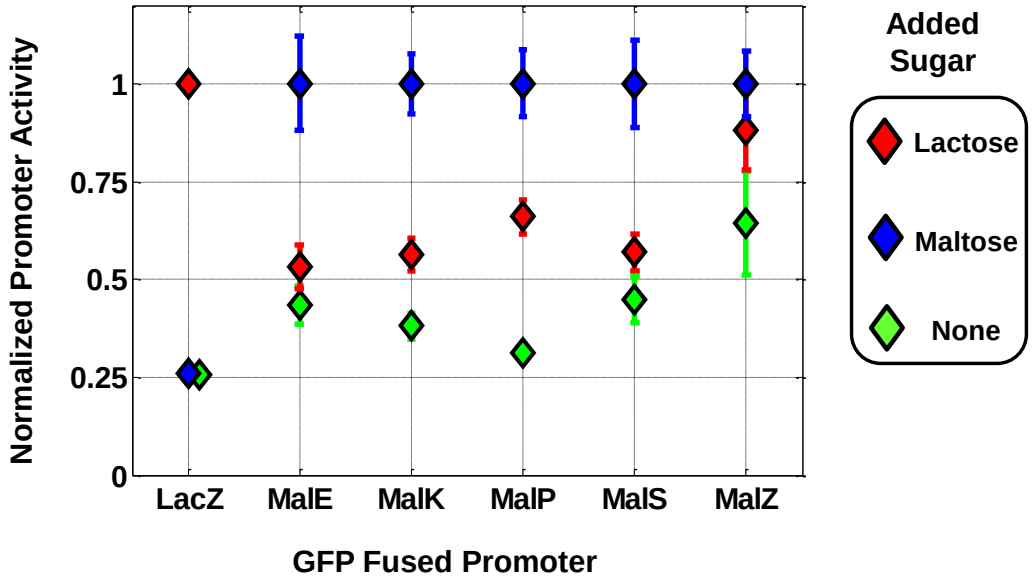


Figure 11. Conditioned response in sugar metabolism of wild-type *E. coli*. The GFP signal of lactose and maltose operons was monitored during exponential growth after addition of either sugars. The normalized promoter activity was calculated by dividing the promoter activities by the maximal promoter activity observed. Bars denote standard deviations of four repetitions. .Promoter activities of four out of five maltose operons are significantly higher under lactose relative to the untreated culture (t-test, $p<0.01$ according to a Bonferroni adjustment).

The asymmetric cross regulation pattern of maltose and lactose operons was further verified using qRT-PCR. Similarly to the results obtained using the GFP reporter library, I observed a pattern of partial induction of maltose operons in response to lactose, as expected by the two phase conditioning response strategy (Figure 12). Reassuringly, the lactose operon remained unaffected by presence of maltose.

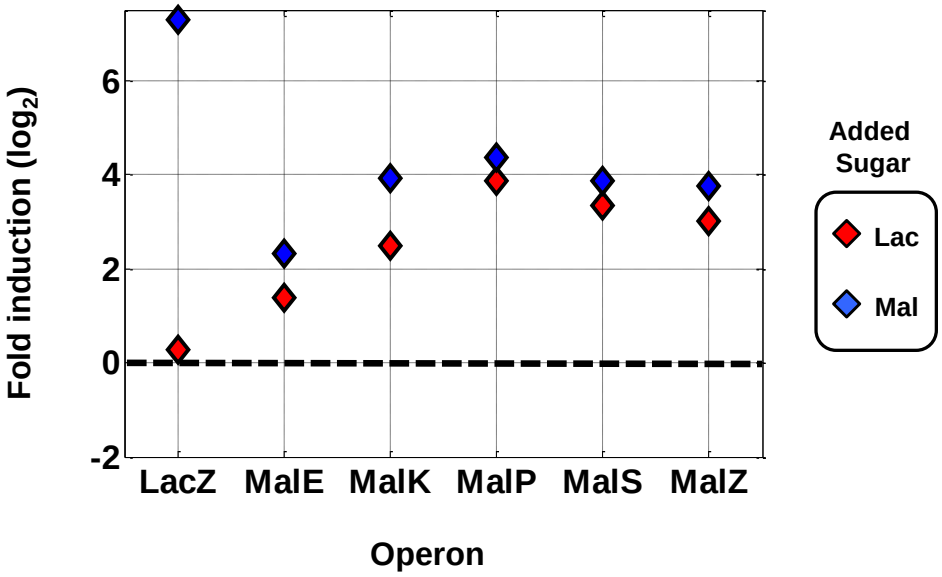


Figure 12. Fold induction (log_2) of lactose and maltose operons in the presence of different sugars as measured by qRT-PCR.

Both in the GFP-fused promoter library and qRT-PCR experiments induction was measured relative to a control population growing exclusively on the background carbon source, glucose (supplemented with cAMP). In order to validate that the response strategy is not affected by the choice of the background carbon source, I repeated the experiments while substituting the glucose by glycerol. Reassuringly, the results similarly indicated on the two-phase conditioned response strategy (Supplementary Figure 4, 5).

3.2.3 Extinction of Adaptive Conditioning in evolved strains

Built into the Classical Conditioning paradigm is the possible extinction of the association – to put in present work terms, it is expected that in an already conditioned strain, repeated exposure to S_1, without consequent arrival of S_2 should select for a weaker connection between S_1 and R_2 due to the futile and costly induction of R_2. Accordingly, I have examined lab-evolved strains, which grew for 500 generations on

high levels of lactose, yet without exposure to maltose[15]. Figure 13A shows a schematic representation of the changing environment: The natural environment is characterized by sequential exposure to lactose and maltose, while the new environment is characterized by daily exposure to lactose only.

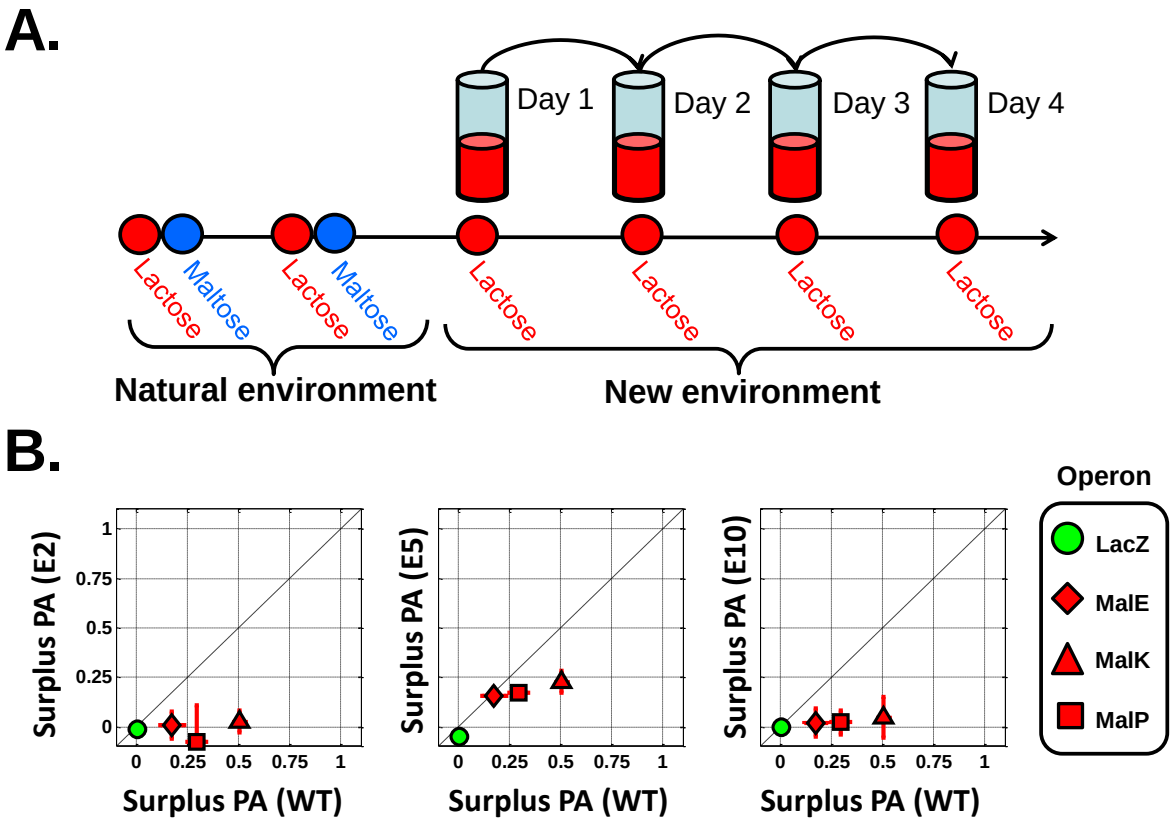


Figure 13. Extinction of conditioned response in evolved strains. (A) Schematic representation of the changing environment: The natural environment is characterized by sequential exposure to lactose and maltose, while the evolution experiment environment is characterized by daily exposure only to lactose. (B) Extinction of conditioning of maltose operons in the evolved strains. Promoter activity (PA) of maltose and lactose operons was measured as previously described. The red shapes mark the surplus promoter activity of the three most highly expressed maltose operons. The surplus promoter activity was calculated by dividing the observed activity on lactose by the activity on maltose after subtracting the basal promoter activity (activity without addition of any sugar to the growth medium). The green circle marks the surplus promoter activity of the lactose operon on maltose. Bars denote standard deviations of four repetitions. The dashed line marks an identical promoter activity of both wild-type and evolved strains.

Strikingly, I found that in all three lines, the maltose operons show markedly reduced activity in response to lactose compared to the wild-type strain (Figure 13B). I

confirmed that the evolved strains preserved both the ability to metabolize maltose (Figure 14A) and promoter responsiveness to maltose itself (Figure 14B). This suggests that only the asymmetrical cross talk between the two sugars was removed during this lab evolution period. These results indicate that the cross talk observed in the wild-type strain is costly and that without a suitable selective force, evolution acts to eliminate the conditioned response. Thus, conditioning emerges as an adaptive trait that is preserved in the wild type strain since the gain exceeds the cost in the natural ecology of the intestinal tract.

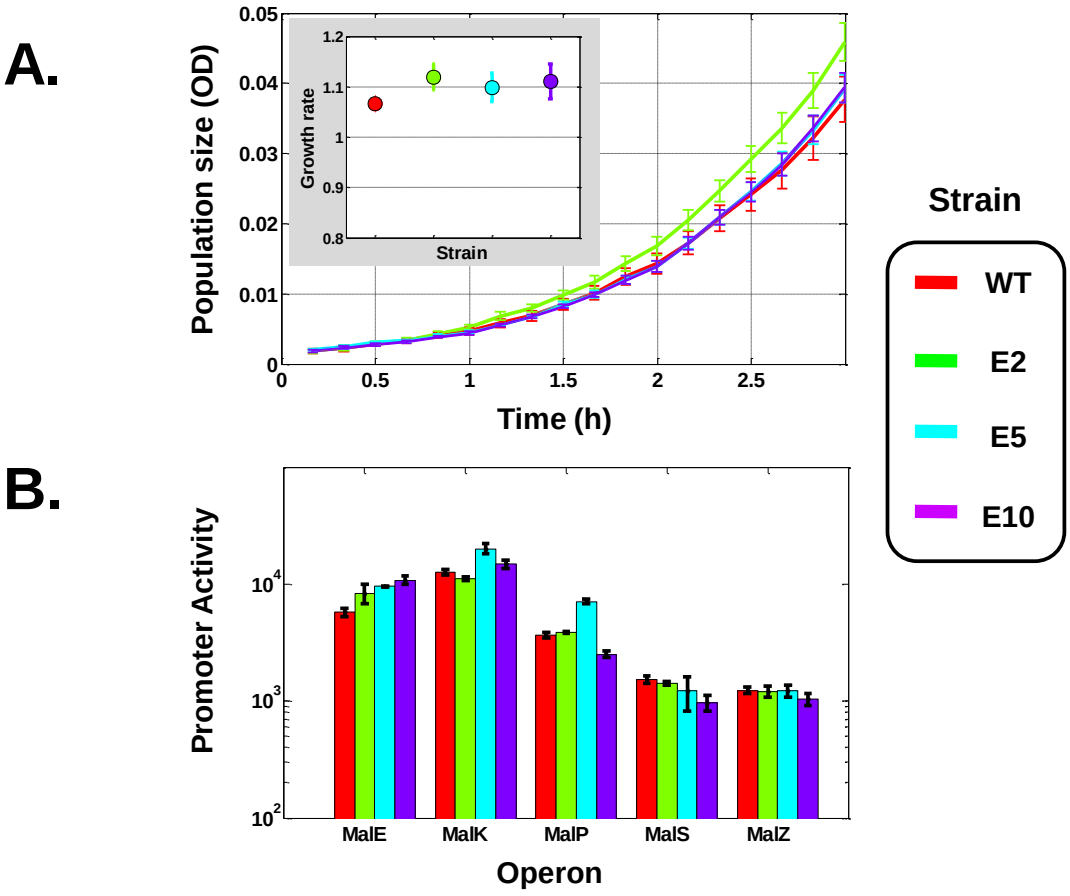


Figure 14. Evidence for an intact maltose response in the evolved strains. (A) Overnight culture of both wild-type (WT) and the evolved strains (E2, E5 and E10) were diluted into fresh media containing maltose as a sole carbon source. OD measurements during 3 hours of logarithmic growth are presented. The inset shows the average growth rate of the different strains during the 3 hour growth. (B) Promoter activity measured for the wild-type and evolved strains are presented for maltose operons.

