Thesis for the degree
Doctor of Philosophy

Submitted to the Scientific Council of the
Weizmann Institute of Science
Rehovot, Israel

עבודת גמר (תזה) לתואר
דוקטור לפילוסופיה

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

במתכונת "רגילה"
In a "Regular" Format

By
Ilya Venger

מאת
איליה ונגר

עמידות שמר האפייה למוטציות מחיקה ותגובה לשינויים בסביבה: מבט דרך עדשת המטבולומיקה

Resilience of *Saccharomyces cerevisiae* to deletion mutations and response to environmental perturbations through the lens of large-scale metabolic profiling

Advisors:
Prof. Yitzhak Pilpel
Dr. Asaph Aharoni

מנחים:
פרופ' יצחק פלפל
ד"ר אסף אהרוני

April, 2010

ניסן תש"ע

Abstract

Cellular phenotype emerges from the interaction between the environment and the genome. Metabolome - the variation in the complement of intracellular metabolites is the most direct manifestation of a cellular phenotype. In my doctorate I set out to investigate the effect of genetic and environmental fluctuations upon the metabolome.
I developed a system for large-scale metabolome analysis in *Saccharomyces cerevisiae*. The method allows simultaneous detection and quantitation of more than one hundred different metabolites including various amino acids, sugars and sugar-phosphates, organic acids and other molecules with mass lower than 1000 Da.
I then utilized the system to collect data from over 500 samples including multiple deletion mutants of genes with paralogs, cultures subjected to various environmental stresses and cultures of a strain evolved to cope with heat shock.
My results show that the metabolome is highly dynamic and responsive to both environmental stresses and genetic changes. High abundance of metabolic changes in deletion mutants of paralogous genes was found to be related to slow rate of evolutionary divergence of paralogs. Three types of reciprocal relationships were found among such gene pairs: (i) absence of a metabolic phenotype in any of the deletion mutants (ii) one-sided response, where deletion of one but not the other paralog exhibits metabolic effect (iii) concerted changes, where deletion of each paralog yields a similar phenotype. The various types of responses are suggested to depend upon the reciprocal transcriptional levels of the paralogous gene pair.
The metabolic response to environmental stresses was shown to be evoked both through direct sensing of the environmental change as well as through an indirect mechanism of a decrease in growth rate of the cultures. The major molecule responsive to tested stresses was trehalose, which showed an elevation of two orders of magnitude during stress. Strains which were evolved towards heat tolerance exhibited high levels of trehalose and other stress-responsive metabolites in steady state even in the absence of stress.
This work constitutes the largest body of study using metabolome analysis approach to functional genomics in *S. cerevisiae*. This is the first time a relationship between evolutionary dynamics and metabolomic variation is shown. I anticipate that future enhancements in analytical techniques will help us better understand the relationship between metabolome, environmental conditions and genetic background of organisms.

תקציר

הפנוטיפ של אורגניזם חי הינו תולדה של האינטראקציה בין הסביבה לבין הגנום. מטבולום – השונות באוסף המטבוליטים התוך-תאיים, הוא הביטוי הישיר ביותר של פנוטיפ תאי. בעבודת הדוקטורט שלי שמתי למטרה לחקור את ההשפעה של שינויים גנטיים וסביבתיים על המטבולום.

פיתחתי מערכת למדידה רחבת היקף של המטבולום בשמר האפייה (*Saccharomyces cerevisiae*). השיטה מאפשרת זיהוי בו-זמני של מעל מאה מטבוליטים שונים, לרבות חומצות אמינו, סוכרים שונים, סוכר-פוספאטים, חומצות אורגניות ומולקולות אחרות עם מסה נמוכה מ- 1000 Da. השתמשתי במערכת לאיסוף נתונים על יותר מ-500 דגימות. הדוגמאות הופקו מזנים עם מוטציות מחיקה של מספר גנים בעלי פאראלוגים בגנום, תרביות שנחשפו לתנאים סביבתיים שונים ותרביות של זן שעבר אבולוציה להתמודדות עם עקת החום.

התוצאות שלי מראות כי המטבולום הינו דינמי ביותר ומגיב הן לסטרסים סביבתיים והן לשינויים גנטיים. ריבוי שינויים מטבוליים בזנים בעלי מחיקה בגנים פאראלוגיים נמצא קשור לקצב נמוך של צבירת שינויים אבולוציוניים בתוך זוגות פאראלוגיים. שלושה סוגים של יחסי הגומלין נמצאו בקרב מונטי מחיקה של זוגות גנים פאראלוגיים כדלהלן: (i) היעדר של פנוטיפ מטבולי בכל אחד מהמוטנטים (ii) תגובה חד צדדית, שבה המחיקה של פאראלוג אחד, אך לא השני מראה השפעה מטבולית (iii) שינויים משותפים, מקרים בהם מחיקת כל אחד מהפאראלוגים מניבה פנוטיפים דומים. אני גורס כי סוגים שונים של תגובות מטאבוליות תלויים ברמות השעתוק ההדדיות של זוגות הגנים הפאראלוגיים.

בהמשך אני מראה כי התגובה המטבולית לשינויים סביבתיים מופעלת הן באמצעות חישה ישירה של השינויים והן על ידי ירידה בקצב הגידול של תרביות. המולקולה המרכזית שמגיבה לסטרסים שנבדקו הינה טרהאלוז, אשר מציגה עלייה של מעל לשני סדרי גודל בתגובה לסטרס. זנים שעברו אדאפטציה אבולוציונית לעמידות לחום הציגו רמות גבוהות של טרהאלוז ומטבוליטים מגיבי-סטרס אחרים גם במצב קבוע - ללא סטרס.

עבודה זו מהווה את המחקר הגדול ביותר שנעשה עד היום, המשתמש במטבולומיקה על מנת לענות על שאלות מתחום הגנומיקה הפונקציונלית ב-*S. cerevisiae*. כמו כן, לראשונה מוצג בעבודה קשר בין תהליכים אבולוציוניים לשינויים המטבוליים שמתלווים אליהם.

אני צופה כי שיפורים עתידיים בשיטות כימיה אנליטית יעזרו לנו להבין טוב יותר את הקשר בין התנאים הסביבתיים למטבולום ולרקע הגנטי של אורגניזם.

Acknowledgements

I would like to thank all the people who contributed to the conception of this work, helped me carry through with it and supported me both professionally and personally.

My supervisors:
Prof. Yitzhak Pilpel, Dr. Asaph Aharoni

People from Aharoni lab for Plant Metabolomics:
Dr. Ilana Rogachev, Sergey Malitsky, Dr. Liron Feldberg and other members of Aharoni lab

People from Pilpel lab for Computational Genomics
Dr. Orna Dahan, Avihu Yona, Dr. Amir Mitchell, Ophir Shalem, Dvir Dahary, Dr. Itay Furman and all other members of the lab, who made this experience very enjoyable.

A special note for special people:
Tzachi, you took me in without much experience in bionfomatics and without me doing a rotation in the lab. You saw the potential, and you kept my fire going. I love your style of thinking.
Asaph, in a casual meeting five years ago you exposed me to new methods and to a new field in biology. From just one talk with you I understood that I want and can be a part of a new developing field. I was fortunate to have you as an advisor, you invested heavily in the project, and were always available for my questions, ideas and thoughts.
Ilana and Sergey thank you for your patience teaching me how to walk the walk and talk the talk of Analytical Chemistry.Ilana –you taught me all I know in analytical chemistry. Without your guidance this project would not be carried out. Sergey – thanks to your suggestions, ideas and instruction I was able to harness the GCMS in our lab, perfect the experimental procedures. Thanks for the countless times you helped me with the instruments in the lab – this saved me many months of work. I really apologize to you two we didn't arrange a good place for you at my wedding ☺.
Orna, your expertise in experimental biology is admirable and your help was immense. I am lucky that my project coincided with you coming to our lab. I would like to

especially thank you for helping me with the submission of the final report. That was very generous of you.
Avihu, thank you for your time, your youthful passion about biology and your generous help allowing me to use data from your evolution experiments. Your persistence pays off – keep on the good work.
Amir and Ophir, I owe many of the ideas for this work to discussions with you. I'm grateful for the opportunity to work alongside you two.
To all people from Pilpel lab past and present – thank you very much for your support, extremely fruitful discussions, comments, ideas and general good spirit.

I would also like to thank several people from outside the institute without whom this journey would be much harder, or would not happen at all.
Michael Elgart - Mishka, having both a good friend and a colleague with a wide heart and sharp thinking is something which most people don't have. I am very lucky.
Special thanks go to Dr. Miriam Veinshtein who really pushed me forward towards completing the final report - your wisdom guided me through.
Most important, I would like to thank my wife, Dana Rubashov and my family for continuous support and love.

Last, but not least, I would like to thank Café Nona, where the thesis was written. The people here and the establishment kept me company in writing, and served coffee and beer all through the long nights.

Table of contents

1 Introduction

Life persists through metabolic processes that occur within the cells of different organisms. While genes and proteins are responsible for carrying out the reactions, metabolites are the subject matter of these reactions. In my thesis I researched the interplay between the two levels – the level of the genes, and the level of the metabolites. Environmental effects exist in-between the two levels, where genes buffer the external effects and modify the complement of intracellular metabolites to allow cells to survive.

1.1 Robustness to deletion mutations

It was previously shown that ~80% of the genes in the yeast *Saccharomyces cerevisiae* are not essential for the survival of the organism when deleted[1,2]. Three different mechanisms were suggested to be responsible for this observed resilience to deletion mutations. Namely, genetic redundancy, distributed robustness and conditional non-functionality. Where: *genetic redundancy* is a case in which a paralog or an analog of the deleted gene may provide functional compensation for the genetic deficiency and restore a viable phenotype[3-5], *distributed robustness* is a mode where the compensation for the lost function is carried out through the rearrangement of portions of the entire metabolic network[4,6], and *conditional non-functionality* is a case when the deleted gene is essential only in a particular subset of conditions that were not tested in the deletion studies; thus, a gene's dispensability is observed only due to lack of an appropriate experiment[5-7].

In general, the complement of the biological molecules inside the cells may be divided to four levels. The genome, the transcriptome, the proteome and the metabolome[8]; the metabolome being the set of all the intra-cellular metabolites in a particular condition. The metabolome can be seen as the uppermost "omics" level since it is affected by the mRNA and protein concentrations, and is the most basic manifestation of the cellular phenotype.

The intracellular metabolites participate in almost all reactions in the living cell. The level of the metabolites is directly affected by deletion of the enzymes that carry out reactions with the metabolites as substrates and products. Therefore, one can expect a

direct connection between the metabolome of the different *S. cerevisiae* strains and their resistance to deletion mutations.

In addition to being used as a model for genetic perturbations yeast has been employed in studies of the effect of various environmental stresses on living cells. Multiple experiments were performed to assess the effect of external stresses on yeast transcriptome and proteome[9-12]. Understanding the manifestation of the stress-response on the level of metabolites is paramount in the understanding of basic cellular processes related to metabolism.

1.2 Metabolomics in Saccharomyces cerevisiae

It was shown that there are about 550 metabolites in *S. cerevisiae*[13,14]. Any single intracellular metabolomics method can identify and quantify 70-80 of these metabolites.

Most of the research up until now focused on the development of methods for the metabolic profiling of yeast strains. There are several research groups that study the metabolomics of *S. cerevisiae*[15-19] both on the level of the intra-cellular metabolome and on the level of the extracellular metabolome. Using metabolomics information it was shown that metabolic profiles discriminate among various yeast strains[20] and growth conditions[15,17]. However, no metabolomic study until today has investigated in a large-scale manner the response of yeast cells to gene deletion or to multiple environmental stresses.

1.3 Research main goals and motivation

In my research I have set out to determine the cellular manifestation on the molecular level of yeast function when faced with either genetic stress (gene deletion) or environmental perturbation as exemplified by several stresses.

I have chosen to analyze how *genetic redundancy* affects the cellular metabolome by analyzing yeast gene deletion strains. Yeast paralogs tend to be dispensable upon deletion more often than singleton genes[1]. Prior work in Pilpel laboratory has shown that paralogous pairs regulate the expression of each other on the transcriptional level[21]. Further studies have outlined the natural selection forces that acted on paralogous genes to retain a certain degree of redundant functionality[22-24].

The metabolome is the most direct measurement of the cellular phenotype. Therefore, I expect natural selection towards partial functional redundancy and differential

transcriptional regulation of paralogs to be manifested when measuring the metabolome – allowing us to better understand resilience to deletion.

To explore yeast stress response in terms of intracellular metabolome I have chosen to look at a few stresses as a model. Environmental stresses are well studied in *S. cerevisiae* and multiple studies dealt with the transcriptional response of yeast to changes in the environment[9,10,25]. Looking at the metabolic response to environmental stresses, we can learn about the relationship between the transcriptome and the metabolome. Additionally, cell exposure to different stress allows identification of metabolites responsive to stress; those that might be helpful in providing the resilience to environmental conditions.

To further assess how yeast cells cope with stresses, I have analyzed the metabolic response to heat of strains evolved for ~1000 generations under laboratory conditions to cope with high heat. This sheds light in greater detail on metabolic state required for best response to environmental stress.

1.4 Hypotheses for metabolic response to paralogous gene deletion

Several potential metabolic responses can be predicted in paralogous deletion mutants.

For mutants in enzymes that provide perfect backup for each other (i.e. carry out the same reaction, such as paralogs that diverged only recently in the evolution) I would expect in some cases no or little difference in the metabolome compared to WT. Yet lack of a metabolic change, would also be expected in the very different cases with respect to deletion of genes which are not active in tested conditions e.g. glucose-repressed genes in glucose rich growth media[26,27]. Deletion of genes from each of the two types above is not expected to have an effect on the metabolic composition of yeast cells, thus telling the two options apart in cases of little metabolic change is not trivial.

There are two other interesting scenarios for paralogous deletion. The first in the case of backup with transcriptional reprogramming, i.e. cases in which the remaining paralog increases its expression level in response to deletion of its counterpart[21-23], and the other is when a continued persistence of both paralogs is sustained to augment gene expression and flux through specific pathways[6,28]. In this case, the expression levels of the genes are not expected to change upon deletion of any of the paralogs.

In the case of backup with transcriptional reprogramming, upon deletion of one of the paralogs the mRNA levels of the other paralog might need to be elevated[23]. This effect might lead to only partial complementation and might exert an effect upon the metabolome. On the other hand, it was also shown that many genes have more than just one function[29]. In duplicate genes this multi-functionality of the common ancestral gene may be responsible for the process leading to specialization and functional divergence of paralogs[30,31]. Therefore, in the case of deletion with reprogramming, cellular effects beyond those resulting from incomplete backup by the upregulated paralog may arise. Minor function unique to the remaining paralog (also termed "moonlighting"[32]) with elevated levels would tend to affect the metabolome. This is another reason I would expect different metabolic profiles of the two deletion mutants.

In some instances, however, transcriptional reprogramming occurs only upon deletion of one of the paralogs, but not upon the deletion of the other[22]. In such special cases a one-sided metabolic response would occur upon the deletion of one paralog but not the other.

In the second case, paralogs which perform similar functions may complement each other to increase the metabolic flux through a reaction. One such example are genes duplicated during the whole genome duplication event in yeasts that increased the overall glycolytic flux within cells[33].

Upon deletion of each one of the paralogs in such case, the concentration of the precursor of the reaction would increase, and may further drive downstream effects. These downstream effects may be similar for both paralogs. Therefore, the metabolic profiles of the deletion mutants relative to WT samples would tend to be similar for both of the paralogs.

The summary of predicted relative metabolic profiles can be found in Table 1.

Hypotheses summary for paralogous backup	
redundancy scheme employed by paralogs	**predicted response to deletion**
full complementation	little effect on metabolic profile
backup with transcriptional response	one-sided or two-sided effect on metabolome
retention of paralogs for flux increase	similar effects for both paralogs when the flux through each of the two genes is similar

Table 1 – Hypotheses summary regarding behavior of mutant strains with deletions of paralogous genes.

2 Materials and Methods

2.1 Strains and cultures

2.1.1 **Standard strains**

Strains used in the study belonged to the large-scale yeast deletion library[34] based on the BY4741 strain (*MATa*; *his3Δ1*; *leu2Δ0*; *met15Δ0*; *ura3Δ0*)

2.1.2 **Heat-adapted strains**

Strains adapted to heat were evolved in several steps by Avihu Yona from Pilpel lab:
Adaptation to high glucose - BY4741 cultures were grown in 24 well plates (1.2 ml) for 73 days in YPD medium with 2 mg/ml doxycyclin at 30°C, using a daily dilution cycle of 1:101.
After that, cells were plated and a single colony was taken for further evolution.
The single colony was grown in a similar manner at 37°C for 28 days.
Days 29-39 at 38°C
Days 40-66 at 39°C
As a control, the single colony after step 1 of evolution was grown for 66 days with similar dilution at 30°C.
The heat-evolved strains showed resilience to heat stress. Avihu showed that control strain held for 90 min. in 45°C survived poorly (26%), but the evolved strains showed significantly higher survival rates (71% survival).
Avihu also showed that after the first period of adaptation to 30°C the cells became diploid, therefore the results from these experiments can't be compared directly to those from the haploid BY4741 strain.

2.2 Culture growth

To minimize the effect of growth conditions on experiments cultures were grown in a controlled fashion to equal cell density. The growth procedure for each experiment was as follows:
Cultures were thawed from a frozen stock to YPD plates.

Twenty four hours before sampling 5 ml starter cultures were grown from single colonies at 30°C in rotary shaker at 100 rpm in mineral medium (see below).
After ~4 hours optical density at 600nm (OD_{600}) was measured for all starters, and cultures were diluted into 50 ml erlenmeyers, such that in ~18 hours they would reach OD_{600} = 0.7. Growth rate estimate of 2.35 hours/generation was used (wild type (WT) growth rate in mineral medium).
Eighteen hours later OD was measured again and actual growth rate was calculated for each mutant. Samples were diluted such that in ~2 hours OD reached 0.75. This step ensured that all strains were harvested at the same growth stage.
After 2 hours samples were harvested for metabolite extraction
The mineral growth medium composition was as follows:
Glucose - 20 g/l
Ammonium sulfate (NH_4SO_4) - 5g/l
KH_2PO_4 (2 g/l),
$MgSO_4 \cdot 7H_20$ (0.55 g/l),
NaCl (0.1 g/l),
$CaCl_2 \cdot 2H_2O$ (0.09 g/l
Uracil (0.02 g/l), L-Histidine (0.02 g/l), L-Leucine (0.1 g/l), Methionine (0.02 g/l)
$ZnSO_4 \cdot 7H_2O$ (0.7×10^{-4} g/l)
$CuSO_4 \cdot 5H_2O$ (0.1×10^{-4} g/l),
H_3BO_3 (0.1×10^{-4} g/l)
KI (0.1×10^{-4} g/l),
$FeCl_3 \cdot 6H_2O$ (0.5×10^{-4} g/l),
inositol (0.12 g/l),
thiamine/HCl (0.014 g/l),
pyridoxine (0.004 g/l),
Ca-pantothenate (0.004 g/l),
biotin (0.0003 g/l)

2.3 Sample quenching and extraction

Quenching and extraction protocol was modified based on of the protocol developed by Castrillo et al[35].

Quenching and Washing

All quenching and washing procedures were carried out at -40°C, temperature was controlled manually using a digitally monitored dry ice-ethanol bath. For each experiment 18 ml of sample were harvested in two batches of 9 ml.

Each batch was pippetted into 50 ml polypropylene Falcon tubes (Falcon) filled with 36 ml pre-cooled 60% quenching solution and shaken vigorously(methanol-water, buffered with 10 mM pH 7.4 Ammonium Acetate (Fluka, ≥99.9999%). Analytical grade methanol and water (Merck) were used). The quenching solution was prepared at most 24 hours before the experiment.

After quenching, the cells were spun for 3 min at 3200 g in pre-cooled (-10°C) centrifuge (Eppendorf) with the buckets cooled to -20°C.

The supernatant was removed from the tubes and 10 ml of fresh quenching solution were added for washing the remaining medium out. Cell pellet was resuspended by vigorous shaking and vortexing. At this stage the two batches taken from each sample were reunited.

Following an additional 3 minute spin at 3200 g the supernatant was removed once again. To normalize for possible loss of samples in the extractions a defined amount of Ribitol (Sigma) was added to each sample (30µl of 0.017mg/ml) as an internal standard[36].

Extraction

Samples were extracted for 3 minutes at 80°C in 5 ml boiling ethanol-water solution (80:20, buffered with 0.05 mM Ammonium Acetate, pH 7.4). The extraction solution was freshly prepared in every experiment day.

To remove cell debris the samples were spun for 15 minutes at 3200 g at room temperature. Supernatant from each sample was collected into three aliquots (Eppendorf, 2 ml, round bottom).

Sample volume was reduced in speedvac (Savant) for 1.5 hrs. The three aliquots for each sample were reunited into one and lyophilized overnight. The dried samples were stored until chemical analysis between two to four weeks at -80°C in a Revco freezer.

2.4 Gas chromatography – mass spectrometry analysis

For chemical analysis of metabolites, gas chromatography – mass spectrometry (GC-MS) was carried out. The GC-MS system was composed of a COMBI PAL autosampler (CTC analytics), a Trace GC Ultra gas chromatograph equipped with a

PTV injector, and a DSQ™ II quadrupole mass spectrometer (ThermoElectron). The protocol for the analysis followed established techniques in Aharoni lab[37-39]. Sample volumes of 1 µl were injected into the GC-MS following methoxymation by Methoxy amine HCl (40 µl of 20 mg/ml solution in pyridine were added to dried samples) and derivatization with N-Methyl trimethyl silyl trifluoro acetamid (MSTFA) – 70 µl / 110 µl[36]. A retention time standard mixture (14 µg/ml each of *n*-dodecane, *n*-pentadecane, *n*-nonadecane, *n*-docosane, *n*-octacosane, *n*-dotriacontane, and *n*-hexatriacontane in pyridine), was injected after each set of samples.

2.5 Quantitation and Assignment of mass signals to metabolites

Pre-processing of multiple chromatograms was performed using xcms package v.1.14[40] for R v.2.8 programming language. Xcms allows multiple chromatogram alignment, signal extraction and quantitation of mass signals.
The following settings were used with xcms:
for xcms peak detection:
method = "matchedFilter", fwhm = 1.8, step = 0.2, steps = 4, mzdiff = 0.7, index = 0, snthresh = 4, max = 10000
Two cycles of grouping and retention time correction were applied.
grouping settings:
bw = 0.8, mzwid = 0.7, max = 500, sleep = 0
retention time correction settings:
method = "linear", span = 0.2
The missed peaks filling was carried out using the default setting

All post-processing normalization and analysis procedures were carried out by custom scripts coded in Matlab v. 7.7 (Mathworks).

2.6 Estimation of OD at the time of quenching

Before quenching, optical density (OD) was measured for each sample at λ=600nm. The OD is proportional to cell density. To calculate the culture opacity at the moment of quenching, the OD for each sample was corrected by the growth rate of the sample and the time that passed from measuring the OD until sampling. $OD_{final} = OD_0 \bullet 2^{t/gr}$

where ***gr*** is the generation time for the sample as measured experimentally and ***t*** is the time that passed from OD_0 measurement until quenching.

2.7 Stress administration

60 minutes stresses were administered to cultures before sampling as follows:
Ethanol stress – Absolute ethanol was added up to 5% v/v in the medium.
Oxidative stress – H_2O_2 was added to 0.3 mM final concentration from a 300 mM stock.
Drug-induced growth arrest – Two independent experiments were carried out using cycloheximide (protein synthesis inhibitor[41]). 350 µl or 500 µl of 700 µg/ml stock in DDW were added to 50 ml cultures for a final concentration of 5 µg/ml and 7 µg/ml in the medium.
Heat stress – cells were transferred to an incubator preheated to 45°C. Temperature of the medium was monitored constantly. Once the cultures reached 36°C (~15 minutes), the cultures were transferred to a 37°C incubator.
Heat stress for heat-adapted strains – A similar procedure was used, however, cells were transferred into an incubator preheated to 50°C. For the final stress of 39°C cells were taken out of the incubator when they reached 38°C (~15 minutes) and transferred to a 39°C incubator. For a final stress of 45°C cells were held until 44°C in the 50°C incubator (~25 minutes), then the temperature was rapidly lowered in the incubator to 45°C.

2.8 Significance of transcriptional changes in trehalose pathway in strains evolved towards heat resilience

Transcriptional changes were measured by Avihu Yona from our lab using an Affymetrix yeast microarray following exposure to 42°C or 45°C stresses.
Trehalose metabolism genes were taken from the Kegg metabolic pathways repository (accessed 10/04/2010)[42-45].
The significance was assessed heuristically:
In every checked condition (60 minute response to 42°C or 45°C relative to 30°C control strain or heat-evolved strain relative to control strain @30°C) the mean upregulation of trehalose biosynthesis genes (Supplementary 3) was calculated. To assess significance of change the mean upregulation level of the trehalose

biosynthesis was compared to the means of 100,000 random sets with the same amounts of genes as in the trehalose biosynthesis pathway (7 genes).

3 Results

3.1 Analytical method development

To analyze the yeast metabolome in a comprehensive manner I required a method that would allow both rapid and stable metabolomic analysis of multiple yeast strains and conditions. Several methods exist that allow the metabolic profiling of *S. cerevisiae*[15,35,46,47(review)], however, none of the published methods in their original form provided satisfactory results for high sample throughput in my experimental settings. The main parameters to control in experiments include sensitivity of the analytical method to a wide spectrum of metabolites and repeatability between biological replicates.

Upon checking many of the existing protocols for usability in a large-scale study I discovered that most current techniques required major modifications for my study. The existing methods were either not scalable enough[15,48] (thus not allowing high sample throughput), lacked in the spectrum of materials that could be reliably extracted from the cells[35,46] or were not reproducible enough[46].

I have thus developed a pipe-line for the cultivation, sampling, extraction and data analysis of yeast cells in a reliable, and high-throughput manner that allowed me to ask questions pertaining to the response of a large part of the yeast metabolome to different conditions and genetic perturbations. The outline of the experimental flow for data acquisition and analysis is depicted in Figure 1. The numbers 1-6 above the different experimental stages are described respectively in subparagraphs 3.1.1 to 3.1.6.

