



TEL AVIV UNIVERSITY

GEORGE S. WISE FACULTY OF LIFE SCIENCES  
THE SMOLARZ FAMILY GRADUATE SCHOOL

אוניברסיטת תל-אביב

הפקולטה למדעי החיים ע"ש ג'ורג' ס. וייז  
המדרשה לתארים מתקדמים ע"ש משפחת סמולרש

# Epigenetic Effects In Evolution

THESIS SUBMITTED FOR THE DEGREE "DOCTOR OF PHILOSOPHY"

by

**Dorit Hizi**

SUBMITTED TO THE SENATE OF TEL-AVIV UNIVERSITY

March 2024



TEL AVIV UNIVERSITY

GEORGE S. WISE FACULTY OF LIFE SCIENCES  
THE SMOLARZ FAMILY GRADUATE SCHOOL

אוניברסיטת תל-אביב

הפקולטה למדעי החיים ע"ש ג'ורג' ס. וייז  
המדרשה לתארים מתקדמים ע"ש משפחת סמולרז

# Epigenetic Effects In Evolution

THESIS SUBMITTED FOR THE DEGREE "DOCTOR OF PHILOSOPHY"

by

**Dorit Hizi**

SUBMITTED TO THE SENATE OF TEL-AVIV UNIVERSITY

March 2024

**This work was carried out under the supervision of Prof. Oded Rechavi, TAU,  
in collaboration with Prof. Yitzhak Pilpel, WIS.**

*Oded Rechavi*

---

Prof. Oded Rechavi

*Yitzhak Pilpel*

---

Prof. Yitzhak Pilpel

## Acknowledgments

I would like to express my gratitude to my supervisors, Prof. Oded Rechavi, and Prof. Tzachi Pilpel for their guidance, support and insights throughout my Ph.D. journey.

My Heartfelt thanks go also to my amazing lab managers, Dr. Orna Dahan and Dr. Sarit Anava, for teaching me the ropes, and repeatedly pulling the impossible while making it all look fun and easy. Your professionalism and friendship were my greatest blessings over the research years.

Additional thanks go to my partner - Dr. Hila Gingold, for her brilliant contributions and help in promoting this research, along with her delightful wit and kindness that helped in promoting this researcher, and to Dr. Yoav Ram for his important collaboration, smart insights, great humor and dubious coffee.

Special thanks go to Prof. Martin Kupiec, for his lasting and invaluable support and mentorship, and his generosity and extreme patience in sharing his knowledge and expertise.

I would also like to thank the members of my committee, Prof. Avigdor Eldar, Prof. Lilach Hadany and Prof. Yaniv Assaf, for their patience and their valuable feedback and insights. I am grateful for the time and effort they put into reviewing and directing my work over the years.

Many thanks go to all my colleagues at the Rechavi and Pilpel labs, who have provided me with a supportive and collaborative environment. I am grateful for the opportunity to have met and worked alongside such talented individuals- It has been a pleasure and a privilege.

To Prof. Maya Schuldiner, Prof. David Bartel and the late Prof. Susan Lindquist and their labs who generously lent me their help and their yeast for this project, and to all the others in WIS and TAU who were always there to lend me their advice and equipment – Thank you from the bottom of my heart.

Finally, I would like to thank my loving friends and family, and especially to my dear parents, Batya and the late Tuvia Hizi and to my late beloved grandparents- Lea and Dov Hizi and Tzila and Moti Swartz for their constant encouragement and support, and for raising me to admire nature and the human quest for knowledge. Their influence and spirit accompanied me throughout my academic ventures.

Thank you all!

I dedicate this dissertation to Dor Dov Yaacov and Alon Tzur- my greatest loves and proudest achievements, and to my beloved Yair- the next Dr. Hizi who nearly beat me to it.

## Table of Contents

1	Abstract.....	6
2	Introduction.....	9
3	Chapter 1- Epigenetic Effects in evolution study .....	11
3.1	Introduction.....	11
3.2	Materials and Methods.....	14
3.3	Results.....	21
3.4	Chapter Summary and Discussion .....	47
4	Chapter 2- A Whole-genome evolvability screen.....	49
4.1	Introduction.....	49
4.2	Materials and Methods.....	50
4.3	Results.....	61
4.4	Chapter Summary and Discussion .....	93
5	Summary.....	95
6	References.....	99

# 1 Abstract

The main focus of my research was an attempt to investigate the underlying mechanisms and effectors that influence the evolutionary process.

Originally, the research was aimed to investigate epigenetic effects in evolution, yet technical challenges that we could not foresee in advance and insights that were gained throughout the process of experimentation forced the project to evolve into an all-genome high-throughput evolvability screen.

The study took place by performing lab evolution with the Baker's yeast *Saccharomyces cerevisiae*, and its two main goals were as stated above-

## 1. Epigenetic Effects in Evolution Study

Here I was aiming to find relevant epigenetic agents that significantly affect evolvability, namely the capacity of biological species to evolve. The main goal was to study in-depth and model the underlying mechanism of adaptation and gain new insights regarding the evolutionary process.

Originally, the method that was chosen was to perform separate lab evolution experiments using strains with tempered/enhanced mechanisms which can be considered "Epigenetic", including HSP-90 (protein chaperone), SUP-35/RNQ prions and RNAi, all of them with a potential capacity to accommodate rapid environmental adaptations via epigenetic mechanisms. Those evolutions took place using a Xylose-based growth media, which presented a mild metabolic challenge (or, perhaps more accurately, an "Opportunity").

Initial results showed that RNAi-deficient strains in which RNAi mechanisms were introduced (by ectopic expression) exhibited different evolutionary dynamics than the parent strain. These results require additional investigation and verification, but the observed max yield and lag time were significantly lower in comparison to the parent strain, even for the strain with only a partial RNAi restoration (i.e. Dicer gene only addition), indicating that reintroducing RNAi could influence evolvability.

Initial analysis also suggested that the HSP-90-related gene deletion strains that were studied have presented some systematically different evolutionary dynamics than the WT (Parent) strain. This was most evident for the HSC homozygote and for the STI mutant strains (HSC-82 deficiency reduces HSP-90 levels and STI-1 deficiency increases it), hence suggesting that HSP-90 indeed affects evolvability. Yet as to the nature and extent of that influence, additional growth tests should be held and analyzed.

As for the SUP-35/RNQ prion strains, initial results did not show any significant difference, and other studies at the lab raised some doubts regarding prion expression conditions.

Additional growth tests were attempted to try and complement the above-mentioned HSP-90 and RNAi data sets, yet all the attempts failed due to equipment reliability, and since these tests are highly labor-intensive and time-consuming and time was of the essence, it was decided to focus on promoting the second project that is described below, using a genome-wide screening approach.

## 2. A Whole-genome evolvability screen

This screen took place by using the yeast deletion project haploid collection and creating a pooled deletion library for all the non-essential deletion mutants. A WT addition was also attempted but was insufficient and therefore ineffective.

This pool was then evolved under five different conditions, including nominal, metabolic, osmotic and temperature stresses, and a metabolic "opportunity" (similar to the one used in the previous chapter), and sequenced under various evolution times and growth points to produce the mutants' relative frequencies.

Elaborate data analysis was used to extract estimated fitness values for each mutant at each evolution time point and to gain an impression of their evolutionary dynamics as depicted by their fitness change over evolution ("acceleration").

From that analysis results, a few promising candidates (with impressive acceleration values) were identified and evolved separately. Initial results for two of them- YDR508C (GNP1) and YGL165C, that were not previously associated with evolvability, and that were evolved over YPD medium with pH 7.6 showed that significant improvement in pH tolerance was achieved during the evolution and that significant differences existed in the extent and nature of that improvement between the WT and candidates, especially at higher pH than the one that was used for evolution.

Therefore, the growth tests for YDR508C (GNP1) and YGL165C, were elaborately repeated for verification (to include additional pH values and evolution times). Once verified, both strains (to include ancestors and evolvants biological repeats) were sequenced and an in-depth analysis took place to better understand the underlying mechanism.

From the analysis, it turns out that for each strain (either YDR508C or YGL165C) all the biological repeats exhibited a set of nonsynonymous point mutations that mapped into different genes belonging to the same pathway. At each strain it was a different pathway, which mostly remained intact for the WT or the other strain, as though the original mutation pushed the evolution dynamic into a specific path, hence obtaining a considerable final phenotypic advantage in its own unique way.

Overall, this research attempted to shed some light on the underlying mechanisms and effectors that influence the evolutionary process, by attempting two complementary and possibly synergetic approaches, based on the lab evolution of the yeast *Saccharomyces cerevisiae* as a model organism. The first, focused approach, was aimed at understanding how epigenetic factors influence evolvability, particularly via the use of strains with "epigenetic" mechanisms such as HSP-90, SUP-35/RNQ prions, and RNAi. Its preliminary findings suggested differences in evolutionary dynamics for RNAi-deficient and HSP-90 strains compared to the wild type, warranting further investigation. The second spanned a wider net by developing a framework and performing a whole-genome high-throughput evolvability screen, where a pooled library of non-essential deletion mutants was evolved under different conditions. Unfortunately, Due to technical reasons, the HSP-90 strains were decimated and their effect could not be mapped in this screen, yet other promising candidates were identified, including YDR508C (GNP1) and YGL165C, which showed significant improvements in pH tolerance during evolution. Detailed analysis revealed

nonsynonymous mutations in specific pathways unique to each strain, suggesting distinct evolutionary paths leading to phenotypic advantages, and facilitating a novel framework for further studies.

The 2018 discovery of YGL-165C as a de novo addition to the yeast genome, implies its role as an "evolutionary cork." The RNAi removal effect (Albi it requires additional validation) seems to be similar in nature, i.e. suggesting that we are facing a gene that if removed, and a mechanism that if added- induces more vigorous adaptations, hence they are added or removed accordingly by the yeast once a sufficient fitness level is achieved. The same may be argued (although again- requires additional verification) regarding the HSP-90 levels, as induced by the different genetic backgrounds, influence on the evolutionary rate.

These findings, and specifically the novel approach and framework that were developed and the "evolution deceleration" genes that were identified invite further research into the molecular mechanisms underlying these effects and the potential for identifying additional evolvability-related genes.

## 2 Introduction

For nearly a century, a highly “Darwinian” approach, “The Modern Synthesis”, dominated the evolutionary paradigm. This approach combines natural selection and the survival of the fittest and complements Mendelian mechanisms of inheritance and population genetics. However, when discussing the actual mechanism that underlies the evolutionary process, it seems that facing the complexity of the discussed phenomena much is there to be discovered yet. This raises a motivation, in the spirit of the long-lasting yeast community tradition, for a whole-genome evolution-effector screen.

In addition, epigenetics (i.e. heritable changes that are not encoded in the DNA sequence) as a whole provides a newly recognized, orthogonal dimension to the concept of heterogeneity in biology, significantly increasing the relevant degrees of freedom and relevant permutations even within a given cell/genome, and trivially, within the population and strain levels. Inspired by recent discoveries in this field, its significance for environmental adaptations is undisputed, yet can it play a role also in the long-term evolutionary process? And if so, what mechanisms are involved in propagating those newly acquired traits across generations?