3.2.4 Fitness advantage of Adaptive Conditioning

Extinction of the conditioned response in strains evolving in a new environment indicates that induction of maltose operons in response to lactose is likely a selectable trait in the natural ecology of *E. coli*. In order to further validate that the observed pre-induction entails an adaptive fitness advantage according to the three criteria outlined in the Introduction section, I monitored the growth of cultures in an alternating sugar environment (Figure 15A) . Note that this system is similar to the setup previously used for the validation of the mathematical model (Figure 8). In these experiments I measured the growth of cells on one sugar (S_2) after an initial treatment with another sugar (S_1). The population size of this conditioned culture was compared to an untreated culture that was not exposed to S_1 prior to S_2 (Figure 15A). Compatible with pre-induction of maltose operons, I observed that the wild-type strain displayed a fitness advantage if growth on maltose was preceded by growth on lactose (black circle, Figure 15B). As required by the asymmetry criterion, no fitness advantage was observed if the order of sugars was reversed (red circle, Figure 15B). Reassuringly, the growth advantage is not observed when testing the evolved strain that features extinction of the conditioned response (cyan circle, Figure 15B). Finally, in line with the specificity criterion, I observed that two alternative chemically close carbon sources, galactose and sucrose, cannot substitute lactose as a preceding signal prior to maltose (yellow circles, Figure 15B). This suggests that the observed anticipation evolved in response to the specific stimuli pair that is presented to the organism in its ecological niche. In summary I conclude that conditioning is an adaptive trait that was likely selected for by the natural ecology of *E. coli*.

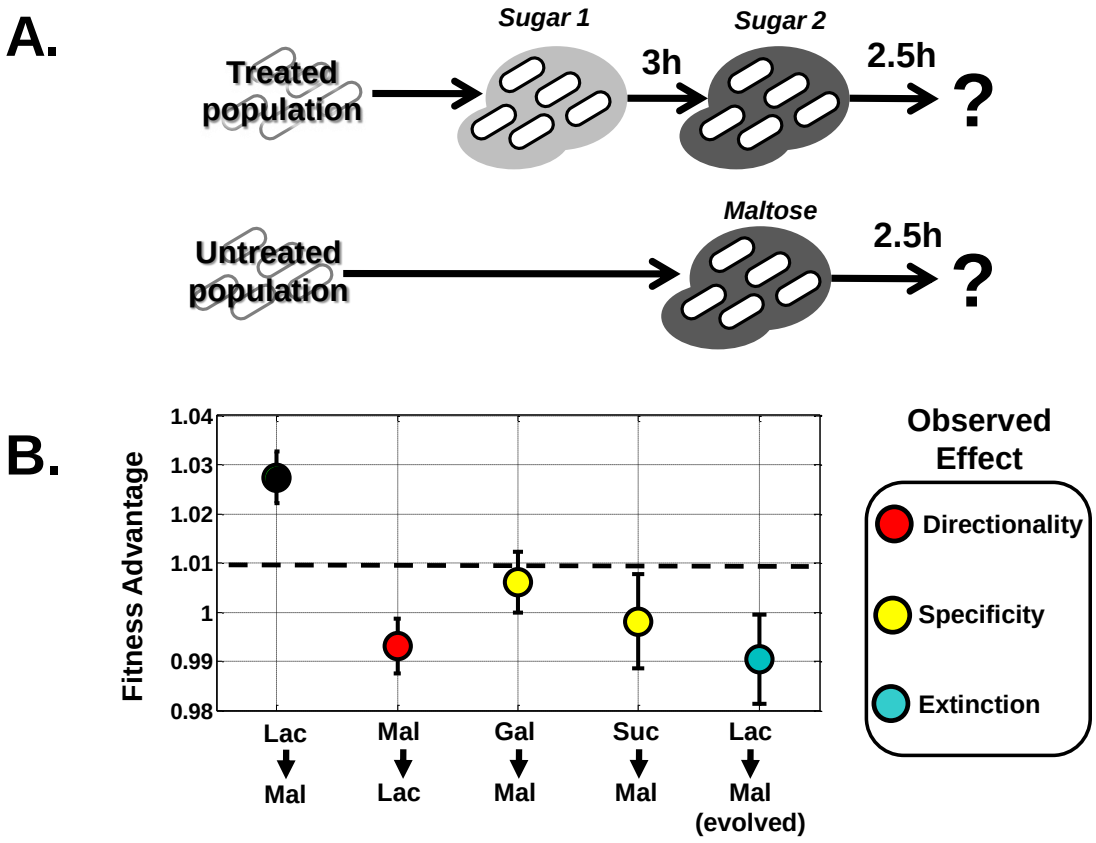


Figure 15. Fitness advantage resulting from pre-induction of maltose operons in wild-type and evolved strains. (A) Experimental setup: In each experiment two isogenic populations were grown on either glycerol supplemented with one sugar - S_1 (treated) or glycerol alone (untreated) for 3 hours. An equal amount of cells were diluted into media containing the second sugar (S_2) as a sole carbon source. The ratio between the population sizes within each experiment (treated/untreated) was taken as the fitness advantage originating from pre-exposure to S_1. (B) Mean fitness advantage according to the three criteria. All ratios are normalized to the ratio measured in a control experiment in which S_2 was only glycerol (Supplementary Results). A significant fitness advantage is observed only in the transfer from lactose to maltose (p=0.02, t-test).

3.3 Adaptive Conditioning in S. cerevisiae and the Wine Ecology

The production process of alcoholic brews exposes the wine yeast, *S. cerevisiae*, to a sequence of stresses in a predictable manner. As traditional wine making relied on yeasts that are naturally found on the grapes and in the winery environment, the process forms an evolutionary cycle that potentially selects for Adaptive Conditioning - cells can use the presence of one stress as a cue for the likely exposure to a subsequent different stress.

The study of existing conditioning in *S. cerevisiae* is presented here in several stages. I start by describing the ecology of wine production as an environment that potentially selects for Adaptive Conditioning. I then continue to focus on core stresses that comprise the diauxic shift, the switch from the preferred catabolic state, fermentation, to the alternative one, respiration. I found that exposure to stresses that occur during fermentation enhances the ability of yeasts to cope with stressful aspects of respiration. This directionality suggests that the yeasts have adapted to the order of stresses typical to the process of wine production. The three criteria for adaptiveness of the conditioned response that were outlined in the Introduction are examined in the context of the increased protection against oxidative stress. Finally, I discuss my study of the transcriptional response as a potential molecular basis of the cross protection between heat shock and oxidative stress.

3.3.1 Wine ecology potentially selects for Adaptive Conditioning

The ecology of fermentation and wine production is reviewed by Pertorius and Bauer[24,1]. Studies reveal that production of alcoholic beverages dates back to 7000 BC. For most part, wine making was carried out by the yeasts that are naturally found on the grapes and in the winery environment. It is important to note that only in very recent times humans started to intentionally intervene with the specific yeast strains carrying out this process e.g., by intentional inoculation. Thus for most history of wine making, alcoholic beverages relied on the natural microflora for a spontaneous fermentation process. As each cycle of fermentation provides a habitat for a vast population of growing yeast cells, it is likely that the evolution of *S. cerevisiae* was greatly influenced

by this process to the extent of selecting for specialized mechanisms under this habitat. Adaptive Environmental Conditioning represents an adaptation strategy that is especially adequate for this type of cyclic life style.

The typical sequence of stresses in the process of wine production is shown in Figure 16A. The initial conditions in the grape must are far from optimal, the starting material has high concentrations of sugar and acid. As fermentation starts the temperature can rise to lethal levels[25,26]. As fermentation progresses, additional stresses appear, rising levels of ethanol, exhaustion of assimiable nitrogen sources and later glucose starvation. When glucose is almost fully consumed the cells go through the diauxic shift and replace fermentation metabolism with respiration which exposes them to reactive oxygen species that originate from active mitochondria[27].

3.3.2 Adaptation to the order of stresses typical to wine production

The wine ecology potentially selects for Adaptive Conditioning between the subsequent stresses. As a potential manifestation of this adaptation I focused on the cross-protection phenotype[28,29,30] between different stress pairs typical to this process. A cross-protection phenotype exists if pre-exposure to one stress improves the survival of cells under a subsequent stress. Under the hypothesis of Adaptive Conditioning I expected an asymmetrical cross-protection phenotype between stress pairs. Particularly, I predict that directionality of cross-protection will be in accordance with the order of events during the transition from fermentation to respiration. For example, as ethanol precedes oxidative stress in wine production, I expect improved survival under oxidative stress if it is preceded by ethanol stress, yet I do not expect cross-protection if the conditions are applied in the reverse order. It is important to note that the protection activity is influenced by the stresses level applied on the cells. Thus, for consistency, each stress type was calibrated to achieve either mild or severe effects on cell survival (Figure 4 and

Table **1**). Cross protection was measured in cultures transferred from mild stress to severe stress.

Figure 16A shows the measured protection phenotype systematically tested for stresses that occur before and after the diauxic shift (see also Supplementary Table 1). In accordance with the asymmetric criterion, I observed that the cross-protection phenotype is indeed extremely asymmetric and most importantly, that the directionality of protection fits the natural temporal context. Specifically, for each stress pair, protection is always stronger in the direction found in the natural environment. Note that activation of the environmental stress response[4,5], namely a set of genes that are transcribed in response to a wide spectrum of stresses, is less likely to explain the observed cross-protection activity. Whereas the environmental stress response is more likely to result in a symmetrical protection, I observed a very asymmetric effect.

I further validated that the protection phenotype is maintained if the yeasts go through a succession of the three stresses. This was done in order to rule out the possibility that exposure to ethanol lowers the protection facilitated by the initial heat shock against oxidative stress. I thus repeated the same kind of cross-protection experiments while exposing cells to mild heat-shock, followed by mild ethanol, and subsequently by a severe oxidative stress (keeping the same time gaps and stress levels as before). The experiments reveal that pre-exposure to heat-shock followed by ethanol gives similar high cross protection efficiency against severe oxidative stress (fold cross protection of 58±2.9).

Focusing on cross protection against oxidative stress revealed the specificity of cross protection. As Figure 16B shows, high cross protection occurs for stresses that typically precede oxidative stress in the natural ecology. This specificity further supports the possibility that the observed cross protection originates from specific adaptation to the temporal order of stresses in wine production rather than from the general stress response.

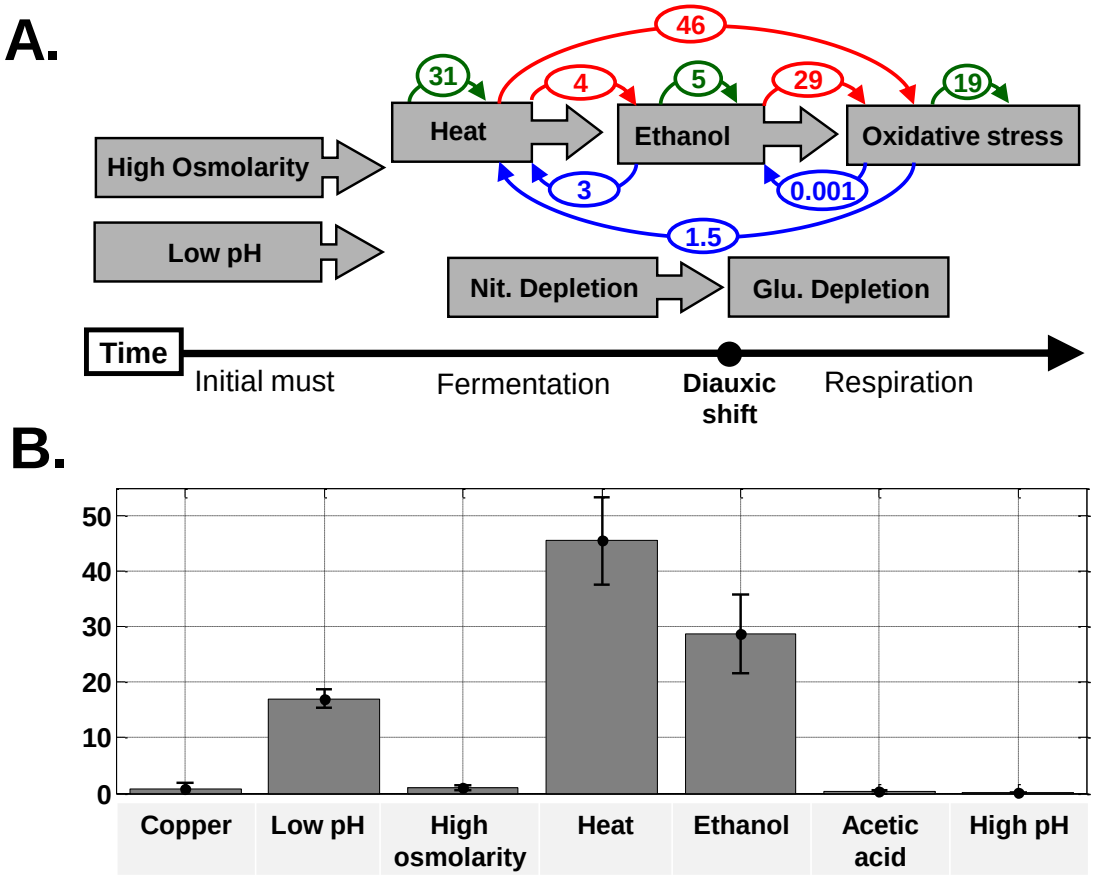


Figure 16. Typical stresses in the ecology of wine production and cross protection phenotype. (A) Temporal sequence of stresses in the process of wine production[24,1] and cross-protection phenotype observed between stresses that occur during the transition from fermentation to respiration. The values on the arrows denote the mean fold protection gained by pre-exposure to a given mild stress followed by a second severe stress. Experiments were done in three independent repeats (see Supplementary Table 1 for standard deviations). Red and blue arrows mark cross protection when stress order is in accordance with the ecology or in reverse order, respectively. Green self-pointing arrows denote cases of auto-protection. (B) Specificity of cross protection against oxidative stress. The columns mark the observed fold cross protection donated by exposure to mild stress prior to severe oxidative stress, bars donate standard deviations.