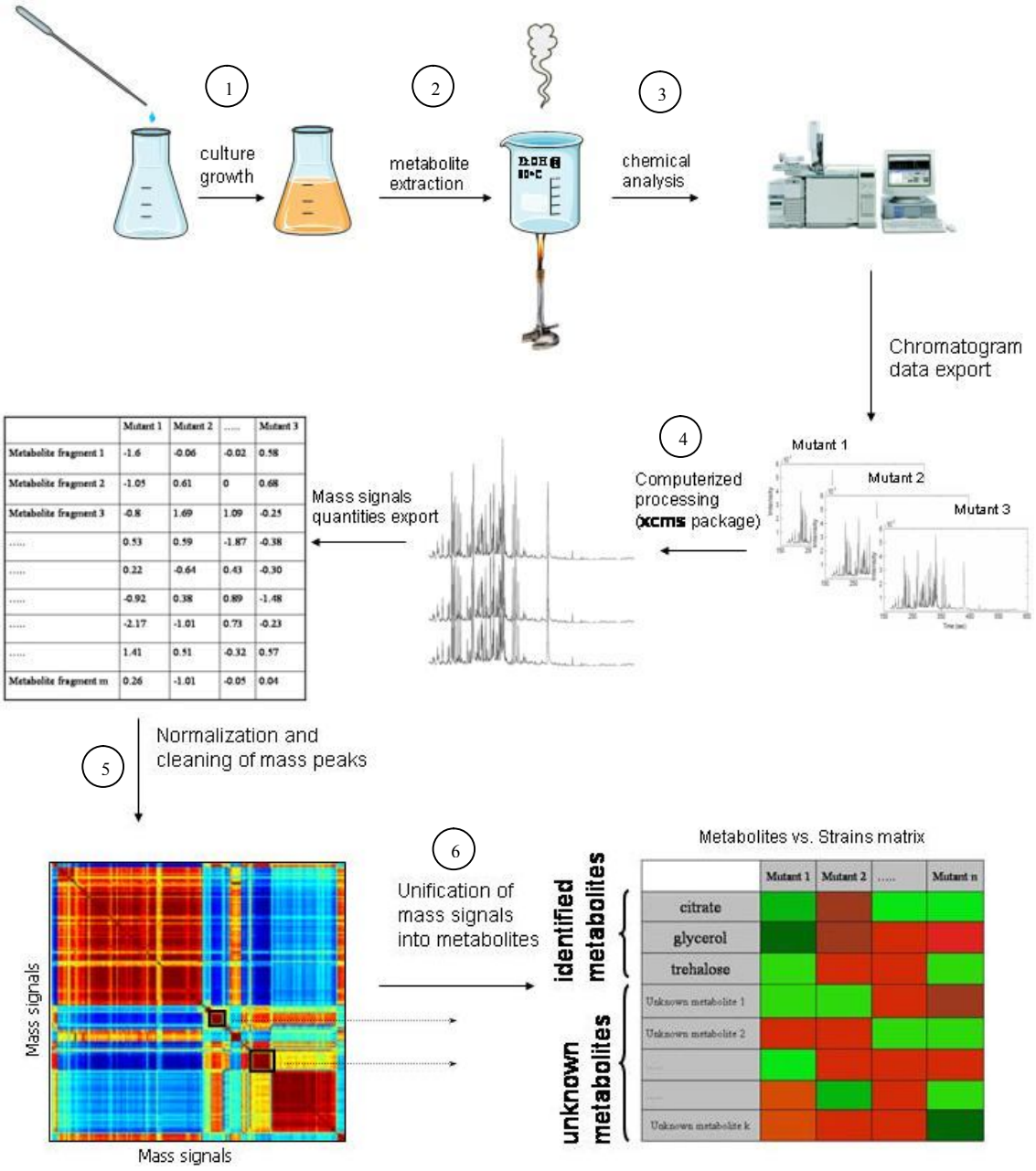

Figure 1 - The workflow for sample preparation, chemical analysis and raw data analysis. Stages marked with numbers 1-6 are described in detail in sub-paragraphs 3.1.1 to 3.1.6 respectively.

3.1.1 Controlled growth of biological samples

Many of the differences in the intracellular metabolome depend upon growth conditions and growth rate[19]. Therefore, uneven growth of cultures is one of the main caveats when analyzing the yeast metabolome. To generate robust results when analyzing multiple samples I developed a protocol to ensure controlled cell growth in

all cultures. The growth conditions I developed are described in detail in Methods section (2.2).

Briefly, the cells were thawed from frozen stock onto YPD plates, and then grown to similar growth stage.

Additionally, there was need to minimize the effect of the culture medium on the results. Many compounds from the standard growth media are also present in the yeast cells. To maximize the amount of metabolites that originate from the cells I used a minimal defined mineral medium[49] supplemented with necessary amino acids for culturing. Ammonium sulfate was used as the nitrogen source and glucose as the carbon source.

As can be observed from Figure 2, the final variation in optical density between the cultures was small (relative standard deviation of ~11%). Further, I applied a correction for the differences in OD at the data normalization stage (see below).

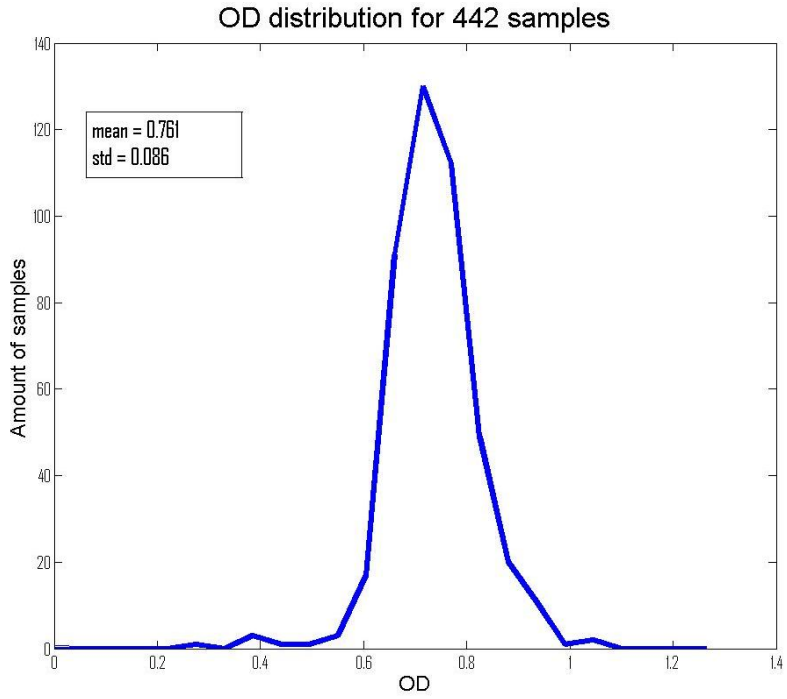

Figure 2 - Distribution of the final OD values before harvesting for 442 samples taken for the analysis of paralogous backup. The final mean optical density over 442 cultures was 0.761±0.086

3.1.2 **Extraction of metabolites from the cells**

Cell quenching and extraction protocol was based on the protocol developed by Castrillo *et al*[35] with the main modification being the pH buffering agent and

sampling volume reduction. The original pH buffering agent (tricine) was seen as an overloaded peak in chemical analysis in my experimental set-up. I evaluated several extraction and quenching conditions differing in pH buffering conditions (tricine buffered, ammonium acetate buffered and non-buffered). Conditions that had the least variability were those in which the quenching solution was buffered with ammonium acetate. Ammonium acetate buffer is volatile, and evaporates during sample drying leaving little impact on the samples. This protocol was recently independently shown[18] to be very effective for large-scale sampling of *S. cerevisiae*. The volume of quenching was reduced to 18ml of yeast culture to allow for higher throughput in the experiments. The final protocol allowed for 12 samples to be harvested at once. This was an improvement over the more laborious protocol for 6 samples employed by Castrillo *et al*[35].

Briefly, cells were quenched in buffered methanol at -40ºC, centrifuged and washed with cold methanol to remove traces of medium. Subsequently, samples were extracted in boiling ethanol, followed by volume reduction in speed-vac dryer (Savant). Finally, samples were freeze-dried in lyophilizer over-night and kept @ -80ºC until chemical analysis.

3.1.3 **Chemical analysis of metabolites**

For chemical analysis of metabolites, gas chromatography – mass spectrometry (GC-MS) was carried out. This method was chosen due to the wide range of the metabolites it allows to detect, high degree of separation between metabolites, the high precision in quantification, the availability of the equipment and the relatively low costs per sample. Other methods for analysis of primary metabolites were considered (specifically liquid chromatography – mass spectrometry), however, the absence of readily-available equipment and protocols has stopped me from pursuing these directions further.

The protocol for the analysis followed established techniques in Aharoni lab[37-39]. Samples were injected into the GC-MS following methoxymation by Methoxy amine HCl and derivatization with N-Methyl trimethyl silyl trifluoro acetamid (MSTFA)[36]. Methoxymation causes the opening of sugar rings, and MSTFA is an effective trimethylsilyl donor which reacts to replace labile hydrogens on a wide range of polar compounds with a $-Si(CH_3)_3$ group. MSTFA produces volatile and thermally stable derivatives.

3.1.4 **Computerized analysis of chromatograms**

After chemical analysis my goal was to automatically analyze the abundance of mass signals in the data. The goals of such an analysis are: robust peak detection in chromatograms, alignment of peaks in different chromatograms and peak integration. I experimented with several software suits (MZmine[50], XCalibur (ThermoFinnigan) and xcms[40] for the R programming language) that performed the above actions. Xcms produced the best results for automatic analysis among the three, as it allows relatively fast processing times, has a very precise quantitation algorithm and highly sensitive detection of compounds in complex matrices (such as used in yeast).
Xcms performs multiple chromatogram alignment, signal extraction and quantitation of mass signals. Mass quantitation of xcms corresponded to manual quantitation (see Figure 3) very well with a fit of R^2=0.99. Manual quantitation for this test was performed in XCalibur v. 1.4.

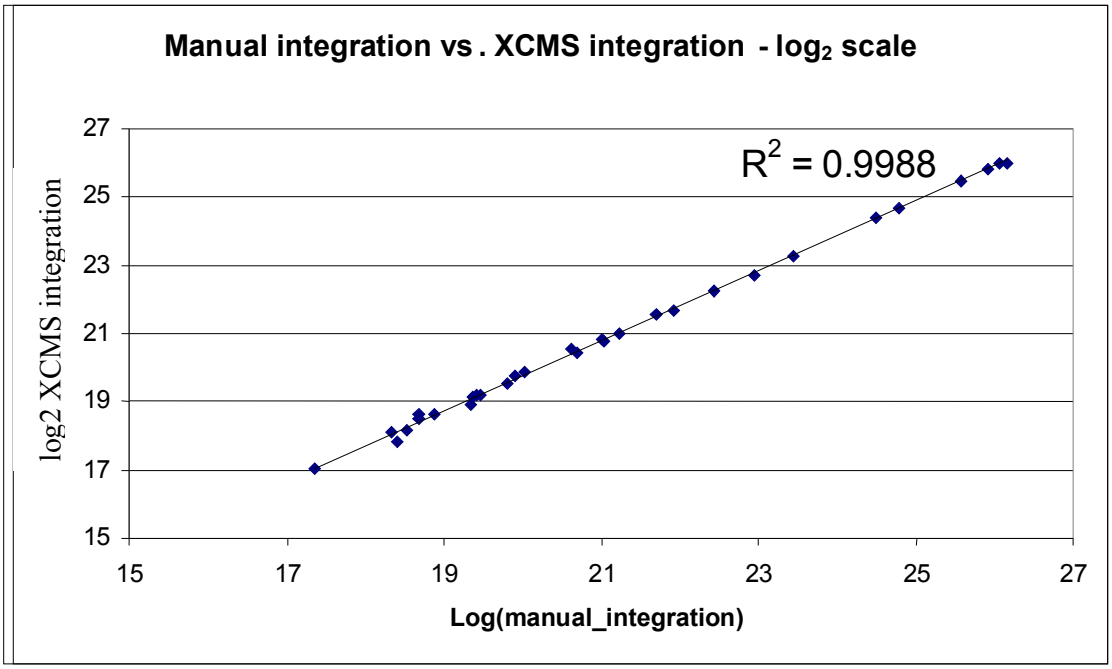

Figure 3 - Comparison of manual quantitation and automatic quantitation by xcms based on 30 different peaks from a standard mixture.

Although xcms presents good results in peak quantitation, manual adjustments must be carried out for proper chromatogram alignment. Chromatograms were re-aligned by time after xcms quantitation and the minimal common set of mass signals identified in all chromatograms was kept for further analysis.

3.1.5 **Data normalizations**

Following my analysis I have encountered several sources of variability in the samples, due to sample loss, growth medium effects, different cell density in the cultures and intrinsic noise due to minute differences in growth conditions. I could compensate for sample loss by careful normalization of data to internal standards and for enhanced cell amounts by normalization to the measurements of cell amounts at sampling time. To cope with other effects I performed statistical analysis of replicates, as detailed below.

3.1.5.1 Normalizing for sample loss during extraction

During the data analysis, for each sample the mass signals were divided by the amounts of ribitol as identified by injection of a standard into the GC-MS.

3.1.5.2 Normalizing the metabolome data for variability in cell density

To account for variability in cell/ml quantities in samples the intensities of mass signals were divided by the estimated OD value (see Methods section 2.6).

3.1.6 **Method for mass signals assignment to metabolites**

During quadruple mass spectrometry analysis metabolites are ionized and broken up to fragments. The original molecule is screened by the abundance of fragments, however, the amounts of each of these fragments (mass signals) is proportional to the original amount of the metabolite. As my aim was to investigate the nature of the metabolic response to environmental perturbations and gene deletions it was important to reassemble the mass signals to metabolites for all the samples. I devised an automatic algorithm that associates mass-signals to metabolites and implemented it in Matlab v. 7.7.

The algorithm developed for mass signal to metabolite assignment utilizes the fact that metabolites vary in their levels across multiple experiments of different biological samples. This variability is due to slight differences in the extraction procedure of samples, as well as biological variation between samples. At the same time, the relative amounts of different mass signals resulting from a single metabolite are expected to be stable across all experiments due to the robust hard ionization of substances in quadrupole mass spectrometer.

The method unifies mass signals based on the correlation between their intensities in multiple conditions as well as on the similarity in chromatographic retention times. Since the ratio between the intensities of two mass signals that belong to the same metabolite is almost constant across different samples, the correlation between them is high across multiple experiments. On the other hand, I expect the correlation between a pair of mass signals that belong to different metabolites because of the variation in relative levels of the two metabolites in different biological samples.
To assign mass signals to metabolites I clustered the results of xcms output derived from multiple samples using a hierarchical clustering algorithm.

3.1.6.1 Distance calculation for clustering

The distance between each two mass signals (Equation 1) was defined as the correlation between them in case the retention time difference between the two mass signals was equal or lower than a user-defined cutoff. When the retention time difference was larger than the threshold, i.e. the two mass signals originated from two different chromatographically separated compounds, the distance between the two mass signals was set to be very large.

$$\begin{cases} \text{If} \quad D_{RTij} \le User\_Threshold, & D_{ij} = 1 - \rho_{ij} \\ \text{Otherwise:} & D_{ij} = 100 \bullet D_{RTij} \end{cases}$$

where:

D_{RTij} is the retention time difference between two mass signals *i* and *j*

User_Threshold is the threshold defined by user of the maximal retention time difference between two mass peaks that can belong to the same compounds. This distance is usually defined based on a set of compounds from a standard sample. In the test data set the retention time distance was set to be 1.5 seconds.

ρ_{ij} is the Spearman's rank correlation coefficient between the intensities of the mass signals across multiple conditions.

D_{ij} is the final distance between the two mass signals that will be used as the distance measure for the clustering procedure.

Equation 1 - Calculation of the distance for the clustering algorithm which unifies mass signals based on Spearman's correlation coefficient and distance in the retention time. If the retention time difference between two mass signals is larger than a user defined threshold, the distance between them is much bigger then between two peaks that are near each other in retention time.

I determined the clustering parameters and method by benchmarking the clustering results vs. a test set of mass signals assigned to metabolites. The benchmark set

included 51 mass signals in 8 groups with overlapping retention times assigned to 17 metabolites. Different clustering methods and cutoff parameters were tested for the automatic assignment of these mass signals to metabolites (Figure 4).

The similarity of automatic clustering results to the manual assignment of mass signals to metabolites was assessed each time by the Jaccard similarity coefficient (Equation 2).

$$Jaccard = \frac{n11}{n11 + n10 + n01}$$

where for each pair of mass signals

n11 is the amount of pairs that were assigned to the same metabolite both automatically and manually

n10 is the amount of pairs that were assigned to the same metabolite manually, but not automatically

n01 is the amount of pair that were assigned to the same metabolite automatically, but do not belong to the same metabolite in the manual assignment.

Equation 2 - Calculation of the Jaccard similarity coefficient. The Jaccard coefficient penalizes both for splitting the same manually identified cluster to smaller clusters, and for combining too many mass signals into one metabolite.

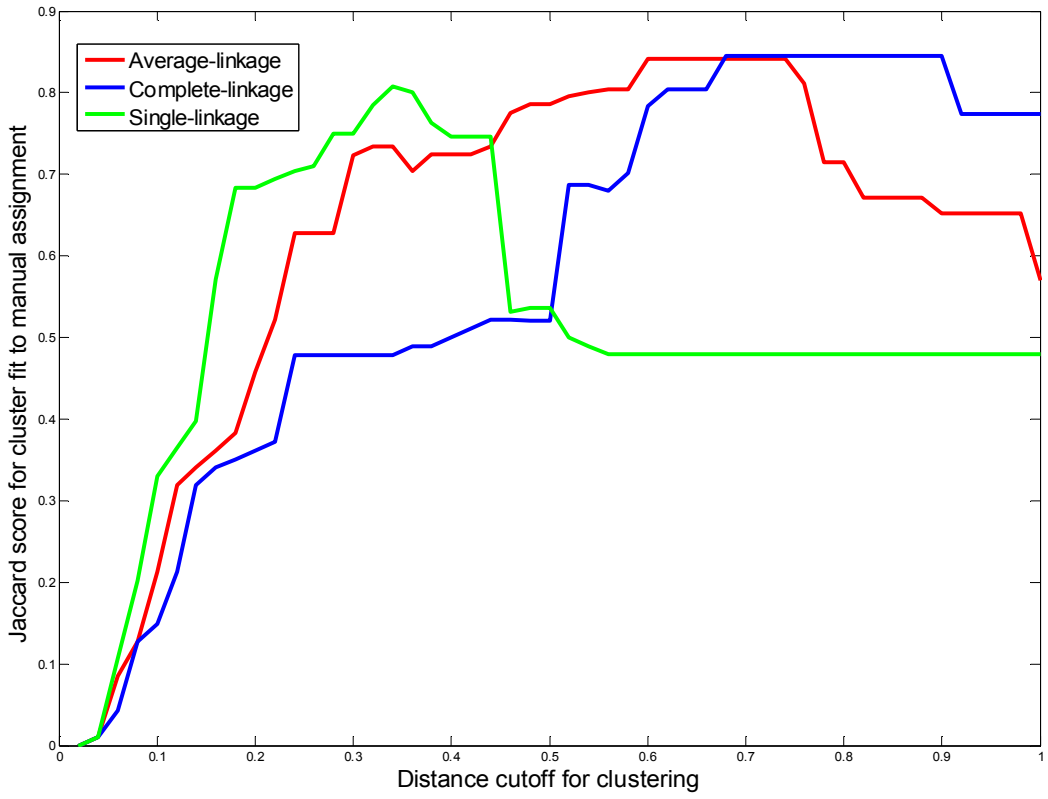

Figure 4 –Jaccard score for the fit of automatic assignment of metabolites to clusters based on a test set for three different clustering methods and different cutoffs for clustering.

Average linkage was the most robust method with respect to the range of parameters. A cutoff of 0.6 gave the highest assignment score for the average linkage method.

3.1.6.2 Manual curation of automatic mass-signal to metabolite assignment and data quality assessment

After clustering of the data I manually curated all resulting clusters by observing in the original chromatograms the masses that clustered together. If the masses did not overlap exactly on the retention time scale in one or more of the chromatograms the masses were separated to different metabolites. Mass signals that appeared in two metabolites with somewhat overlapping retention times were removed from further analysis due to possible mistakes in quantitation by xcms (see example in Figure 5).

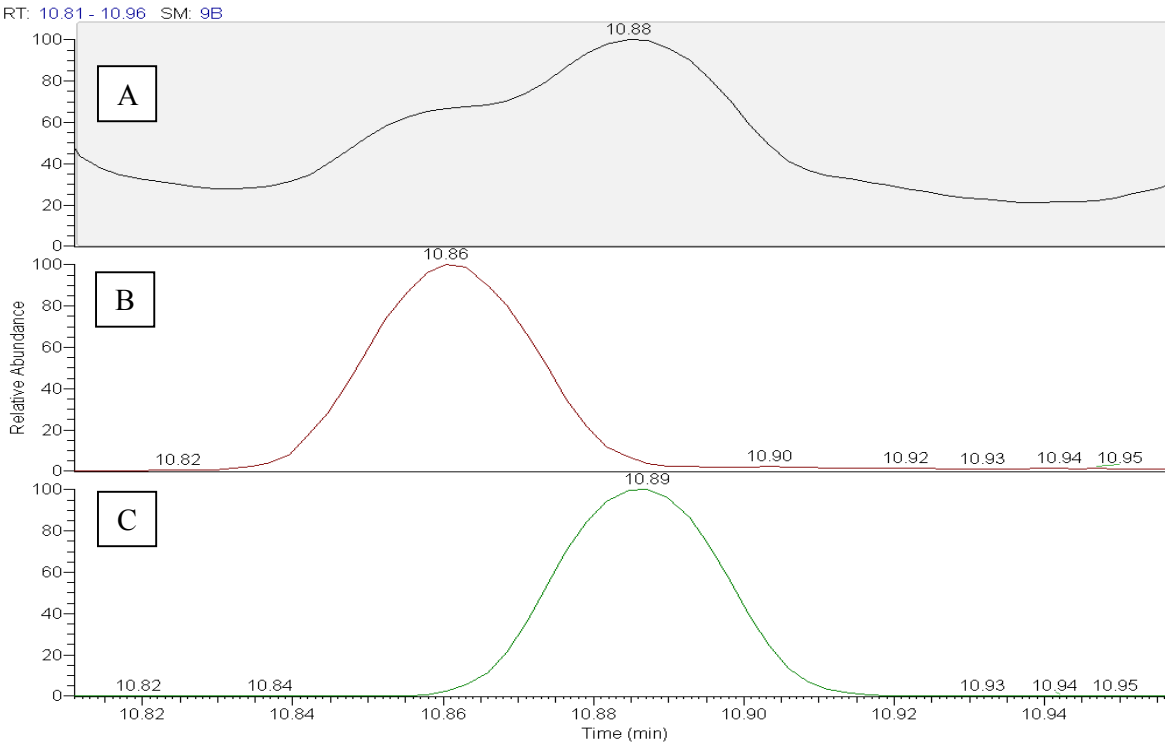

Figure 5 – An example of manual post-processing of chromatograms. A screenshot from a sample chromatogram in XCalibur with single ions 73, 245 and 263 (A, B and C respectively). The mass in section A. (mz=73), exists both at retention time 10.86 min. and at 10.89 min. xcms is prone to error in quantitation of such masses, therefore this mass signal and similar instances within other samples were removed from further analysis. The two compounds were quantified by their unique respective masses at mz=245 (B) and mz=263 (C).

The above pre-processing steps resulted in a list of masses, their quantities in each sample (peak areas) and their assignment to metabolites.

3.1.6.3 Identification of metabolites (mass-signal clusters)

All previous analysis allowed me to assign mass signals to metabolites, yet it does not supply with the identity of the metabolites. To get more biological understanding from metabolomics results I needed to identify as many metabolites as possible in the samples.

Compounds were putatively identified by comparison of their retention index and mass spectrum with those generated for authentic standards analyzed on my instrument[39]. When the corresponding standards were not available, compounds were putatively identified by comparison of their retention index and mass spectrum with those present in the mass spectra library of the Max-Planck-Institute for Plant Physiology (Q_MSRI_ID; http://csbdb.mpimp-golm.mpg.de/csbdb/gmd/msri/gmd_msri.html) and the commercial mass spectra library NIST (www.nist.gov). MS-Search v.2.0d software (NIST) was used for matching spectra of metabolites to standards. Manual comparison of retention indices of standards to the detected metabolites was used to filter the hits from MS-Search software. Retention time indices calculations were based on a mix of alkane chains injected into the GC-MS during every run.

My method allowed the detection of 136 compounds in yeast cells. Forty three (43) out of them fit to the retention time and mass spectrum of known compounds and, therefore, could be identified. The list of identified compounds can be found in Supplementary 1.

3.1.6.3.1 Metabolic network coverage by identified metabolites.

In the next step I wanted to observe how well the metabolic network of *S. cerevisiae* was covered by the identified metabolites. Major classes of the identified metabolites included organic acids, amino acids, sugar-phosphates, sugar-alcohols and others.

I plot the metabolites that I could identify on the metabolic network representation of yeast (Figure 6– in red). Overall, I see that the coverage of the metabolic network is quite uniform apart from the ergosterol/lipid biosynthesis pathways and vitamins/co-factors pathways which are underrepresented in my set of identified metabolites. This is mainly due to detection limits of the GC-MS apparatus I used.

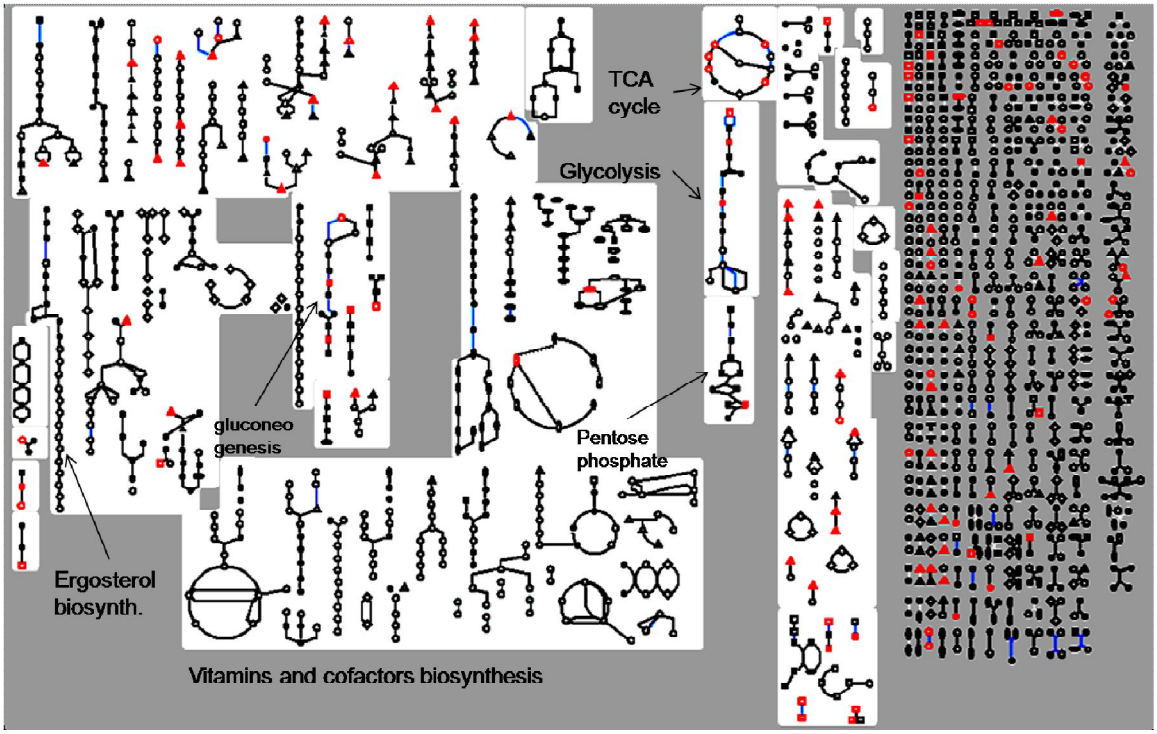

Figure 6 - Representation on the metabolic network of *S. cerevisiae* of reactions that were perturbed (by deletion mutations – see paragraph 3.2.1) in cyan and metabolites that could be identified in red. The names near each section represent a cluster of similar pathways.

3.1.7 Statistical analysis of metabolome measurements of different strains and treatments vs. control

3.1.7.1 Experimental design for sampling and subsequent replicates analysis

To allow for multiple replicates I analyzed 12 samples in every day of experiments. As a general rule, different strains or treatments were analyzed in four replicates along with four controls (wild type (WT) or untreated quadruplicate) within each day. A major goal was to reduce variation in sampling conditions between the replicates. The standard daily setup included 3 experiment sets of 4 samples, each containing:

4 wild type/untreated replicates.

4 replicates of one mutant or treatment.

4 replicates of a second mutant or treatment.

Time between quenching of the first and the last samples in every experiment day was 15 minutes at most. After extraction and drying, samples were kept at -80ºC for a minimal period of two weeks, but no longer than one month to minimize the variability in the treatment of different samples.
The order of the injections into GC-MS was randomized within every day of chemical analysis. Samples were prepared every day, such that the time between injection into GC-MS of the first and last sample was not more than 24 hours.