Better understanding the evolutionary mechanisms or even establishing the existence of alternative evolutionary mechanisms that allow faster environmental adaptations via epigenetic means [13,15] and that will comply more with the "soft inheritance" of acquired traits [1], may shift that existing paradigm to a new and exciting point of view.

Therefore, heretically, I was interested in studying whether tempering either with the genome or with its epigenetic fidelity (i.e. - the organism’s capability to maintain its epigenetic landscape) can enhance or direct evolution and evolvability under certain circumstances.

This latter concept is somewhat illusive and requires some refinement and quantification to evaluate results. Possible dimensions/parameters for that may be:

- \* Speed - how fast will a strain find a solution to a carefully selected challenge?
- \* Solution "quality" - evolved strains' fitness and solution stability (for example, point mutations rather than duplicating a chromosome [30])
- \* Solution variance - a highly evolvable strain may produce a multitude of good solutions to a given challenge, although due to sample size limitations- this dimension will probably remain undetected.
- \* Versatility and Generality- The ability of the original (non-evolved) strain to evolve under various challenges, and the performance of the evolved strain (for a specific challenge) under conditions different from the ones it was evolved to.

Where the above-mentioned "solution finding" will consist of:

- \* Generating the solution
- \* Solution taking over or fixating in the population.

Once such an impact on evolvability is established, if it is epigenetic by nature, two possible mechanisms for alternative evolutionary effects may be conceptualized: (a) The more conservative one - enabling selective genome accessibility for “regular” evolution (via chromatin modulations, for example [31]) or (b) Perpetuating an epigenetic response for an environmental influence [15, 36,37,38].

The approach for this study was based on lab evolution of the entire yeast (*S. cerevisiae*) deletion library, and in addition, of yeast strains with tempered/enhanced epigenetic mechanisms, applied to cells that are evolving towards some carefully selected metabolic challenges.

Populating the intersection group of these two studies, the HSP-90 deletion mutations set was originally planned as a possible control, being a known epigenetic agent based on single (in some of the cases) ORF deletions, albeit inherent technical issues resulted in insufficient signal regarding this subject, initial results suggest it does merit additional future research.

### 3 Chapter 1- Epigenetic Effects in evolution study

#### 3.1 Introduction

The underlying question of this part of my research dealt with the question of whether tempering with the epigenetic fidelity of a genome (i.e. - the genome's capability of maintaining its epigenetic landscape) can enhance evolution and evolvability under certain circumstances.

There are different definitions of epigenetics. As per our discussion, epigenetic inheritance is an inheritance that occurs independently of changes to the DNA sequence. As was mentioned above, an epigenetic change can also be regarded as a factor that significantly increases the relevant degrees of freedom and solution span in a given scenario, hence enabling a wide range of possible fast environmental adaptations [46, 47]. This part of my research dealt with the question of whether it can play a role also in the long-term evolutionary process, and if so, what mechanisms may be involved in propagating the newly acquired traits across generations.

“Epigenetics” encapsulates a collection of diverse phenomena, including chromatin-based mechanisms (such as DNA methylation, histone modifications, chromatin remodeling, etc.), noncoding and coding RNA (such as mRNA stability, RNAi, etc.), structural templating (such as prions and chaperones unleashing cryptic genetic variation, etc.), self-sustaining feedback loops, microbiotic effects and metabolites [46]

For this study, three epigenetic mechanisms were chosen out of this substantial span, including HSP-90 (protein chaperone), SUP-35/RNQ prions and RNAi, hence providing both variance and relative technical assay maturity.

#### **HSP-90**

HSP-90 (heat shock protein 90) is a major chaperone protein and one of the most common heat-related proteins. It is found in bacteria and all branches of Eukarya, but it is apparently absent in archaea [2]. It is among the most highly expressed cellular proteins across all species [3]. Whereas cytoplasmic Hsp90p is highly conserved and essential for viability under all conditions in eukaryotes [4], the bacterial homolog HtpG is dispensable under non-heat stress conditions [5].

HSP-90 provides a molecular mechanism for buffering genetic variability and facilitating its release in response to environmental stress, hence acting as a capacitor, capable of inducing epigenetic inheritance that profoundly influences evolutionary trajectories.

Several studies in the field provide multiple experimental evidence of the inheritance of phenotypic effects upon HSP-90 inhibition and the possible molecular mechanisms responsible for the observed inheritance patterns. In the cavefish *Astyanax mexicanus*, for example, HSP-90 conceals cryptic eye-size variation in surface populations. Inhibition of HSP-90 unmasks this variation, which can be selected for, leading to a reduced-eye phenotype even in the presence of normal HSP-90 activity [12]. Similar findings arose in *Drosophila melanogaster*, where inhibiting HSP-90 during development resulted in morphological alterations, with subsequent generations retaining these phenotypic changes through epigenetic and genetic

mechanisms, including genetic selection and heritably altered chromatin states [57,59], and in *Arabidopsis thaliana*, where reduction of HSP-90 function yields a spectrum of morphological phenotypes dependent on underlying genetic variation [58].

As was mentioned above, beyond merely unleashing phenotypic diversity from pre-existing genetic pools within populations, HSP-90 depletion has been proposed to impact evolution also through diverse avenues such as inducing heritably altered chromatin states [59], transposon mobilization [14,23], the induction of aneuploidy [13] and interacting with various DNA repair pathways. These findings underscore HSP-90's role as a capacitor for morphological evolution through both epigenetic and genetic mechanisms. [59].

Furthermore, it's imperative to note that after several generations produced by crossings, the phenotypic inheritance patterns become independent of the HSP-90 levels. This observation underscores the importance of considering HSP-90's role in designing long-term evolutionary experiments, as it highlights the dynamic nature of the interplay between HSP-90 and phenotypic inheritance over successive generations.

As was mentioned, the model organism for this study was the budding yeast *Saccharomyces cerevisiae*. In this organism, HSP-90 exists as two isoforms, encoded by the genes HSC82 and HSP82 [6]. An HSP90 family member activity is required for folding a specific set of difficult-to-fold proteins from nascent polypeptides into biologically active structures as well as for the refolding of denatured proteins back into native conformations [4]. Although most cellular proteins do not require Hsp82p/Hsc82p chaperone activity for correct folding under normal conditions, Hsp82p and Hsc82p are required for the activation of many key cellular regulatory and signaling proteins, like kinases and transcription factors [7,8,9].

HSP82 and HSC82 share ~97% sequence identity and together, the encoded proteins comprise 1-2% of all the proteins in the cytosol. While HSC82 is expressed constitutively at high levels and only slightly induced by heat shock, HSP82's transcription is strongly induced by both heat and other stresses (such as exposure to ethanol, anoxia or heavy-metal ions) [10].

All members of the HSP90 family function as dimers and protein folding is driven by an ATPase activity of the chaperone [11]. They associate with many co-chaperones which both positively and negatively regulate Hsp82p/Hsc82p function. One such major co-chaperone is Sti1p, which regulates Hsp82p/Hsc82p activity through the inhibition of the above ATP hydrolysis.

## **Prions**

Prions (proteinaceous infectious particles) are infectious self-propagating protein isoforms. Initially proposed to explain certain mammalian diseases, it is now clear that yeast also has a significant and various span of recognized prions- heritable elements that are transmitted via protein [56].

[PSI+], the prion state of Sup35p can be conceptualized as an epigenetic switch that confers both phenotypic plasticity and evolvability by unleashing Cryptic Genetic Variation (CGV) induced by ribosomes reading into regions that have not recently been subject to selective pressures and that are divergent in different yeast genetic backgrounds [15]. In evolutionary simulations, loss-of-function mutations that reveal phenotypic variation facilitate more rapid evolution to new adaptive peaks [21]. Therefore, Evolutionary

Capacitors such as prions that mimic loss-of-function mutants can potentially function as evolutionary capacitors similarly to Hsp90 [22,36].

The essential yeast SUP-35 gene encodes the translation termination factor eRF3 that has a role in mRNA de-adenylation and decay. An altered protein conformation creates the [PSI<sup>+</sup>] prion determinant that causes nonsense suppressor phenotype due to a reduced function of the translation termination factor Sup35p (eRF3) that is polymerized into amyloid fibrils [15,16,17,18,19] That can be phenotypically inherited [36].

A prion state of the Rnq1 protein, [PIN<sup>+</sup>], is required for the [PSI<sup>+</sup>] de novo generation, but not propagation. Yeast [psi<sup>-</sup>] [PIN<sup>+</sup>] cells overproducing Sup35p can exhibit nonsense suppression without the generation of a stable [PSI<sup>+</sup>]. In such cells, most of Sup35p is present in amyloid polymers, though the remaining Sup35p monomer is sufficient for normal translation termination. The presence of these polymers strictly depends on [PIN<sup>+</sup>], suggesting that their maintenance relies on efficient generation de novo, rather than inheritance. Sup35p polymers contain Rnq1p, thus confirming that Rnq1p polymers seed Sup35p polymerization [15].

### **RNAi**

Epigenetic inheritance of small RNAs that mediate RNA interference is well established in different animal models and can confer protection against foreign genome parasites [39,40]. Previous works at the Rechavi lab [37] also demonstrated transgenerational epigenetic inheritance of responses to different environmental challenges, for example, a starvation-induced pattern of gene silencing caused by endogenous siRNAs and resulting in increased longevity in the third-generation progeny of *C. elegans*. Combined with recent findings in prokaryotes (e.g. CRISPR), these results suggest that Lamarckian-type inheritance of acquired traits is an exciting evolutionary phenomenon [38]. Having said that, It is important to distinguish what is now called “Lamarckian inheritance” from what Lamarck originally wrote, of course. He, a product of his time, was wrong in many other ways and was ignorant of current knowledge regarding the real mechanisms of inheritance and their limitations.

RNA-silencing pathways contribute to transposon silencing, viral defense, DNA elimination, heterochromatin formation, and posttranscriptional repression of cellular genes [24,25]. In the simplest form of silencing, known as RNA interference (RNAi), the RNaseIII endonuclease Dicer successively cleaves double-stranded RNA (dsRNA) into small interfering RNAs (siRNAs), which are loaded onto the effector protein Argonaute to guide the cleavage of target transcripts [24,26]. RNAi arose in an early eukaryotic ancestor and appears to have been conserved throughout most of the fungal kingdom [27,28]. A prominent exception is *Saccharomyces cerevisiae*, the budding yeast that lacks recognizable homologs of Argonaute, Dicer, and RNA-dependent RNA polymerase (RdRP), which in some RNAi pathways produces dsRNA. However, a 2009 ground-breaking work by Bartel *et al.* [29] restored this pathway in *Saccharomyces cerevisiae* (by introducing Dicer and Argonaute from *S. castellii*), thus creating an additional powerful tool in this species.