3.3.3 Adaptive Conditioning in the transcriptional response to stress

I next turned to investigate a potential molecular mechanism that might account for the observed cross-protection capacity. I focused on the pair of stresses that exhibits

the strongest cross-protection phenotype, heat shock and oxidative stress, and measured genome-wide transcriptional response under these stresses both when they are introduced in isolation or in continuation. As a first stage I restricted the analysis to genes that are essential for the stress response, that is, genes that when mutated cause a fitness reduction under the particular stress conditions. Specifically, I defined gene sets that are annotated in the Proteome BioKnowledge Library (a manually curated, proteome-wide compilation of scientific literature) as essential exclusively, for one of the stresses, heat shock or oxidative stress. This strict definition allowed discarding key genes that are essential for the general stress response while defining non-overlapping essential responses for each one of the stresses. I found that the observed directionality of cross-protection is largely recapitulated in the pattern of gene expression (Figure 17). Particularly, I observed that in response to heat shock, both heat shock exclusive genes and also the oxidative stress exclusive genes are over-expressed. As predicted by the Adaptive Conditioning hypothesis, the results also showed that this response is rather asymmetrical, i.e., in response to oxidative stress the oxidative exclusive genes are over-expressed, while the induction of heat-shock exclusive genes is not statistically significant (Figure 17). Furthermore, I noted that the response of the oxidative exclusive genes to heat-shock is significantly higher than the response of the heat-shock exclusive genes to oxidative stress ($p = 0.043$, using Fisher's exact test).

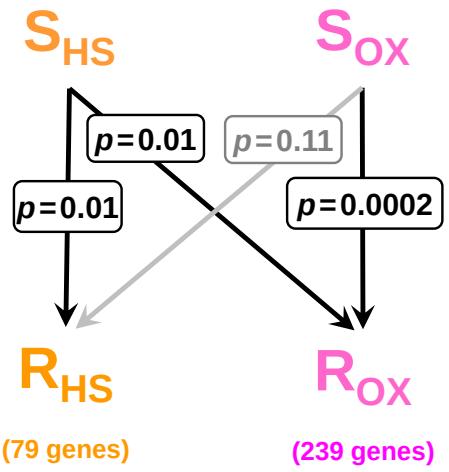


Figure 17. Asymmetrical transcriptional response to heat shock and oxidative stress measured by microarrays. A binomial p-value was used to examine whether sets of exclusive essential genes are over-expressed under each of the two stresses

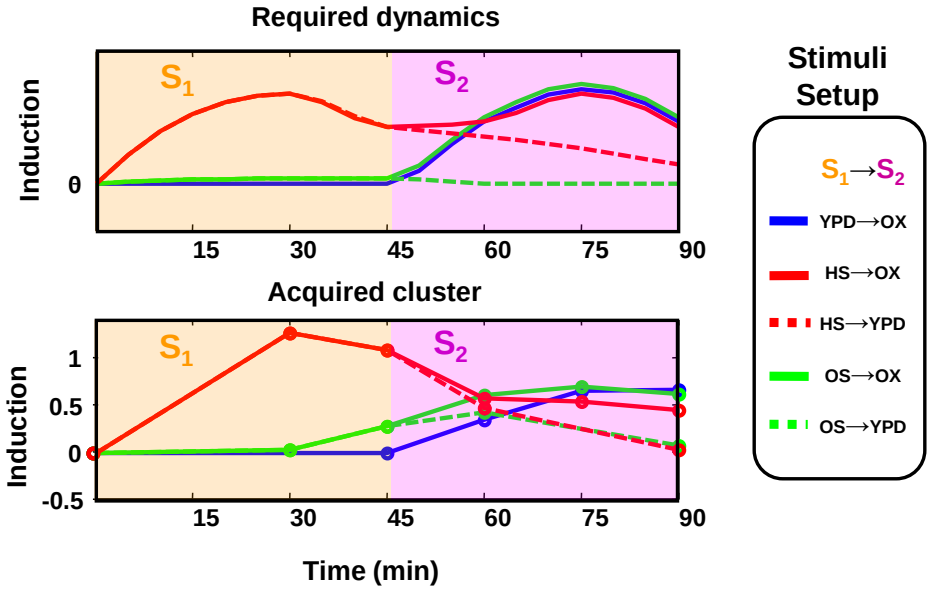


Figure 18. Identifying candidate genes that may underlie the unidirectional cross-protection phenotype between heat shock and oxidative stress. Genome-wide transcriptional response under two sequential stresses was measured in response to different stimuli series. The upper panel marks the expression

dynamics of an ideal gene that may facilitate cross-protection against oxidative stress under these different stimuli setups. The lower panel marks the mean profile of 300 genes matching the desired expression pattern. The ideal gene profiles were defined based on the following criteria in the stimuli setups: 1) transfer from rich medium (YPD) into an oxidative stress (blue line): the ideal gene is induced following the transition. 2) Heat shock followed by oxidative stress (red line): the gene is induced already under heat-shock and induction is sustained later in the oxidative stress. 3) Heat shock followed by rich medium (red dashed line): the ideal gene is induced under heat shock and decays when cells are moved to the rich medium. 4) An osmotic stress followed by oxidative stress (green line): the gene is only induced under the oxidative stress and not in the control osmotic stress. 5) Osmotic stress followed by rich medium (green dashed line): the gene does not respond at all.

Essential for the understanding of Adaptive Conditioning is to monitor gene expression when stresses are applied in a sequential order. Since I observed that heat-shock activates oxidative exclusive genes, I next carried out experiments in which application of heat shock (S_1) was followed by an oxidative stress (S_2). As a negative control I also examined the response to oxidative stress preceded by an osmotic stress, a condition found not to protect significantly from oxidative stress (fold cross protection of 1.06). I used these series of stresses to define an ideal expression profile of genes that can underlie the cross-protection phenotype and ultimately conditioning (Figure 18, upper panel). Basically the profile describes genes pre-induced in heat shock, induced in oxidative stress and unaffected by osmotic stress. In a supervised search I identified a cluster of 300 genes following the required dynamics (Figure 18, lower panel). Notably, this set shows a significant overlap with the oxidative exclusive genes (p=0.013) but not with heat shock exclusive genes (p=0.2).

In order to further clarify the role of the acquired gene cluster I tested for functional enrichment using the Gene Ontology (GO) annotations[31]. The analysis revealed a significant enrichment of oxidative stress categories (Table 2), such as oxidoreductase activities, and many mitochondria related categories, consistent with a known role of mitochondria in oxidative stress in yeasts[32]. Furthermore, no heat shock related categories were found to be enriched in this set.

Table 2– GO enrichment in the gene set matching the ideal profile (300 genes)

p-value	Description	Category
2.11E-25	mitochondrial matrix	GO:0005759
2.11E-25	mitochondrial lumen	GO:0031980
8.49E-13	Mitochondrion	GO:0005739
8.49E-13	mitochondrial part	GO:0044429
2.22E-12	mitochondrial ribosome	GO:0005761
2.22E-12	organellar ribosome	GO:0000313
3.79E-11	organellar large ribosomal subunit	GO:0000315
3.79E-11	mitochondrial large ribosomal subunit	GO:0005762
2.33E-05	oxidoreductase activity	GO:0016491
5.00E-04	large ribosomal subunit	GO:0015934
0.000468	structural constituent of ribosome	GO:0003735
0.00124	ribosomal subunit	GO:0033279
0.00387	Cytoplasm	GO:0005737
0.00453	oxidoreductase activity, single donors	GO:0016701
0.00453	dioxygenase activity	GO:0051213

Reassuringly, the genes that corresponded to the control profile (in which I still used oxidative stress as an S_2, yet with an un-protective S_1) did not yield similar enrichment for oxidative related activities (Table 3). This indicates that while an S_1 stimulus that provides cross-protection elevates in advance genes relevant for the R_2 response, a control S_1 signal does not affect these R_2 genes.

Table 3 - GO enrichment in the gene set matching the control profile (494 genes)

p-value	Description	Category
0.00316	pre-autophagosomal structure	GO:0000407
0.00362	cellular iron ion homeostasis	GO:0006879

0.00362	iron ion homeostasis	GO:0055072
0.00537	protein catabolic process	GO:0030163
0.00555	protein targeting to vacuole	GO:0006623
0.00686	Sporulation	GO:0030437
0.00686	reproductive sporulation	GO:0048622
0.00919	Proteolysis	GO:0006508
0.00919	cellular cation homeostasis	GO:0030003

The analysis of gene annotations in the conditioned cluster, both according to GO categories and by essential genes sets, suggests that this response arm that is induced under heat shock might not be required for coping with the heat stress itself. Thus it seems that this cluster is indeed neutral upon encounter with heat-shock, in line with the third criterion of adaptiveness. However, in order to test this notion directly I designed another experimental system. Towards this aim I collected 29 strains[18], each mutated in a gene from the cluster described above and tested their sensitivity to heat shock and oxidative stress. Analysis of survival ratios reveal that while 65% of the strains show increased sensitivity to oxidative stress, none shows a significantly increased sensitivity to heat shock (Figure 19, Supplementary Table 2). This is a direct indication that the expression of many of the heat-induced genes is at least neutral, if not maladaptive, under heat-shock. Taken together with the fact that these genes are essential under oxidative stress, this experiment corroborates the claim that the early preparation to oxidative stress during heat-sock is adaptive in this system.

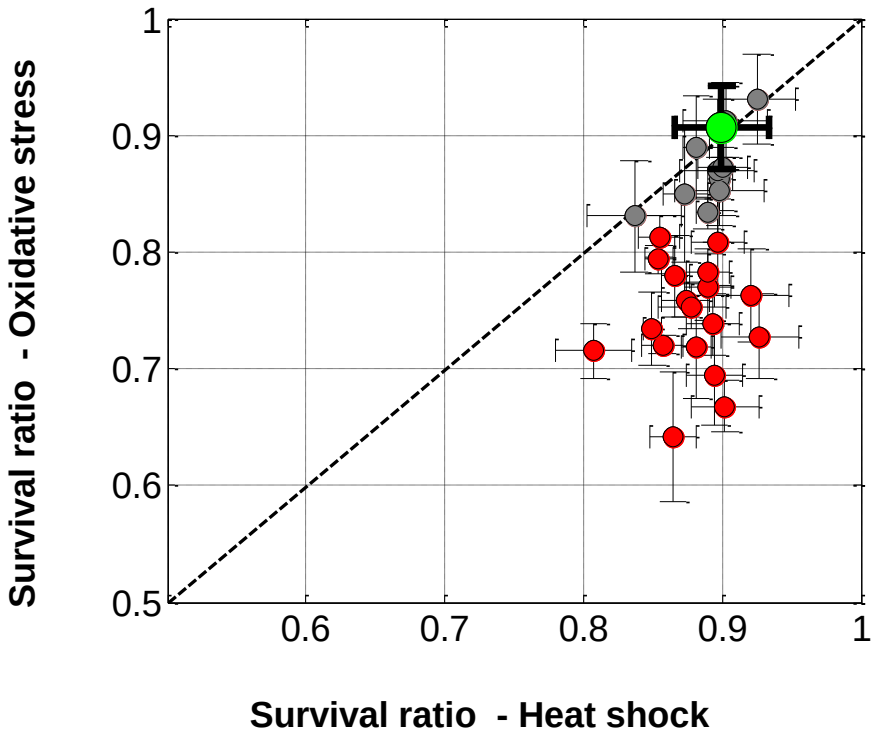


Figure 19. Survival ratios of wild type and deletion strains under heat shock and oxidative stress. The survival ratio was calculated by dividing the OD of a stress treated culture by the OD of the untreated culture. The survival ratio of each strain was measured in heat shock and oxidative stress in four independent repeats, bars denote the standard errors. The green circle marks the survival of the wild type. A t-test was used to determine whether the difference in the survival ratio of a deletion strain and the wild type is statistically significant in a given stress (FDR=0.05). Red circles mark deletion strains that show significantly reduced survival under oxidative stress. Grey circles mark deletion strains with similar sensitivity to oxidative stress as the wild type strain. Note that all deletion strain show similar sensitivity to heat shock as the wild type strains. See Supplementary Table 2 for further details.