3.1.7.2 Relative quantitation and analysis of significance of replicate measurements via randomization

The first stage in analysis of a large set of samples consisted of removal from further analysis of outlier samples detected by manual inspection of the data.
Due to the high variability in instrument sensitivity across days it was not possible to directly compare the signal intensity between different strains or treatments. Therefore, for each metabolite within each of the samples I needed to normalize the signal to the WT/untreated samples. The mean of ratios of experiments (r_h in Equation 3) to control samples within the same day were treated as the metric for a change.

$$r_h = \frac{\left(1/n_h\right) \cdot \sum_{j=1}^{n_h} a_{h_j}}{\left(1/n_w\right) \cdot \sum_{i=1}^{n_w} W_i}$$

Equation 3 – For every metabolite, for every day of experiments for each experiment set h within that day, the ratio between the mean of the values of experiments was divided by the mean of the WT values within that day. n_h is the amount of replicates in an experimental set. n_w is the amount of wild type/untreated control samples within a day. a_{h_j} to $a_{h_{n_h}}$ are the measured metabolite values of each experimental set within a day. W_i to W_{n_w} are the measured metabolite values of wild type/untreated samples within a day.

To determine the significance of the ratio score I constructed a null distribution for each metabolite in the following manner:
For every set of replicates h out of a total of m sets of experiments (including WTs/non-treated sets), and every replicate h_i out of total n_h replicates within a set

having metabolite levels a_{h_i} to $a_{h_{n_h}}$ corresponding values α_{h_i} to $\alpha_{h_{n_h}}$ were calculated for the null model according to Equation 4:

$$\alpha_{h_i} = \frac{a_{h_i}}{\left(\frac{1}{n_h - 1}\right) \cdot \left(\sum_{j=1}^{n_h} a_{h_j} - a_{h_i}\right)}$$

Equation 4 – Formula for calculation of metabolite-specific null values for significance analysis of sample ratios to control. *h* is a particular set of replicates, n_h is the amount of replicates for set *h*. a_{h_i} is the intensity of metabolite in an instance *i* of set *h*.

This means that every value was normalized by the mean of the other replicates, thus representing the variability within each sample group and creating a null distribution which is has the same variance or higher than the variance in the original samples. I calculated the p-value for each mean of ratios of set *h* normalized to the mean of WT (r_h in Equation 3) by randomization of the null model set. To mimic the fact that I averaged ratios across n_h replicates I drew random sets of the same size from the null set and averaged them. This procedure was carried out 1,000,000 times for each metabolite and for each ratio r_h.

I defined the p-value of a given ratio score as the fraction of random sets that had a similar or higher average than the ratio score r_h.

I corrected for multiple hypotheses testing using the FDR method[51] with a q-value of 5%. All ratios of metabolites in all the mutant strains/treatments that did not pass the resulting cutoff of 5% were set to 1 to abstain from further analysis of non-significant results.

3.1.8 Reproducibility of data for deletion mutants

Metabolomics results are known to be highly unstable and may vary greatly even between replicate samples within the same day[52]. To assess long-term reproducibility of my method I carried out complete and independent profiling of 8 deletion mutants with a difference of 10 months between samplings.

First, I analyzed changes that were labeled significant in both studies. The correlation coefficient between metabolite changes between the two samples was 0.82, across 33

metabolites ($p < 10^{-8}$). Looking at the dot plot (Figure 7), I can observe that only three metabolites out of the 33 change in opposite directions in the two studies, while all the other metabolites change in the same direction in both studies (same-direction changes are 90% of all significant changes).

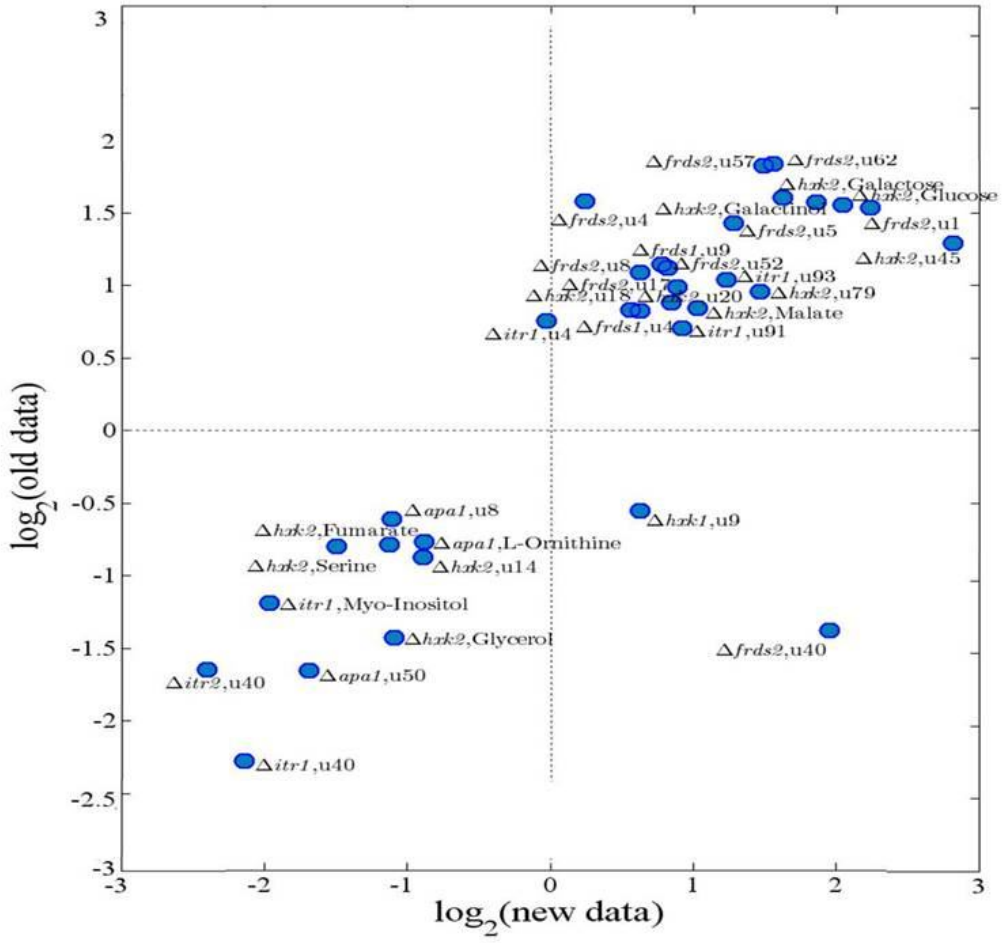

Figure 7 – Dot plot of (mutants/WT) on $\log_2$ scale for metabolites significantly changing in two replicate studies of 8 deletion mutants with 10 months difference between sampling dates. Each label designates the metabolite that significantly changed, and the mutant in which this change occurred. Unidentified metabolites are marked with "u" and a number following the designation.

However, when I performed such a comparison on the combination of significant changes in the two studies the correlation dropped markedly to r=0.23 ($p<10^{-4}$). Nevertheless, same-direction changes (i.e. instances in which a metabolite's concentration changed in the same direction in both studies) represented 73% (213/290) of overall significant changes in at least one of the experimental sets. To assess the significance of this extent of agreement between replicates I performed a

shuffling permutation analysis and derived the null distribution of expected same-direction changes. With 100,000 permutations I have not even once reached 213 same-direction calls in shuffled data (Figure 8) (p-val<10^{-5}).

old data / new data	↑ significant	↑ non-significant	↓ non – significant	↓ significant
↑ significant	20	75	35	2
↑ non-significant	91	359	153	11
↓ non-significant	15	198	70	8
↓ significant	1	14	11	10

Table 2 – Distribution of amounts of metabolites increase/decrease in two independent experiment sets (sampled 10 months apart) of 8 single gene deletion mutants. For each category of metabolites (significantly changing or non-significantly fluctuating upwards/downwards as signified by arrows) in the old set (columns) the table contains in rows the distribution between the categories for the new set of experiments.

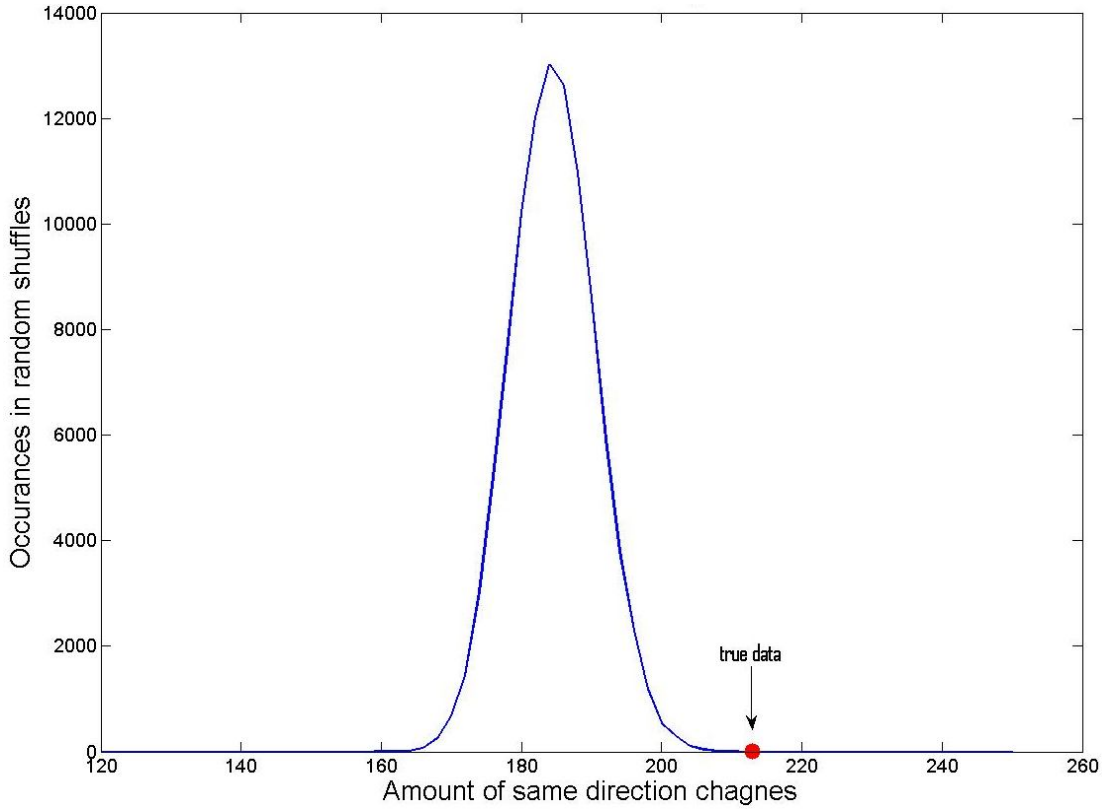

Figure 8 – Distribution of randomly shuffled same-direction changes for old and new extraction experiments of deletion mutants. At random ~185 same-direction hits are expected in the data, this is much less than the observed 213 in true data (marked with a red dot).

Therefore I can deduce that the direction of change is overall very significantly repeatable.

3.2 Metabolic profiling of S. cerevisiae paralogous deletion mutants

After the development of the analytical method I have set out to explore the phenomenon of backup through genetic redundancy that is carried out by paralogs in the *S. cerevisiae* genome. By analyzing deletion mutants that have a paralog elsewhere in the genome I expected to uncover the metabolic component underlying genetic redundancy and transcriptional reprogramming.

3.2.1 **Analysis of paralogous single mutants**

To check the hypotheses outlined in Table 1 in the introductory section 1.4 I have carried out metabolic profiling of 78 single gene deletions that represent 39 pairs of paralogous enzymes from diverse gene families metabolic network of *S. cerevisiae*. This set represents more than a third of the total of 105 duplicate gene families in the metabolic network of *S. cerevisiae*. Mutant list can be found in Supplementary 2. Through mining these metabolic profiles I will gain additional understanding of the deletion phenotype of genes in *S. cerevisiae*.

3.2.1.1 A view on co-regulation of different compounds

First, I wanted to look at the similarity of response of different compounds to gene deletions. I plotted a correlation matrix of fluctuations for the different compounds. For additional insight into the closeness of different compounds in the space of the mutants I performed clustering of metabolites and the data was sorted according to the clusters.

Two observations are evident when looking at the correlation matrix (Figure 9). One is that many compounds are co-regulated. The other is that some clusters are dominated by a certain metabolite type. For example the amino-acids cluster 1 (marked on plot by 1), contains exclusively amino-acids (ornithine, lysine, arginine and glutamate). Moreover, I can probably explain the reason for these amino acids to occur in the same cluster. Ornithine, glutamate and arginine all originate in the same biosynthetic pathway in *S. cerevisiae*; glutamate in turn is a major substrate in the production of lysine. It is highly possible that the similar fluctuations due to amino acids control or availability of substrates gave rise to this cluster. Other clusters with similar compounds included the sugar-phosphate cluster (cluster 2) and a second amino acid cluster (cluster 3). Unfortunately, my strength to elucidate pathways or

relationships between compounds is severely limited by the inability to identify more metabolites. It does seem however, that relative metabolite levels are related to the levels of other metabolites from the same biosynthetic pathways.

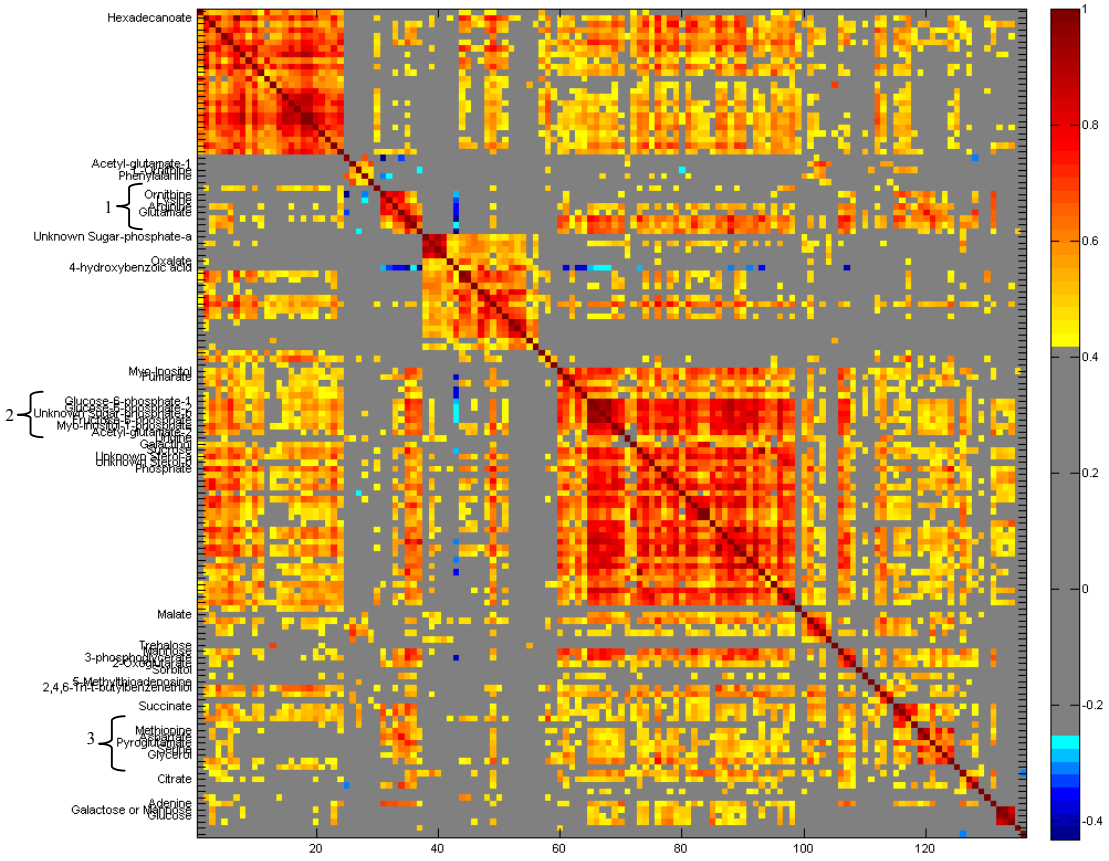

Figure 9 – Correlation matrix of different compounds across multiple deletion mutants. Data was $\log_2$ scaled, and metabolites sorted by simple hierarchical clustering. Numbers of groups relate to: 1) amino acid cluster 1; 2) sugar-phosphates and 3)amino-acids cluster 2

3.2.1.2 Observing metabolic response in the Δaco1 and Δaco2 mutants

As a primary check for my results I wanted to focus on the behavior of one specific pair of paralogous genes with well studied function. For this I chose *ACO1* and *ACO2* and analyzed the metabolic profile obtained for their deletion mutants. *ACO1* encodes the enzyme aconitase which is part of the tri-carboxylic acid cycle (TCA cycle - Figure 10) and converts citrate to iso-citrate via the intermediate aconitate. *ACO2* has an amino acid sequence similarity of ~52% to *ACO1* and has a putative aconitase activity[53].

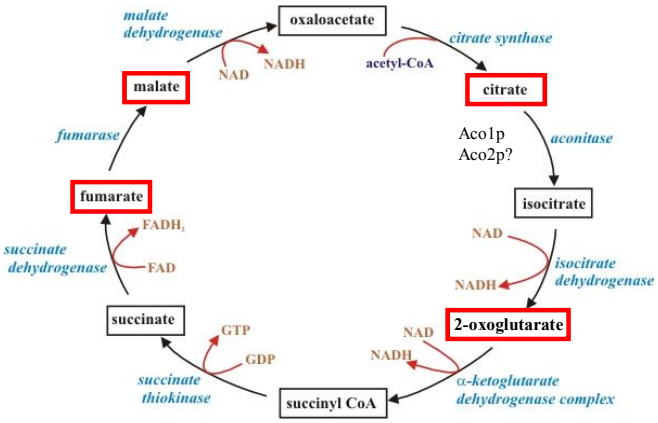

Figure 10 - TCA cycle, Aco1p and Aco2p enzymes are noted on the plot. Metabolites which were measured and changed significantly upon deletion of either *ACO1* or *ACO2* are marked in red.

In Figure 11 I plotted the changes in metabolites for both mutants relative to the WT. I can see that Δ*aco1* mutant had more metabolites significantly changing in response to the deletion than Δ*aco2* (27 in Δ*aco1* and 19 in Δ*aco2*, with mean fold change of 1.98). The *ACO1* gene is known to be the major isoenzyme responsible for the aconitase activity[54], it was also shown to be one of the factors for mitochondrial genome maintenance[55]. As could be expected, metabolites from the TCA cycle that I measured (marked with red color in Figure 10 and in Figure 11A) were up-regulated in Δ*aco1*, but less so in Δ*aco2* possibly due to its lower enzymatic activity and lower flux change in the cycle. Thirteen (13) metabolites exhibited significant changes in both mutants with similar directionality of the change. Interestingly, two of the identified common metabolites are sterols of unidentified structure. Sterol biosynthesis in yeast cells are occurs only in aerobic conditions and multiple mitochondria genes were suggested to function in sterol biosynthesis[56]. Deficiency in mitochondrial function by deletion of aconitase may have inadvertently led to accumulation of particular sterols in the cells.

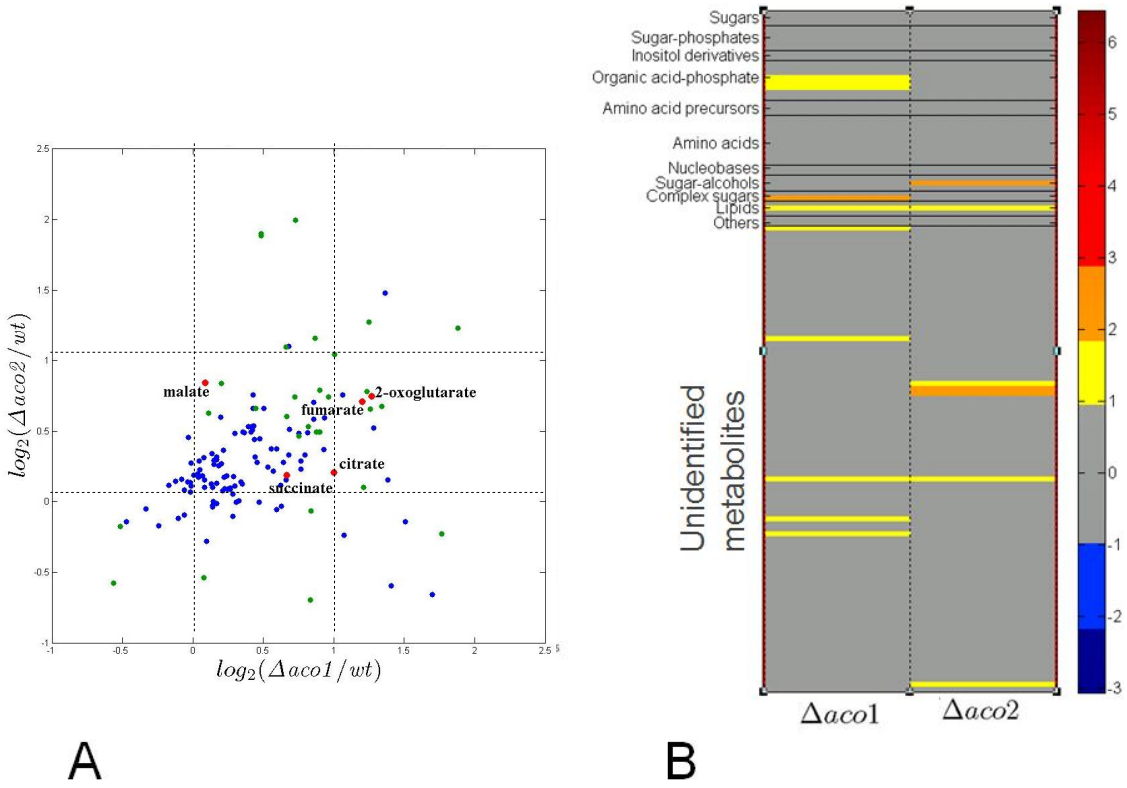

Figure 11 - A. Dot plot on a log2 scale describing the ratios of metabolites in Δ*aco1* and Δ*aco2* mutants to WT. Metabolites changing significantly in one of the mutants are marked in green, metabolites from the TCA cycle are labeled marked in red. Metabolites that do not exhibit a significant change in any of the mutants are marked in blue. B. Relative metabolite changes in the two mutants on log2 scale. Color bar on the right presents colors associated with two-fold change steps. All metabolites not changing significantly or having less than two-fold change are grey.

3.2.1.3 Observing the metabolic effect of Δnth1 and Δnth2 mutants

Looking at the response of the metabolome in deletion mutants of an additional paralogous gene pairs - the *NTH1* and *NTH2* genes, few metabolites exhibit changes (Figure 12). However, when observing known metabolites in the Δ*nth1* mutant (Figure 13), elevated levels (~100 fold increase) of trehalose are evident. Nth1p is responsible for the degradation of trehalose to glucose; therefore, its elevated levels could be expected. The *NTH2* gene is 77% similar to *NTH1*, however its expression is silenced during exponential growth in *S. cerevisiae*[53]. The Δ*nth2* mutant does not show the same elevated trehalose levels as its counterpart, probably due to the experimental conditions which I used; in my studies the cells were harvested during exponential growth, when *NTH2* is silenced.

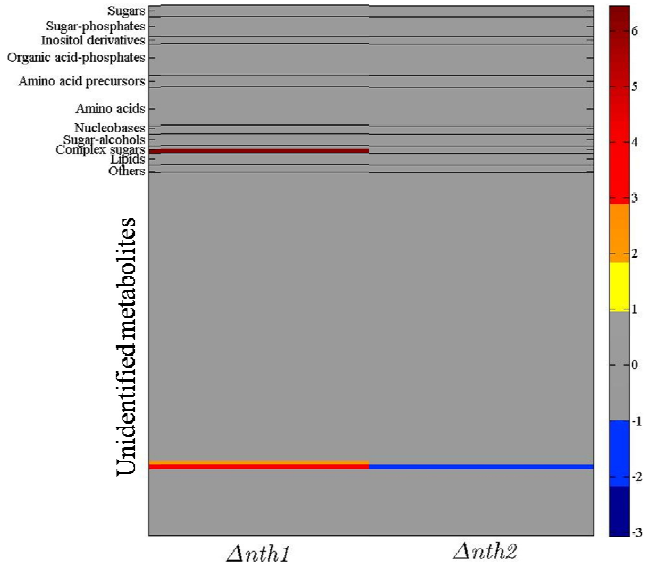

Figure 12 - Levels of metabolites in Δ*nth1* and Δ*nth2* mutants. The data is on a $\log_2$ scale with the color bar representing the fold change on the right. Grey metabolites do not exhibit significant or strong changes. Colorscale is similar to Figure 11B. Zoom-in on known metabolites (upper part of the plot) is in Figure 13.

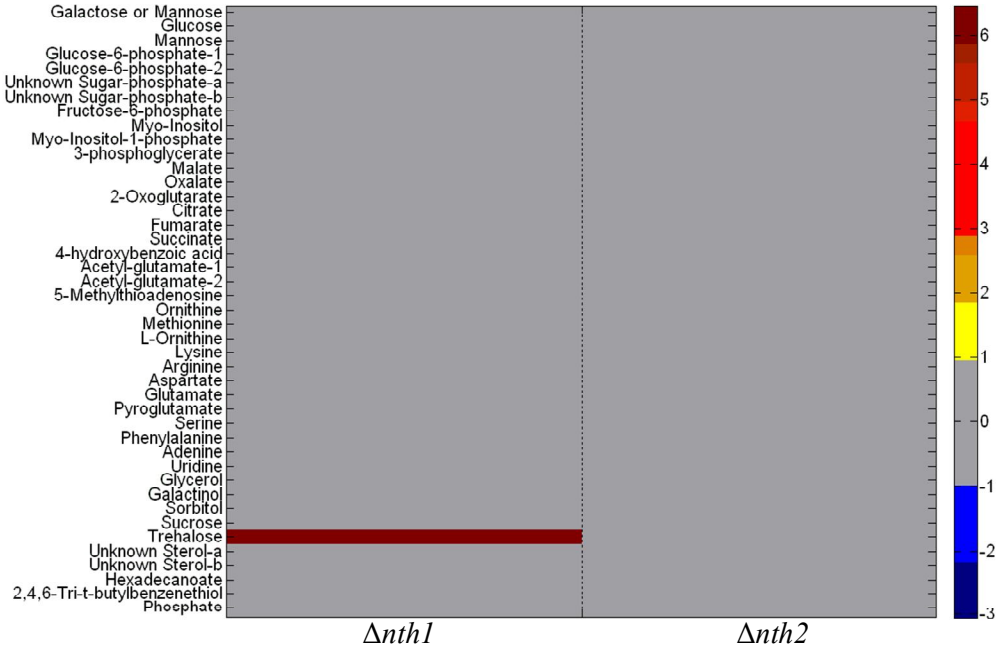

Figure 13 – Levels of known metabolites in Δ*nth1* and Δ*nth2* mutants. The data is on a $\log_2$ scale with the color bar representing the fold change on the right. Grey metabolites do not exhibit significant or strong changes. Color scale is similar to Figure 11B.

Additional discussion regarding the Δ*nth1* mutant can be found in paragraph 4.2.2 (p. 77).

3.2.1.4 Reciprocal metabolic response of paralogous gene deletion mutants

After the analysis of the particular pairs of paralogs I have set out to obtain a large-scale view of paralogous backup. In order to notice the most prominent effects I primarily limited my analysis to 2-fold changes in the mutants relative to the WT controls (Figure 14).