## 3.2 Materials and Methods

### Experimental design

Overall, the test methodology consisted of taking strains with tempered/enhanced epigenetic mechanisms and evolving them in the lab under a carefully selected challenge, to explore whether tempering either with the genome or with its epigenetic fidelity can enhance or direct evolution and evolvability under certain circumstances.

As was mentioned in the introduction section above, evolvability consists of speed (to find a solution), solution quality (resulting strains' fitness and stability), and solution generality (evolved strains' fitness under various conditions). Additional parameters consist of solution variance and versatility (the ability to evolve under various challenges), but they are out of scope for this design.

Assuming that the above-mentioned "solution finding" consists of generating the solution followed by the solution taking over, or fixating on the population, the actual detection that a solution was accomplished and the assessment of its quality took place by performing periodic growth tests (as fitness indicators) all through the evolution, and mapping their results vs time to establish the dynamics.

Once such an impact on evolvability was established, and a difference in the evolutionary processes and dynamics was indeed observed between strains with different epigenetic capacities, determining whether it is genetic (a "classical" genetic trait acquisition) or epigenetic by nature, and if epigenetic- whether it was accomplished by enabling selective genome accessibility for "regular" evolution or by perpetuating an epigenetic response for an environmental influence (for example- HSP-90 inactivation can influence the rate and outcome of evolution by exposing phenotypic diversity rooted in the pre-existing genetic variation or de novo mutations accumulating in the evolving yeast population) would be of an essence.

This was to be accomplished by comparatively studying the evolutions' final and interim products (in comparison to the ancestral strains, to determine the underlying molecular mechanisms.

To enable that, proper viable samples were extracted on a weekly basis and carefully frozen for all the strains and repeats throughout the evolution, to enable future analysis by implementing the full current toolkit of high throughput sequencing, comparative genomics, functional validation, proteomics, small RNA sequencing, etc.

This method allows us also to evaluate the dominant mechanisms at each stage, since epigenetic inheritance is notoriously unstable sometimes, and phenotypic inheritance may be replaced by genetic determinants already in the early stages of the evolutionary process (for example- in drosophila, a mutation in the Hsp83 gene that encodes Hsp90 protein causes a heritable eye phenotype that became independent of it after only 6 generations [57]).

### The medium (challenge) and evolution

As was mentioned above, the selected strains with tempered/enhanced epigenetic mechanisms were evolved in the lab under a carefully selected challenge.

This challenge was a metabolic one and was considered to be a borderline "opportunity" rather than a stress. The motivation for providing an "opportunity" or a milder challenge and not a stress in the classic sense of the word, as is usually done in lab evolution experiments, was that upon evolution to stress the yeast reaction may be strongly affected by the known stress-related functionalities that some of the candidates' strains possessed. This reaction may mask and bias any desired evolvability effect and is therefore less desirable as the selected challenge.

An additional dilemma concerned the selection of a fixed challenge contrary to a fluctuating selection regime for uncovering the role of epigenetic inheritance. Although epigenetic inheritance is known to increase evolvability under fluctuating environments [60 61], and the technical implementation of such a regime, although more complicated, error-prone and labor-intensive is definitely feasible, some major concerns regarding our ability to perform effective parameters separation in assessing the evolutionary dynamics and analyzing the underlying mechanisms swayed us from this course of action, favoring the more robust, fixed alternative.

Xylulose ( $C_5H_{10}O_5$ ) is a ketopentose, a monosaccharide containing five carbon atoms, and a ketone functional group. It is fermentative and does enable yeast growth, though it is much less efficient than other sugars such as glucose for example [35], hence presenting a mild metabolic challenge. Since there was no reason to expect it would favor any of the strains that were produced/used in this work, a Xylulose-based medium was selected as the evolutionary challenge.

To prepare the growth medium, a mixture of xylulose and xylose was produced in the lab, following the method described in [48,49]. Briefly, 350 gr of xylose (Sigma-Aldrich) was dissolved in 500 ml water and 20 gr of immobilized Xylose Isomerase (Sigma-Aldrich) was added. This mixture was incubated at 60°C for 24 hours with agitation (300 rpm). The enzyme was inactivated by heating to 100°C and then filtered off. The resulting xylose:xylulose ratio was determined by HPLC analysis. This was done with the Barkai lab kind assistance, using an Agilent 1200 series high-performance liquid chromatography system equipped with an anion exchange Bio-Rad HPX-87H column (Bio-Rad, Hercules, CA). The column was eluted with 5 mM sulfuric acid at a flow rate of 0.6 mL/min at 45°C.

The actual evolution growth medium consisted of YP medium (Yeast Extract Peptone- [(1 L) 10 g yeast extract, 20 g peptone and DDW (Double distilled water)]) supplemented with the xylose/xylulose mixture described above, to a final concentration of 2% xylulose. Finally, a 30µg/ml doxycycline was added.

The lab evolution itself consisted of growing the strains' samples (a minimal number of four independent biological repeats/ independent lines per strain was maintained) to the stationary stage (yet not allowing them to go too deep), and performing daily dilutions for at least ~500 generations (which based on past experiments [30,33], should be sufficient to see if there is any significant effect worth exploring). Evolution samples for each repeat were frozen every ~6 days (~42 generations).

During this evolution, the strains were grown in 24-well plates, containing 1.2ml of evolution media until entering the stationary phase as was mentioned above ("Day"/~7 generations), and then diluted at a 1:120 ratio by transferring 10µl of the old culture into a well filled with fresh media.

## Growth tests and analysis

The yeast strains performance and adaptation were evaluated by growth experiments, at which the tested strains were inoculated into 5 ml of YPD and incubated over two nights at 30°C until reaching the deep stationary phase. These starters were then diluted 1:50 into the relevant (evolution) media and loaded into 96-well plates. Each plate contained two strains (assay+ reference) and some non-inoculated medium wells as blank references. The plates were incubated at 30°C, and OD measurements for each culture (=well) were taken every hour, hence creating a growth curve. Wells data were compiled using MATLAB, to create the average growth data for each tested strain.

Since direct comparison of growth curves is problematic, the “Curveball” method [41] was used (with Dr. Y. Ram kind assistance), to extract relevant parameters such as max growth (2<sup>nd</sup> derivative of the Optical Density (OD) vs. time curve), growth rate (1<sup>st</sup> derivative), and growth yield (O.D. asymptote),

Those parameters were then used as a relative basis for comparison and relative fitness inference, by comparing them to those of the relevant ancestors and control strains, (including WT).

## Ploidity

All the strains constructed as part of this study (i.e.- the HSP-90 strains that will be presented below) were deliberately in the diploid state, mainly to avoid known ploidity effects, i.e. the yeast's natural inclination to switch toward the diploid state [32,33].

Additional design considerations that played a role in that decision were

- How to optimize to the predicted population size - exploration is more efficient in diploids and fixation in haploids. Therefore, haploids are optimal for bigger populations ( $>10^7$ ) but lose their advantage in small ones.
- How to allow for better growth rates.
- How to allow more degrees of freedom in creating the strains – dosage compensation is negligible in yeast [34], and heterozygosity may enrich the array of tested strains and help to establish observed trends.

However, strains that were received (rather than constructed as part of this study), i.e. the RNAi and SUP strains were all haploid.

## The strains used

As was mentioned above, three categories of possible epigenetic mechanisms were studied:

### HSP-90 (protein chaperone)

These strains were constructed as part of this study. They were all based on [Euroscarf's](#) (EUROpean *Saccharomyces Cerevisiae* ARchive for Functional Analysis) three BY-4743 Heterozygote deletion strains (Heterozygote was selected to minimize the risk for background suppressory mutations) for HSC-82, HSP-82 and STI-1.

These strains were sporulated, tetrad-dissected and mated to produce a variety of deletion combinations, and properly verified by their auxotrophic markers (via replica plating) and by PCR with proper primers (for deletion cassette size and location verification).

For strain tracking convenience, the following 3-digit nomenclature system was used, in which each digit marks the copy number of each of the three genes as presented in Table 1.

# of functional HSC-82 alleles	# of functional HSP-82 alleles	# of functional STI-82 alleles
0-2	0-2	0-2

**Table 1 - HSP-90 strains nomenclature system-** Was used to keep track of the HSP strains that were constructed for this study. Each strain is designated by a 3-digit number, where each digit marks the # of functional alleles for each of the three OPRFs of interest- HSC-82, HSP-82 and STI-82

For example- 222 is WT, 122 is  $\Delta$ HSC-82 heterozygote and 220 is  $\Delta$ STI-1 homozygote.

The original and constructed strains are detailed in Table 2 below.

#	Description	Genotype
1 <b>122</b>	ΔHSC-82 Euroscarf Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSC82::kanMX4/HSC82
2 <b>212</b>	ΔHSP-82 Euroscarf Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSP82::kanMX4/HSP82
3 <b>221</b>	ΔSTI-1 Euroscarf Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; STI1 ::kanMX4/STI1
4 <b>222</b>	Euroscarf WT	BY-4743 WT
DH1 <b>022</b>	ΔHSC-82 Homozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSC82::kanMX4/HSC82::kanMX4
DH2 <b>122</b>	ΔHSC-82 Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSC82::kanMX4/HSC82
DH3 <b>202</b>	ΔHSP-82 Homozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSP82::kanMX4/HSP82::kanMX4
DH4 <b>212</b>	ΔHSP-82 Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSP82::kanMX4/HSP82
DH5 <b>112</b>	ΔHSC-82 + ΔHSP-82 Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSC82::kanMX4/HSC82; HSP82::kanMX4/HSP82
DH6 <b>220</b>	ΔSTI-1 Homozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; STI1 ::kanMX4/STI1 ::kanMX4
DH7 <b>221</b>	ΔSTI-1 Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; STI1 ::kanMX4/STI1
DH8 <b>121</b>	ΔHSC-82 + ΔSTI-1 Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSC82::kanMX4/HSC82; STI1::kanMX4/ STI1
DH9 <b>211</b>	ΔHSP-82 + ΔSTI-1 Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSP82::kanMX4/HSP82; STI1::kanMX4/ STI1
DH10 <b>111</b>	ΔHSC-82 + ΔHSP-82 + ΔSTI-1 Heterozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSC82::kanMX4/HSC82; HSP82::kanMX4/HSP82; STI1::Hph/STI1
DH11 <b>120</b>	ΔHSC-82 Heterozygote+ ΔSTI-1 Homozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSC82::kanMX4/HSC82; STI1::kanMX4/ STI1::Hph
DH12 <b>210</b>	ΔHSP-82 Heterozygote+ ΔSTI-1 Homozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSP82::kanMX4/HSP82; STI1::kanMX4/ STI1::Hph
DH13 <b>012</b>	ΔHSP-82 Heterozygote+ ΔHSC-82 Homozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSC82::kanMX4/HSC82::Hph; HSP82::kanMX4/HSP82
DH14 <b>201</b>	ΔSTI-1 Heterozygote+ ΔHSP-82 Homozygote	Mat a/α; his3Δ1/his3Δ1; leu2Δ0/leu2Δ0; lys2Δ0/LYS2; MET15/met15Δ0; ura3Δ0/ura3Δ0; HSP82::kanMX4/HSP82::Hph; STI1::kanMX4/ STI1

**Table 2 - HSP-90 strains** – Details the HSP-90 strains accomplished/ constructed for this study. Strains 1-4 are the original Euroscarf strains, and the rest of the strains (marked DH1-DH14) were constructed from them via sporulation, tetrad dissection and mating to produce a variety of relevant functional alleles combinations. These combinations are marked for each strain according to Table 1 methodology.