The conditioned cluster of 300 genes obtained by a supervised search follows the proposed dynamics of an ideal conditioned gene. I next turned to test whether a similar cluster can be obtained through an unsupervised approach. Towards this aim I identified all genes that showed induction in response to oxidative stress (a total of 1700 genes) and used standard clustering methods to partition them into co-regulated clusters. As Figure 20A shows, hierarchical clustering yields three major clusters with distinct induction profiles under the different experimental setups. The biggest cluster (denoted ESR -

environmental stress response in Figure 20), represents a core response common to the three stresses. Next I observed an oxidative unique cluster (denoted OX in Figure 20) which consists of genes that are expressed only upon encounter with the oxidative stress. The last cluster (denoted CP – Cross protection in Figure 20) consists of genes that are pre-induced in heat shock, induced in oxidative stress and remain unaffected by osmotic stress. The mean expression profile of this cluster is very similar to the ideal dynamics defined previously under the supervised approach (Figure 18, lower panel). Accordingly, this cluster is also enriched in mitochondria and oxidoreductase related categories (results not shown).

Focus on the mean expression profiles in the ESR cluster point to another intriguing effect. Although this gene set is induced under all experimental setups, the mean induction of the cluster is considerably lower under oxidative stress that was primed by heat shock. Importantly, a similar effect is not observed in the other two clusters (CP and OX). The decreased activity of the ESR cluster in cells benefiting from cross-protection might indicate that this cluster reflects the stress level experienced by the cells. This internal stress level can be viewed as the "subjective" stress, differing from the "objective" stress – the level we chemically apply on the cells. The fact the two other clusters do not feature this reduced effect might indicate that they react to the "objective" stress level.

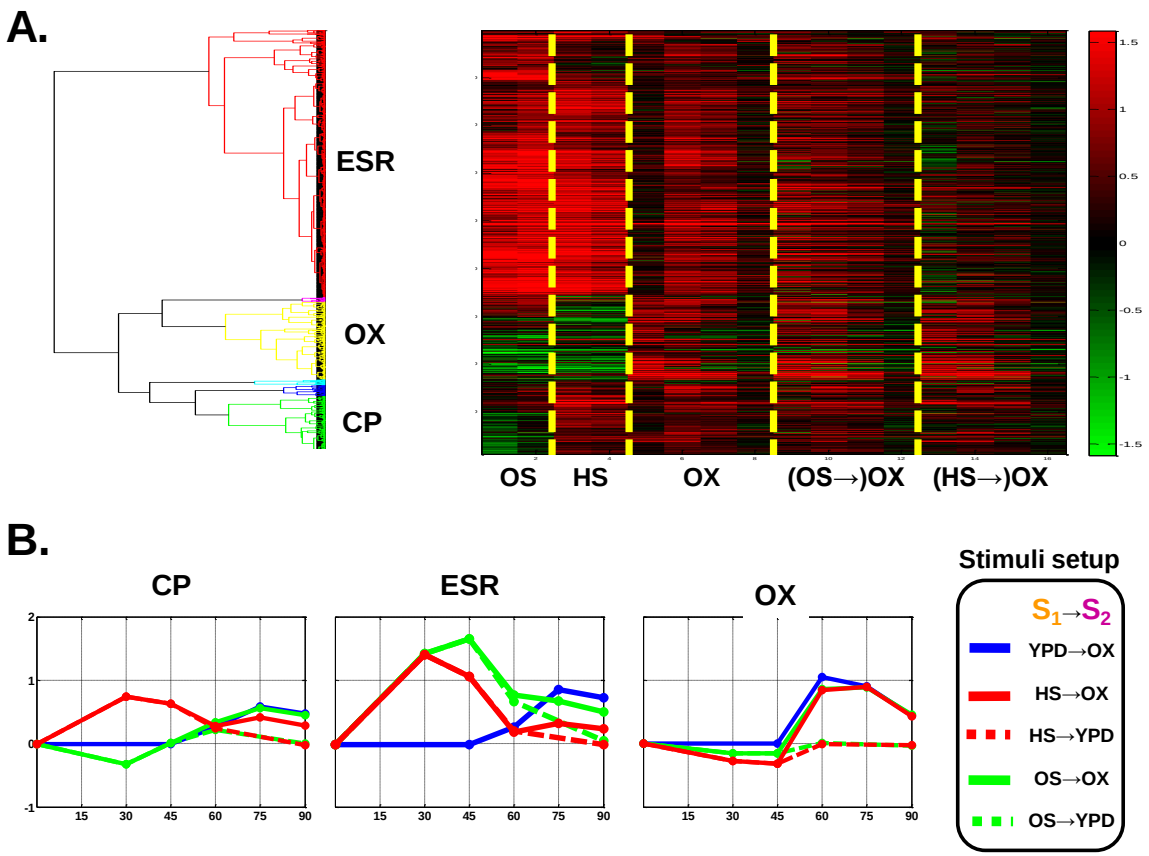


Figure 20. Unsupervised clustering of oxidative responsive genes. (A) Main clusters obtained by hierarchical clustering of expression data under osmotic stress (OS), heat shock (HS), oxidative stress (OX), oxidative stress that was primed by osmotic stress ((OS→) OX) or oxidative stress that was primed by heat shock ((HS→) OX). (B) Mean expression profile of the three major clusters obtained.

3.4 Directed Evolution towards Adaptive Conditioning

Adaptive conditioning is proposed to be an advantageous evolvable trait under explicit selective circumstances – a consistent environment of recurring stimuli. I thus reasoned that Conditioning can be created, *de novo*, in a directed evolution experiment given the appropriate selective pressure. In this line of experiments *E. coli* cultures were diluted daily into a new growth medium and were sequentially exposed to two stimuli to eventually select for lineages of cells that condition between the two. Control cultures were exposed daily to identical stimuli but simultaneously rather than sequentially.

Two parallel experiments were conducted, each one designed to select for conditioning between the appearance of maltose with future arrival of either stress (heat shock) or a superior carbon source (lactose). The serial transfer experiment was conducted for nearly four months, with a total of 113 cycles of exposure to the stimuli pair. Unfortunately, there was no phenotypic indication of conditioning in the evolved strains. Culture evolving on maltose followed by lactose were assayed by plating them on agar plates containing either lactose or maltose as a sole carbon sources while assaying the activity of the β-galactosidase enzyme (by x-gal coloring). Results showed no elevation in β-galactosidase activity when evolved cultures grew on maltose (relative to the ancestral wild type strain). This observation was taken as evidence that the lactose pathway was not conditioned to respond to maltose in the evolved strains.

Cultures evolving on maltose followed by heat shock were assayed by testing whether addition of maltose prior to heat shock enhances their survival (relative to cultures untreated with maltose). Results showed a similar survival rate of cultures irrespectively to prior addition of maltose, indicating the heat shock response was not conditioned to maltose in the evolved strains. Similar results were also observed in the evolved control cultures, these cultures were exposed daily to maltose and heat shock simultaneously. It is important to note however, that the phenotypic assays used to test for a conditioned response might not have been accurate enough to detect subtle cases of conditioning. For example, it is possible that although a subset of the heat shock response genes was elevated in response to maltose, the increase in survival was undetectable.

Although a conditioned response was not observed in the evolved strains, other adaptations to the experimental setup were fixated. Specifically, all lineages improved their growth on the background carbon source used during the evolution experiment (glycerol). In order to identify the mutations that might underlie this improvement, genomic loci from the coding sequence of enzymes in the metabolic pathway of glycerol were sequenced. These sequences were previously shown to affect the growth rate on glycerol after an evolution experiment that selected for increased growth rate on this carbon source[33]. In total, six different loci were sequenced in eight evolved strains and in

the wild-type ancestral strain. Analysis pointed to a single missense mutation in the coding sequence of glycerol kinase in all the evolved strains but not in the wild type strain. This single mutation leads to a substitution of Arg→Leu at position 237 of the protein sequence. As Figure 21 shows, imposing the substituted residue on the three dimensional structure of the protein reveals that the mutation is not located in spatial proximity to the active site of the enzyme. However, the mutation is located near two key residues previously shown to considerably improve the kinetic properties of the enzyme, one of which can support an increase of 60% in the growth rate on glycerol[33]. I note that the observation of the exact same mutation in all the evolved strains points to a possible uncontrolled transfer between the independent lines of the evolution experiment.

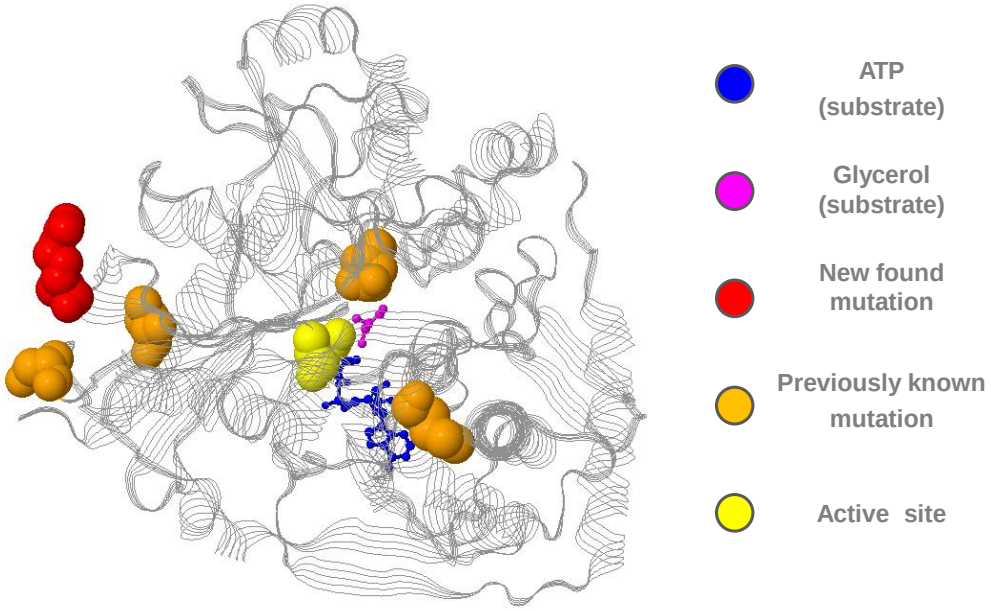


Figure 21. Three dimensional structure of the *E. coli* glycerol kinase (PDB structure 1bo5). The mutation found in the evolved strains (red) is located in spatial proximity to two positions previously known to considerably improve the kinetic properties of the protein[33].

Increased resistance to heat shock was another prominent adaptation observed in the all strains daily exposed to heat. The stress level used in the experiment was calibrated to achieve a survival rate of 50% in the ancestral wild type strain. Figure 22 shows increased survival rates measured for the all evolved strains. As noted before, the increased resistance was not influenced by prior addition of maltose to the growth medium and was also observed in the control cultures that were exposed daily to maltose and heat shock simultaneously.

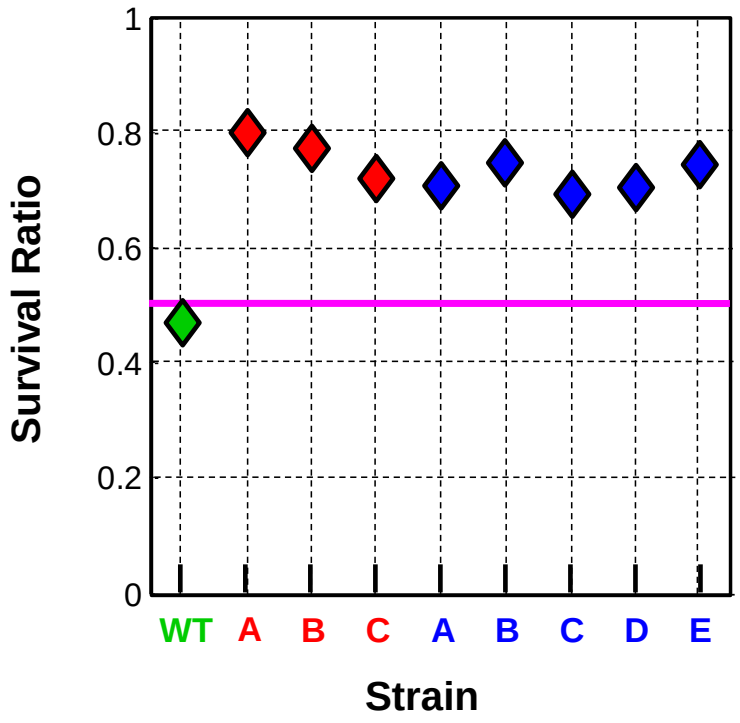


Figure 22. Heat resistance in wild type and evolved strains. The diamonds mark the measured survival of ancestral (green), control (red), and conditioned (blue) strains. The heat shock used in the evolution experiment was calibrated to achieve a survival rate of 50% in the wild type strain.

4 Discussion

The discussion section follows the structure of the Results section. In each of the subjects I briefly review the results and discuss my conclusions and possible future work. I then continue to discuss the conclusions that arise from my hypothesis as a whole.

4.1 A Mathematical Model of Adaptive Conditioning

My study begins with the development of a mathematical model as a tool to understand the key forces that affect selection of this response strategy. The model sets to predict the fitness advantage gained by Adaptive Conditioning over the Direct Response strategy in a given environment. The Direct Response strategy is taken as a *null model*, the simplest alternative strategy that is typically assumed to exist between a stimulus and the cellular response. The described model was intended to remain simple and intuitive while relying on a minimal set of biological assumptions. Specifically, note that the key elements, such as the cost, gain and response dynamics are introduced into the model separately. This modularity allows the model to be adjusted and further developed to represent other, more complex systems. The most direct example for this ability to elaborate the model is the incorporation of an additional cost parameter required in the specific case of pre-induction of the lactose operon (Figure 8).