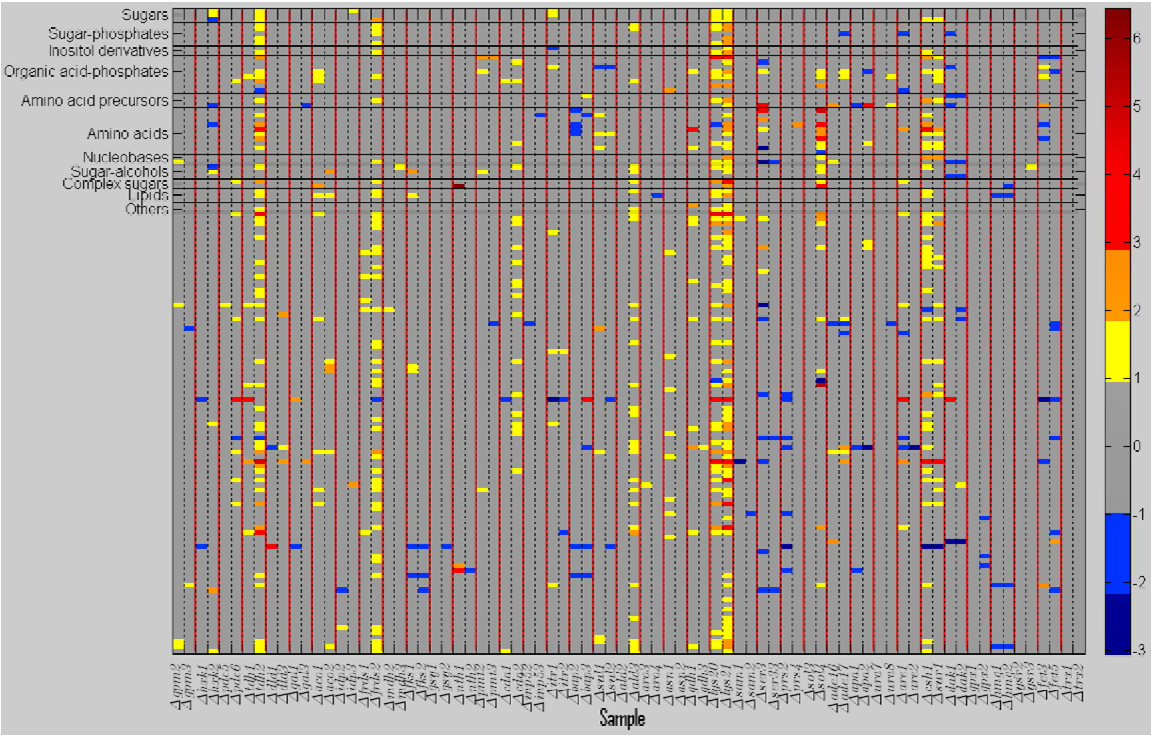

Figure 14 – Metabolite changes in paralogs relative to WT. Rows represent metabolites' relative intensity. Columns mark different mutants. Both members of each paralogous pair are found near each other and the different pairs are separated from one another by vertical red lines. The data is on a $\log_2$ scale with the color bar representing the fold change on the right. Grey metabolites do not exhibit significant or strong changes. In the upper part of the figure appear metabolites that could be identified, different groups of metabolites are separated by black horizontal lines. The groups are according to structural (chemical) classes of compounds, detailed in Supplementary 1.

Three different types of responses depicted in Figure 15 were identified when comparing reciprocal metabolic response within deletions of pairs of paralogs.

Little or no metabolic response upon the deletion of each of the paralogs.

Concerted changes in metabolites upon the deletions of the paralogs, i.e. the deletion mutants shared significantly up- or down-regulated metabolites.

One-sided response. I observe pairs, in which only of one of the deletants responds strongly to the deletion, while the metabolic response to the deletion of the paralogous counterpart is very mild.

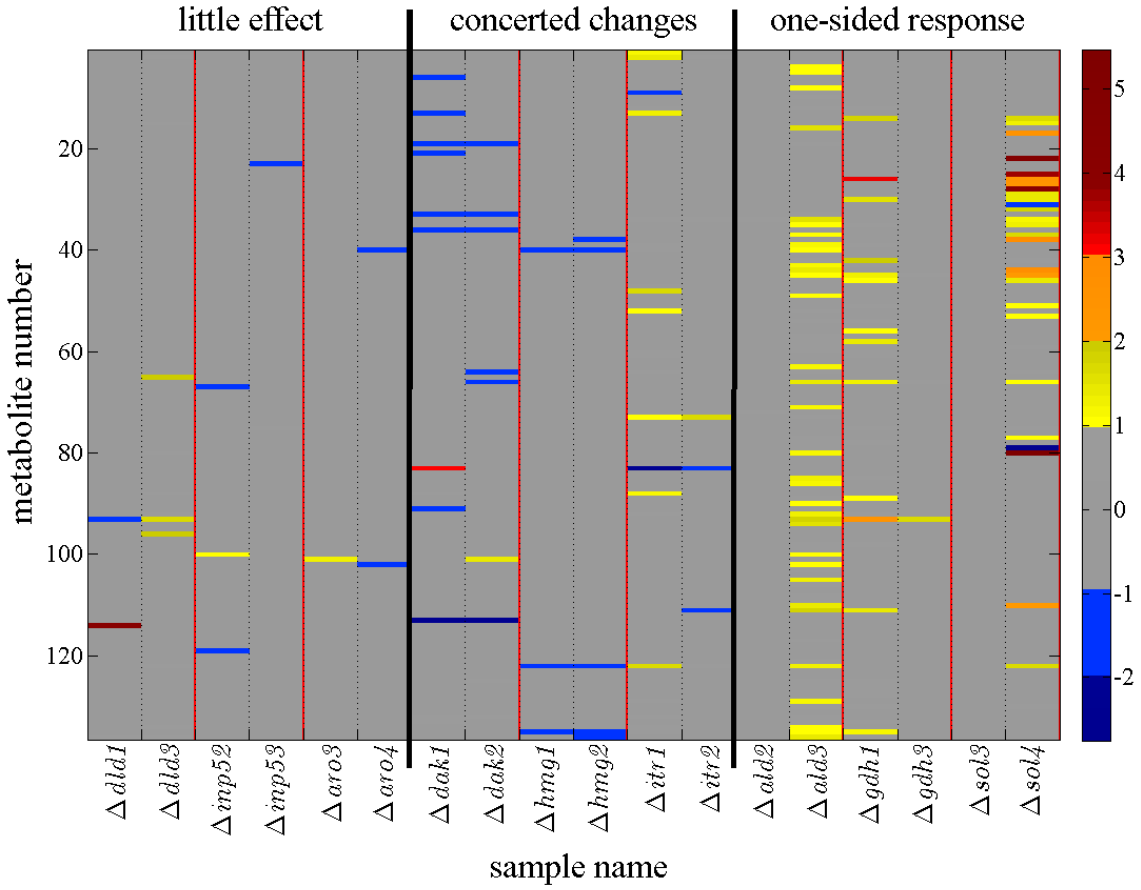

Figure 15 - Nine examples of different types of metabolic responses (from left to right – little metabolic effect, multiple concerted changes and one-sided response to deletion). Color scale and metabolite order are as in Figure 14.

Several explanations may be offered for the observed phenomena. I will go over the different types one by one.

3.2.1.4.1 Evolutionary correlates for the extent of metabolic response to deletion

To better understand the relationship between paralogs with little metabolic response, or rather to understand the significance of the amount of changes in the metabolome, I assessed the rate of purifying selection on the paralogs. I have employed analysis of rate of synonymous (Ks) and non-synonymous (Ka) substitutions - Ka/Ks – a well excepted measure of rate of evolution of paralogs[57,58]. The rate of mutation of sequences is not similar for all positions in genes. Mutations which do not change

amino acid composition of a gene (synonymous mutations) are considered evolutionary neutral, while mutations which do result in amino acid substitution are likely to cause detrimental changes in protein function[58]. Therefore, the rate of fixation of non-synonymous mutation is slower. It is possible to assess the rate of synonymous (Ks) and non-synonymous (Ka) substitutions in a pairs of genes[21,59]. Gene pairs with a lower Ka/Ks ratio are usually thought to be under purifying selection[57,58], i.e. these genes are under selective pressure to weed out non-synonymous mutations which disrupt gene function.

I analyzed the Ka/Ks ratios in my gene pairs' sequences and compared them with the mean amount of changed metabolites for each pair (Figure 16).

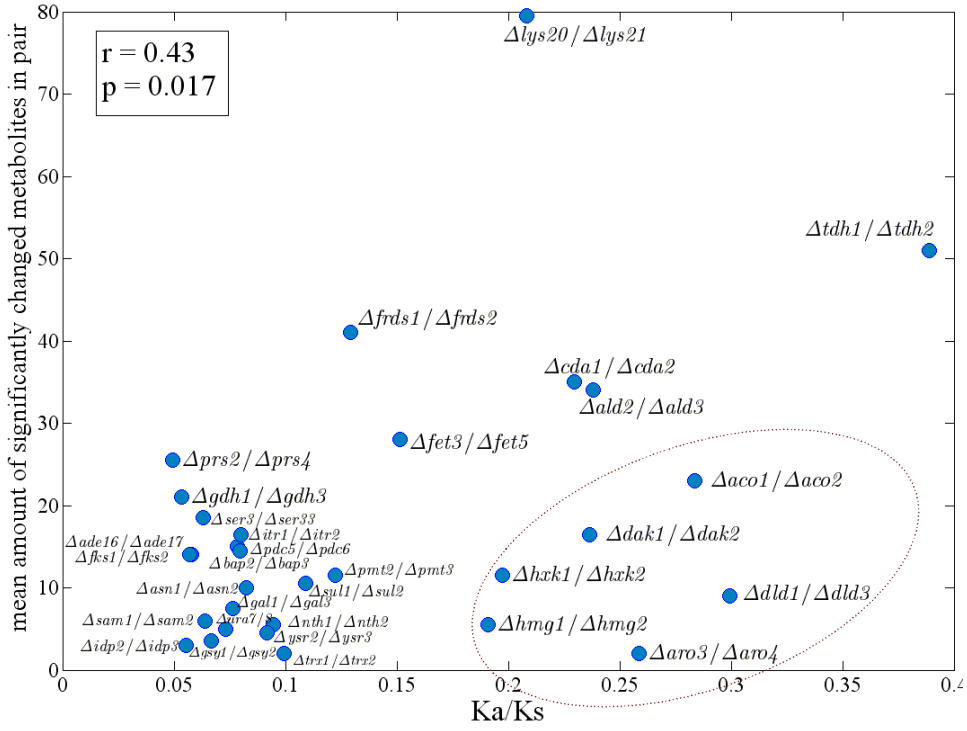

Figure 16 – Ka/Ks ratio vs. the mean amount of significantly changed metabolites in each paralogous deletion pair. Red ellipse marks gene pairs that are outliers to the main correlation trend.

I observed a significant positive correlation between the mean amount of changed metabolites in a pair and its Ka/Ks (r=0.43, p=0.017). However, I have additionally found that there is a negative correlation between Ks and the mean amount of changes in metabolites (r=-0.45). To verify that the positive correlation with Ka/Ks is not solely due to the Ks values, but also due to the Ka component, I performed a partial

correlation analysis of Ka/Ks vs. the mean amount of metabolic changes controlling for Ks values. The resulting partial correlation was still significantly positive (r=0.34, p=0.03).

Looking beyond the correlations two groups of gene pairs can be discerned in Figure 16. I identify specific groups of paralogous pairs which exhibit a deviation from the overall pattern (see red circled gene pairs in Figure 16). These pairs exhibit a relatively high Ka/Ks ratio (and thus lower purifying selection), while exhibiting relatively low amounts of metabolic changes upon deletion.

Further discussion in paragraph 4.1.1

3.2.1.4.2 Pairs exhibiting "concerted" changes

Another type of metabolic phenotype is the one displaying "concertedness", as exemplified by the pairs of paralogs appearing in Figure 17.

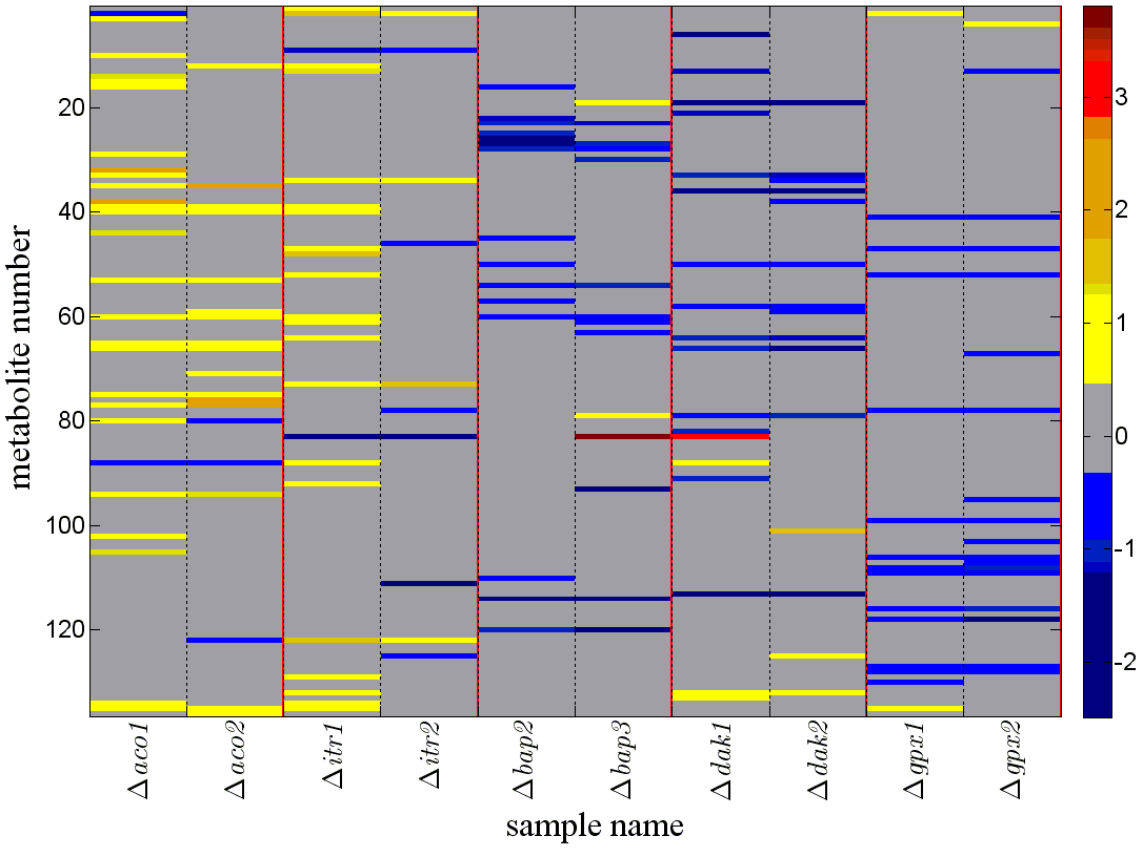

Figure 17 – Five mutant pairs with high concertedness scores as calculated by random permutation analysis. Color scale and metabolite order are is as in Figure 14. Grey metabolites are not significantly changing.

Pairs with concerted changes are these in which similar metabolic changes relative to the WT strain arise upon deletion of any of the two paralogs.

In order to achieve a deeper insight into paralogous pairs with concerted response I first needed to devise a score for the level of "concertedness". I have counted for each pair of deletion mutants the amount of same-direction changes in significantly responding metabolites in both mutants and subtracted from it the amount of opposite direction changes in significantly responding metabolites.

To assess the significance of this overlap I performed $2 \cdot 10^7$ shuffles for each pair of paralogous deletants such that the labels of significantly changing metabolites were distributed between all measured metabolites for each of the deletion mutants and then calculated the overlap score for each of the shuffles. For the significance of metabolic response "concertedness" for each of the pairs I calculated the proportion of times shuffled data had "concertedness" score ≥ true "concertedness" score. The final "concertedness" measure was derived by taking $-\log_{10}$ of that fraction (Figure 18).

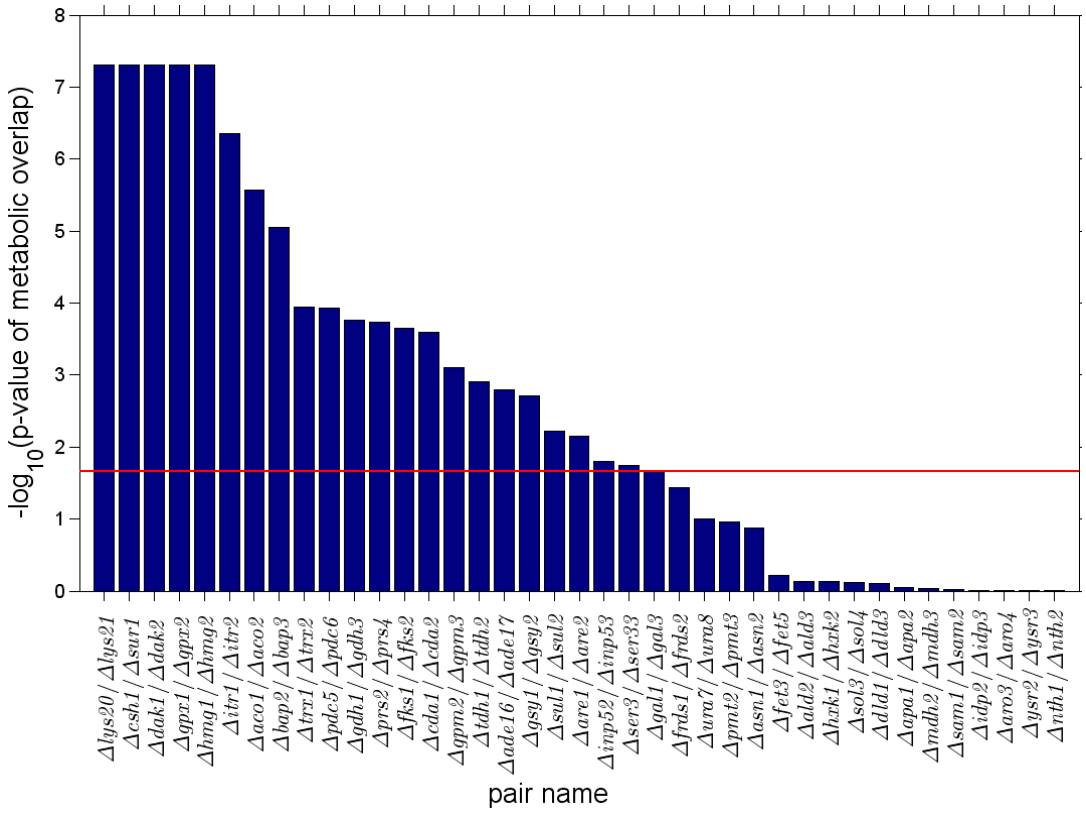

Figure 18 – $\log_{10}$(p-value) estimating the "concertedness" for different paralogous gene deletion pairs. Values above the red line (p-value≤0.02) passed significance testing controlling for multiple hypotheses with FDR q-value of 0.05.

Further discussion in paragraph 4.1.2.

3.2.1.4.3 Pairs exhibiting one-sided response to gene deletion

The third phenotype that could be discerned among the relative metabolic profiles of paralogous deletion pairs was termed “one-sided response”: a situation in which deletion of one of the paralogs evokes a strong metabolic response, while the deletion of the second paralog causes little or no metabolic phenotype. Several examples of such a behavior can be found in Figure 19.

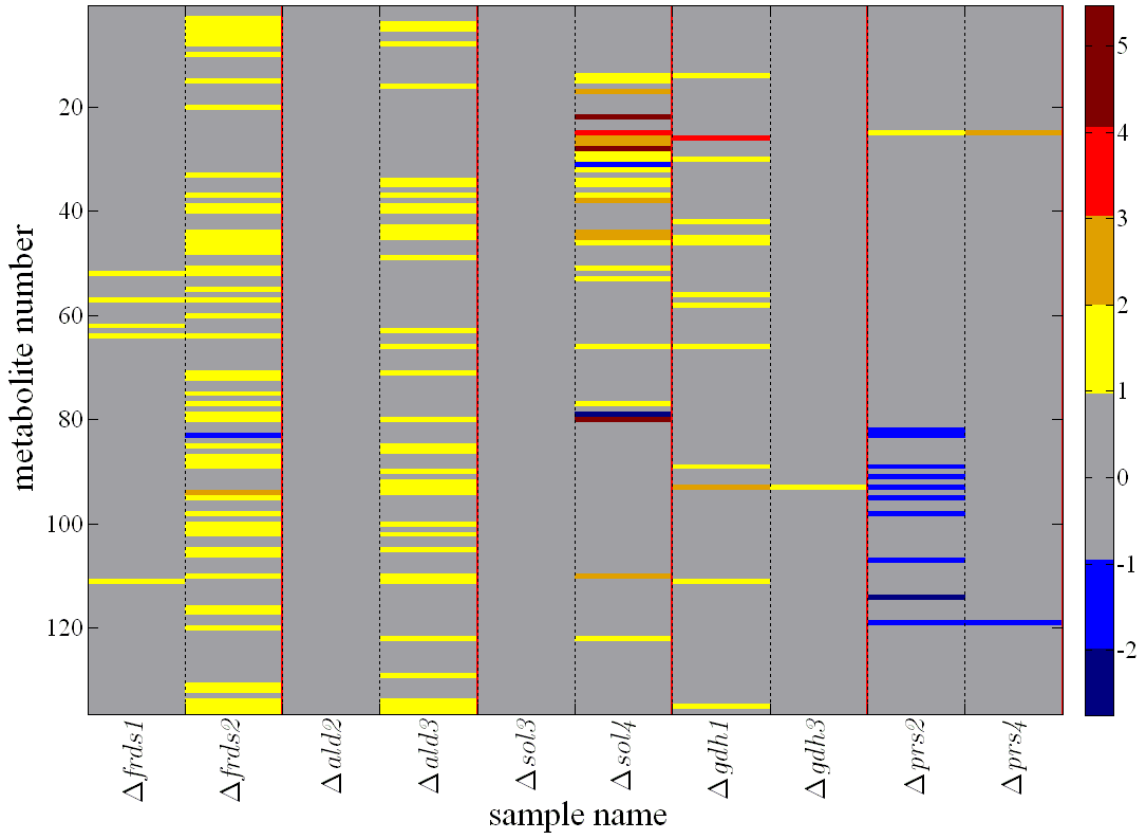

Figure 19 - Five mutant pairs with high one-sidedness scores as calculated by random permutation analysis. Color scale and metabolite order are is as in Figure 14.

I quantified the basic measure for one-sidedness of response of paralogous pairs to deletion by calculating the value given in Equation 5. I measured the difference between the amounts of significantly changing metabolites in each of the samples, subtracted the overlap in changed metabolites and normalized by the total amount of changing metabolites.

$$O_k = \frac{\left|N_i - N_j\right| - N_{common_{ij}}}{N_i + N_j}$$

Where O_k is the one-sidedness measure for each pair of single gene deletion mutants. N_i and N_j are the amount of significant changes in each of the deletion mutants' metabolic profiles. $N_{common_{ij}}$ is the amount of metabolites changed to the same direction in both mutants.

Equation 5 – Concertedness measure calculation for each pair of gene deletion mutants.

To assess the significance of one-sidedness I performed 10,000 shuffles for each pair of paralogous deletants such that the labels of significantly changing metabolites were randomly distributed between all measured metabolites in both deletion mutants. For the p-value of "one-sidedness" of response for each of the pairs I calculated the proportion of times shuffled data had one-sidedness score ≥ true one-sidedness score. The final "one-sidedness" score was derived by taking $-\log_{10}$ of the p-value of one-sidedness measure. This scoring method captures well apparent one-sidedness, when many metabolites are perturbed upon deletion, but with few changes in a pair (such as in the Δ*nth1*/Δ*nth2* pair) it is harder to achieve significant one-sidedness; since by random chance the few changes can appear in one, but not the other deletion mutant. I can see that only 8 pairs of mutants had a significant one-sided response to deletion mutation (Figure 20).

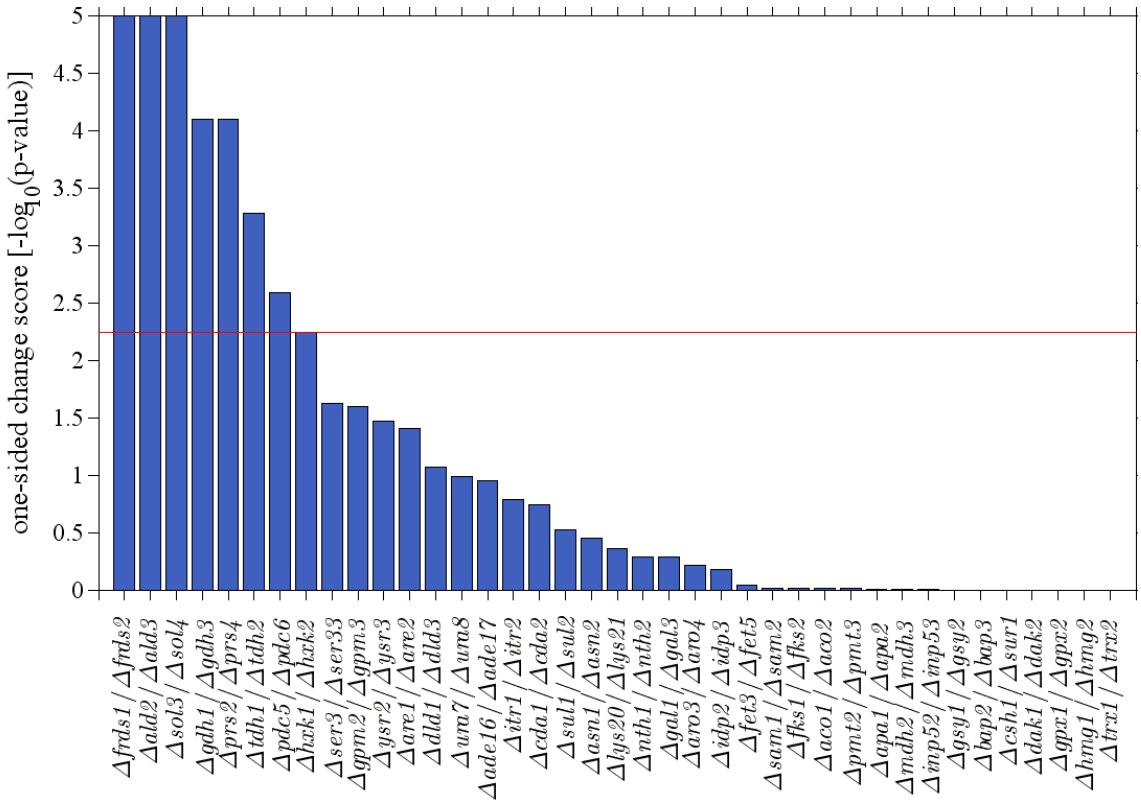

Figure 20 – "one-sidedness" levels of different paralogous gene deletion pairs. The score was constructed as noted above. Values on the red line and above (p-value≤0.0057) passed significance testing control for multiple hypotheses with FDR q-value of 0.05.

Further discussion in paragraph 4.1.3.

3.2.1.5 A potential metabolic basis for mutants' growth defects

Among the mutants I have profiled, some differed in their growth rate on minimal medium relative to the wild type strain. Metabolites both regulate the growth rate of yeast cells and are affected by it. Therefore, I have set out to check which metabolites are either positively or negatively correlated with the growth rate or fitness of cells. For each mutant I measured the growth rate relative to the WT grown in the same day. As a proxy to fitness I used the inverse of the relative growth rate (1/relative growth rate).

First, I wanted to examine whether the sheer amount of changed metabolites in mutants was indicative of deviation from the wild type fitness levels.

I can observe that there is no direct correlation between the count of changed metabolites and the fitness of mutants (Figure 21). However, when looking only at

mutants with relative fitness lower than wild type I see a modest, yet significant negative correlation (r=-0.43, p-value = 0.02) between relative fitness and the amount of changed metabolites.