Once constructed, the above HSP-90 strains were characterized and calibrated by performing growth rate tests under the following conditions:

1. YPD, 30°C- To provide a general idea regarding the strain viability and possible growth defects.
2. YPD, 39°C- To quantify the strain resistance to heat shock, which may also indicate the HSP-90 levels.
3. Xylulose (evolution medium), 30°C- A calibration test for the evolution. Used to infer the initial growth rates (in order to time the dilutions), select the final strains and document the starting point.

Table 3 summarizes the calibration results for nine of Table 2 strains.

Test conditions ↓	Strain DH1	Strain DH2	Strain DH3	Strain DH4	Strain DH5	Strain DH 6	Strain DH7	Strain DH8	Strain DH9
	ΔHSC Homozygote	ΔHSC Heterozygote	ΔHSP Homozygote	ΔHSP Heterozygote	ΔHSC ΔHSP Heterozygote	ΔSTI Homozygote	ΔSTI Heterozygote	ΔHSC ΔSTI Heterozygote	ΔHSP ΔSTI Heterozygote
	<b>022</b>	<b>122</b>	<b>202</b>	<b>212</b>	<b>112</b>	<b>220</b>	<b>221</b>	<b>121</b>	<b>211</b>
YPD, 39°C	0	0	0	-	-	+/-0	+	-	-
YPD, 30°C	0	0	-	-	0	0	0	+	+
Xylulose, 30°C	-/0	-/0	-	-/0	-	-/0	0	-	-

(0) No difference between mutant and WT  
 (+) Mutant grows better than WT  
 (-) WT grows better than Mutant

**Table 3 - HSP90 Ancestors mutants' growth data summary**– Calibration growth test data for nine of Table 2 strains. Strains were compared to the WT while grown on nominal conditions and evolution medium to detect initial growth defects and calibrated dilution times, and at 39°C to quantify resistance to heat shock as an indication of HSP-90 levels. Functional allele combinations are marked for each strain according to Table 1 methodology.

Resulting of these calibration tests, the following strains were selected, while considering as was presented above both their effect on HSP-90 levels, as indirectly measured by their ability to grow at 39°C (which should preferably be maximal, yet established only for DH6 and DH7) and their initial growth defect on Xylulose (which should preferably be minimal, as was established for all to exclude DH3):

1. WT (strain 4)
2. ΔHSC-82 Homozygote (DH1, 022)
3. ΔHSC-82 Heterozygote (DH2, 122)
4. ΔHSP-82 Homozygote (DH3, 202)
5. ΔSTI-1 Homozygote (DH6, 220)
6. ΔSTI-1 Heterozygote (DH7, 221)

As was mentioned above, DH3 presented a different starting point on the evolution medium (some growth defect in comparison to the WT). It was evolved nevertheless to get a more comprehensive view, and its initial growth defect was addressed mathematically, to properly evaluate its adaptation.

The above-selected HSP strains were evolved in the Xylulose medium for 50 days (~350 generations). 4 biological repeats were evolved for each mutant strain and 6 biological repeats were evolved for the WT.

### SUP-35/RNQ prions

14 different Prion strains were kindly received from the Lindquist lab. After initial calibration tests, including growth rate tests on the evolution medium (at a nominal 30°C temperature) the strains that were selected for the evolution are described in Table 4. They consisted of a parent strain and two strains that were constructed from it, to significantly enhance or prevent prionicity, thus enabling the comparison of the PSI prion effect. These three strains had almost identical initial growth characteristics on the evolution medium, hence providing an ideal starting point.

#	Description	Genotype
LQ8 yA3080 (406)	74D strain with Sup35 prion domain expansion (significantly enhances PSI prion formation-converts to a PSI+ state about 100-fold more frequently than wild-type Sup35)	Mata, ade1-14(UGA), trp1-289(UAG), his3?-200, ura3-52, leu2-3,112, sup35R2E2,URA3
LQ9 yA3083 (407)	74D strain with Sup35 partial deletion (impossible or almost impossible to form the PSI prion)	Mata, ade1-14(UGA), trp1-289(UAG), his3?-200, ura3-52, leu2-3,112, sup35-R?2-5
LQ10 2119 (408)	74D parent strain from which the previous two were constructed	Mata, ade1-14(UGA), trp1-289(UAG), his3?-200, ura3-52, leu2-3,112

**Table 4 – The selected prion strains** – Details the Lindquist lab strains that were evolved in this study. The basic strain was a 74D Mat a haploid, and the two other strains were constructed from it to significantly enhance or prevent PSI prionicity. For each strain, the table details its nomenclature for this study (LQ8-10), its original Lindquist lab name and the Pilpel lab revco number assigned to it (406-408).

The above prion strains were evolved for 540-560 generations, with 4 biological repeats for each strain.

### RNAi

The RNAi strains were kindly received from the Bartel lab [29]. As is detailed in Table 5 below, they consist of the parent strain, a strain with a partial RNAi mechanism (dicer only) and a strain with a fully restored RNAi mechanism (Dicer + Ago). Each of these strains was evolved for 610-650 generations, in 4 independent biological repeats.

#	Description	Genotype
DPB 249 (444)	parent strain	MAT α leu2-3,112 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15 / S. cerevisiae W303-1B
DPB 258 (448)	W/S. <i>castellii</i> Dicer and Argonaute	MAT α LEU2::pTEF-Dcr1 TRP1::pTEF-Ago1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15 / S. cerevisiae W303-1B
DPB 255 (446)	W/S. <i>castellii</i> Dicer	MAT α LEU2::pTEF-Dcr1 trp1-1 can1-100 ura3::EGFP(S65T)-KanMX6 ade2-1 his3-11,15 / S. cerevisiae W303-1B

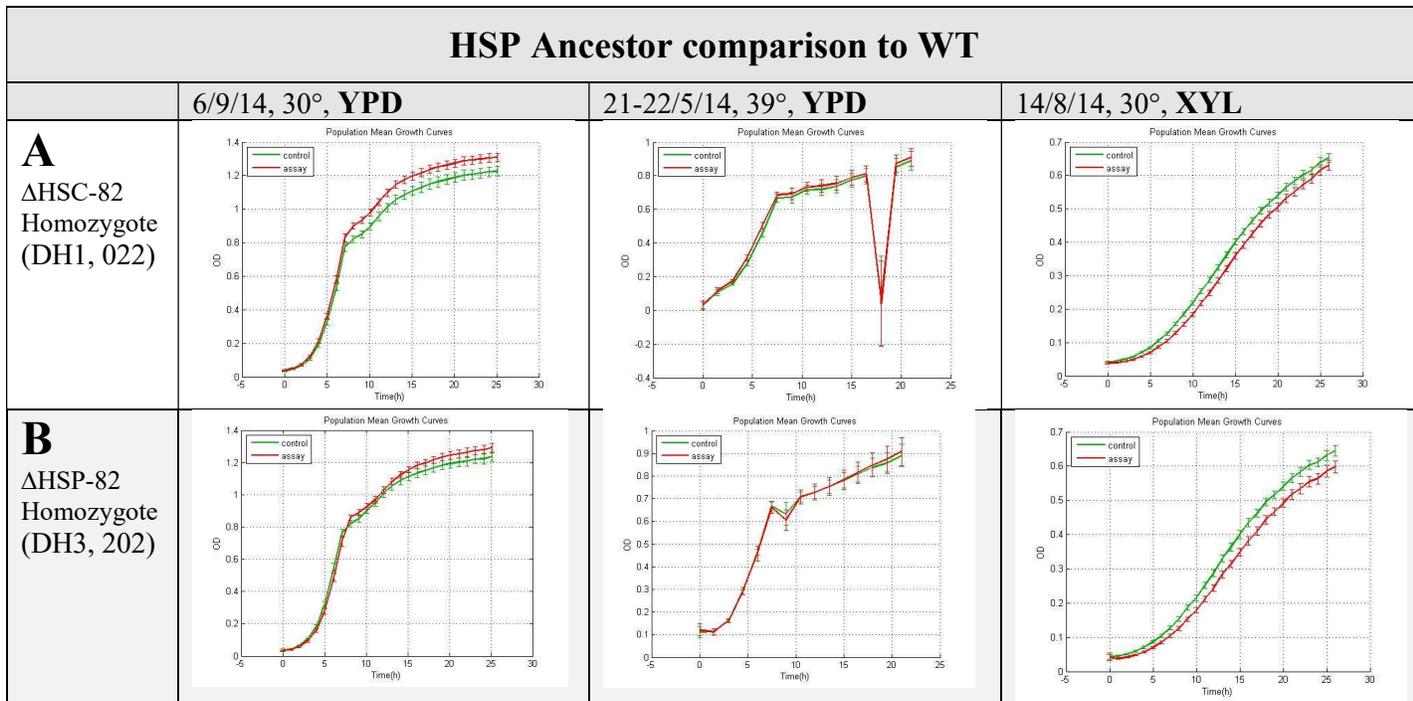
**Table 5 - RNAi strains** – Details the Bartel lab strains that were evolved in this study. The basic strain was a W303-1B Mat  $\alpha$  haploid and the two other stains were constructed from it to include partial (Dicer only) and full (Dicer+Argonaute) RNAi functionality. For each strain the table details its original Bartel lab name (DPB249-255) and the Pilpel lab revco number assigned to it (444-446).