I have also designed an experimental setup to test the accuracy of the model predictions in a specific system, induction of the *E. coli* lactose operon prior to lactose exposure. The good agreement between the model's predictions and the observed results (Figure 8B) indicates that the even this simplistic model can be used to estimate the fitness advantage of conditioning given basic parameters of the biological and environmental system. Thus, the model can be used as a predictive tool to identify ecological systems that potentially select for Adaptive Conditioning (Figure 10C). Indeed, I observed that the model successfully predicted the two-phase conditioned strategy used by *E. coli* in the regulation of the lactose and maltose pathways (Figure 11 and 12). It is noteworthy to add that the successful prediction was done although some values of biological parameters were unknown. This indicates that the model might be utilized to study even ecological systems that are not fully characterized.

The mathematical model can be further expanded to incorporate more realistic assumptions about the environment and the biological system under study. For example, consider the simplistic assumptions regarding the typical time constants of the environment (e.g., Δt has one single value). In a faithful emulation, the value of each of the time parameters will be taken from a range of possible values according to an assumed probability distribution.

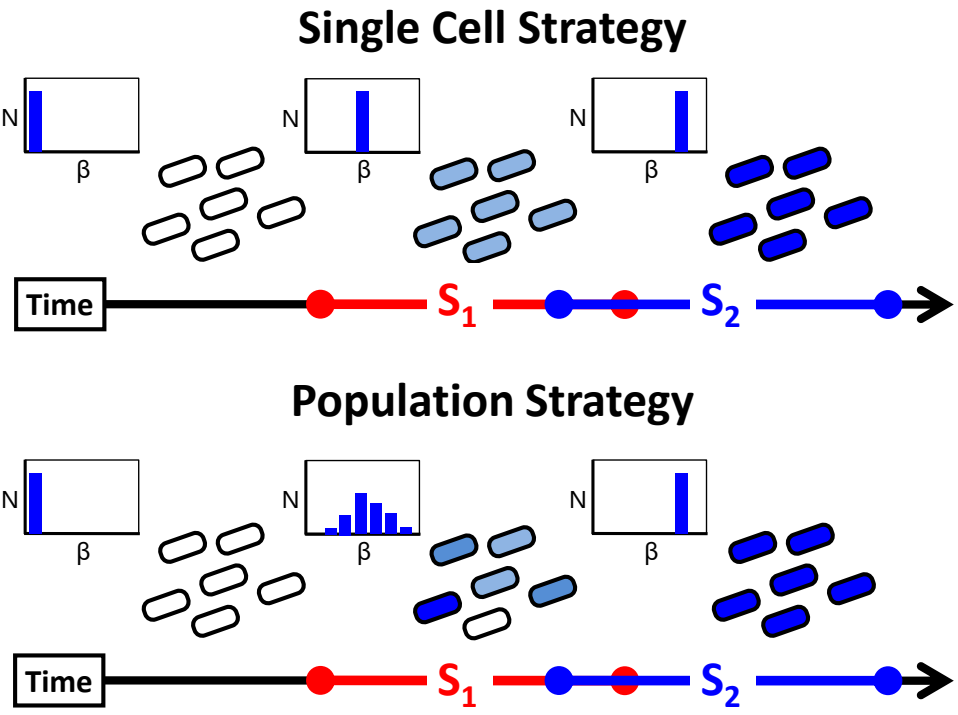


Figure 23. Different implementations of Adaptive Conditioning. The insets mark the distribution of R_2 production rate (β). Note that in the population model the response level upon encounter with S_1 varies in different individuals.

In the scope of this discussion I want to propose another, more exciting, possibility for expansion of the existing model. I suggest implementation of Adaptive Conditioning as a population response strategy rather than a mere single cell strategy (Figure 23). Under this expansion, upon encounter with the conditioning stimulus, S_1, the target population can create a wide heterogeneity in the conditioned response R_2. This heterogeneity can be viewed as an economical stochastic response strategy (Figure 2).

The heterogeneity does not exist under normal growth conditions but is only triggered upon encounter with the predictive signal. The level of heterogeneity can also be modeled using a probability distribution allowing a whole range of population dynamics to be explored (see inset in Figure 23).

4.2 Adaptive Conditioning in E. coli and the Intestinal Ecology

After developing the mathematical model I turned to explore contemporary living organisms and their ecologies in search for potential existing cases of Adaptive Conditioning. I focused on the known ecology *E. coli* and the digestive tract as an environment that may have selected for adaptive Conditioning in regard to sugar metabolism.

The analysis of this ecological system started with the use of the mathematical model as a predictive tool. Although the values of some parameters are unknown in the environment, the model predicts that a two-phase conditioned strategy will be the most advantageous strategy for a wide range of values (Figure 10C). This observation can be taken as evidence that the intestinal ecology selects for Adaptive Conditioning irrespectively of the precise values of these parameters. However, another, more interesting, conclusion can further be drawn: Conditioning can be viewed as a robust response strategy that is advantageous under a wide range of biological and environmental circumstances that in fact exist in nature. Consider for example, the time estimate of a three hour delay between lactose and maltose[11]. This single value can probably not account for all mammalian digestive tracts. However, according to the model's predications conditioning emerges as the most adventurous strategy if the delay periods in nature range from almost zero to five hours (Figure 10C).

The experimental research of the *E. coli* system started with the study of the transcriptional response of lactose and maltose operons in order to reveal the response strategy used by the wild type *E. coli*. Results were consistent with the model's prediction and indicated an asymmetric regulation pattern that captures the temporal order of sugars in the digestive tract - maltose operon are inducted to an intermediate level in response to

lactose while the lactose operon remains unaffected by maltose (Figure 11). The analysis was repeated using different methods and background carbons sources to ensure that the findings are not an artificial result of the experimental system used (Figure 12, Supplementary Figure 4 and 5).

I next continued to monitor the transcriptional response of the relevant operons in *E. coli* strains that evolved under a new environment. The strains grew for 500 generations in a lactose rich environment without any exposure to maltose. I reasoned that the environment in this experiment could select for extinction of the conditioned response as lactose, the natural preceding stimulus (S_1) can no longer serve as a predictor for maltose, the natural subsequent stimulus (S_2). Indeed as expected, I observed extinction of the conditioned response in three independent evolved strains. This finding points to unwiring of the regulatory network that left the cells with a direct response strategy. Furthermore, as measurement of promoter activity was performed using the naïve promoters (fused-GFP plasmids) that were introduced to the evolved cells, it is likely that the loss of the regulatory connection was done in *trans* – the original intact promoters remain unaffected by lactose. It is important to note however that this observation does not exclude the possibility that extinction also took place in *cis*.

Finally, I examined the three criteria outlined in the Introduction section to test whether the observed asymmetric regulation pattern forms an adaptive trait that could be selected for during evolution. Towards this aim, I designed an experimental system that can measure subtle fitness advantages in an alternating sugar environment (Figure 15A). In line with the first criterion, I observed a unidirectional fitness advantage in the transfer between lactose and maltose but not if the order of sugars is reversed (Figure 15B). Note that this directionality fits the order of sugars in the intestinal tract. Furthermore, this advantage is specific to pre-exposure to lactose and not to other sugars with similar break down products, galactose and sucrose (Figure 15B). In line with the extinction observed in the evolved strains at the transcriptional level, pre-exposure to lactose does not increase fitness upon subsequent growth on maltose in these strains (Figure 15B).

To conclude, the observed asymmetrical response strategy in the wild type *E. coli* fits the hypothesis of Adaptive Conditioning. Furthermore, testing the three criteria outlined in the Introduction indicates that the conditioned response is most likely an adaptive trait that has been selected for during evolution. A question remaining open in this system regards the implementation of the conditioned response at the single cell level. As proposed in the previous section, conditioning can be implemented as a response strategy at the population level through creation of heterogeneity between individuals (recall Figure 23). According to this hypothesis, in response to lactose, individual cells will induce the maltose operon to different levels. Since the promoter activity was measured by monitoring the GFP signal of a culture, the dynamics at the cellular level remain unknown. I note that monitoring the response of individual cells is feasible with the use of a fluorescent microscope or flow cytometry in a FACS machine.

4.3 Adaptive Conditioning in S. cerevisiae and the Wine Ecology

Studying the ecology of wine production reveals another habitat that can potentially select for Adaptive Conditioning, this time in the case of the yeast stress response. However, as the stress response is quite complex and involves numerous genes the mathematical model cannot be easily applied in this system. I focused on the core sequence of three stresses that form the diauxic shift and systematically tested the cross protection phenotype between each pair. As expected under the hypothesis, I observed that the directionality of cross protection fits the order of stresses in the wine production ecology (Figure 16A). Additionally, focusing on oxidative stress revealed the high specificity of cross protection as only stresses that naturally precede oxidative stress facilitate a protection phenotype (Figure 16A).

I next turned to study the transcriptional response in order to test whether the observed cross protection can be explained by a conditioned response strategy. Indeed, the analysis revealed that during encounter with heat, *S. cerevisiae* induces a seemingly unnecessary gene set, in addition to the heat required cluster, that is enriched mainly in anti-oxidative related functions (Figure 17, 18, and 20). This conditioned gene set is not induced under the osmotic stress, the control stress that does not facilitate cross

protection against oxidative stress. Finally, in order to test whether the conditioned cluster is indeed unnecessary during heat shock but not under oxidative stress, I measured the sensitivity of strains deleted for genes of the cluster. In accordance with the conditioning hypothesis, the analysis revealed that almost 65% of the 29 tested strains are sensitive to oxidative stress while none was significantly sensitive to heat shock (Figure 19).

To conclude, the observed asymmetrical response strategy in the wild type *S. cerevisiae* fits the hypothesis of Adaptive Conditioning. Furthermore, testing the three criteria outlined in the introduction indicates that the conditioned response is most likely an adaptive trait that has been selected for during evolution.

The currently prevalent perspective in the study of the stress response is based on the notion that upon encounter, cells induce genes that are required for coping with the current environmental perturbation. The observation that some non-essential genes are also induced is attributed to either: (1) failure of the experimental system to detect the gene essentiality or by (2) neutrality of the induction, i.e., the noisy induction is not selected against. The hypothesis of Adaptive Conditioning offers an alternative explanation that challenges this paradigm – *some* of the seemingly irrelevant genes expressed in a given stress are induced due to adaptation to the typical order of stresses in the natural ecology. The genes are induced as a preventive measure against another stress that typically follows the given stress in organism's habitat. Figure 24 shows a schematic representation of the two alternative explanations under the assumption of two subsequent stresses ($S_1 \rightarrow S_2$). Note that lineages that evolved in a new environment and that feature extinction of this overlapping response represent a strong indication for the dependency between the stimuli in the natural environment, e.g., extinction of maltose operons responsiveness to lactose (Figure 13) and uncoupling of the cellular responses to temperature elevation and reduction in oxygen availability previously shown in *E. coli*[9].

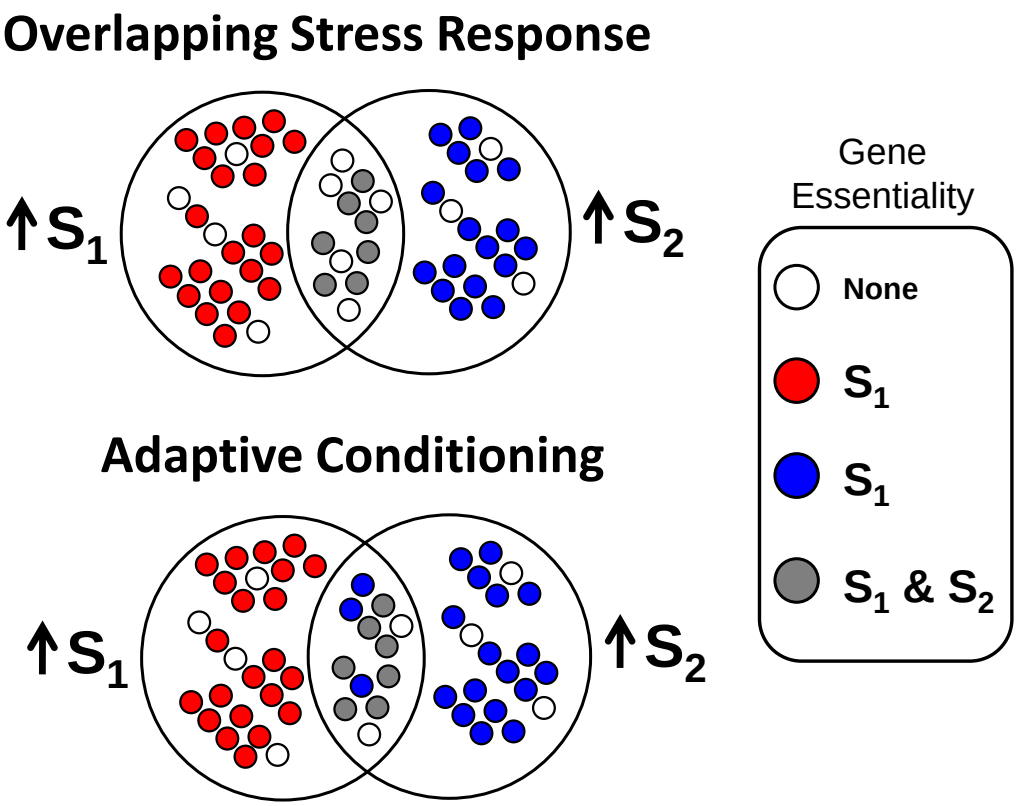


Figure 24. Alternative explanations for the overlapping transcriptional response to two stresses.