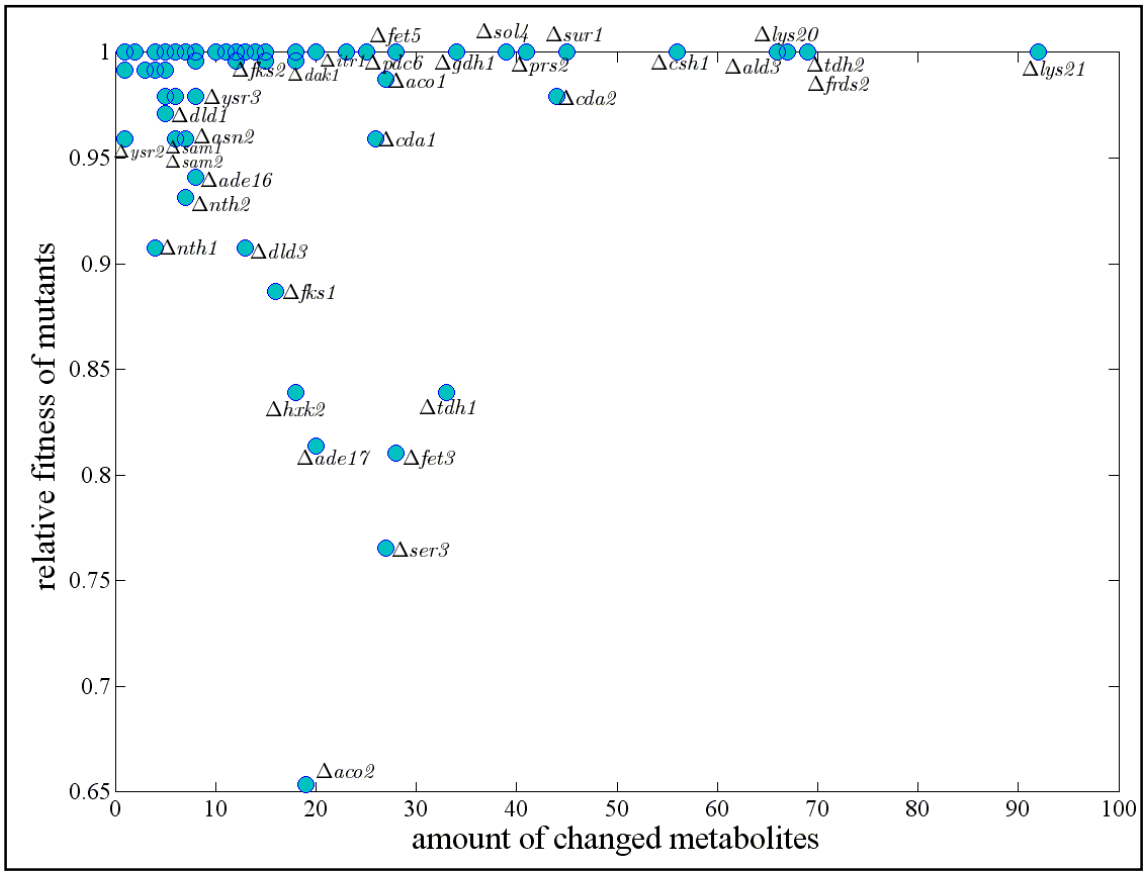

Figure 21 – Relative fitness of paralogous single mutants as a function of amount of significantly changed metabolites in a mutant. Labels of select mutants are given on the plot.

I further performed a similar analysis splitting between metabolites whose concentrations increased in mutants and those whose concentrations decreased (Figure 22).

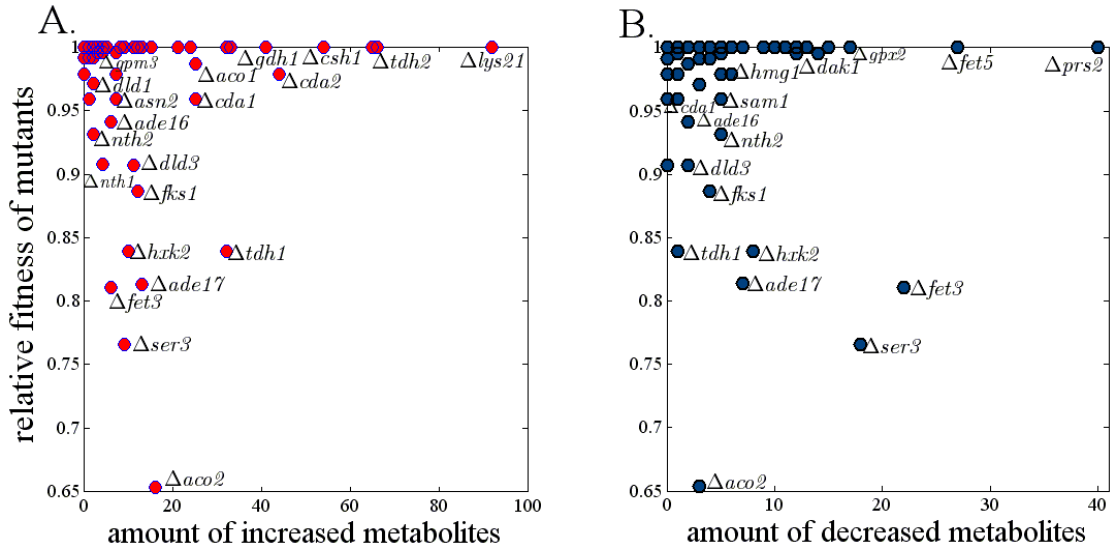

Figure 22 - A. Amount of metabolites increased after deletion of each mutant vs. relative fitness of the mutant (mutant growth rate/wt growth rate). B. Amount of metabolites decreased after deletion of each mutant vs. relative fitness of the mutant. Labels of select mutants are given on the plots.

The correlation between the amount of changes to each of the sides and the relative fitness of mutants different than WT is still negative (-0.1 and -0.3 for amounts of increased and decreased metabolites respectively) but not significant given the sample size. This suggests, then, that the association of fitness is with the overall metabolic change, rather than the directionality of the change.

To check which specific metabolite's levels are associated with the fitness of mutant strains I performed a correlation analysis between the two variables. Yet, no individual metabolite showed significant correlation with the mutant fitness. Perhaps, however, due to some higher level interactions between metabolites the combined information from multiple metabolic profiles of mutants can be used to predict fitness defects?

To tackle this question I used a simple artificial neural network which learned to predict the fitness of all mutants from metabolic profiles. The predictor network was constructed with just one perceptron neuron. The perceptron was trained on a randomly chosen set of metabolic profiles of mutants containing 70% of the data and each time validated on the remaining 30%. The proportion of training and the validation set contained the same amounts of data with growth defects. The target was a vector with designation whether a mutant had or did not have a growth defect. For

each such perceptron, as a control I trained a similar perceptron on the same data with a shuffled target vector. The comparison between performance on shuffled labels and true data was repeated 500 times. Figure 23 depicts the distributions of the amount of correctly classified mutants in the validation set of shuffled and true data.

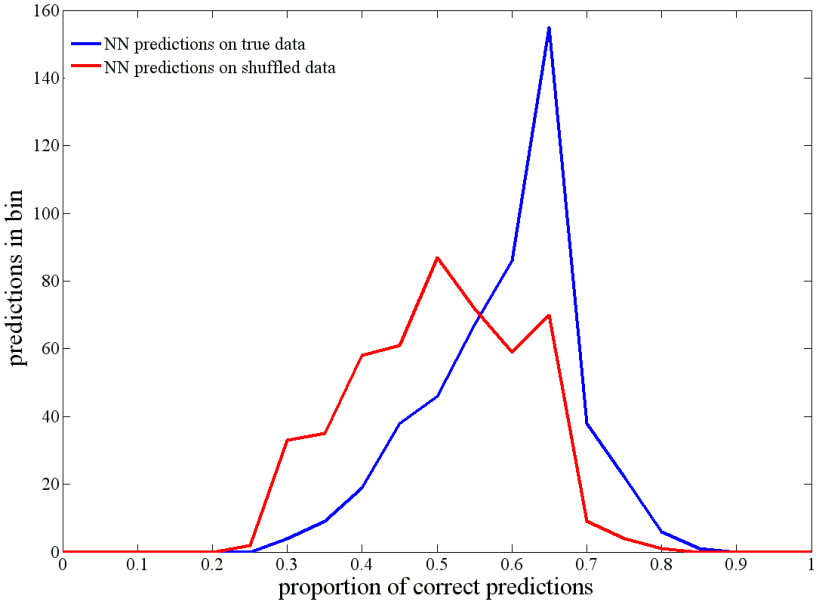

Figure 23 – Based on 500 splits of data to training and validation set. In blue is the distribution of proportion of correctly classified growth defects phenotypes from validation sets as predicted by perceptrons trained on true labels. In red is a control: the distribution of proportion of correctly predicted growth defect phenotypes from the validation sets predicted by perceptrons trained on shuffled labels.

I can see a difference in the two distributions. The mean and the median proportion of correct predictions using the shuffled labels was as expected 50%. Based on the true data, however, correct predictions existed on average in 60% of the cases. This difference was highly significant (Wilcoxon rank sum test p-value$<10^{-32}$). I also observed that in successful predictions some metabolites (e.g. ornithine) repeatedly exhibited high weights in the classifying neuron.
Further discussion in paragraph 4.1.4.

3.2.1.6 Prediction of subcellular localization of proteins based on the metabolic profile of deletion mutants

Deletion mutants that I analyzed differed in their subcellular localization. I observed that metabolites known to exist in a specific subcellular location change in response to

the deletion of genes that are localized to that location. Specifically the mutants with deletion in the *ACO1* and *ACO2* genes (which are part of the TCA cycle genes localizing to the mitochondria) exhibited significant changes in metabolites that participate in mitochondrial respiration (Figure 11, section 3.2.1.2).

I wanted to check further whether this behavior is a general property of deletion mutants. If so, then metabolomics data from deletion mutants might be used to predict the subcellular localization of genes.

I have employed a strategy similar to that described in section 3.2.1.5. I constructed a dataset with the annotation of subcellular localization according to GO[60] (Saccharomyces Genome Database accessed June, 2009) of proteins whose deletion mutants I profiled. Subsequently, I divided the genes into 4 groups; mitochondrial only, cytoplasmic and mitochondrial, cytoplasmic only and belonging to other subcellular localizations (peroxisome, endopasmic reticulum *etc.*). A simple two-layer probabilistic neural network was constructed, using the "newpnn" built-in Matlab program, to predict subcellular localization of a mutant by its metabolic profile. The data was divided once again to 70% training set mutants and 30% validation set. The proportion of genes belonging to each of the four subcellular localization groups was kept constant in the training and the validation sets. As a control I trained a similar neural network using shuffled labels of the subcellular localization of the mutants. The process was repeated 5000 times. Figure 24 depicts the distributions of the proportion of correctly classified mutants in the validation sets of shuffled and true data.

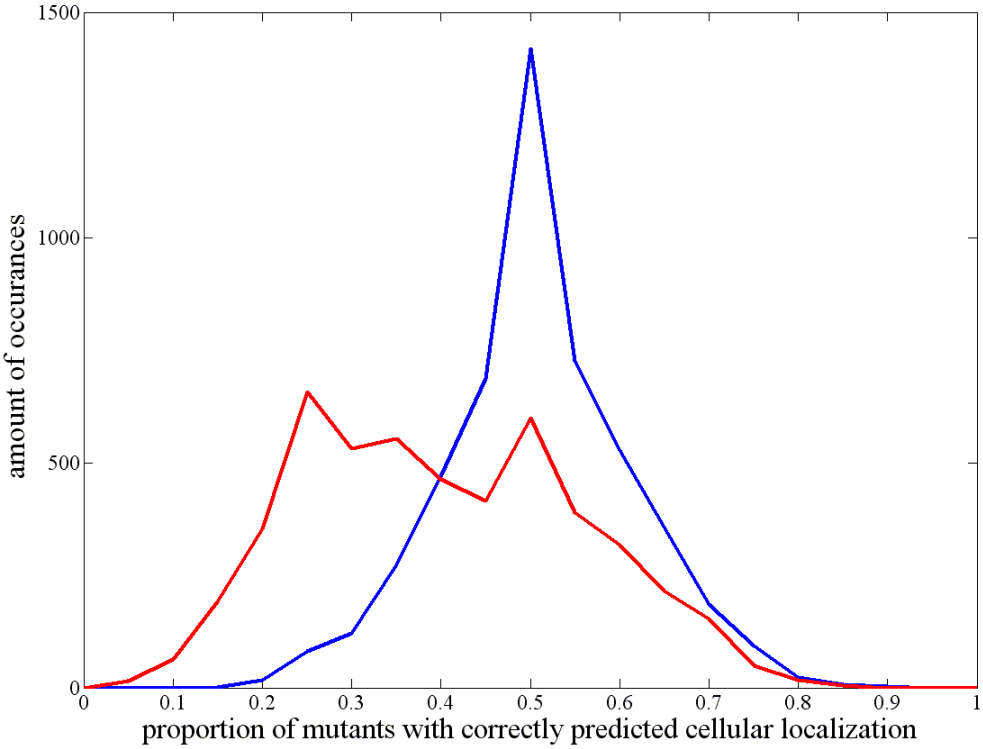

Figure 24 - Distribution of proportion of correctly predicted cellular localizations of mutants based on the metabolic profiles of deletants – 5000 randomizations. In blue is the distribution of correct predictions in the validation sets based on neural networks trained on true designations of subcellular localizations to mutants, in red is the distribution of proportion of correct predictions in validation sets classified based on neural networks trained on data with shuffled designations of subcellular localization.

I can see a difference in the two distributions. The mean and the median proportion of correct predictions in the control was 43%. Whereas correct predictions ratio based on the true data rises to 53% of the cases on average. This difference was highly significant (Wilcoxon rank sum test p-value$<10^{-200}$).

Further discussion of the results in paragraph 4.1.5.

3.2.2 Metabolic profiling of yeast double mutants in paralogous genes

Four different paralogous pairs whose double mutant was viable in the minimal medium used in the study were chosen for the analysis of double-mutations (*APA1, APA2, FRDS1, FRDS2, HXK1, HXK2, ITR1* and *ITR2*).

Since I mainly attempted to understand the relationship between single gene deletions and double deletions, I was especially interested in the metabolic manifestation of negative epistasis in yeast. Epistasis is a phenomenon in which a gene either masks or

augments the effect on the phenotype of another gene. In the case of gene deletions and their effect on fitness one can define epistasis (ε) as a case when the fitness of a double mutant is different than the expected product of the fitness of the single mutants ($\varepsilon = W_{x'x''} - W_{x'} \bullet W_{x''}$, where $W_{x'}$ is the fitness of the deletion mutant in the first gene, $W_{x''}$ is the fitness of the second deletion mutant and $W_{x'x''}$ is the measured fitness of the double mutant). Negative epistasis is, therefore, a situation in which the fitness drop of the double mutant relative to the WT is more significant than the product of fitness decreases in each of the single mutants.

I obtained epistasis measurements from the study of DeLuna *et al.*[61] for pairs of yeast paralogs. In the DeLuna study epistasis was experimentally determined for each pair in a set of multiple pairs of paralogous genes including the four genes used in my study. However, it is important to notice that the epistasis measurements of DeLuna were carried out in rich medium, while my study was carried out in minimal medium. The setup of the experiments in the current study slightly differed from that of the single gene deletion experiments. Each day of experiments contained triplicates of WT, mutant 1, mutant 2 and double mutant denoted "1/2".

Figure 25 depicts the metabolite changes in the single and double mutants following the standard normalization procedures. To facilitate viewing of the results, all metabolites, that showed insignificant, or less than 3-fold change relative to the WT were grayed out.

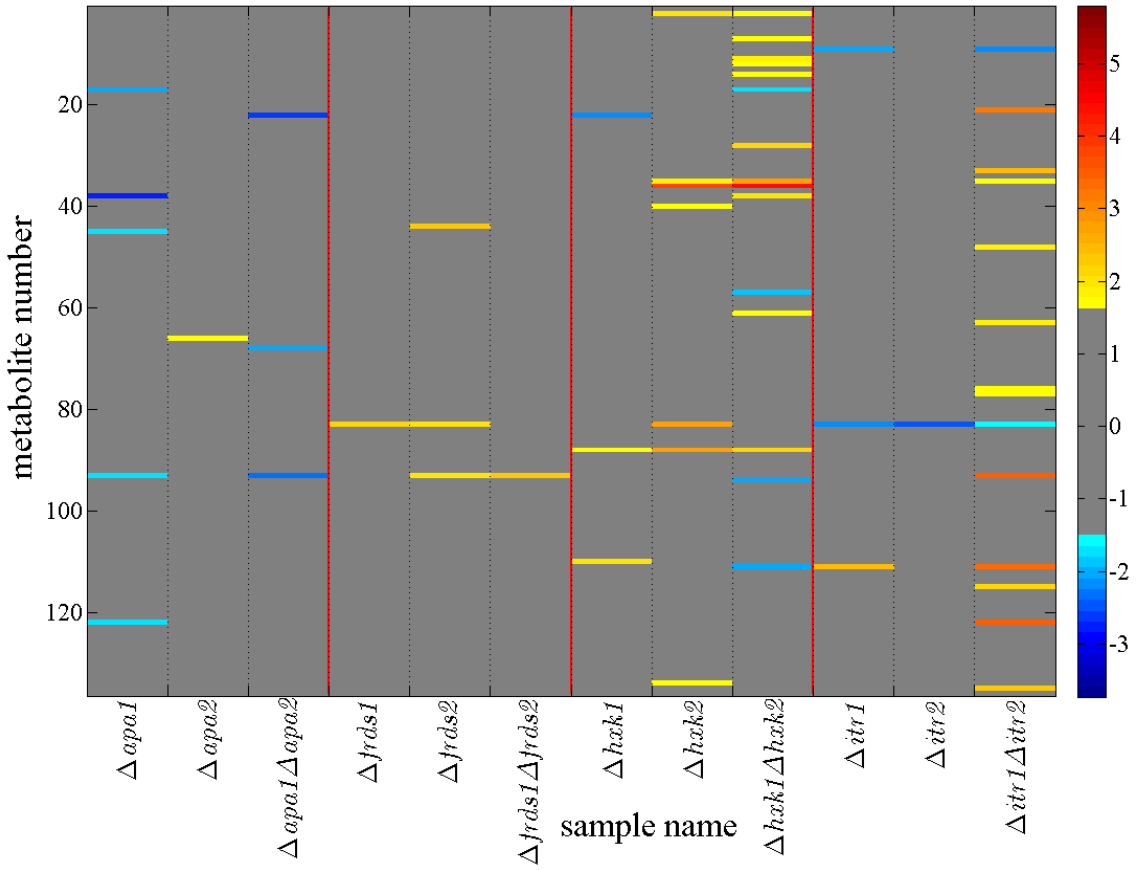

Figure 25 – Metabolite changes in single and double mutants in paralogs relative to WT. Rows represent metabolites' relative intensity. Columns mark different mutants. Metabolomics data for each single mutant and the double mutant from each paralogous pair are found near each other. Different pairs are separated by vertical red lines. Color scale and metabolite order as in Figure 14. To facilitate visual analysis metabolites that either show a response of less than 3-fold, or metabolites that are not significantly changing are depicted in grey.

One can see that there are marked differences between the response of each pair of genes to deletion and their double mutant. Both double mutants Δ*apa1*Δ*apa2* and Δ*frds1*Δ*frds2* are quite similar to each of their single mutants in the magnitude of their metabolic response to deletion; in contrast, Δ*hxk1*Δ*hxk2* and Δ*itr1*Δ*itr2* are exhibiting very strong metabolic changes relative to their respective single mutants. In the next step of analysis I looked at the relationship between each of the mutants in the mutant set and the double mutant with regard to the amounts and identity of all significantly changed metabolites (Figure 26 – note, that the results taken, also

include metabolites that showed less than 3-fold change, therefore the amount of metabolites is larger than what was shown in Figure 25).

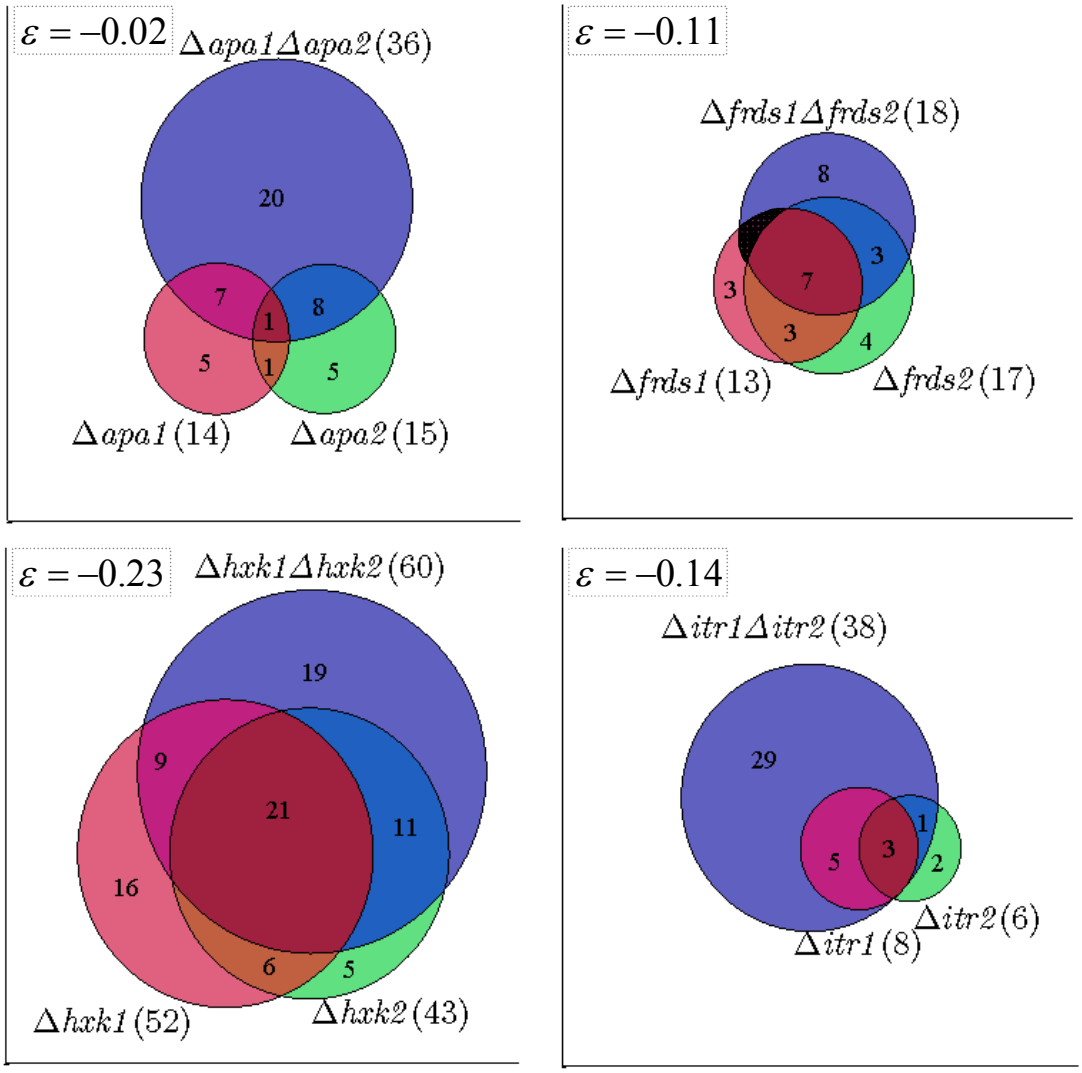

Figure 26 – A Venn diagram describing the amounts of significantly changed metabolites in each of the mutant sets. The labels of the mutants are adjacent to the circles. Circle size and intersections are approximately proportional to the amount of metabolites changed in each group – total amount of significantly varying metabolites in each mutant is given in parentheses (including those with less than two fold change). Epistasis scores (ε) from the DeLuna study for each gene pair is given in the upper left part of each subplot.

If there were no epistasis, the basic expectation for double mutants would be to show response only among metabolites that were changed in the single deletion mutants. I, however, observed an interesting phenomenon. The set of metabolites changed in the double mutants in all cases contained most of the metabolites that were changed in the single mutants and additionally had a relatively very large set of metabolites changing uniquely in the double mutants. The Δ*hxk1*, Δ*hxk2* and Δ*hxk1*Δ*hxk2* strains present a

special case. In these three mutants the overlap between metabolic changes is very large. This can be explained by the fact that each of the single mutants by itself exhibits significantly slow growth, and the double mutant adds upon that phenotype a synergistic effect resulting in a markedly pronounced growth defect (as indicated by the most negative epistasis coefficient ε =-0.23).

Δ*frds1* and Δ*frds2* on the contrary, exhibiting lower epistasis (ε =-0.11), did not exhibit any growth defect in my medium, and that could be the reason for the absence of a strong metabolic response or a marked increase in the amount of changes in the double mutant.

3.3 Metabolic profiling of S. cerevisiae response to various stresses

So far I examined the metabolic response of yeast to genetic perturbations, i.e. single or double deletion mutants. I next wanted to examine the metabolic effect of another type of perturbation – environmental stresses. For that I measured the metabolic response of *Saccharomyces cerevisiae* cells to several different environmental stresses.

The following stresses were chosen for evaluation: ethanol stress (5%), heat shock (37°C), oxidative stress (H_2O_2 0.3 mM), and drug-mediated growth arrest via protein synthesis inhibition using cycloheximide (7 μg/ml)[*]. The exposure to the stress lasted for 60 min. after which the samples were harvested.

The experiments were carried out in a manner similar to the single gene deletion set of experiments. Figure 27 depicts the metabolite changes in different environmental conditions following the standard normalization procedures.

[*] Repeat ethanol and cycloheximide 5 μg/ml experiments were also carried out and showed qualitatively similar results (data not shown).

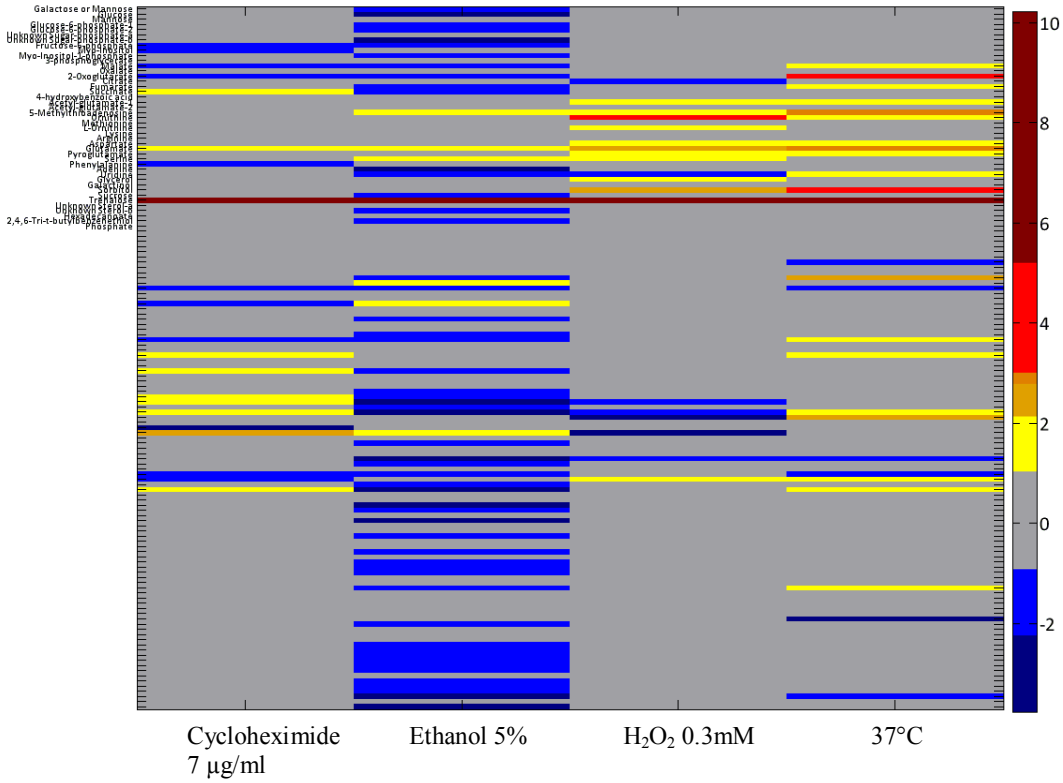

Figure 27 – Cellular response to a 60 minute exposure to different environmental stresses relative to the standard non-stressed conditions. Color scale (log2 scale) and metabolite order are as in Figure 14. Zoom in on known metabolites is in Figure 28

A zoom-in on known metabolites (the upper portion of Figure 27) clarifies the signal (Figure 28).