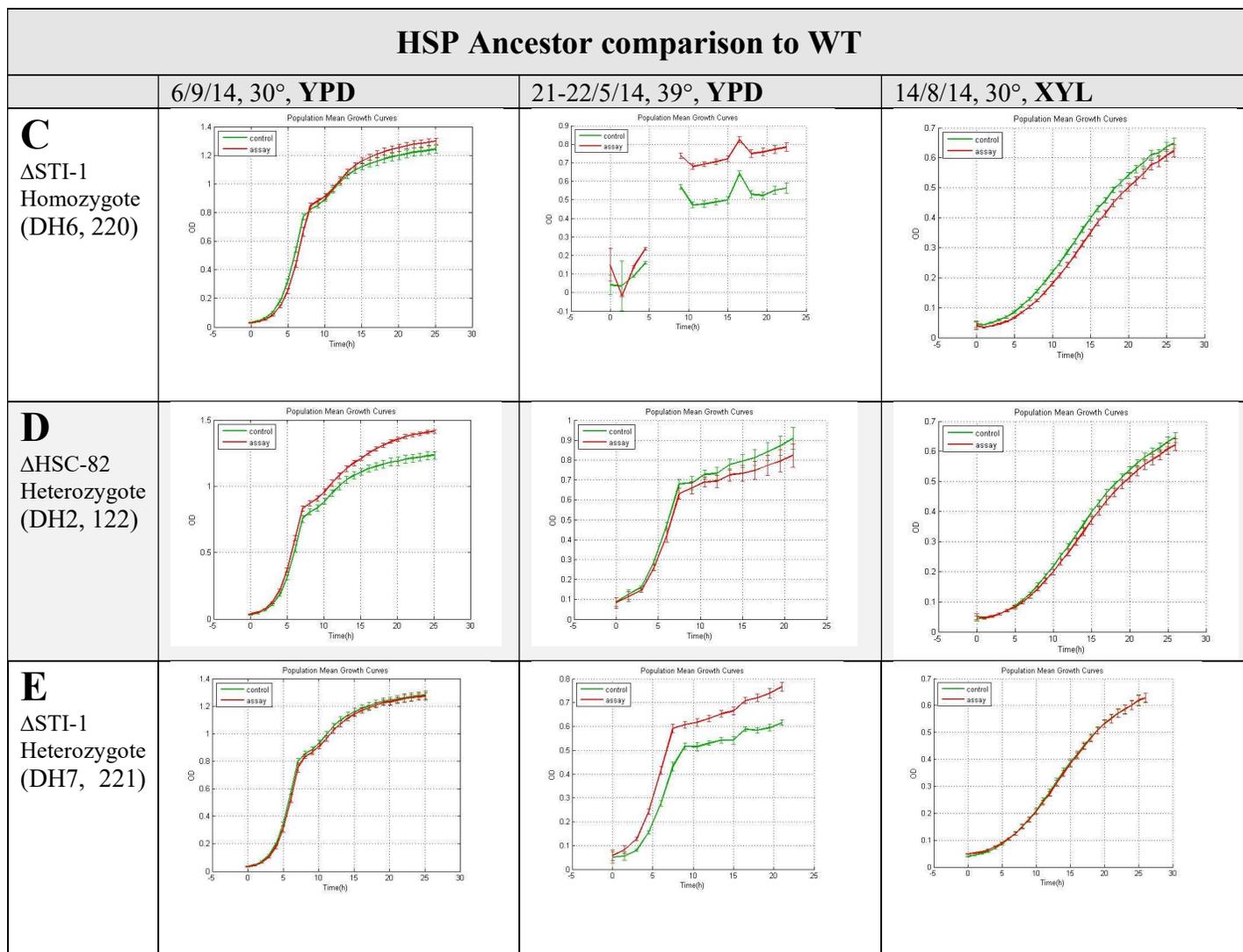
### 3.3 Results

As mentioned above, evolution results were assessed by measuring and parametrizing growth curves. Growth assays were conducted by using the Hamilton evolution robot and the new Pilpel lab robot. Both robots suffered from considerable technical issues at the time, resulting in incomplete data sets. Yet, even from these partial results, some interesting trends were obtained.

Inherently, the growth assays that were used enabled the comparison of two strains per assay.

Comparing the various strains/ mutants to the strain from which they were constructed (parent) or to the WT before evolution started enabled the quantification of the initial growth defect. These results are presented in Figure 1 For HSP-90, Figure 2 for SUP-35 and Figure 3 for RNAi.

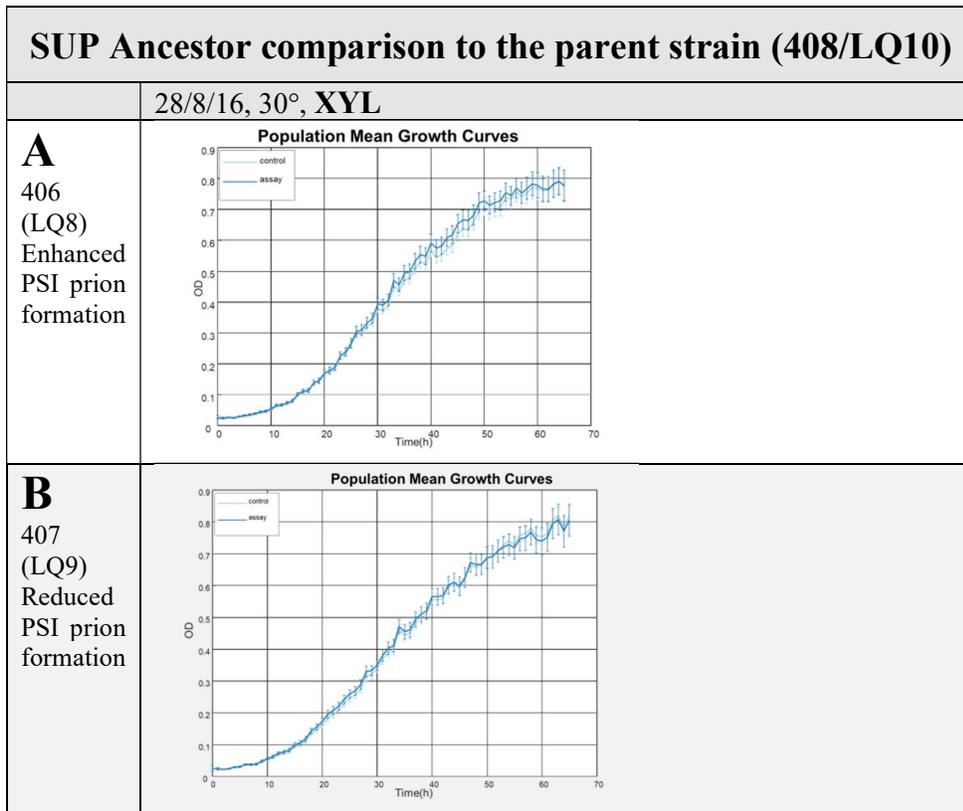




**Figure 1- HSP 90 ancestor strains comparison to WT on YPD and XYL** – The following growth curves were produced by a MATLAB analysis of comparative growth experiments that took place using the Hamilton evolution robot. Depicting OD over time, each of the five ancestor mutant strains was compared to the WT while grown over YPD at 30°C (nominal conditions, the test took place on 6/9/2014), YPD at 39°C (heat shock, indicative of Hsp-90p levels, the test took place on 21-22/5/2014. this test was repeated thrice due to technical difficulties. Results presented here are from tests that were not technically perfect, yet they were complemented with other tests' partial results- not presented) and Xylulose mixture at 30°C (evolution conditions, this test was repeated many times due to technical difficulties. Results presented here are from the 14/8/14 test that was not technically perfect, yet its results were consistent with all the other tests' partial results), to include (A) ΔHSC-82 Homozygote (DH1, 022); (B) ΔHSP-82 Homozygote (DH3, 202); (C) ΔSTI-1 Homozygote (DH6, 220); (D) ΔHSC-82 Heterozygote (DH2, 122); (E) ΔSTI-1 Heterozygote (DH7, 221). Note - The ΔSTI-1 Homozygote 39°C tests were of very low technical quality. They are complemented by additional (low quality) results yet should be treated cautiously.

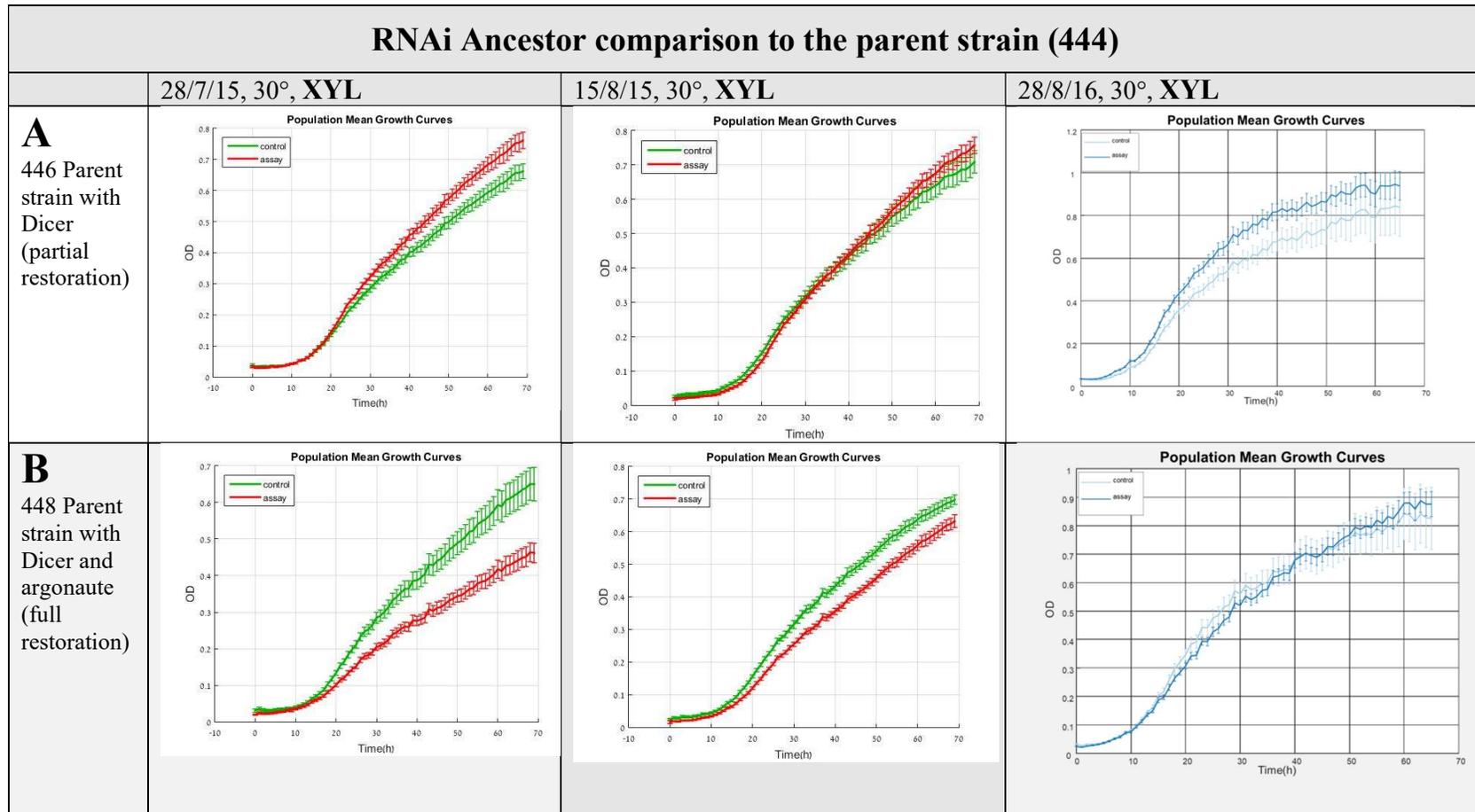
From the above results it is evident that the HSP mutants have no growth defect on YPD, and in fact- they grow even better than the WT on nominal conditions (slightly better for all, significantly better for ΔHSC-82 homozygote and heterozygote. This makes sense since HSP production requires resources, yet its main

utility is in times of stress, which is not relevant to this experiment's nominal conditions). On the evolution medium, a slight advantage for the WT is observed, yet it may be considered significant (although small) only for the  $\Delta$ HSP-82 Homozygote. Hence for all practical reasons- they may be considered as having the same starting point. As to the 39°C growth data as an indication of the Hsp-90p levels, we can see no significant difference for the HSP and HSC mutants, and a small advantage for the STI mutants (the STI homozygote results comes from low quality test and should be treated cautiously), which stands to reason with the known mechanisms of STI roll as an inhibitor, and HSP/HSC backing up each other. **Table 3** summarizes these results.