The discovery of Adaptive Conditioning in the yeast stress response can motivate several stimulating paths for future research. Further analysis of the genome-wide transcriptional response seems like the most straightforward next goal. For example, initial analysis identified unexpected induction dynamics of the common stress response (Figure 20B, ESR cluster). This gene set shows a reduced induction level in cells featuring cross protection, a reduction that was not observed in other highly induced gene clusters (Figure 20B, OX and CP clusters). Thus it is possible that the ESR induction level reflects the "subjective" stress level experienced by the cells, while the induction of other clusters reflects the "objective" stress level applied on the cells. Another path for analysis of the transcriptional data should test whether the observed expression patterns can be associated with the activity of specific transcription factors. Ideally, the analysis of

the promoter regions of the co-expressed conditioned cluster might point to a key underlying transcription factor.

The study of Adaptive Conditioning in yeasts can also continue in the form of comparative research. It would be interesting to characterize the stress response in close strains or species in light of the newly discovered conditioned response in the laboratory *S. cerevisiae* strain. For example, some strains are used for unique alcoholic fermentations that might feature a different set of stresses (e.g., fermentation of champagne done at low temperatures and reaching high ethanol concentrations). Evolution in these environments might select for reshaping of the conditioned response to match the new habitat.

4.4 Directed Evolution towards Adaptive Conditioning

Adaptive conditioning is proposed to be an advantageous evolvable trait in a consistent environment of recurring stimuli. I thus tried, as a proof of concept, to evolve a lineage of *E. coli* cells in a laboratory evolution experiment to condition between previously unconnected stimuli. Unfortunately, although the results indicated that an evolutionary process did take place (Figure 21 and 22) conditioning did not evolve within the time frame of the experiment. A possible explanation for this result might be that the extent of change required to achieve conditioning is considerable and therefore was not reached. As the fitness improvement in asexual populations is expected to follow a stepwise increase with diminishing returns[34], it seems that the adaptations that were fixated may thus be considerably more beneficial than conditioning and were therefore selected at the early stages at the expense of Adaptive Conditioning.

In light of the observation of existing Adaptive Conditioning in the *E. coli* and *S. cerevisiae* systems, new evolution experiments should be considered. In *S. cerevisiae,* the study revealed a conditioned response between heat shock and oxidative stress. Thus, an evolution experiment can be performed in order to select for extinction of this conditioned response. The evolved strain can then be tested in the original environment of alcoholic brew production in order to estimate the importance of the extinct

conditioned response. In the *E. coli* system I already identified evolved strains that feature extinction of the conditioned response between lactose and maltose. In this system it can be extremely interesting to evolve the strains to re-acquire the conditioned response (*reinstatement* in cognitive studies terminology). As it is possible that the strains did not completely extinct the conditioned response, reinstatement experiments might have a good chance of succeeding.

4.5 Closing Conclusions

In my thesis I proposed a new paradigm for the response to environmental stimuli in micro-organisms. The hypothesis of Adaptive Conditioning was developed and studied on fundamentally different levels and using a diversity of research tools. A considerable portion of my thesis is dedicated to the identification of Adaptive Conditioning in contemporary living organisms. Specially, I focused on two micro-organisms, *E. coli* and *S. cerevisiae*, and their known ecologies to uncover evidence both on the phenotypic level and on the molecular level that Adaptive Conditioning exists. The significance of this observation is noteworthy especially since the two are model species that have been extensively studied over the years.

As Adaptive Conditioning appears to have evolved independently both in prokaryotes (*E. coli*) and eukaryotes (*S. cerevisiae*), and has proven beneficial under remote systems such as sugar metabolism and stress response, it is possible that it is in fact a prevalent trait used by diverse species throughout the tree of life. A prominent ecology that potentially selects for this trait might be found in the habitat of pathogens infecting the human body - as infection progresses these organisms sequentially transition from one tissue or environment to another in a predictable order. For example, *Candida albicans* is commensal yeast that usually resides harmlessly in human mucosal tissues. However, sometimes, these yeasts can become aggressive pathogens that penetrate the blood stream and spread throughout the host's body[35]. Once inside, these cells are within reach of host's immune system. Did *C. albicans* evolve Adaptive Conditioning to prepare in advance for the potentially lethal encounter with host's phagocitic cells? Some studies provide evidence supporting this hypothesis, Specifically, Martchenko et. al.[36] have

observed that superoxide dismutase is induced in the transition to a hyphal growth pattern usually associated with exposure to blood plasma. The authors further raised the possibility, that the function of this antioxidant enzyme is important to cope with the future oxidative burst that will be generated in the phagosome of a host's immune cells.

Throughout my thesis Adaptive Conditioning was discussed as a beneficial response strategy that might be selected for by specific ecological niches. As a proof of concept for this reasoning I have attempted to evolve *E. coli* cells to condition between previously unconnected stimuli. I note however that a similar goal might be achieved through genetic engineering: the regulatory network of cells can be deliberately modified to condition between previously unconnected stimuli. Thus a compelling biotechnological application that arises is to harness Adaptive Conditioning as a *conceptual principle* when synthetically designing micro-organisms that will be used for specific multi-stage process. Under this paradigm, the sensing and response circuits will be wired so that the modified organism will feature early preparation, taking advantage of the temporal order of stimuli in the underlying industrial process.

Consider, for example the prevalent industrial process of bio-ethanol production that is often carried out by *S. cerevisiae*[37]. Although many aspects of this process are similar to the natural ecology of wine production, some differences do exist, such as different nutrient availability and the controlled temperature in the industrial process. It is important to note that, unlike traditional wine production, bio-ethanol production does not form an evolutionary cycle (the yeasts used are inoculated into the medium and are discarded at the end). Furthermore, attempts to redesign strains used for industrial process often focus on improving their adaptation under a specific isolated aspect of the environment (e.g.,[38]). Under the hypothesis of Adaptive Conditioning another complementing layer of improvement can be applied to yeasts in the bio-ethanol industry: Perfecting the conditioned stress response so it would better fit the specific requirements of bio-ethanol production. Application of Adaptive Conditioning is even more compelling if the organism used to carry out the industrial process originally evolved in

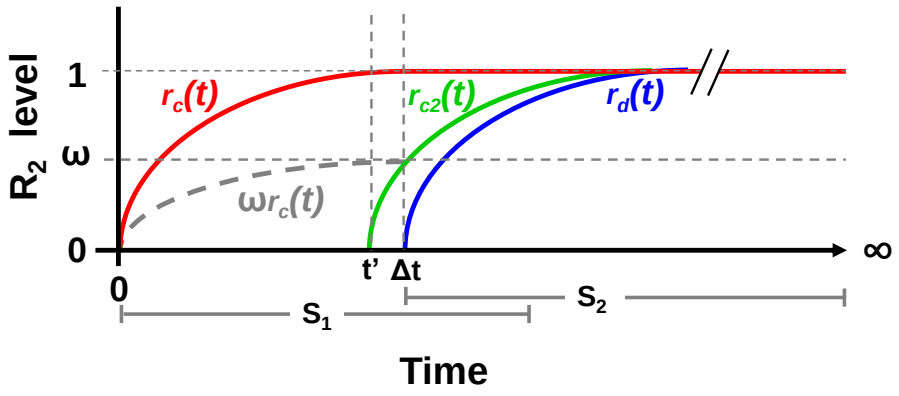
an entirely different ecological niche. For example in the use *E. coli* cells genetically modified to carry out the process of bio-ethanol production[37].

5 Supplementary Results

5.1 A Mathematical Model of Adaptive Conditioning

5.1.1 Integral of gain and cost functions

In order to calculate the accumulative effects of all changes on the basal growth rate I integrated over the gain and cost functions during the relevant time intervals. Under both direct and a conditioned response the accumulative benefit is calculated by integrating over the gain functions during the period of S_2 exposure. This is in fact integrating over the response function and multiplying by the system specific gain parameter κ (Supplementary Figure 1). Note that the response functions are expected to join if S_2 is sufficiently long. Thus, for simplicity in integration I use the upper limit $t=\infty$. Evidently the integrals can be adjusted for an environment characterized by different S_2 periods such that the response functions are far from the joining point.



Supplementary Figure 1. Response functions under different regulation strategies. The function $r_d(t)$, $r_c(t)$ and $r_{c2}(t)$ describe the relative R_2 level in a given environment under direct response, conditioned response, and 2-phase conditioned response strategies, respectively.

The integral for Direct Regulation response function, $r_d(t)$:

$$R_d = \int_{\Delta t}^{\infty} 1 - e^{-\alpha(t-\Delta t)} dt$$

$$R_d = t\Big|_{\Delta t}^{\infty} - e^{\alpha\Delta t} \cdot \int_{\Delta t}^{\infty} e^{-\alpha t} dt = t\Big|_{\Delta t}^{\infty} - e^{\alpha\Delta t} \cdot \left(\frac{e^{-\alpha t}}{-\alpha} \Big|_{\Delta t}^{\infty} \right) = t\Big|_{\Delta t}^{\infty} + \left(\frac{e^{\alpha\Delta t - \alpha t}}{\alpha} \Big|_{\Delta t}^{\infty} \right) = t\Big|_{\Delta t}^{\infty} - \frac{1}{\alpha}$$

The integral for one-phase conditioning response function, $r_c(t)$:

$$R_c = \int_{\Delta t}^{\infty} 1 - e^{-\alpha t} dt$$

$$R_c = \int_{0}^{\infty} 1 - e^{-\alpha t} dt - \int_{0}^{\Delta t} 1 - e^{-\alpha t} dt = t\Big|_{\Delta t}^{\infty} + \frac{e^{-\alpha t}}{\alpha}\Big|_{\Delta t}^{\infty} = t\Big|_{\Delta t}^{\infty} - \frac{e^{-\alpha\Delta t}}{\alpha}$$

Unlike the accumulating gain, the accumulating cost is calculated for a different time interval in each of the response strategies. Under Direct Regulation, the cost function is integrated during exposure to S_2 while under one-phase conditioning the function is integrated from 0 (the beginning of S_1) to the end of S_2. As the cost function is constant given the relative production rate in these cases, the integral is a β dependent constant multiplied by the response period t:

$$C = \int_{t} \eta \cdot dt = t\eta$$

5.1.2 Two-phase conditioning model

The response function of the two-phase conditioning strategy, $r_{c2}(t)$ is shown in Supplementary Figure 1.

(S1) $$r_{c2}(t) = 1 - e^{-\alpha(t-t')}$$

Where I define a time point t' ($t' \leq \Delta t$) such that a protein accumulating from this point, under full induction ($\beta=1$), will reach, at time point Δt, a level identical to that obtained by a protein accumulating from t=0 at an intermediate rate β ($0 \leq \beta \leq 1$):

(S2) $$r_{c2}(\Delta t) = \beta r_c(\Delta t)$$

This equality defines an entire family of two-phase response functions differing from one another by the intermediate level of induction β. This equation family is restricted by the two boundary cases β=0 and β=1 which represent Direct Regulation and one-phase conditioning strategies, respectively.

The gain function described previously holds for two-phase conditioning. Here I define the integral on the response function of two-phase conditioning. First I use equation (S2) to develop a term containing t' which is required for the integral

$$r_{c2}(\Delta t) = \beta r_c(\Delta t)$$
$$1 - e^{-\alpha(\Delta t - t')} = \beta(1 - e^{-\alpha \Delta t})$$
$$1 - \beta(1 - e^{-\alpha \Delta t}) = e^{-\alpha \Delta t} e^{-\alpha t'}$$
$$e^{-\alpha t'} = \frac{1 - \beta(1 - e^{-\alpha \Delta t})}{e^{-\alpha \Delta t}}$$

Using this equation I can calculate the integral for two-phase conditioning response function, $r_{c2}(t)$:

$$R_{c2} = \int_{\Delta t}^{\infty} 1 - e^{-\alpha(t - t')} dt$$
$$R_{c2} = t \Big|_{\Delta t}^{\infty} - e^{-\alpha t'} \int_{\Delta t}^{\infty} e^{-\alpha t} dt = t \Big|_{\Delta t}^{\infty} - e^{-\alpha t'} \left(\frac{e^{-\alpha t}}{-\alpha} \Big|_{\Delta t}^{\infty} \right) = t \Big|_{\Delta t}^{\infty} - e^{-\alpha t'} \left(0 - \frac{e^{-\alpha \Delta t}}{-\alpha} \right) = t \Big|_{\Delta t}^{\infty} - \underbrace{e^{-\alpha t'}}_{t'term} \cdot \frac{e^{-\alpha \Delta t}}{\alpha}$$
$$= \Big|_{\Delta t}^{\infty} - \frac{1 - \beta(1 - e^{-\alpha \Delta t})}{e^{-\alpha \Delta t}} \cdot \frac{e^{-\alpha \Delta t}}{\alpha} = \Big|_{\Delta t}^{\infty} - \frac{1 - \beta(1 - e^{-\alpha \Delta t})}{\alpha}$$

The binary cost function described previously cannot be used for the two-phase regulation strategy since the production rate can now span an entire range (0≤β≤1). I thus generalize the cost function deduced previously by Dekel and Alon for the lactose operon[17] in order to describe the relationship between the relative production rate

(normalized to the maximal production rate) and the relative decrease in the basal growth rate.