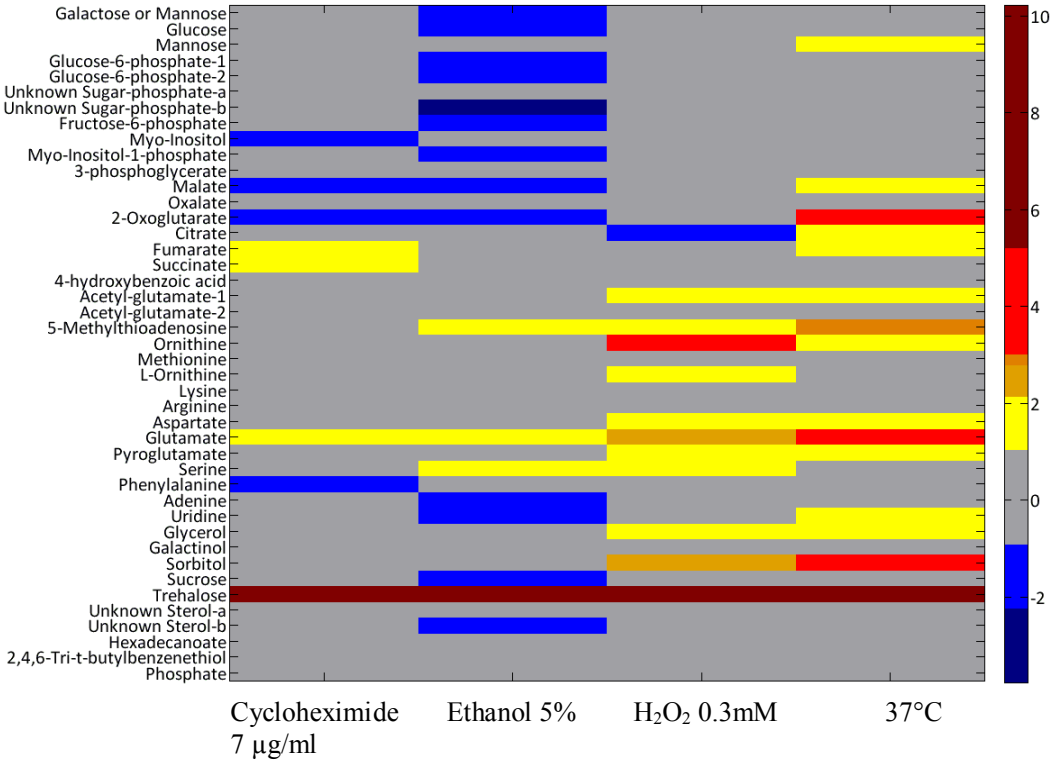

Figure 28 – Cellular response to a 60 minute exposure to different environmental stress relative to the standard non-stressed conditions – Identified metabolites. Color scale (log2) is similar to Figure 14.

Four observations are apparent from the results. One is that the environmental stress response has a wide effect on the metabolome (14% - 35% of metabolites exhibit more than 2-fold change relative to unperturbed strains).

From the correlation matrix of the stress responses (Figure 29) it appears that the cycloheximide stress response is least similar to other metabolic phenotypes.

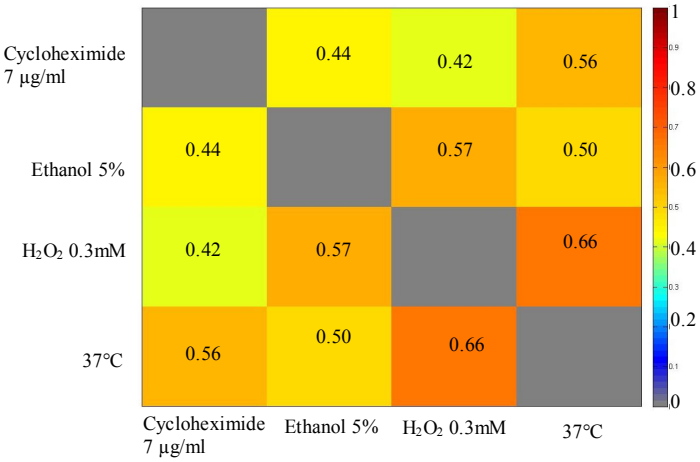

Figure 29 – Pearson correlation coefficient matrix for fold change relative to unperturbed culture in environmental stresses

A third, striking result is the highly elevated level of trehalose (at least 30 fold increase in all applied stresses). Glutamate also shows response in all stresses, albeit to a lesser extent. Serine and 5-methylthioadenosine show an increase in three of the four stresses.

The fourth apparent result is related to the directionality of change in metabolites (the tendency of a stress to cause accumulation of metabolites or depletion thereof).

	Cycloheximide 7 µg/ml	Ethanol 5%	H_2O_2 0.3 mM	37°C
Number of metabolites Increased ↑	12	8	16	28
Number of metabolites Decreased ↓	7	40	5	3

Table 3 – Count of increased or decreased metabolites (at least 2 fold change) in different environmental stresses – including unidentified metabolites

From Table 3 two opposite patterns emerge. On one hand the ethanol exposure caused depletion of multiple metabolites, on the other hand in response to both heat shock and H_2O_2 more metabolites increase rather than decrease. This pattern may be related to growth decrease mediated by ethanol, and to growth increase caused by temperature increase.

See section 4.2 for additional discussion of the environmental response results.

3.3.1 Ethanol stress analysis

In Figure 28, several specific metabolites changing in the 60 minute ethanol stress can be seen. However, I additionally wanted to observe the development of the response to ethanol. Metabolic response can occur either on the level of the proteins already existing in the cell, or via new protein biosynthesis. Metabolites changing first in response to stress are expected to be controlled by existing protein pool, while those responding later are expected to be controlled by stress responsive newly synthesized proteins.

In general, in the 60 minute ethanol-exposed cultures a metabolome-wide decrease in most metabolites can be seen, however, changes in two metabolites are of special interest – a decrease in Glucose-6-phosphate (G6P) levels, and extremely high

trehalose levels (Figure 28); G6P levels decreased ~3 fold, while trehalose levels increased ~226 fold (see section 4.2.1 for further discussion of the role of trehalose and Glucose-6-phosphate).

To analyze the speed of the response of cells to ethanol additional, independent, experiments were carried out with similar levels of ethanol stress, but different durations (10 minutes, 30 minutes and 60 minutes). The 60 minute replicate shows qualitatively similar results to the first 60 min experiment.

Results in Figure 30 and Figure 31 show that, while limited, a metabolic response occurs within the first 10 minutes of exposure to 5% ethanol. In particular (looking at identified metabolites in Figure 31), a decrease in G6P levels is evident across all timepoints. Trehalose levels in the first two experiments remain similar to the unperturbed strain, and rise only towards 60 minutes. Therefore, it would be logical to assume that the decrease in G6P is mediated mainly by existing proteins, while an increase in trehalose is mediated via protein biosynthesis.

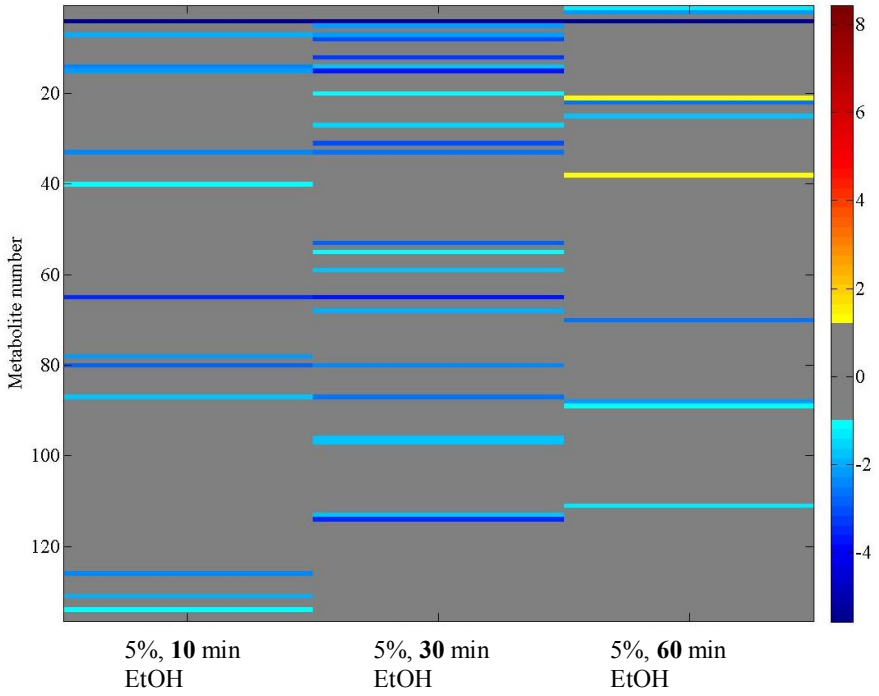

Figure 30 – Cellular response to ethanol 5% exposure of 10 minutes and 30 minutes relative to the metabolic profile of an unperturbed strain. Color scale (log2) and metabolite order are similar to Figure 14. (Results in 60 minutes stress may differ somewhat from those in Figure 27, since the two experiments were carried out with ~1 year difference)

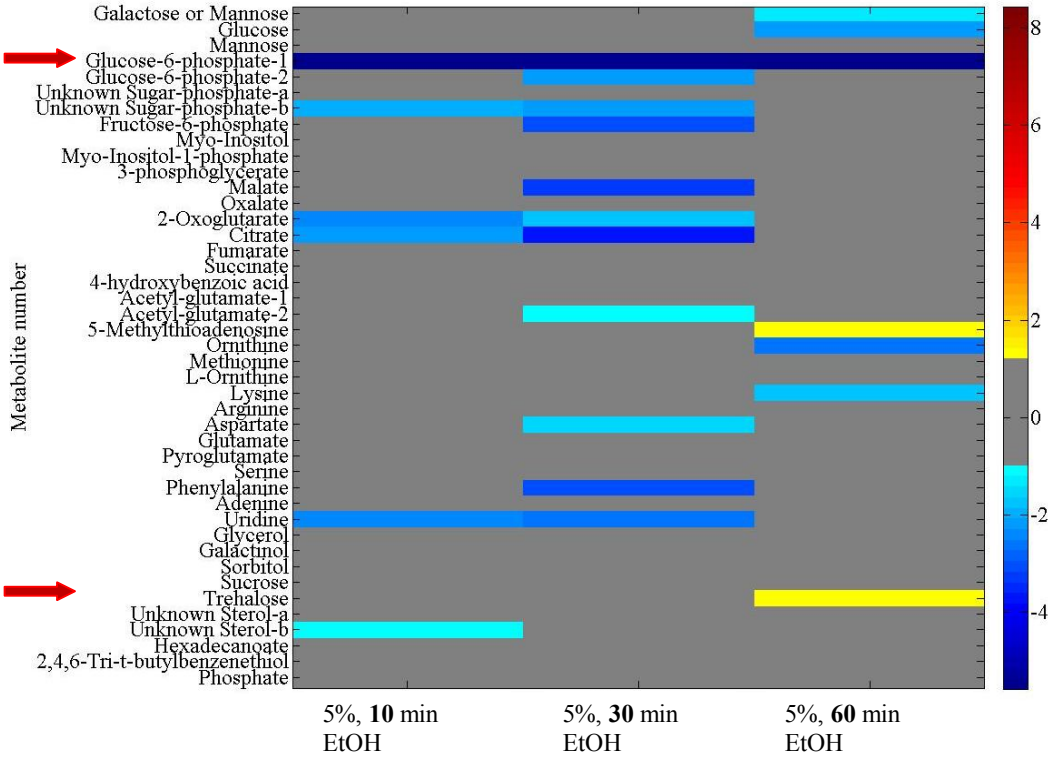

Figure 31 - Cellular response to ethanol 5% exposure of 10 minutes and 30 minutes – Identified metabolites. Glucose-6-phosphate and trehalose are marked with red arrows. Color scale (log2) is similar to Figure 14.

Further discussion of regarding ethanol response can be found in section 4.2.1.

3.4 Metabolic profiling of S. cerevisiae strains evolved to cope with high intensity heat stress

To further analyze the response of yeast cells to stresses I have profiled a strain of *S. cerevisiae* that was evolved to cope with increased heat (see methods section 2.1.2). The metabolic response of evolved cultures and control cultures (that evolved under same conditions, yet permissive temperature of 30°C) to heat stress was assessed. Two levels of stress were used; 39°C exposure that is extremely stressful for non-evolved cultures, while the evolved cultures routinely cope with this stress and exhibit normal growth rates. To check the response of evolved strains to a temperature that represents a heat shock for them – a treatment of 45°C was used too. Additionally, to assess the baseline changes in the metabolome of the evolved strains I profiled the evolved strain at 30°C. Results of the experiments are presented in Figure 32 and Figure 33.

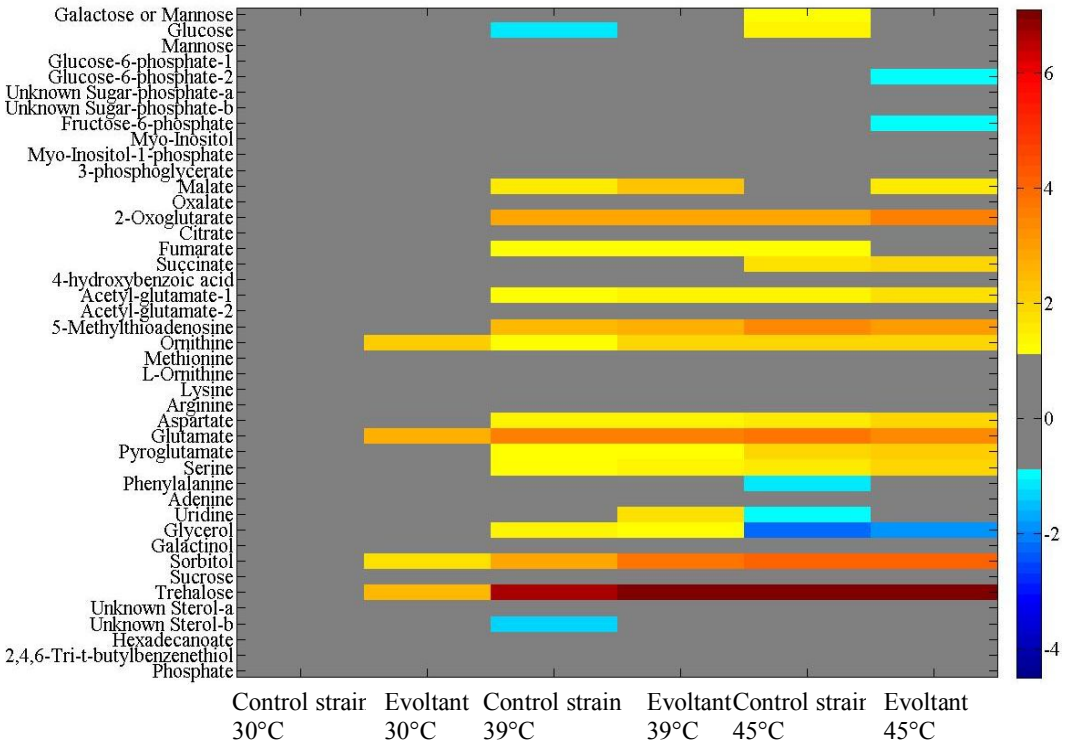

Figure 32 - Cellular response to heat shock of evolved and control strain. Data is given on a log2 scale relative to the values of control strain in 30°C. Color scale (log2 scale) and metabolite order are as in Figure 14. Since the control strain at 30° is normalized to itself, it appears in gray

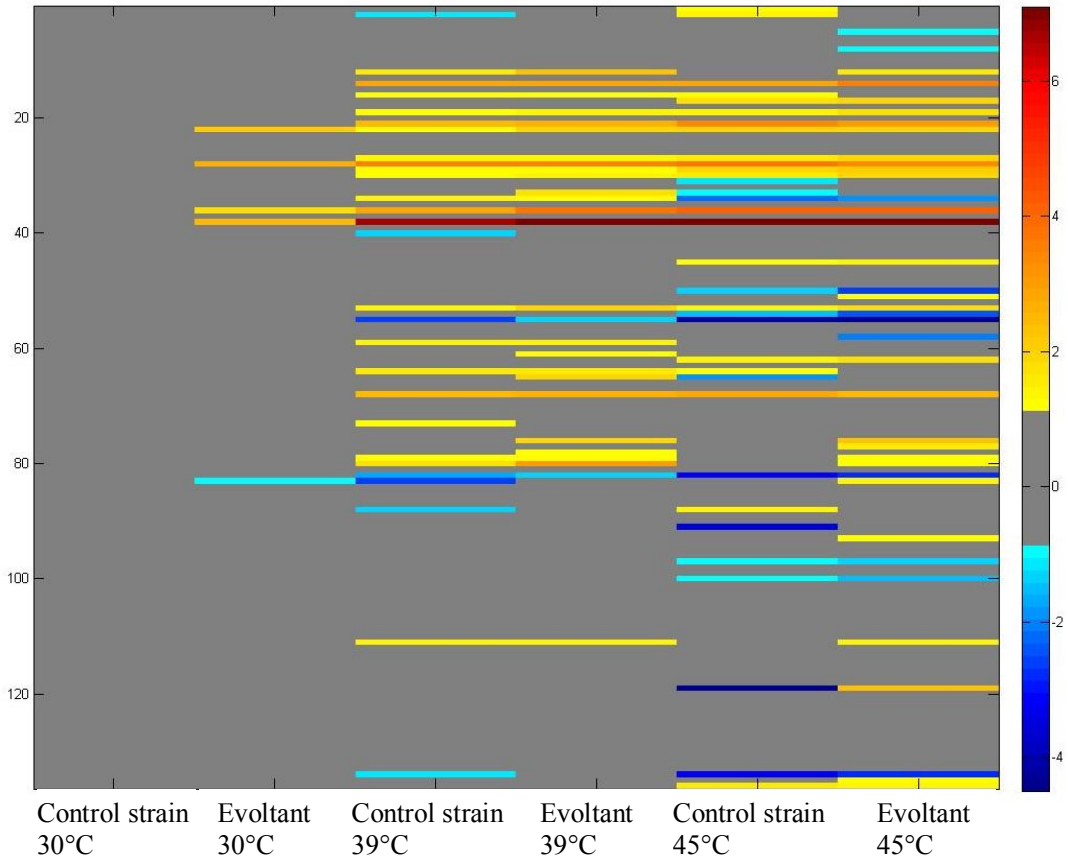

Figure 33 - Cellular response to heat shock of evolved and control strain. Data is given on a log2 scale relative to the values of control strain in 30°C – only identified metabolites. Color scale (log2) is similar to Figure 14. Since the control strain at 30° is normalized to itself, all metabolites in this strain appear in gray

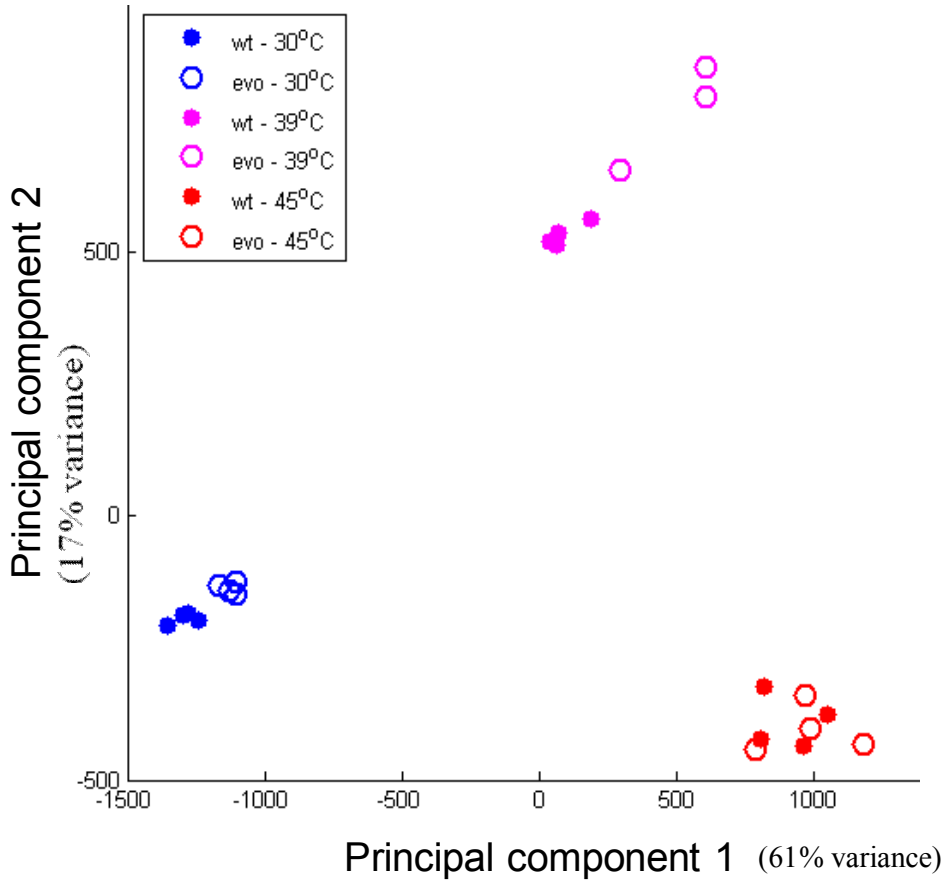

Figure 34 – Plot of the mass fragment signals (markers) projection on the first two principal components, following a principal component analysis of the evolved strain (in blank circles) and the control strain (full circles) response to varying heat stress. Mass signals were log2-transformed before PCA. The first principal component captures 61% of variance in the data, while the second captures 17%.

Looking at the results, several phenomena can be noted.

The growth temperature is the dominant factor that affects the metabolome. At each temperature the evolved strain and the control strain are more similar to one another than to the same strain at a different temperature

Judging from the distance in the space of the first two principal components between control strain at 30°C and the evolved strains (Figure 34), as could be expected, the response of both the evolved and the control strain to higher temperature (45°C) is more pronounced than the response to 39°C

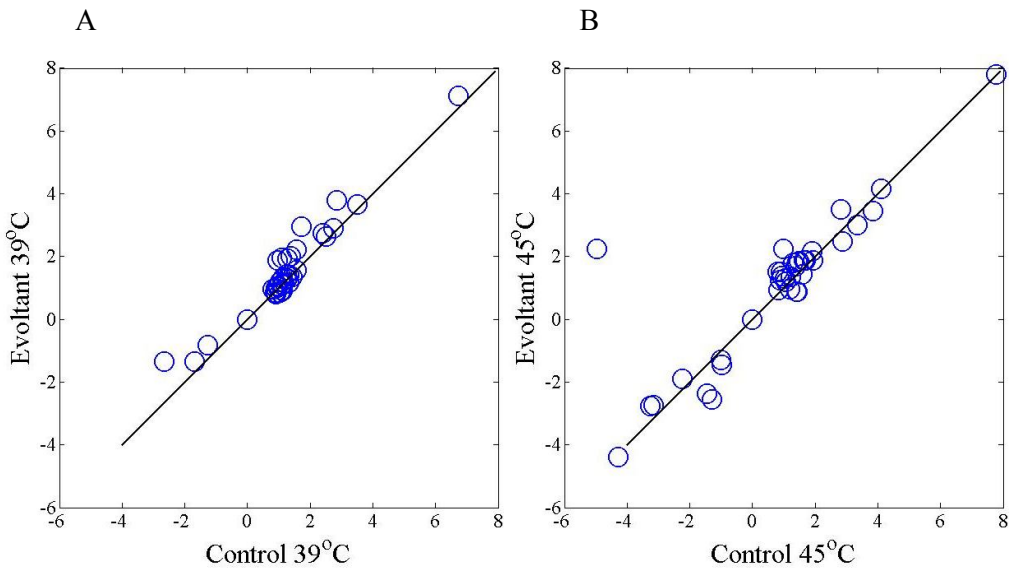

Figure 35 –Dot plot for metabolite changes relative to control in 30°C in control and evolved strains at 39°C (plot A) and 45°C (plot B). Ratios were log2 transformed and non significantly changing metabolites were set to 0. The diagonal (y=x) is plotted in black.

The response of both strains to 45°C exposure is similar (in Figure 35B the metabolites in both strains appear on the diagonal, and the location in the PCA space of the evoltant and the control strain in 45°C is similar in Figure 34), while the exposure to 39°C degrees (Figure 35A) in the evolved strain brings about a stronger response than in the control strain. Most metabolites that increase upon exposure to 39°C exhibit a higher response in the evoltant than in the control.

The baseline metabolome (at 30°C) of the evolved strain is mostly similar to that of the control strain. However, when changes do occur most of them are in metabolites that are strongly responsive to heat stress (Figure 32, Figure 33).

When looking only at identified metabolites (Figure 33), it can be seen that changes in the evolved strain under non-stressed conditions (30°C) are in the following compounds: Ornithine, Glutamate, Sorbitol and Trehalose.

In addition to the metabolome, a transcriptional profiling of the evolved (denoted evo39) and the control (denoted evo30) strains was carried out. Avihu Yona from our lab, who carried out the evolution experiments, subjected the evolved and control cultures to 42°C and 45°C stresses and measured mRNA levels across several timepoints. I analyzed the control strain data after 60 minutes exposure to 42°C and 45°C stresses relative to no heat exposure, as well as two replicates of the strain

evolved under heat shock relative to control strain at 30 degrees (t=0). Results of transcriptome analysis with respect to metabolic genes (as outlined in the iLL672 model[62]) appear in Figure 36.

When comparing the evoltant to the control strain (Figure 36A) it appears that no single gene related to the metabolic changes that I observe is up- or down-regulated in the evoltant (see genes and gene names marked in red on Figure 36A). However, when looking at genes which are related specifically to trehalose metabolism (genes marked with magenta full circles in Figure 36 A-C, Supplementary 3) is appears that as a whole they are significantly (see Methods 2.8 for methodology) up-regulated in the evoltant. Additionally it appears, that those same genes are also significantly up-regulated in the control strain after exposure to heat stress of 42°C and 45°C (Figure 36B and Figure 36C respectively). This leads to the conclusion that in the evolved strain, genes related to a pathway necessary to cope with heat stress have higher levels to begin with. i.e. even before the induction of stress response.