**Figure 2- SUP35 ancestor strains comparison to parent ancestor strain** – The following growth curves were produced by a MATLAB analysis of comparative growth experiments that took place using the Hamilton evolution robot. Depicting OD over time, the two ancestor mutant strains were compared to the parent strain from which they were constructed. The comparison took place by growing the strains on the Xylulose mixture at 30°C (evolution conditions). This test was repeated many times due to technical difficulties. Results presented here are from the 28/8/16 final test that was both complete and consistent with previous partial results. The presented results consist of **(A)** 406 (LQ8)- 74D with enhanced PSI prion formation; **(B)** 407 (LQ9) - 74D reduced PSI prion formation); both compared to 408 (LQ10)- the parent 74D strain from which the previous two were constructed.

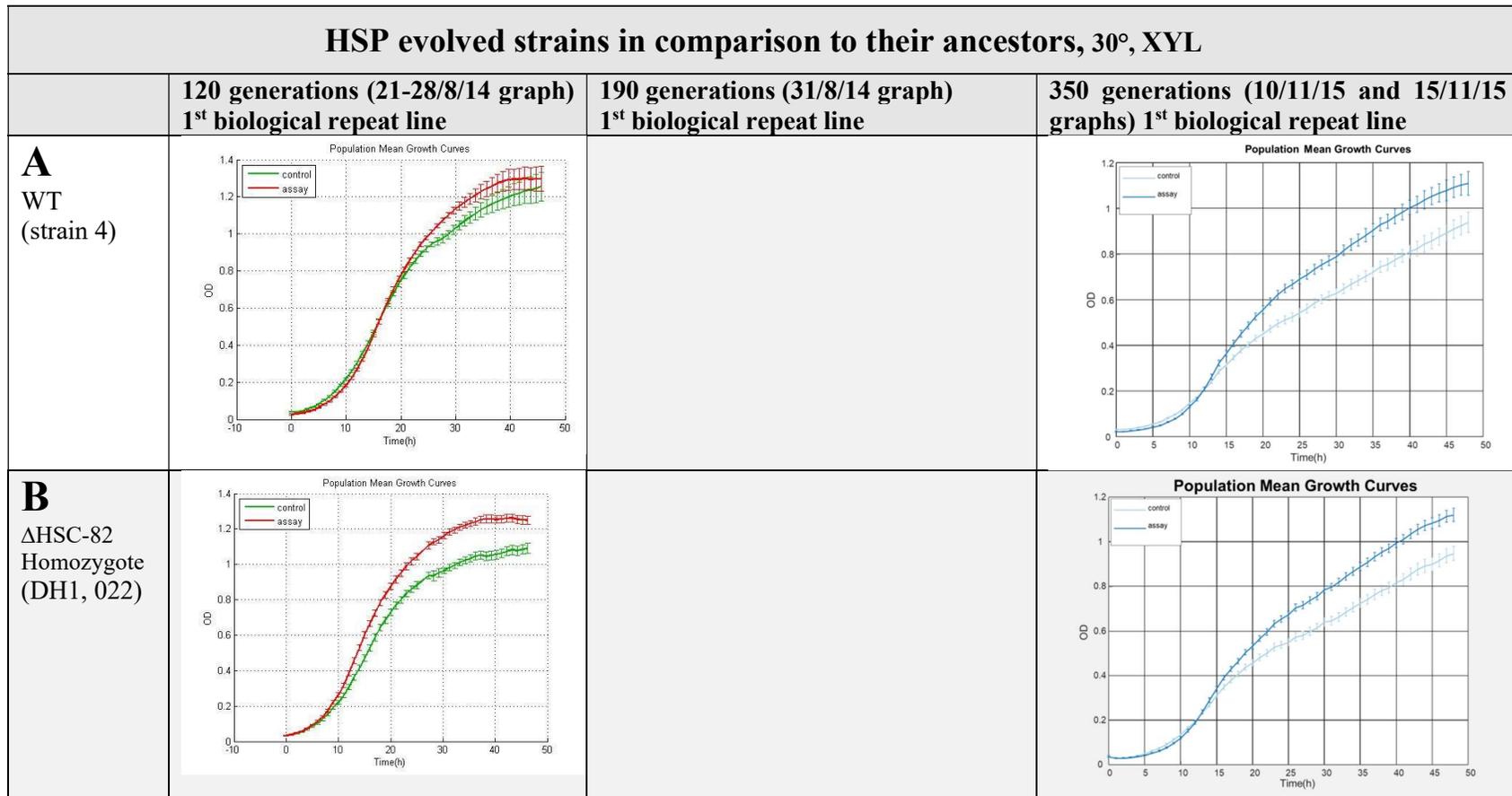
As can be seen, while grown on the evolution medium, the enhanced/reduced prionicity strains showed no growth defect in comparison to the parent strain and hence possessed the same starting point.



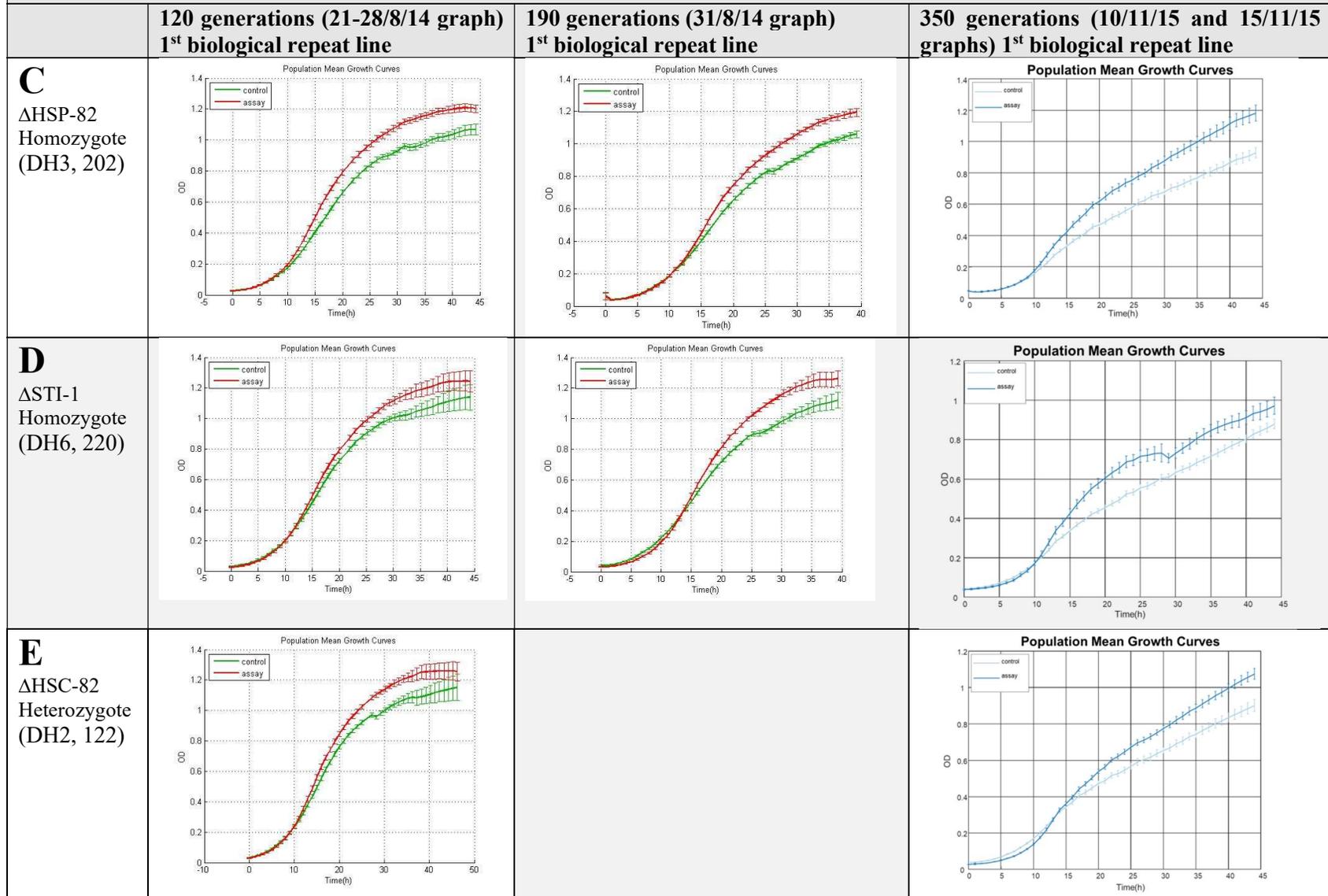
**Figure 3- RNAi ancestor strains comparison to parent ancestor strain** – The following growth curves were produced by a MATLAB analysis of comparative growth experiments that took place using the Hamilton evolution robot. Depicting OD over time, the two ancestor mutant strains were compared to the parent strain from which they were constructed. The comparison took place by growing the strains on the Xylulose mixture at 30°C (evolution conditions), this test was repeated several times to verify consistency. Results presented here are from the 28/7/15, 15/8/15 and 28/8/16 tests, to include **(A)** 446- Parent strain with dicer added (partial RNAi restoration); **(B)** 448- Parent strain with dicer and argonaute added (full RNAi restoration); both compared to 444- the parent strain from which they were both constructed. **Please mind the different OD scales used in this figure.**