(S3) $$c(t,\beta) = \eta \cdot \frac{0.44 \cdot \beta}{1-(\beta/1.8)}$$

where η is the same system specific cost scaling parameter and the cost function described before at equation (Equation 6a) is a specific case of the new cost function (with β=1).

Note that as I assume that the degradation rate, α, does not change the steady state protein level reached is only β dependent (Y_{st}=β/α). Hence, β marks both the relative production rate and the relative steady state attained by R_2. The integral on the new cost function is:

$$C(\beta) = \int_t \eta \cdot \frac{0.44 \cdot \beta}{1-(\beta/1.8)} dt = t \cdot \eta \cdot \frac{0.44 \cdot \beta}{1-(\beta/1.8)}$$

I can now rewrite the fitness difference between conditioning and direct regulation (equation 1a from the main text) using the new integrals and the dependency on β. This equation now accommodates also two-phase conditioning:

(S4) $$\Delta F_{c2-d} = p\left(\kappa \frac{\beta(1-e^{-\delta\Delta t})}{\delta} - \Delta t \frac{0.44\beta\eta}{1-(\beta/1.8)}\right) - (1-p)t_{S1}\frac{0.44\beta\eta}{1-(\beta/1.8)}$$

This expression can be calculated only after assignment of β. However, one can assume the evolution selects for the optimal β in a given environment. Hence I can derive equation (S4) to find a unique level (β_o) that maximizes the fitness difference. The gain and cost terms can be derived independently. The derivative of the gain term in is straightforward and derivative of the cost function is:

$$\frac{dC}{d\beta} = \frac{0.44t\eta}{1-(\beta/1.8)} + \frac{0.44t\eta\beta}{1.8(1-\beta/1.8)^2}$$

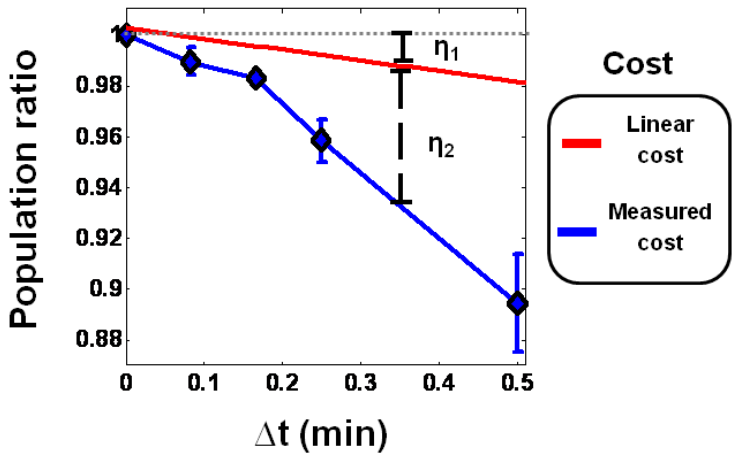
Thus the derivative of the entire equation is with respect to β:

$$\frac{d(\Delta F)}{d\beta} = p\left(\frac{\kappa(1-e^{-\delta\Delta t})}{\delta}\right) - p\left(\frac{0.44\eta\Delta t}{1-(\beta/1.8)} + \frac{0.44\eta\beta\Delta t}{1.8(1-\beta/1.8)^2}\right) - (1-p)\left(\frac{0.44\eta t_{S1}}{1-(\beta/1.8)} + \frac{0.44\eta\beta t_{S1}}{1.8(1-\beta/1.8)^2}\right)$$

Note that the finding the optimal β_o in the entire range ($0 \leq \beta_o \leq 1$) is in fact equivalent to finding the optimal regulation strategy in a given environment: $\beta_o=0$ is Direct Regulation, $\beta_o=1$ is one-phase conditioning, and the entire range $0<\beta_o<1$ indicates on optimality of two-phase conditioning. Thus equation S4 encapsulates the modeling of all response strategies under study. Moreover, the equation allows us to analytically find the best strategy given an environment and a biological system.

5.1.3 Additional cost in the *E. coli* lactose system

In order to quantify this cost parameter, which represents an instantiation of the "additional cost" parameter in the model (η^* in equation (6a) of the main text) in the lactose system I used a β-galactosidase mutant strain that can transport lactose but cannot metabolize it[13]. As the mutant is unable to metabolize lactose I used an identical experimental setup as described in the Results section (Figure 8A) to measure only the cost factors of conditioning. As the production cost η is already known in this system I could have subtracted it from the total measured cost in the mutant strain and accurately quantified the additional cost for a range of delay times between operon pre-induction and exposure to lactose (Supplementary Figure 2). I have confirmed that the production cost in the mutant strain is identical to that of the wild type strain we use (data not shown). It is important to note since the additional cost exists only if pre-induction is followed by arrival lactose, the two cost parameters C_{cp} and C_{uc} in our model are not identical as only C_{cp} include the additional cost.



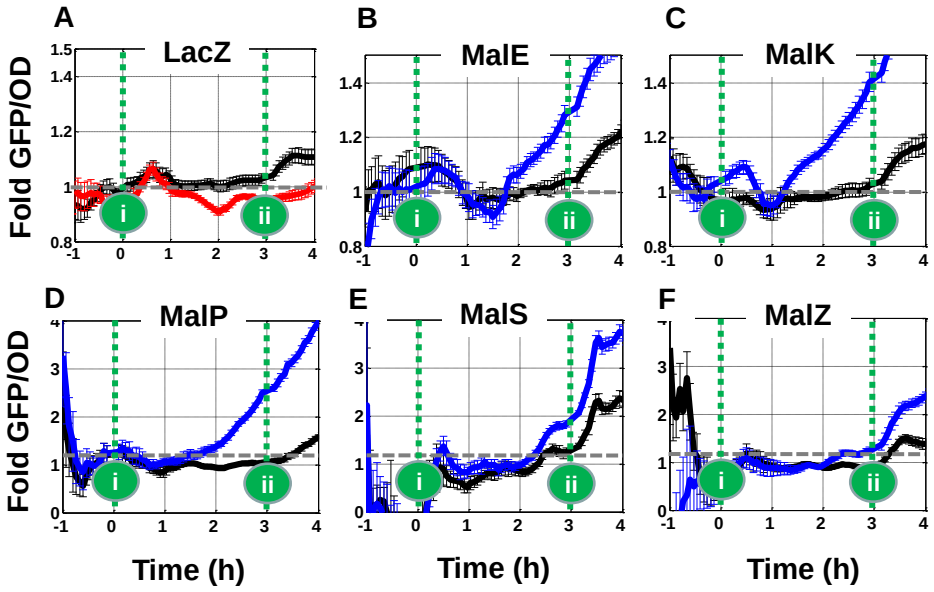
Supplementary Figure 2. Measured cost in β-galactosidase mutant strain. Experiment setup is described in the Results section (Figure 8A). The blue graph marks the experimentally measured cost in the lacZ mutant strain at different time intervals between operon pre-induction and lactose exposure, bars denote the standard errors. The observed cost reflects both the linear cost (η_1) and the additional cost (η_2). The red graph marks only the linear cost (η_1).

5.2 Adaptive Conditioning in E. coli and the Intestinal Ecology

5.2.1 Response of operons to sequential addition of sugars

In order to systematically test the expression profiles of all relevant operons in the response to sequential addition of sugars I used the promtor-fused GFP library previously described[14]. I monitored the GFP levels in cells while adding maltose and lactose to the growth medium either sequentially or simultaneously. The experiments revealed the asymmetrical regulation pattern expected by the conditioning hypothesis (similarly to Figure 11). The sequential setup allows us to observe the dynamics of protein level as a result of the two-phase conditioned strategy. In order to exclude the possibility that the asymmetry in activation by the two sugars arises from difference in their effective inducing concentration I repeated the experiments using a fivefold higher maltose concentration. Even under these conditions, the inability of maltose to induce the lactose operon is maintained (data not shown). In addition, as a negative control I verified that the asymmetry shown here is specific to the operons discussed and does not reflect a widespread effect of lactose as an inducer on unrelated genes. Toward this aim, I monitored the activity of a constitutive promoter with a σ70 binding site after addition of

lactose and maltose. I observed no difference in the induction capacity of the two sugars (data not shown). Note that the induction levels measured using the GFP reporter library are lower than those obtained using the qRT-PCR method. This observation is in agreement with previous reports using this GFP library[39].

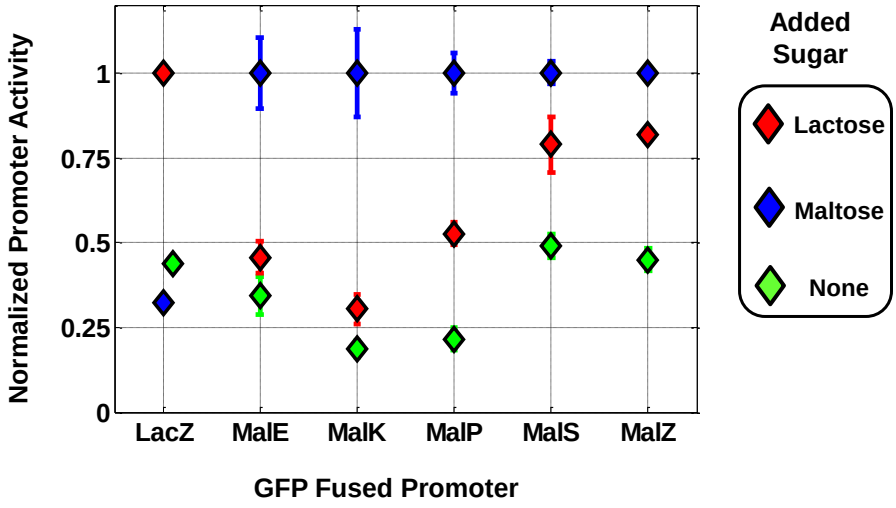


Supplementary Figure 3. Fold GFP levels per cell reflecting transcriptional activity of the lactose and maltose operons in response to addition of sugars. Cultures growing on M9-Glu medium where treated with various sugars and the ratio of cell normalized GFP measured for a treated culture over an untreated culture was taken as fold protein level per cell. Sugars were added either simultaneously at time point ii (black graph) or sequentially with a 3 hour delay (blue graph – lactose followed by maltose, red graph – maltose followed by lactose). (A) The lactose operon and (B-F) the maltose operons.

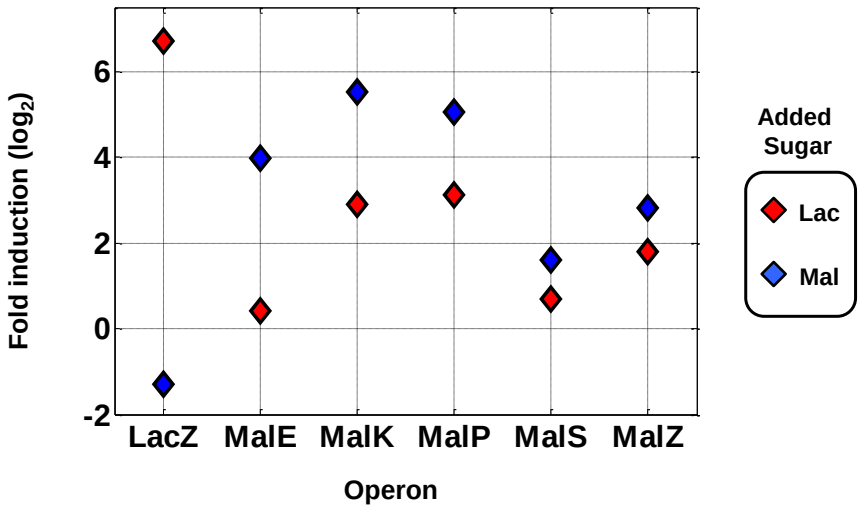
5.2.2 Adaptive conditioning under different background carbon source

In order to control for possible residual catabolite repression by glucose on the studied promoters, I measured promoter activity while using glycerol as background carbon source instead of glucose (supplemented with cAMP). The experiments were done similarly to those described in the main text. Similarly to the results in the presence of glucose, I observed a pattern of partial induction of maltose operons in response to lactose (Supplementary Figure 4, 5). Reassuringly, the lactose operon remained

unaffected by presence of maltose. It is important to note that the use of glycerol is less desirable when using the promoter-fused GFP library since treatment with additional carbon sources gives rise to different growth rate (an effect that is avoided when glucose is used as a background sugar) This point is essential since the basal promoter activity in a cell is correlated with the rate of cell growth. Thus I control for this effect by normalizing the promoter activity to the measured growth rate.



Supplementary Figure 4. Relative promoter activity of lactose and maltose operons in response to addition of different sugars. Bars denote standard deviations of four repetitions. The experiments were done using glycerol as a background carbon source.



Supplementary Figure 5. Fold induction (log_2) of lactose and maltose operons in the presence of different sugars as measured by qRT-PCR. The experiments were done using glycerol as a background carbon source.

5.2.3 Control fitness experiments

As described in the Results section (Figure 15), I observed that pre-exposure to lactose increases cells fitness when grown on maltose. In order to control for the contribution of lactose as a carbon source, rather than a signaling molecule, to the fitness increase I transferred an equal number of cells from S_1 to S_2 in both treated and untreated cultures (see Materials and Methods section). In addition, I preformed a control experiment to fully address this concern. In this experiment, overnight cultures were diluted either into M9-Gly+lactose (treatment) or M9-Gly (untreated) and then diluted into M9-Gly (rather than into a medium containing maltose). The results clearly showed a very small effect on fitness on glycerol due to prior exposure to lactose (1%). In comparison the effect of lactose as an anticipation signal before maltose is much higher (4%). This is a clear indication that the growth advantage observed when maltose was preceded with lactose is, for the most part, not due to a potential role of lactose as a fuel molecule, but rather due to its role as a signal (used by the cells to prepare for maltose). I

thus normalized all fitness measurements of pre-induction by lactose to the small effect measured in the control experiment.