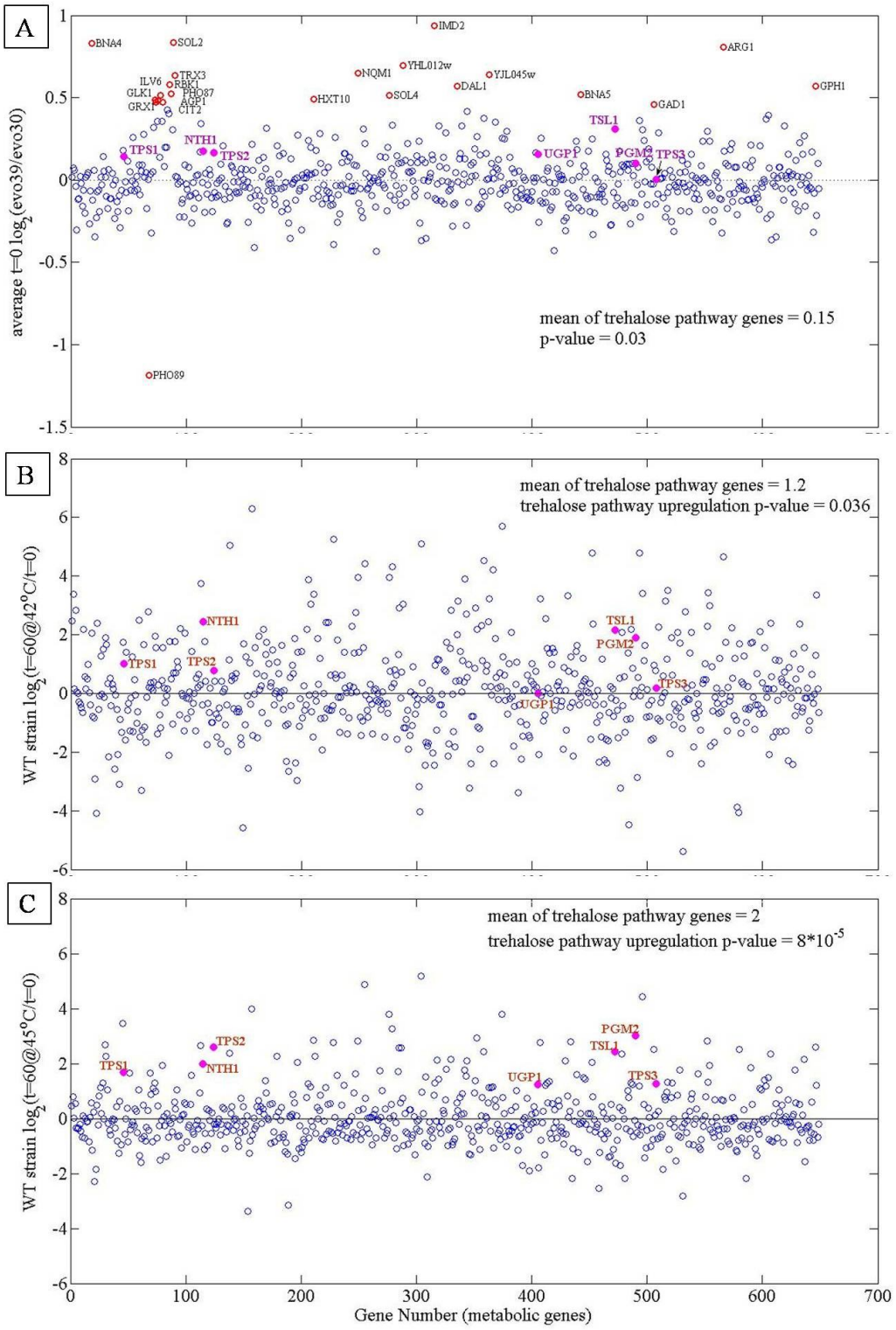

Figure 36 – Transcriptional response of metabolic genes in evolution. A. The average fold change (on log2 scale) of different transcripts in the evolved strain relative to WT. In red circles are marked (and labeled) genes which fluctuate more than 2.5 standard deviations from the mean of all genes expression levels.

Magenta full circles are genes related to trehalose biosynthesis pathway (for full list of genes see Supplementary 3). B. Relative transcript levels @42^{o}C exposure for 60 minutes of control strain. In magenta circles are marked genes related to trehalose biosynthesis. C. Relative transcript levels @45^{o}C exposure for 60 minutes of control strain. In magenta circles are marked genes related to trehalose biosynthesis. On all the figures the mean increase in genes related to trehalose pathway is marked, as well as the p-value of its significance given the number of genes in the pathway (see Methods section 2.8)

Further discussion with respect to the comparison of strains evolved under heat stress regime can be found in Section 4.3.

4 Discussion

According to the Central Dogma, the genome is the blueprint for cellular functions, the cellular functionality is mostly carried out via proteins, with the transcriptome serving as the mediator between the two levels. Ultimately, the function of many of the proteins, predominantly of enzymes and transporters, is to maintain cellular metabolites at desired levels and to ensure cell proliferation.

The metabolome, i.e. the complement of all the small intracellular molecules is affected both by the environment and by the genetically encoded proteins. Therefore, it is the first and the most direct manifestation of the cellular phenotype; the complex of the traits which is determined at the meeting point of the environment and the genotype.

In my thesis I have set out to study on one hand the effect genetic changes on the most basic manifestation of the phenotype (the metabolome), and on the other hand the effect of the environment on the metabolome.

To carry out the analysis of yeast metabolome I developed a robust method which allows multiple sample analysis in relatively short periods of time (Results Section 0). The method showed repeatability and relative stability of experimental results over long periods of time.

4.1 Analysis of deletion mutants of paralogous genes

One of the commonly occurring evolutionary processes in genomes is gene duplication. The effect of gene duplication on cell survival and transcriptome is quite will studied, however, the effect on the metabolome has not been explored. I applied

the newly developed method for the analysis of 78 deletion mutants representing 39 pairs of paralogous genes, which are known to be related to metabolism in yeast.
To begin with, I conducted analysis of several specific deletion mutants mainly as a control for the experimental method. Two observations were apparent from the analysis of *Δaco1*, *Δaco2*, Δ*nth1* and Δ*nth2*. First, the metabolites that significantly change upon deletion of these genes are often related to the pathway in which the deletion was carried out. Second, as could be expected, the deletion of the more active paralog in the checked conditions (*ACO1* and *NTH1*) brought about more changes than the deletion of the less active counterpart.
Next step was to analyze the group of paralogous gene deletions as a whole. Three major types of relative metabolic changes could be noted – small changes upon deletion of each of the two paralogs, "concerted" changes and one-sided changes. In the latter, deletion of one paralog causes many metabolic changes, while the deletion of the other leads to a few changes.

4.1.1 **Paralogous gene pairs that exhibit low or no metabolic response upon deletion**

As could be seen from Section 3.2.1.4 (p. 40, Figure 15), one of the phenotypes to deletion of paralogs was that little that or no significant metabolic changes occurred in response to the deletion. There are two possible explanations for such pairs. One is that the genes that were deleted have no effect on the metabolome, at least in the checked conditions, e.g. because of repression (for instance *DLD1* and *DLD3* which both play a function in lactate metabolism, but are repressed upon growth on glucose). The second explanation for no apparent effect can be that paralogs provide good backup to each other, i.e. that the genes function in a similar manner.
A definitive test between the two hypotheses is true, would be to measure the fitness of double mutants. According to the first hypothesis the double mutant should be viable, while, if the second hypothesis is true the double mutant would suffer from the absence of backup, and exhibit low fitness. Since creating double mutant strains is a laborious procedure and would be out of scope of current dissertation, a possible correlative key for deciding between the two hypotheses may be found in the rate of evolutionary divergence of paralogous gene pairs. Selective pressure toward similar functionality of genes (purifying selection) that exhibit little metabolic change would

favor the second hypothesis, as non-functional genes are expected to be evolving independently of the amount of metabolic changes.

In paragraph 3.2.1.4.1 (Figure 16) a positive correlation was found between the Ka/Ks ratio of paralogous pairs and the mean amount of metabolic changes upon the deletion of each of the members. This positive correlation implies that the amount of changed metabolites is associated with purifying selection towards keeping paralogous pairs similar. Upon deletion of genes that show slow functional divergence from one another there are less metabolites changing. This finding supports the second hypothesis – regarding the similarity of the genes. I propose that the effect of little or no metabolic change might be due to well carried out backup between paralogs, which were selected to maintain similar functionality, rather than due to the occurrence of non-functional genes. The above is also in agreement with Kafri and colleagues[22-24] who suggested that paralogous gene pairs are selected to retain some degree of similarity in their activity to allow for better functioning of the cells e.g. with respect to molecular noise.

As to the pairs that showed relatively low purifying selection with few metabolic changes upon deletion (red circled gene pairs in Figure 16) I can hypothesize that genes in this part of plot are not strongly active in the conditions I checked (high glucose).

Indeed, when looking at their identity I notice that the *DLD1* (which is repressed by glucose) and *DLD3* (whose expression is driven by mitochondrial damage) gene pair is in that group. Another example is the *DAK1* and *DAK2* paralogous pair responsible for the dihydroxyacetone kinase activity. These genes are too mostly inactive in glucose rich unstressful conditions (BIOBASE knowledge library, accessed 31/08/2009); the deletion mutants for these two genes can be found too in the red circled region. Given their relatively high Ka/Ks ratio it is possible that in stressful conditions or upon growth on glycerol deletion of these genes will produce a stronger metabolic response.

4.1.2 **Paralogous gene pairs that exhibit "concerted" metabolic response upon deletion**

Concerted changes appear to be encountered quite often upon the deletion of paralogs (see section 3.2.1.4.2). This phenotype was significantly shown in 23 out of the 39 gene pairs I have analyzed.

At least two possible reasons may contribute to the observed phenotypes. The first is the dosage effect which assumes that the two paralogs perform a common reaction and both are required to provide the full flux through the reaction and synthesize the full amount of the downstream product. In case of deletion of any of these paralogs the flux through the reaction will be diminished and the downstream effects will be similar for both of the paralogs (mainly increase in upstream metabolites). Another possible explanation is that the two paralogous genes belong to the same protein complex. Without either of them the complex is destabilized or its formation happens with lower probability. Therefore the flux through the reaction is diminished and the downstream effects are similar. However, as most of the checked pairs are not part of known complexes this second hypothesis is rather unlikely.

Unfortunately, not enough data exists to unequivocally support the first hypothesis either. Direct experimental measurements of metabolic fluxes would be very helpful.

4.1.3 **Paralogous gene pairs exhibiting one-sided metabolic response upon deletion**

One sided metabolic response in my data is a situation in which one of the deleted paralogs exhibits significantly more changes than its peer. My results show that 8 out of 39 paralogous pairs exhibit significant one-sided metabolic response (results section 3.2.1.4.3)

I can offer two plausible hypotheses explaining the observed one-sided response of paralogs. One is existence of good backup by transcriptional reprogramming of one paralog, but not the other[22]. Another explanation may lie once again in non-functionality of one of the checked genes in my experimental conditions. For example, the glutamate dehydrogenase gene *GDH3* is known to be repressed by glucose rich conditions, while its paralog (*GDH1*) is active in glucose rich conditions[63]. This is in good correspondence with the absence of metabolic response I observed in the Δ*gdh3* mutant, and a strong response upon deletion of Δ*gdh1*.

Moreover, it was experimentally shown that deletion of *GDH3* does not cause any change in the levels of *GDH1* (Michael Springer, personal communication). Such a change could have caused an increased flux through the reaction and might have resulted in some change in the metabolic phenotype.
Another pair with one-sided phenotype is the *HXK1* and *HXK2* pair. Double mutants Δ*hxk1*Δ*hxk2* cannot ferment glucose, i.e. the two genes provide backup for each other[64]. *HXK2* encodes the major isoform of hexokinase that is required for glycolysis. Its expression represses *HXK1*. Upon deletion of *HXK2* though, *HXK1* is derepressed[65]. My results show that indeed under glucose rich conditions the mutant Δ*hxk1* did not exhibit a strong metabolic phenotype (second mutant pair in Figure 14) probably due to its repression in my experimental conditions. *HXK2* deletion on the other hand, produced a few metabolic changes. This phenomenon might be explained by non-perfect backup of Δ*hxk2* by the upregulated *HXK1*. Interestingly, this pair of mutants has a Ka/Ks ratio of 0.2 (Figure 16). I can hypothesize that non-full backup of the Δ*hxk2* mutant is among other things due to a relatively high divergence rate between the two paralogs.

4.1.4 **The state of the metabolome is a predictor of growth rate decline**

Some deletion mutants that I profiled showed slowed growth rates (results section 3.2.1.5). I conjecture that the metabolome on one hand affects the growth rate of mutant, and on the other hand is affected by the growth rate. Therefore, it was logical to expect that the metabolome carries information regarding the growth rate.
No specific metabolite was found to correlate with growth rate of deletion mutants. However, I could observe that growth rate could be predicted to a limited extent based on the combined levels of all the metabolites. Several metabolites, such as ornithine, often appeared in the successful predictor solutions. Such metabolites might be those that mediate the relationship between metabolite levels and growth rates in *S. cerevisiae* mutants. However, the significance of this finding was hard to assess statistically. This finding requires future study and analysis to establish the causality of the phenomenon to pinpoint the role of specific metabolites in determining the growth rate.

4.1.5 **Metabolic response to deletion is a predictor of the cellular localization of the deleted gene**

It is plausible to assume that deletion of a gene in a particular subcellular location e.g. mitochondria will cause accumulation or decrease in metabolites that belong to that specific organelle. Indeed both anecdotal observations and a large-scale prediction of cellular localization based on the metabolic level (section 3.2.1.6), show that metabolites are connected and have predictive power with respect to the cellular location of the deleted gene. While the quality of subcellular location prediction using metabomolome changes is not very high (as compared to random), it nevertheless allows enrichment of the annotation of deleted genes.

While in *S. cerevisiae*, high-throughput studies have already pinpointed the cellular localization of most genes[66], metabolomics opens a possibility to uncover the localization of genes in other species.

4.1.6 **Analysis of double mutants in *S. cerevisiae***

In order to further elucidate the behavior of paralogous gene pairs in *S. cerevisiae*, I carried out metabolic profiling on selected double mutants in paralogous genes. Overall I could see that metabolomic analysis of double mutants carries some potential for understanding the mechanisms underlying the epistasis of genes, however, a fuller, more complete, understanding and decryption of metabolic profiles requires more accurate results and larger data sets. It is also possible, that a clearer signal would be available, if I had growth rate measurements in mineral medium, the conditions in which all my metabolomics experiments were performed. Even though the DeLuna data set is based on minimal medium, the yeast nitrogen base used for that study, is more complex than the defined mineral basis for the mineral medium.

4.2 The yeast metabolome is highly responsive to environmental stimuli

After gaining an insight into the response of the metabolome to genetic perturbations. I set out to address the second variable affecting phenotype – the environment. To that end I profiled the metabolome after exposure to several different environmental stresses.

It was shown that yeast exhibit a similar transcriptional response to an array of stresses[9,10]. Several transcription factors were shown to be responsible for the observed similarity, namely the YAP gene family[25,67].

The results from my experiments (Paragraph 3.3) lead to several conclusions; some are basic, while others are not as apparent. First of all, the results showed that the environmental perturbations lead to an extensive response on the level of metabolites. The high responsiveness of the metabolome to the environment allows cells to achieve high plasticity. A simultaneous modification of the levels of multiple metabolites permits cells to achieve optimal response to external conditions.

Another observation was that there are stresses which are similar to each other in terms of response to the environmental perturbation (specifically the oxidative stress and heat shock). Interestingly, it was previously shown that hydrogen peroxide and heat shock invoke very similar transcriptional stress response in yeast cells and that similar signaling molecules mediate the information needed for the response[9,68,69].

High elevation in trehalose levels will be discussed below, and can perhaps be better understood by a more particular ethanol stress analysis (Paragraphs 3.3.1 and 4.2.1).

4.2.1 Ethanol stress in *S. cerevisiae* induces high levels of trehalose to cope with substrate-accelerated toxicity

Ethanol stress that was used in my study is a well-studied stress to which cells are routinely exposed. During yeast fermentation ethanol is a major by-product of energy metabolism. Ethanol is toxic for most organisms, as it disrupts cellular membranes, exerts osmotic stress on cells and may inhibit enzyme function within the cells. While yeasts can better cope with ethanol stress than most other organisms, they too suffer from its adverse effects[70-72].

Cells respond to ethanol first by growth arrest, due to either cell volume shrinkage or loss of turgor pressure due to water leakage through osmosis. This process is very fast and is usually completed within a minute. After that, cells develop a secondary response to stress by synthesis of membrane-stabilizing compounds, osmo-protective compounds and heat shock proteins. One of the dangers of growth arrest in yeast cells, is that it may cause adverse effects due to substrate accelerated toxicity[73]. Substrate accelerated toxicity in glycolysis is caused by an imbalance between the production and degradation of hexose-phosphates. Production of glucose-6-phosphate and fructose-1,6-bisphosphate from glucose in the early stages of glycolysis utilizes

ATP. Insufficient degradation of these sugar-phosphates due to inhibited growth and "stuck" downstream pathways may lead to accumulation of the two intermediates, sequestering of phosphate and its depletion[74]. This paradoxically causes eventual cell death by starvation.

Yeast response to ethanol was investigated to some extent on the level of the transcriptome[12]. Several works were presented on the level of the metabolome[75,76] in which the authors mainly looked at processes associated with long term fermentation by *S. cerevisiae*. In these studies the changes in the extracellular environment are gradual, and only the late fermentation stages are subject to a significant ethanol stress. It is not possible to examine through such experiments the short-term response of *S. cerevisiae* to ethanol, which is associated with fast cellular accommodation to significant pressure from the environment.

It was shown that the transcriptional response to low concentrations of ethanol[12] and to other stresses[9,10] peaks ~30 minutes after exposure to stress. It was also shown that ~7% of the yeast transcriptome exhibits significant changes after ethanol exposure[12], whereas I show that the metabolome exhibits a much wider effect in response to similar levels of ethanol stress (~40% of the metabolome changes in response to the stress).

Noticeably, in my experiments the levels of most metabolites are lower in stress than in standard conditions. This might be the result of decrease in overall metabolism of yeast cells, and the growth arrest that is the result of stress.

The exception to this rule, however, is the extremely elevated level of trehalose (~230 fold increase in ethanol-exposed cells relative to standard conditions).

It is known that trehalose and genes related to its metabolism increase in ethanol stress[12] among other stresses[77]. It was also suggested that biosynthesis of trehalose benefits yeast cells in coping with external stress not only by reducing the direct effects of stress (e.g. osmotic pressure), but mainly by introduction of an energetic futile cycle to cope with the substrate accelerated toxicity[74,78] mentioned above. Trehalose biosynthesis constitutes a futile cycle beginning with glucose-6-phosphate and ending with glucose (Figure 37). It was also shown that all genes that constitute this cycle are upregulated upon exposure to ethanol[12].

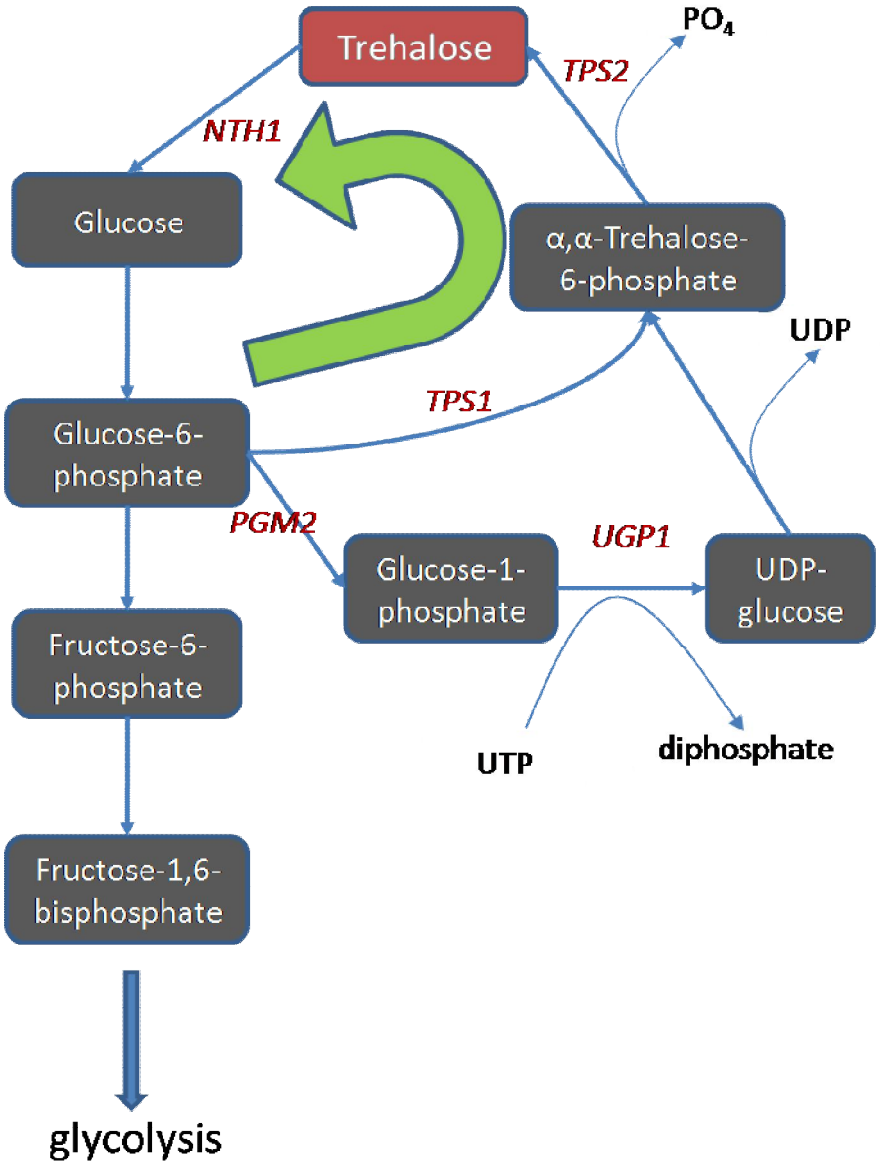

Figure 37– Metabolic pathways near the beginning of glycolysis in yeast cells, along with existing and putative futile cycles. Boxes depict metabolites, arrows depict enzymes. The names near the arrows are names of genes that encode relevant proteins. All the genes that appear on the plot were shown to be upregulated after 30 minutes of 7% ethanol stress[12]. Trehalose (elevated levels) is marked in red. Glucose 6-phosphate is marked in blue. The wide green arrow depicts the futile cycle that was suggested to function for sequestering of phosphate upon osmotic stress.

This cycle depletes the intracellular stocks of glucose-6-phosphate and allows the regeneration of phosphate molecules.

Indeed the metabolome results showed a decrease in the levels of glucose-6-phosphate upon exposure to stress.

Fast biosynthesis of trehalose with lagging catabolism thereof, may lead to such an effect. The relative concentrations of the two molecules are interdependent as expected. The correlation between the transcript elevation and the changing metabolite levels elucidates the regulatory mechanism behind the futile cycle. It is logical to suggest that the regulation of the cycle's proteins is mostly on the transcriptional level rather than on the post-transcriptional levels. However, the very fast metabolic response to ethanol hints, that some changes are effected even before the initiation of the transcriptional response.

To settle this issue unambiguously, however, detailed analysis of the dynamics of response is needed, involving higher frequency of measurements, flux analysis and simultaneous transcriptome analysis.

4.2.2 Trehalose increase can be mediated both by the sensing of the extracellular environment and by deficient cell growth

The stresses that I tested vary in the mechanisms of cellular sensing as well as in their effect on the growth rate (Table 4). While both ethanol and H_2O_2 lead to a decline in growth rate and are sensed by cells, the cycloheximide stress greatly decreases growth rate[79] (repeated in our lab - data not shown), but is not expected to be sensed by yeast directly and evoke an environmental stress response. The 37°C stress on the other hand, actually increases growth rate by ~10%[80] (and data not shown), while being sensed and invoking an environmental stress response[9,81].

	Growth arrest	Extracellular stress sensing
Cycloheximide 7 µg/ml	**Yes**	**No**
Ethanol 5%	**Yes**	**Yes**
H_2O_2 0.3 Mm	**Yes**	**Yes**
37°C	**No**	**Yes**

Table 4 – Division of stresses to those with growth arrest and with apparent sensing of the stress in the environment.

Therefore, the elevated trehalose levels found in all stresses (Paragraph 3.3) constitute a surprising finding. The trehalose response is invoked in all cases, whether by the

sensing of the extracellular signal or by the decline in the growth rate. This suggests that trehalose is a major stress-responsive molecule regulated on multiple levels – both by sensing the environment **and** by growth rate decrease.

Observing trehalose levels across the full experimental dataset (Figure 38) it is apparent that the mutant with the largest positive change (>70 fold change) in trehalose is Δ*nth1* which was already mentioned earlier (paragraph 3.2.1.3 on p. 37).

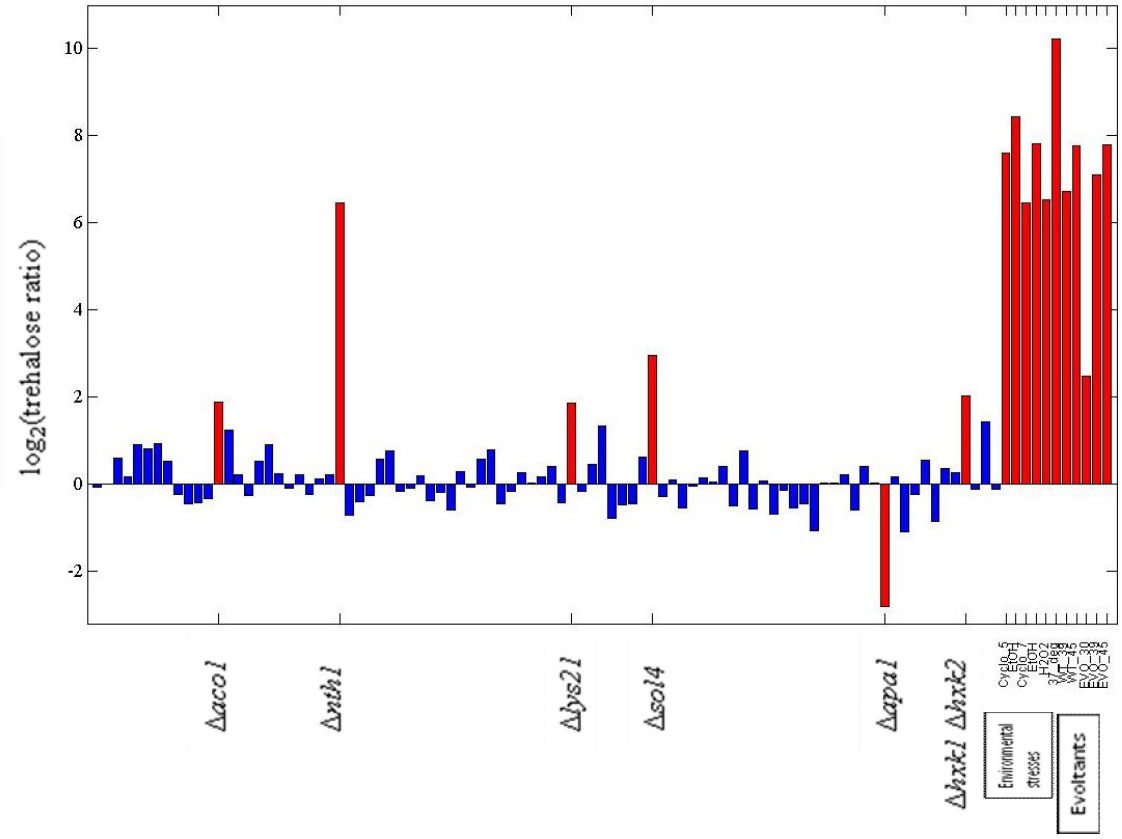

Figure 38 – Trehalose changes relative to control (on log2 scale) across all the collected experimental datasets. To avoid cluttering of the x-axis only mutants/conditions in which trehalose exhibited more than three-fold change are labeled and the respective bars colored in red. On the right side of the plot appear the environmental stresses and stress exposure experiments of heat-adapted evoltant strains

Despite the increased trehalose levels in Δ*nth1*, it was previously shown that this mutant has a growth deficiency when exposed to heat shock[82]. This leads to a conclusion that trehalose cannot function by itself as a protective compound in stress, but its pathway must be activated in stress response. Thus additional support is

provided to the role of trehalose in the metabolic futile cycle which is needed to cope with substrate-mediated toxicity.

In contrast to Δ*nth1*, Δ*apa1* deletion mutant exhibits a decrease in intracellular trehalose. Apa1p is responsible for the catabolism of polyphosphates[83]. Both the polypohosphate hydrolysis and trehalose pathway activation are responsible for regulation of intracellular phosphate levels. Therefore, it is possible that the decrease in polyphosphate degradation is linked to trehalose biosynthesis to regulate the levels of phosphate.

4.3 Evolution towards heat resilience elevates the steady-state levels of metabolites needed to cope with the stress

To complement the study of the genetic and the environmental effect on the metabolic phenotype I have carried out profiling of a strain that was evolved in laboratory conditions towards 39°C heat shock resilience (see Paragraph 3.4).

Several interesting observations could be noted.