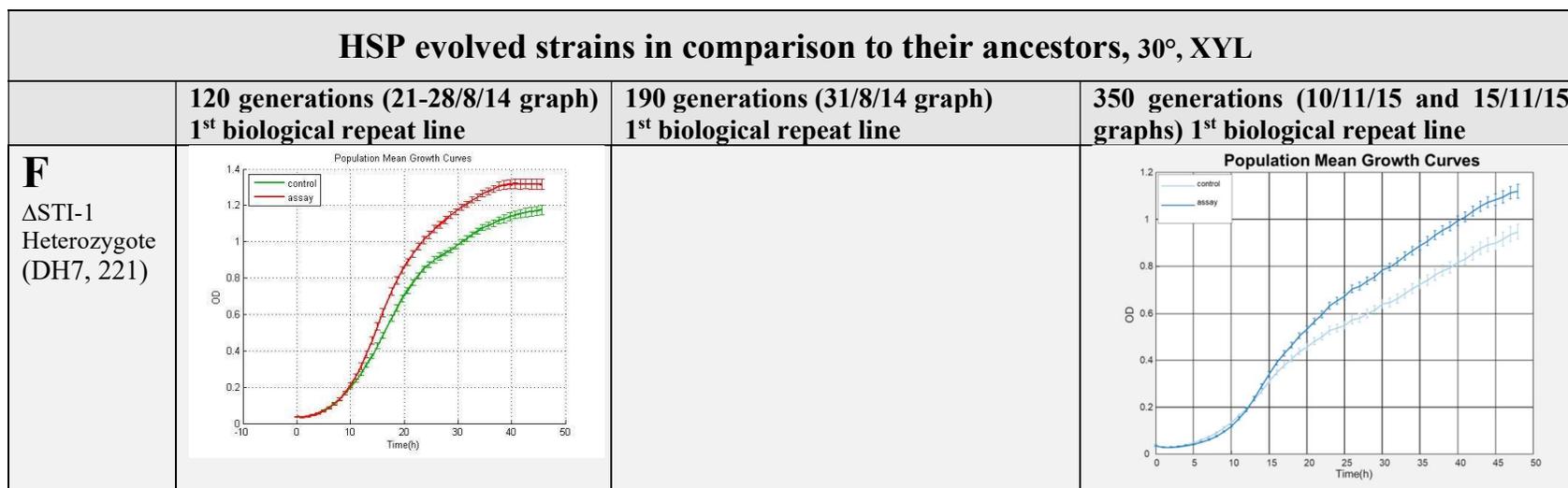
As can be seen, while grown on the evolution medium, the partly restored RNAi strain (446 with dicer) exhibits what seems like a small yet significant advantage over the parent strain, while the fully restored RNAi strain (448 with dicer and argonaute) shows a small yet significant growth defect. These initial fitness differences were calculated and accounted for in analyzing the evolution results.

After completing the starting point calibration, lab evolution commenced. Comparing the different evoltants to their ancestors showed that significant improvement was accomplished for all the strains (at all the biological repeats). A few representative examples are presented in Figure 4, Figure 5 and Figure 6 below.



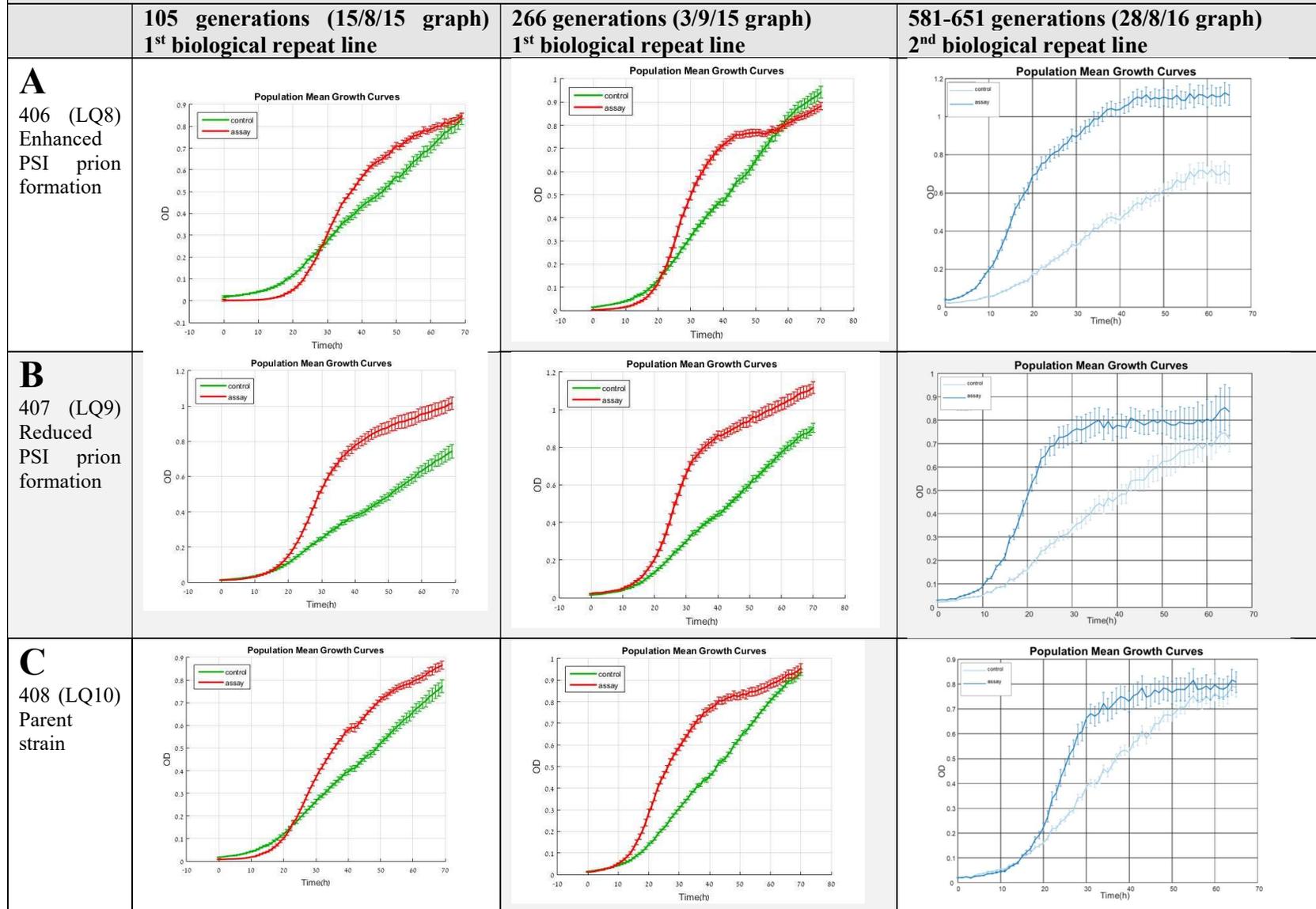
## HSP evolved strains in comparison to their ancestors, 30°, XYL