5.3 Adaptive Conditioning in S. cerevisiae and the Wine Ecology

5.3.1 Standard deviations in cross protection experiments

All cross protection experiments were done in three independent repeats. Supplementary Table 1details the mean and standard deviations measured in the cross protection experiments between each stress pair.

Supplementary Table 1 - Fold cross protection between stress pairs

S_2 / S_1	**Heat shock**	**Ethanol**	**Oxidative stress**
Heat shock	31.4 ± 5.8	4.0 ± 0.3	45.5 ± 7.9
Ethanol	3.1 ± 0.35	5.3 ± 0.3	28.7 ± 7.1
Oxidative stress	1.5 ± 0.3	<0.1 ± 0	18.4 ± 2.3

5.3.2 Fitness contribution of conditioned genes under stresses

Supplementary Table 2 details the strains used for the experiment described at Figure 19 and their respective survival ratios under oxidative stress and heat shock.

Supplementary Table 2 - Survival ratios of deletion strains

	Oxidative stress		Heat Shock	
Strain	**p-value**	**Survival ratio**	**p-value**	**Survival ratio**
BY4741	**-**	**0.91**	**-**	**0.90**
Δ*YJL079C*	1.53E-12	0.40	0.271	0.86
Δ*YLR346C*	1.76E-06	0.67	0.945	0.90
Δ*YBR185C*	2.36E-06	0.64	0.311	0.86
Δ*YJL071W*	2.49E-05	0.70	0.875	0.89
Δ*YJR151C*	3.82E-05	0.72	0.218	0.86
Δ*YPL250C*	3.86E-05	0.72	0.011	0.81
Δ*YGL090W*	1.07E-04	0.72	0.592	0.88
Δ*YNL241C*	1.39E-04	0.73	0.437	0.93
Δ*YMR225C*	1.75E-04	0.73	0.146	0.85
Δ*YOL162W*	1.89E-04	0.74	0.847	0.89
Δ*YPR015C*	4.19E-04	0.75	0.533	0.88
Δ*YML101C*	5.73E-04	0.76	0.456	0.87
Δ*YHR179W*	1.40E-03	0.76	0.542	0.92
Δ*YDL243C*	1.57E-03	0.77	0.773	0.89
Δ*YPR093C*	2.89E-03	0.78	0.769	0.89
Δ*YKL071W*	3.32E-03	0.78	0.332	0.87
Δ*YNL239W*	0.014	0.79	0.190	0.85
Δ*YKL137W*	0.019	0.81	0.194	0.85
Δ*YKR061W*	0.020	0.81	0.934	0.90
Δ*YHL036W*	0.058	0.83	0.759	0.89
Δ*YGR232W*	0.076	0.83	0.080	0.84
Δ*YLL056C*	0.164	0.85	0.960	0.90
Δ*YEL066W*	0.199	0.85	0.502	0.87
Δ*YMR095C*	0.250	0.86	0.971	0.90
Δ*YLR024C*	0.401	0.87	0.986	0.90
Δ*YFL055W*	0.410	0.87	0.920	0.90
Δ*YOL158C*	0.554	0.93	0.469	0.93
Δ*YOR382W*	0.670	0.89	0.591	0.88
Δ*YOL163W*	0.885	0.91	0.937	0.90

6 Literature

1 Bauer, F. F., Pretorius, I. S., *Yeast stress response and fermentation efficiency: How to survive the making of wine - a review*. (2000).

2 Maris, A. F. et al., Diauxic shift-induced stress resistance against hydroperoxides in *Saccharomyces cerevisiae* is not an adaptive stress response and does not depend on functional mitochondria. *Current Genetics* **39** (3), 137 (2001).

3 Pavlov, Ivan Petrovich, *Conditioned Reflexes*. (Oxford University Press, London, 1927).

4 Gasch, A. P. et al., Genomic expression programs in the response of yeast cells to environmental changes. *Molecular Biology of the Cell* **11** (12), 4241 (2000).

5 Causton, H. C. et al., Remodeling of yeast genome expression in response to environmental changes. *Molecular Biology of the Cell* **12** (2), 323 (2001).

6 Acar, M., Mettetal, J. T., and van Oudenaarden, A., Stochastic switching as a survival strategy in fluctuating environments. *Nat Genet* **40** (4), 471 (2008).

7 Balaban, N. Q. et al., Bacterial persistence as a phenotypic switch. *Science* **305** (5690), 1622 (2004).

8 Kussell, E. and Leibler, S., Phenotypic diversity, population growth, and information in fluctuating environments. *Science* **309** (5743), 2075 (2005).

9 Tagkopoulos, I., Liu, Y. C., and Tavazoie, S., Predictive behavior within microbial genetic networks. *Science* **320** (5881), 1313 (2008).

10 Savageau, M. A., *Escherichia-coli* Habitats, Cell-Types, and Molecular Mechanisms of Gene-Control. *American Naturalist* **122** (6), 732 (1983).

11 Savageau, M. A., Demand theory of gene regulation. II. Quantitative application to the lactose and maltose operons of *Escherichia coli*. *Genetics* **149** (4), 1677 (1998).

12 Mortimer, R. K., Evolution and variation of the yeast (Saccharomyces) genome. *Genome Research* **10** (4), 403 (2000).

13 Cook, A. and Lederberg, J., Recombination studies of lactose nonfermenting mutants of *Escherichia coli* K-12. *Genetics* **47**, 1335 (1962).

14 Zaslaver, A. et al., A comprehensive library of fluorescent transcriptional reporters for *Escherichia coli*. *Nat Methods* **3** (8), 623 (2006).

15 Oxman, E., Alon, U., and Dekel, E., Defined order of evolutionary adaptations: experimental evidence. *Evolution* **62** (7), 1547 (2008).

16 Kaplan, S. et al., Diverse two-dimensional input functions control bacterial sugar genes. *Mol Cell* **29** (6), 786 (2008).

17 Dekel, E. and Alon, U., Optimality and evolutionary tuning of the expression level of a protein. *Nature* **436** (7050), 588 (2005).

18 Winzeler, E. A. et al., Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science* **285** (5429), 901 (1999).

19 Alon, Uri, *An Introduction to Systems Biology: Design Principles of Biological Circuits*. (Chapman & Hall/CRC, 2006).

[20] Stoebel, D. M., Dean, A. M., and Dykhuizen, D. E., The Cost of Expression of *Escherichia coli* lac Operon Proteins Is in the Process, Not in the Products. *Genetics* **178** (3), 1653 (2008).

[21] Dykhuizen, D. and Hartl, D., Transport by Lactose Permease of *Escherichia-Coli* as Basis of Lactose Killing. *Journal of Bacteriology* **135** (3), 876 (1978).

[22] Caugant, D. A., Levin, B. R., and Selander, R. K., Genetic diversity and temporal variation in the *E. coli* population of a human host. *Genetics* **98** (3), 467 (1981).

[23] Soupene, E. et al., Physiological studies of *Escherichia coli* strain MG1655: Growth defects and apparent cross-regulation of gene expression. *Journal of Bacteriology* **185** (18), 5611 (2003).

[24] Pretorius, I. S., Tailoring wine yeast for the new millennium: novel approaches to the ancient art of winemaking. *Yeast* **16** (8), 675 (2000).

[25] Sablayrolles, J. M. and Barre, P., Kinetics of alcoholic fermentation under anisothermal enological conditions. *Am. J. Enol. Vitic.* **44** (2), 127 (1993).

[26] Kunkee, RE., Amerine, MA., *The Yeasts*. (Academic Press, 1970).

[27] Brauer, M. J., Saldanha, A. J., Dolinski, K., and Botstein, D., Homeostatic adjustment and metabolic remodeling in glucose-limited yeast cultures. *Molecular Biology of the Cell* **16** (5), 2503 (2005).

[28] Lewis, J. G., Learmonth, R. P., and Watson, K., Induction of Heat, Freezing and Salt Tolerance by Heat and Salt Shock in *Saccharomyces-Cerevisiae*. *Microbiology-Uk* **141**, 687 (1995).

[29] Piper, P. W., The Heat-Shock and Ethanol Stress Responses of Yeast Exhibit Extensive Similarity and Functional Overlap. *Fems Microbiology Letters* **134** (2-3), 121 (1995).

[30] Sanchez, Y., Taulien, J., Borkovich, K. A., and Lindquist, S., Hsp104 Is Required for Tolerance to Many Forms of Stress. *Embo Journal* **11** (6), 2357 (1992).

[31] Dwight, S. S. et al., Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Research* **30** (1), 69 (2002).

[32] Grant, C. M., MacIver, F. H., and Dawes, I. W., Mitochondrial function is required for resistance to oxidative stress in the yeast *Saccharomyces cerevisiae*. *FEBS Lett* **410** (2-3), 219 (1997).

[33] Herring, C. D. et al., Comparative genome sequencing of *Escherichia coli* allows observation of bacterial evolution on a laboratory timescale. *Nat Genet* **38** (12), 1406 (2006).

[34] Elena, S. F. and Lenski, R. E., Evolution experiments with Micro-organisms: the dynamics and genetic bases of adaptation. *Nat Rev Genet* **4** (6), 457 (2003).

[35] Naglik, J., Albrecht, A., Bader, O., and Hube, B., *Candida albicans* proteinases and host/pathogen interactions. *Cell Microbiol* **6** (10), 915 (2004).

[36] Martchenko, M., Alarco, A. M., Harcus, D., and Whiteway, M., Superoxide dismutases in *Candida albicans*: Transcriptional regulation and functional characterization of the hyphal-induced SOD5 gene. *Molecular Biology of the Cell* **15** (2), 456 (2004).

[37] Hahn-Hagerdal, B. et al., Bio-ethanol - the fuel of tomorrow from the residues of today. *Trends in Biotechnology* **24** (12), 549 (2006).

[38] Alper, H. et al., Engineering yeast transcription machinery for improved ethanol tolerance and production. *Science* **314** (5805), 1565 (2006).

[39] Setty, Y., Mayo, A. E., Surette, M. G., and Alon, U., Detailed map of a cis-regulatory input function. *Proc Natl Acad Sci U S A* **100** (13), 7702 (2003).

7 Independent Efforts and Collaborations

I hereby declare that this thesis summarizes my independent efforts. Two projects were performed in collaboration as detailed bellow:

Cross protection experiments measuring the survival of *S. cerevisiae* in response to various stress types were preformed as joint work with Gal Hagit Romano, a fellow PhD student, from Martin Kupiec's laboratory at Tel Aviv University.

Identification of mutated genome loci contributing to the increase in growth rate of evolved *E. coli* strains was performed jointly with Avihu Yona, a fellow PhD student, from Yitzhak Pilpel's laboratory at Weizmann Institute of Science.

8 Acknowledgements

One small thing before we start, if you reached this part and the term “Adaptive Conditioning” doesn’t even *ring a bell*, you probably skipped one section too many.

Yet another small thing, please ignore the order of names. Consider the rolling credits at the end of an (excellent) Hollywood movie, so many names of super stars, each one doing his part to an Oscar level and one poor bastard left to print the titles without stepping on anyone’s toes.

First, I’d like to thank my PhD advisor Tzachi Pilpel. The first time I heard him lecture was way back when, at a small seminar in Tel Aviv University. Looking back, I can still recall it was a last minute decision, I almost didn’t go. One lecture that reminded me why is Biology the Science I choose to focus my life around. One lecture that made clear where it is exactly I want to go. Soon after, I found myself joining his young lab.

Tzachi, from the day I joined, you gave me a *carte blanch* to find my own way. For the first time in my academic life, I was the one who had to search deep down inside and ask: what do **I** hypothesis**?**

The question I came back with, after almost a year, was so fundamental and crude, yet since then WE have refined it to the best research there is. I don’t regret it one bit, it was one hell of a ride - so many ups, so many downs, an overdose of exciting Science and you foremost to thank for!

I’d like to thank Orna Dahan. I was a clueless, reckless, pipette tips chewing experimentalist before you showed me how experimental Biology should be exercised, you are the best research associate we could have asked for. Thank you for being a friend and an advisor.

Avihu, from day one I had a hunch you’re made from the right stuff to join the crusade. From the minute I saw you looking into systematic error of multi-pipettes I knew we are both obsessive about the exact same stuff. We have always shared the

conviction that Adaptive Conditioning is a necessary truth and that our task is merely to show it at work – I couldn't have hoped for better partner to join my work.

Ophir, we fought side by side in the trenches of Tzachi's lab from day one, we jumped together to the heart of darkness (the wet lab) and came back for more with pipettes between our teeth. We shared our ideas and thoughts on what's next all along the way. If only we would have founded the Science Triplet club earlier, I sure we would have co-authored a manuscript or two.

Thank you all present and past pilpels, you have made the lab a second home. Thank you Furman Itay for your patient help with the endless equations. Thank you Lapidot Michal for making me feel immediately at home in a new lab. Thank you Bella Groisman for rtPCRing again and again. Thank you Zohar for always willing to lend an extra hand. Thanks you all Pilpel Lab members.