When a higher response is needed, the acquired higher metabolic response to stress does not scale up. The similarity in the response of the control and of the evoltant to 45°C vs. higher response of the evoltant at 39°C suggests that the evolution on the metabolome is directed towards a particular environmental response – towards a particular temperature.

In the evoltant - the higher levels in steady state of metabolites that are later increased in stress and the higher levels of these compounds in evoltant in the 39°C stress hints to the necessity of "early preparedness" to a stress as a mechanism to cope with it.

Transcriptional state of genes related to the biosynthesis of stress-related compounds (specifically trehalose) exhibits the same phenotype – a steady state higher level of the transcripts necessary for heat response in the evolved strain.

An alternative mechanism, which theoretically can be more beneficial, would be a faster response in needed molecules in strains that were evolved towards resilience to stress. However, probably rewiring of the transcriptional/translational modules in such a way is less feasible from an evolutionary point of view.

Amusingly, an essentially similar mechanism has evolved in human metabolism. Caucasians are more adapted to ethanol than other populations[84,85]. The basal levels of alcohol dehydrogenase in Caucasians is can be higher than in other populations due to

increased gene expression[86-88]. Perhaps, over relatively short term evolutionary scale the easiest solution is to increase basal activity of metabolic enzymes (such as *ADH* in humans or trehalose pathway enzymes in *S. cerevisiae*).

4.3.1 Stress-related compounds are elevated in a strain evolved to cope with heat

As noted in paragraph 4.2 the rise in trehalose concentration which exists in the evolved strain is common to many stresses. The elevated levels of the other compounds, on the other hand, are quite interesting. Sorbitol was shown to often function similarly to trehalose in yeast cells[89]. Genes in the Glutamate pathway were shown to increase transcription upon exposure of cells to mild heat shock[90]. It is quite possible that ornithine levels subserve this mechanism – Ornithine can be converted to Glutamate by Car2p and Put2p in *S. cerevisiae* via Glutamate-5-semialdehide[53].

Once again the coordination between metabolic response and transcriptional response suggests that response to environmental stress is via a transcriptional mechanism rather than major post-transcriptional effects.

Summary

In my thesis I strived to explore the effect of the genotype and the environment on the most basic manifestation of the cellular phenotype – the metabolome. The phenotype is shaped from the interaction between the environment and the genotype. Therefore, I investigated the effect of different genotype and environmental perturbations on the cellular metabolome.

First, I developed an analytical method that allowed simultaneous quantitation of over one hundred metabolites in multiple experiments in *S. cerevisiae*. The method showed sufficient reproducibility for the experimental needs.

As a first biological control for the method I profiled the deletion mutants of genes *NTH1*, *NTH2* and *ACO1*, *ACO2* genes. The metabolome analysis of the pairs of mutants produced changes in pathways pertaining to the respective genes.

Afterwards, I analyzed multiple deletion mutants in genes with paralogous pairs (78 deletion mutants representing 39 pairs of genes). Several observations could be noticed. Among different deletion mutants the amount of significantly changing metabolites relative to the WT strain varied widely. Few metabolic changes were observed when the genes were non-functional or not expressed in the conditions of the study. Paralogous pairs with multiple changes upon deletion showed relatively high levels of purifying selection (i.e. selective pressure to maintain similar functionality of the genes). This result suggests that relatively to other genes, paralogs with high metabolic impact are slower to evolve. Possibly multiple metabolic changes are detrimental and interfere with protein evolution.

When comparing the response to deletion of the different paralogs two phenotypes emerge. The metabolic response to deletion of each of the paralogs may be one-sided and accompanied by complementation of one but not the other paralog, or it can be "concerted" and lead to a similar metabolic impact upon deletion of each of the two paralogs. Concerted changes may occur in cases when paralogs are retained in evolution to increase flux through a specific reaction, therefore deletion of each of them brings about a similar phenotype. One sided changes on the other hand, occur either when one of the paralogs is non-functional in experimental conditions or is not able to provide backup for its peer due to high functional divergence.

In general, the metabolic phenotype of deletion mutations is somewhat predictive of the cellular localization of the deleted gene and can be related to the decrease in growth rate.

The metabolome of double mutants, as could be expected, showed more changes relative to each of the single mutants. However, the relationship between the metabolic profiles of single or double deletion mutants and the epistasis associated with the affected pathway could not be inferred unequivocally from the experimental results. Additional experiments are needed to elucidate this intricate association.

To execute the second part of the study of the metabolic phenotype, I have next set out to observe the impact of the environment on the metabolome. I analyzed the effect of several environmental conditions on the metabolic profile of yeast cells.

All environmental stresses showed a consistent increase in several compounds, trehalose being the most notable of them. The change in multiple compounds in response to environmental stress may point to a general mechanism of cell protection from environmental changes. When analyzing particularly the metabolome in response to ethanol, it is plausible that the increase in trehalose is a mechanism to cope with substrate accelerated toxicity. The dynamics of ethanol stress response development suggests there are some metabolites that react to stress via the complement of the proteins already available in the cell. The major metabolic stress response (including trehalose increase), though, is executed after a time period sufficient for protein synthesis. Judging by the diversity of stresses, a response can be induced both via an active sensing of hostile environment and via passive sensing of cellular growth arrest.

In the last set of experiments I set out to explore the relationship between the genome and the environment. Results from examination of a strain evolved to cope with high temperature stress suggest that in the short term evolution acts to increase the steady state of metabolites that serve as protective compounds in stress response, this can be carried out perhaps by a new transcriptional steady-state of genes necessary for coping with the stress.

Overall, this doctorate contains the largest metabolome study available to-date in *Saccharomyces cerevisiae*, one of the major model organisms in functional genomics. Protocols and results from this study can be used for future investigation of the metabolome and genomic studies.

Supplementary data

1) The following compounds could be identified in metabolomics experiments and classified to different compound classes:

Compound	Class
Galactose	Sugar
Glucose	Sugar
Mannose	Sugar
Glucose-6-phosphate	Sugar-phosphate
Unknown Sugar-phosphate-a	Sugar-phosphate
Unknown Sugar-phosphate-b	Sugar-phosphate
Fructose-6-phosphate	Sugar-phosphate
Myo-Inositol	Inositol derivatives
Myo-Inositol-1-phosphate	Inositol derivatives
3-phosphoglycerate	Organic acids and organic acid phosphates
Malate	Organic acids and organic acid phosphates
Oxalate	Organic acids and organic acid phosphates
2-Oxoglutarate	Organic acids and organic acid phosphates
Citrate	Organic acids and organic acid phosphates
Fumarate	Organic acids and organic acid phosphates
Succinate	Organic acids and organic acid phosphates
4-hydroxybenzoic acid	Amino acid precursors
Acetyl-glutamate	Amino acid precursors
5-Methylthioadenosine	Amino acid precursors
Ornithine	Amino acids
Methionine	Amino acids
Lysine	Amino acids

Arginine	Amino acids
Aspartate	Amino acids
Glutamate	Amino acids
Pyroglutamate	Amino acids
Serine	Amino acids
Phenylalanine	Amino acids
Adenine	Nucleobases
Uridine	Nucleobases
Glycerol	Sugar Alcohols
Galactinol	Sugar Alcohols
Sorbitol	Sugar Alcohols
Sucrose	Complex sugars
Trehalose	Complex sugars
Unknown Sterol-a	Sterols
Unknown Sterol-b	Sterols
Hexadecanoate	Fatty acids
2,4,6-Tri-t-butylbenzenethiol	Other
Phosphate	Other

2) The following mutants were used in the study of paralogous deletions:

GPM2 GPM1

HXK1 HXK2

PDC5 PDC6

TDH1 TDH2

DLD1 DLD3

GAL1 GAL3

ACO1 ACO2

IDP2 IDP3

FRDS1 FRDS2

MDH2 MDH3

FKS1 FKS2

GSY1 GSY2

NTH1 NTH2

PMT2 PMT3

CDA1 CDA2

INP52 INP53

ITR1 ITR2

BAP2 BAP3

SUL1 SUL2

ALD2 ALD3

ARO3 ARO4

ASN1 ASN2

GDH1 GDH3

LYS20 LYS21

SAM1 SAM2

SER3 SER33

PRS2 PRS4

SOL3 SOL4

ADE16 ADE17

APA1 APA2

URA7 URA8

ARE1 ARE2

CSH1 SUR1

DAK1 DAK2

GPX1 GPX2

HMG1 HMG2

YSR2 YSR3

FET3 FET5

TRX1 TRX2

3) The following genes are related to trehalose biosynthesis reactions:

Genes: TPS1, TPS2, TPS3, TSL1

Reaction: UDP-Glucose+Glucose-6-phosphate =>Trehalose-6-phosphate+UDP

Gene: TPS2

Reaction: Trehalose-6-phosphate => trehalose + PO_4

Gene: PGM2

Reaction: Glucose-6-phosphate<=>Glucose-1-phosphate

Gene: UGP1

Reaction: Glucose-1-phosphate+UTP=>UDP-Glucose+2Pi

Gene:NTH1

Reaction: Trehalose=>Glucose

References

1. Giaever, G. et al. Functional profiling of the Saccharomyces cerevisiae genome. *Nature* **418**, 387-91 (2002).
2. Steinmetz, L.M. et al. Systematic screen for human disease genes in yeast. *Nat Genet* **31**, 400-4 (2002).
3. Wagner, A. Robustness against mutations in genetic networks of yeast. *Nat Genet* **24**, 355-61 (2000).
4. Wagner, A. Distributed robustness versus redundancy as causes of mutational robustness. *Bioessays* **27**, 176-88 (2005).
5. Gu, Z. et al. Role of duplicate genes in genetic robustness against null mutations. *Nature* **421**, 63-6 (2003).
6. Papp, B., Pal, C. & Hurst, L.D. Metabolic network analysis of the causes and evolution of enzyme dispensability in yeast. *Nature* **429**, 661-4 (2004).
7. Thatcher, J.W., Shaw, J.M. & Dickinson, W.J. Marginal fitness contributions of nonessential genes in yeast. *Proc Natl Acad Sci U S A* **95**, 253-7 (1998).
8. Nielsen, J. & Oliver, S. The next wave in metabolome analysis. *Trends Biotechnol* **23**, 544-6 (2005).
9. Gasch, A.P. et al. Genomic expression programs in the response of yeast cells to environmental changes. *Mol Biol Cell* **11**, 4241-57 (2000).
10. Causton, H.C. et al. Remodeling of yeast genome expression in response to environmental changes. *Mol Biol Cell* **12**, 323-37 (2001).
11. Godon, C. et al. The H2O2 stimulon in Saccharomyces cerevisiae. *J Biol Chem* **273**, 22480-9 (1998).
12. Alexandre, H., Ansanay-Galeote, V., Dequin, S. & Blondin, B. Global gene expression during short-term ethanol stress in Saccharomyces cerevisiae. *FEBS Lett* **498**, 98-103 (2001).
13. Duarte, N.C., Herrgard, M.J. & Palsson, B.O. Reconstruction and Validation of Saccharomyces cerevisiae iND750, a Fully Compartmentalized Genome-Scale Metabolic Model. *Genome Res.*, 2250904 (2004).
14. Herrgard, M.J. et al. A consensus yeast metabolic network reconstruction obtained from a community approach to systems biology. *Nat Biotechnol* **26**, 1155-60 (2008).
15. Villas-Boas, S.G., Moxley, J.F., Akesson, M., Stephanopoulos, G. & Nielsen, J. High-throughput metabolic state analysis: the missing link in integrated functional genomics of yeasts. *Biochem J* **388**, 669-77 (2005).
16. Allen, J. et al. High-throughput classification of yeast mutants for functional genomics using metabolic footprinting. *Nat Biotechnol* **21**, 692-6 (2003).
17. Martins, A.M. et al. A systems biology study of two distinct growth phases of Saccharomyces cerevisiae cultures. *Current Genomics* **5**, 649-663 (2004).
18. Ewald, J.C., Heux, S. & Zamboni, N. High-throughput quantitative metabolomics: workflow for cultivation, quenching, and analysis of yeast in a multiwell format. *Anal Chem* **81**, 3623-9 (2009).
19. Castrillo, J.I. et al. Growth control of the eukaryote cell: a systems biology study in yeast. *J Biol* **6**, 4 (2007).
20. Raamsdonk, L.M. et al. A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nat Biotechnol* **19**, 45-50 (2001).

21. Kafri, R., Bar-Even, A. & Pilpel, Y. Transcription control reprogramming in genetic backup circuits. *Nat Genet* **37**, 295-9 (2005).
22. Kafri, R., Levy, M. & Pilpel, Y. The regulatory utilization of genetic redundancy through responsive backup circuits. *Proc Natl Acad Sci U S A* **103**, 11653-8 (2006).
23. Kafri, R., Springer, M. & Pilpel, Y. Genetic redundancy: new tricks for old genes. *Cell* **136**, 389-92 (2009).
24. Kafri, R., Dahan, O., Levy, J. & Pilpel, Y. Preferential protection of protein interaction network hubs in yeast: evolved functionality of genetic redundancy. *Proc Natl Acad Sci U S A* **105**, 1243-8 (2008).
25. Rodrigues-Pousada, C., Menezes, R.A. & Pimentel, C. The Yap family and its role in stress response. *Yeast* (2010).
26. Westergaard, S.L., Oliveira, A.P., Bro, C., Olsson, L. & Nielsen, J. A systems biology approach to study glucose repression in the yeast Saccharomyces cerevisiae. *Biotechnol Bioeng* **96**, 134-45 (2007).
27. Macquillan, A.M. & Halvorson, H.O. Physiological changes occurring in yeast undergoing glucose repression. *J Bacteriol* **84**, 31-6 (1962).
28. Seoighe, C. & Wolfe, K.H. Yeast genome evolution in the post-genome era. *Curr Opin Microbiol* **2**, 548-54 (1999).
29. Khersonsky, O., Roodveldt, C. & Tawfik, D.S. Enzyme promiscuity: evolutionary and mechanistic aspects. *Curr Opin Chem Biol* **10**, 498-508 (2006).
30. Nobeli, I., Favia, A.D. & Thornton, J.M. Protein promiscuity and its implications for biotechnology. *Nat Biotechnol* **27**, 157-67 (2009).
31. Jensen, R.A. Enzyme recruitment in evolution of new function. *Annu Rev Microbiol* **30**, 409-25 (1976).
32. Jeffery, C.J. Moonlighting proteins. *Trends Biochem Sci* **24**, 8-11 (1999).
33. Conant, G.C. & Wolfe, K.H. Increased glycolytic flux as an outcome of whole-genome duplication in yeast. *Mol Syst Biol* **3**, 129 (2007).
34. Winzeler, E.A. et al. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science* **285**, 901-6 (1999).
35. Castrillo, J.I., Hayes, A., Mohammed, S., Gaskell, S.J. & Oliver, S.G. An optimized protocol for metabolome analysis in yeast using direct infusion electrospray mass spectrometry. *Phytochemistry* **62**, 929-37 (2003).
36. Roessner, U. et al. Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* **13**, 11-29 (2001).
37. Fernie, A.R., Trethewey, R.N., Krotzky, A.J. & Willmitzer, L. Metabolite profiling: from diagnostics to systems biology. *Nat Rev Mol Cell Biol* **5**, 763-9 (2004).
38. Mintz-Oron, S. et al. Gene expression and metabolism in tomato fruit surface tissues. *Plant Physiol* **147**, 823-51 (2008).
39. Malitsky, S. et al. The transcript and metabolite networks affected by the two clades of Arabidopsis glucosinolate biosynthesis regulators. *Plant Physiol* **148**, 2021-49 (2008).
40. Smith, C.A., Want, E.J., O'Maille, G., Abagyan, R. & Siuzdak, G. XCMS: processing mass spectrometry data for metabolite profiling using nonlinear peak alignment, matching, and identification. *Anal Chem* **78**, 779-87 (2006).

41. Cooper, T.G. & Bossinger, J. Selective inhibition of protein synthesis initiation in Saccharomyces cerevisiae by low concentrations of cycloheximide. *J Biol Chem* **251**, 7278-80 (1976).
42. Kanehisa, M. Representation and analysis of molecular networks involving diseases and drugs. *Genome Inform* **23**, 212-3 (2009).
43. Kanehisa, M., Goto, S., Furumichi, M., Tanabe, M. & Hirakawa, M. KEGG for representation and analysis of molecular networks involving diseases and drugs. *Nucleic Acids Res* **38**, D355-60.
44. Kanehisa, M. The KEGG database. *Novartis Found Symp* **247**, 91-101; discussion 101-3, 119-28, 244-52 (2002).
45. Kanehisa, M. et al. From genomics to chemical genomics: new developments in KEGG. *Nucleic Acids Res* **34**, D354-7 (2006).
46. Villas-Boas, S.G., Hojer-Pedersen, J., Akesson, M., Smedsgaard, J. & Nielsen, J. Global metabolite analysis of yeast: evaluation of sample preparation methods. *Yeast* **22**, 1155-69 (2005).
47. Canelas, A.B. et al. Quantitative Evaluation of Intracellular Metabolite Extraction Techniques for Yeast Metabolomics. *Anal Chem* (2009).
48. de Koning, W. & van Dam, K. A method for the determination of changes of glycolytic metabolites in yeast on a subsecond time scale using extraction at neutral pH. *Anal Biochem* **204**, 118-23 (1992).
49. Baganz, F., Hayes, A., Marren, D., Gardner, D.C. & Oliver, S.G. Suitability of replacement markers for functional analysis studies in Saccharomyces cerevisiae. *Yeast* **13**, 1563-73 (1997).
50. Katajamaa, M. & Oresic, M. Processing methods for differential analysis of LC/MS profile data. *BMC Bioinformatics* **6**, 179 (2005).
51. Benjamini, Y. & Hochberg, Y. Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B-Methodological* **57**, 289-300 (1995).
52. Steuer, R. On the analysis and interpretation of correlations in metabolomic data. *Briefings in Bioinformatics* **7**, 151-158 (2006).
53. SGD. SGD project. "Saccharomyces Genome Database". (SGD project. "Saccharomyces Genome Database", 2009).
54. Gangloff, S.P., Marguet, D. & Lauquin, G.J. Molecular cloning of the yeast mitochondrial aconitase gene (ACO1) and evidence of a synergistic regulation of expression by glucose plus glutamate. *Mol Cell Biol* **10**, 3551-61 (1990).
55. Chen, X.J., Wang, X., Kaufman, B.A. & Butow, R.A. Aconitase couples metabolic regulation to mitochondrial DNA maintenance. *Science* **307**, 714-7 (2005).
56. Schneiter, R. Intracellular sterol transport in eukaryotes, a connection to mitochondrial function? *Biochimie* **89**, 255-9 (2007).
57. Hughes, A.L. *Adaptive evolution of genes and genomes*, 235 (Oxford University Press, 1999).
58. Kimura, M. Preponderance of synonymous changes as evidence for the neutral theory of molecular evolution. *Nature* **267**, 275-6 (1977).
59. Goldman, N. & Yang, Z. A codon-based model of nucleotide substitution for protein-coding DNA sequences. *Mol Biol Evol* **11**, 725-36 (1994).
60. Dwight, S.S. et al. Saccharomyces Genome Database (SGD) provides secondary gene annotation using the Gene Ontology (GO). *Nucleic Acids Res* **30**, 69-72 (2002).

61. DeLuna, A. et al. Exposing the fitness contribution of duplicated genes. *Nat Genet* **40**, 676-81 (2008).
62. Kuepfer, L., Sauer, U. & Blank, L.M. Metabolic functions of duplicate genes in Saccharomyces cerevisiae. *Genome Res* **15**, 1421-30 (2005).
63. DeLuna, A., Avendano, A., Riego, L. & Gonzalez, A. NADP-glutamate dehydrogenase isoenzymes of Saccharomyces cerevisiae. Purification, kinetic properties, and physiological roles. *J Biol Chem* **276**, 43775-83 (2001).
64. Ma, H. & Botstein, D. Effects of null mutations in the hexokinase genes of Saccharomyces cerevisiae on catabolite repression. *Mol Cell Biol* **6**, 4046-52 (1986).
65. Rodriguez, A., De La Cera, T., Herrero, P. & Moreno, F. The hexokinase 2 protein regulates the expression of the GLK1, HXK1 and HXK2 genes of Saccharomyces cerevisiae. *Biochem J* **355**, 625-31 (2001).
66. Huh, W.K. et al. Global analysis of protein localization in budding yeast. *Nature* **425**, 686-91 (2003).
67. Fernandes, L., Rodrigues-Pousada, C. & Struhl, K. Yap, a novel family of eight bZIP proteins in Saccharomyces cerevisiae with distinct biological functions. *Mol Cell Biol* **17**, 6982-93 (1997).
68. Wu, W.S. & Li, W.H. Identifying gene regulatory modules of heat shock response in yeast. *BMC Genomics* **9**, 439 (2008).
69. Estruch, F. Stress-controlled transcription factors, stress-induced genes and stress tolerance in budding yeast. *FEMS Microbiol Rev* **24**, 469-86 (2000).
70. Mager, W.H. & Ferreira, P.M. Stress response of yeast. *Biochem J* **290 (Pt 1)**, 1-13 (1993).
71. Alexandre, H. & Charpentier, C. Biochemical aspects of stuck and sluggish fermentation in grape must. *Journal of Industrial Microbiology & Biotechnology* **20**, 20-27 (1998).
72. Bisson, L.F. Stuck and sluggish fermentations. *American Journal of Enology and Viticulture* **50**, 107-119 (1999).
73. Norbeck, J. & Blomberg, A. The level of cAMP-dependent protein kinase A activity strongly affects osmotolerance and osmo-instigated gene expression changes in Saccharomyces cerevisiae. *Yeast* **16**, 121-37 (2000).
74. Teusink, B., Walsh, M.C., van Dam, K. & Westerhoff, H.V. The danger of metabolic pathways with turbo design. *Trends Biochem Sci* **23**, 162-9 (1998).
75. Ding, M.Z., Cheng, J.S., Xiao, W.H., Qiao, B. & Yuan, Y.J. Comparative metabolomic analysis on industrial continuous and batch ethanol fermentation processes by GC-TOF-MS. *Metabolomics* **5**, 229-238 (2009).
76. Devantier, R., Scheithauer, B., Villas-Boas, S.G., Pedersen, S. & Olsson, L. Metabolite profiling for analysis of yeast stress response during very high gravity ethanol fermentations. *Biotechnology and Bioengineering* **90**, 703-714 (2005).
77. Voit, E.O. Biochemical and genomic regulation of the trehalose cycle in yeast: review of observations and canonical model analysis. *J Theor Biol* **223**, 55-78 (2003).
78. Blomberg, A. Metabolic surprises in Saccharomyces cerevisiae during adaptation to saline conditions: questions, some answers and a model. *Fems Microbiology Letters* **182**, 1-8 (2000).
79. Shilo, B., Riddle, V.G. & Pardee, A.B. Protein turnover and cell-cycle initiation in yeast. *Exp Cell Res* **123**, 221-7 (1979).

80. Mensonides, F.I., Schuurmans, J.M., Teixeira de Mattos, M.J., Hellingwerf, K.J. & Brul, S. The metabolic response of Saccharomyces cerevisiae to continuous heat stress. *Mol Biol Rep* **29**, 103-6 (2002).
81. Mensonides, F.I., Brul, S., Klis, F.M., Hellingwerf, K.J. & Teixeira de Mattos, M.J. Activation of the protein kinase C1 pathway upon continuous heat stress in Saccharomyces cerevisiae is triggered by an intracellular increase in osmolarity due to trehalose accumulation. *Appl Environ Microbiol* **71**, 4531-8 (2005).
82. Nwaka, S., Kopp, M. & Holzer, H. Expression and function of the trehalase genes NTH1 and YBR0106 in Saccharomyces cerevisiae. *J Biol Chem* **270**, 10193-8 (1995).
83. Booth, J.W. & Guidotti, G. An alleged yeast polyphosphate kinase is actually diadenosine-5', 5'"-P1,P4-tetraphosphate alpha,beta-phosphorylase. *J Biol Chem* **270**, 19377-82 (1995).
84. Dohmen, K. et al. Ethnic differences in gastric sigma-alcohol dehydrogenase activity and ethanol first-pass metabolism. *Alcohol Clin Exp Res* **20**, 1569-76 (1996).
85. Parlesak, A., Billinger, M.H., Bode, C. & Bode, J.C. Gastric alcohol dehydrogenase activity in man: influence of gender, age, alcohol consumption and smoking in a caucasian population. *Alcohol Alcohol* **37**, 388-93 (2002).
86. Birley, A.J. et al. ADH single nucleotide polymorphism associations with alcohol metabolism in vivo. *Hum Mol Genet* **18**, 1533-42 (2009).
87. Chen, H.J., Tian, H. & Edenberg, H.J. Natural haplotypes in the regulatory sequences affect human alcohol dehydrogenase 1C (ADH1C) gene expression. *Hum Mutat* **25**, 150-5 (2005).
88. Edenberg, H.J. Regulation of the mammalian alcohol dehydrogenase genes. *Prog Nucleic Acid Res Mol Biol* **64**, 295-341 (2000).
89. Shen, B., Hohmann, S., Jensen, R.G. & Bohnert, H. Roles of sugar alcohols in osmotic stress adaptation. Replacement of glycerol by mannitol and sorbitol in yeast. *Plant Physiol* **121**, 45-52 (1999).
90. Sakaki, K., Tashiro, K., Kuhara, S. & Mihara, K. Response of genes associated with mitochondrial function to mild heat stress in yeast Saccharomyces cerevisiae. *J Biochem* **134**, 373-84 (2003).

List of Publications

1. Fruit-surface flavonoid accumulation in tomato is controlled by a SlMYB12-regulated transcriptional network.
   Adato A, Mandel T, Mintz-Oron S, Venger I, Levy D, Yativ M, Domínguez E, Wang Z, De Vos RC, Jetter R, Schreiber L, Heredia A, Rogachev I, Aharoni A.
   PLoS Genet. 2009 Dec;5(12):e1000777. Epub 2009 Dec 18.

2. Dual labeling of metabolites for metabolome analysis (DLEMMA): A new approach for the identification and relative quantification of metabolites by means of dual isotope labeling and liquid chromatography-mass spectrometry.
   Feldberg L*, Venger I*, Malitsky S, Rogachev I, Aharoni A.
   Anal Chem. 2009 Nov 15;81(22):9257-66.
   *Equal Contribution

3. The transcript and metabolite networks affected by the two clades of Arabidopsis glucosinolate biosynthesis regulators.
   Malitsky S, Blum E, Less H, Venger I, Elbaz M, Morin S, Eshed Y, Aharoni A.
   Plant Physiol. 2008 Dec;148(4):2021-49. Epub 2008 Oct 1.

4. Non-targeted analysis of spatial metabolite composition in strawberry (Fragariaxananassa) flowers.
   Hanhineva K, Rogachev I, Kokko H, Mintz-Oron S, Venger I, Kärenlampi S, Aharoni A.
   Phytochemistry. 2008 Oct;69(13):2463-81. Epub 2008 Sep 4.

5. Gene expression and metabolism in tomato fruit surface tissues.
   Mintz-Oron S, Mandel T, Rogachev I, Feldberg L, Lotan O, Yativ M, Wang Z, Jetter R, Venger I, Adato A, Aharoni A.
   Plant Physiol. 2008 Jun;147(2):823-51. Epub 2008 Apr 25.

6. Network analysis of protein structures identifies functional residues.
   Amitai G, Shemesh A, Sitbon E, Shklar M, Netanely D, Venger I, Pietrokovski S.
   J Mol Biol. 2004 Dec 3;344(4):1135-46.

Statement of Independent Contribution

The majority of this work is my independent effort.

Parts of the work pertaining to the analysis of evolutionary adaptation to heat stress in *S. cerevisiae* were carried out by Avihu Yona, a fellow PhD student in Pilpel lab. Avihu has carried out the evolution of the strains and performed experimental transcriptomics analysis on the mutant strains after exposure to heat stress.

I have performed all the metabolomics analysis procedures, as well as the analysis of the resulting data and analysis of the relationship between metabolomics and transcriptomics.