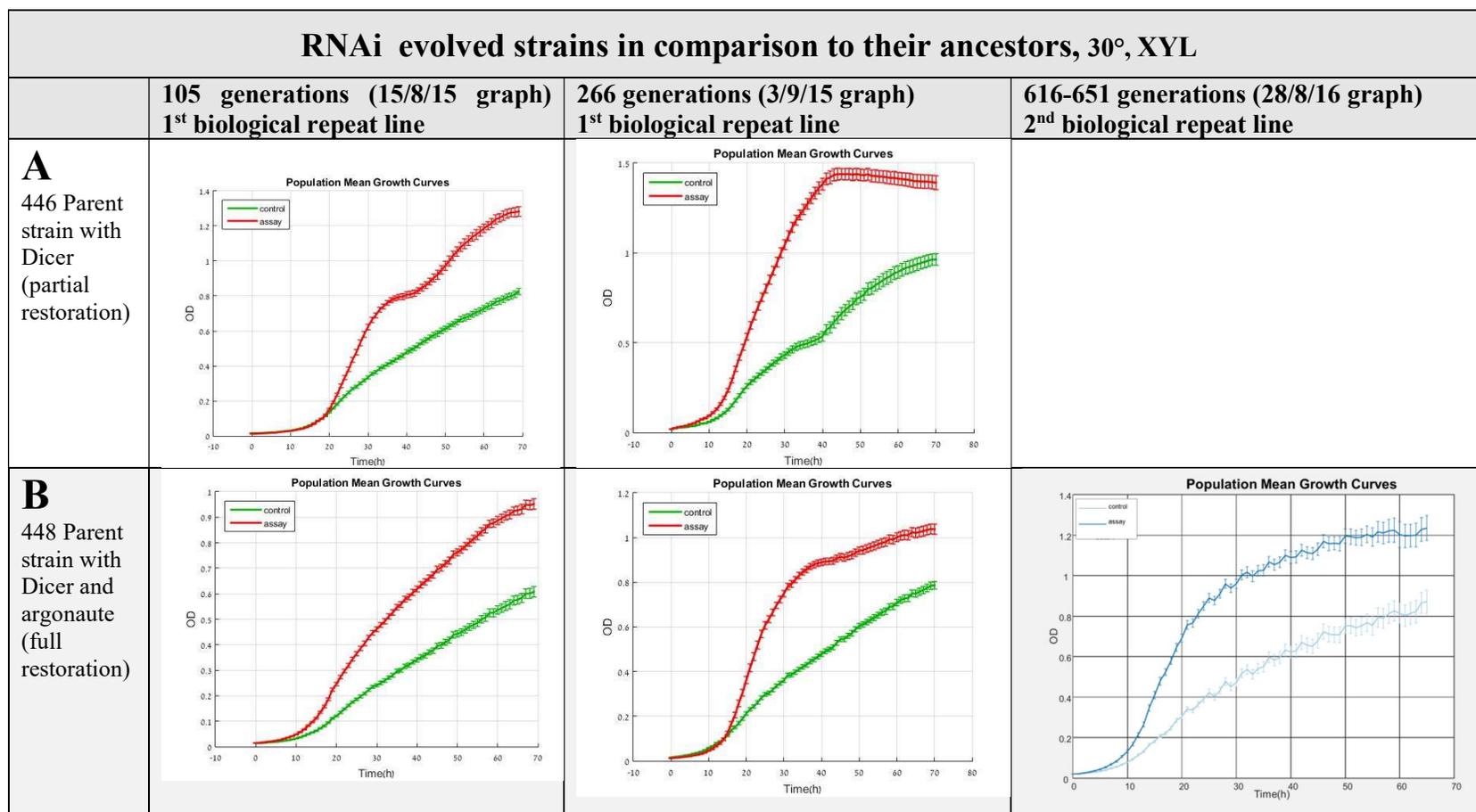


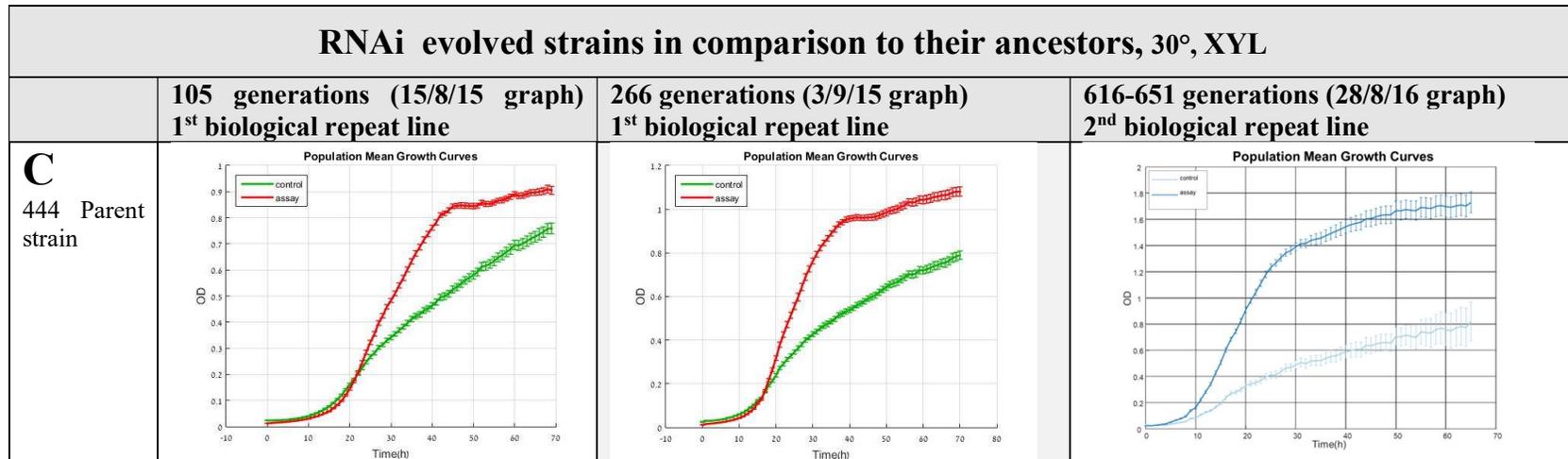
**Figure 4- HSP evolved strains comparison to their ancestor strains**– The following growth curves were produced by a MATLAB analysis of comparative growth experiments that took place using the Hamilton evolution robot. Depicting OD over time, the WT and mutant strains were compared to their ancestor strains at three evolutionary time points- after 120, 190 and 350 generations. The comparison took place by growing the strains on the Xylulose mixture at 30°C (evolution conditions). Results presented here are from the 21-28/8/14, 31/8/14 and 10-15/11/15 growth tests (not to be confused with evolution dates). The last evolution point (350 generations measured at 10-15/11/15) was of lower technical quality resulting in lower OD, yet the trend was consistent and the results, therefore, were included for complicity. The evolved and strains included **(A)** ΔHSC-82 Homozygote (DH1, 022); **(B)** ΔHSP-82 Homozygote (DH3, 202); **(C)** ΔSTI-1 Homozygote (DH6, 220); **(D)** ΔHSC-82 Heterozygote (DH2, 122); **(E)** ΔSTI-1 Heterozygote (DH7, 221). For all the strains the first biological repeat line evoltants data are presented. **Please mind the different OD scales used in this figure.**

## SUP evolved strains in comparison to their ancestors, 30°, XYL



**Figure 5- SUP evolved strains comparison to their ancestor strains**– The following growth curves were produced by a MATLAB analysis of comparative growth experiments that took place using the Hamilton evolution robot. Depicting OD over time, the parent and mutant strains were compared to their ancestor strains at three evolutionary time points- after 105, 266 and 581-651 (pending on the strain) generations. The comparison took place by growing the strains on the Xylulose mixture at 30°C (evolution conditions). Results presented here are from the 15/8/15, 3/9/15 and 28/8/16 growth tests (not to be confused with evolution dates), to include (A) 406 (LQ8)- 74D with enhanced PSI prion formation; (B) 407 (LQ9) - 74D reduced PSI prion formation); (C) 408 (LQ10)- the parent 74D strain. For all the strains the first biological repeat line evoltants data are presented at 105 and 266 generations, and the 2<sup>nd</sup> biological repeat line evoltants data are presented at 651 (for A), 609 (for B) and 581(for C) generations. **Please mind the different OD scales used in this figure.**





**Figure 6- RNAi evolved strains comparison to their ancestor strains**– The following growth curves were produced by a MATLAB analysis of comparative growth experiments that took place using the Hamilton evolution robot. Depicting OD over time, the parent and mutant strains were compared to their ancestor strains at three evolutionary time points- after 105, 266 and 616-651 (pending on the strain) generations. The comparison took place by growing the strains on the Xylulose mixture at 30°c (evolution conditions). Results presented here are from the 15/8/15, 3/9/15 and 28/8/16 growth tests (not to be confused with evolution dates), to include (A) 446- The parent strain with dicer added (partial RNAi restoration); (B) 448- The parent strain with dicer and argonaute added (full RNAi restoration); (C) 444- the parent strain from which the other two were constructed. For all the strains the first biological repeat line evoltants data are presented at 105 and 266 generations, and the 2<sup>nd</sup> biological repeat line evoltants data are presented at 651 (for A,) and 616(for C) generations. **Please mind the different OD scales used in this figure.**

As was demonstrated above, in the course of the lab evolution, significant improvement was accomplished for all the strains. This improvement due to evolution was consistent for all the independent biological repeats, as was evident during the daily dilutions, and as was verified in spot growth tests. An example of such verification is demonstrated in

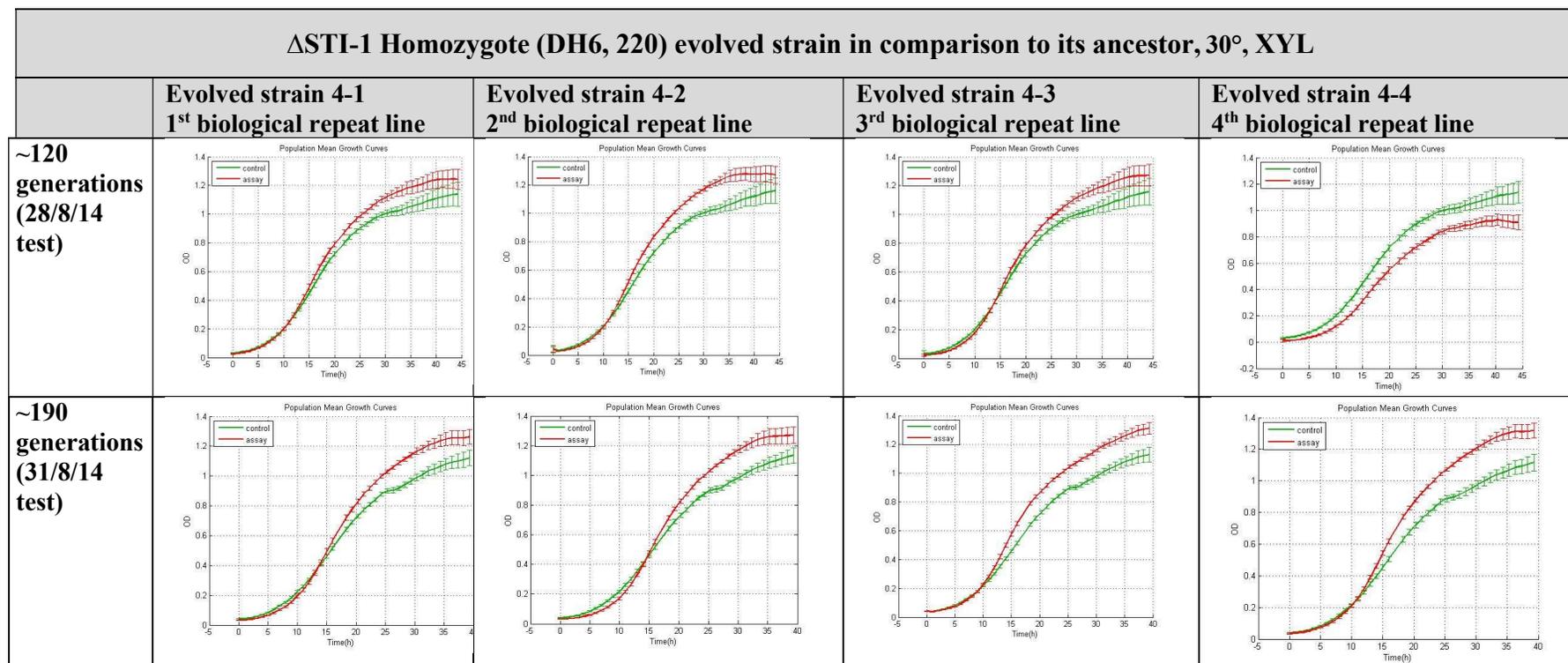
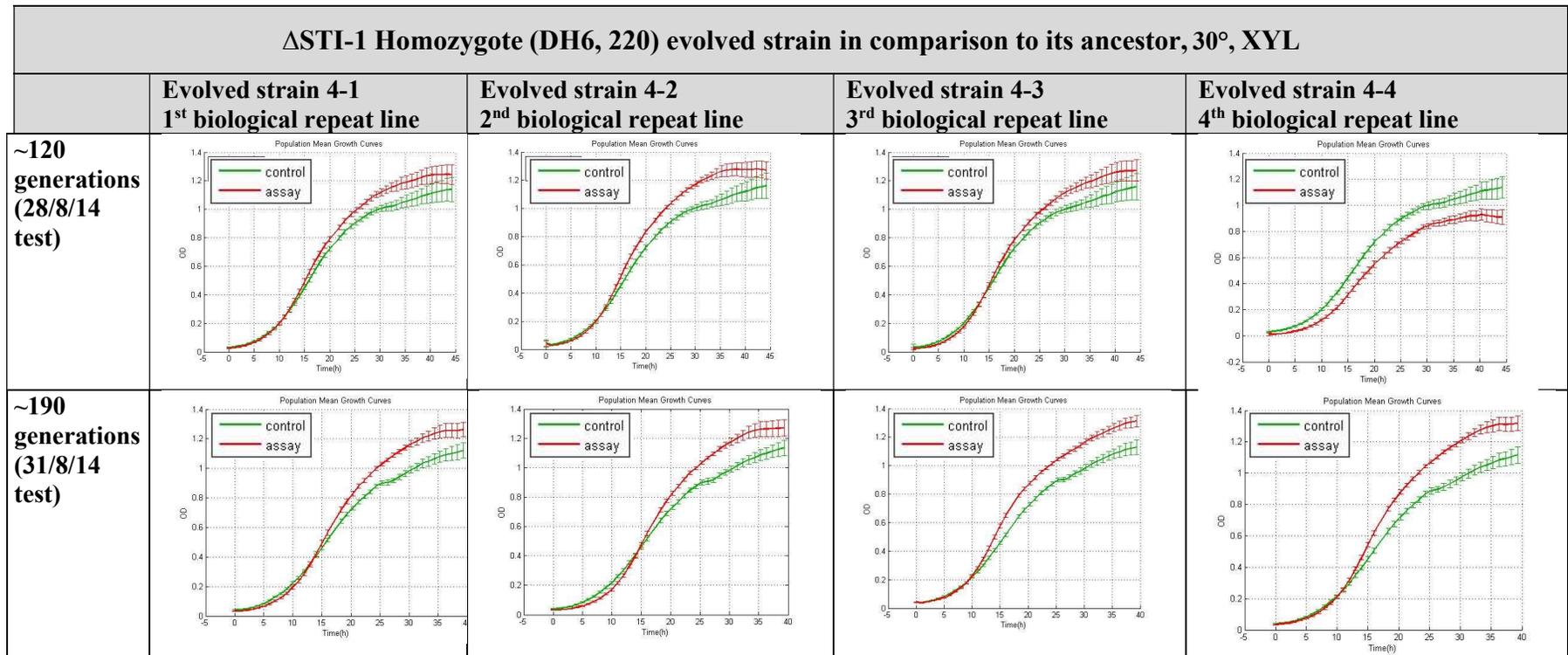
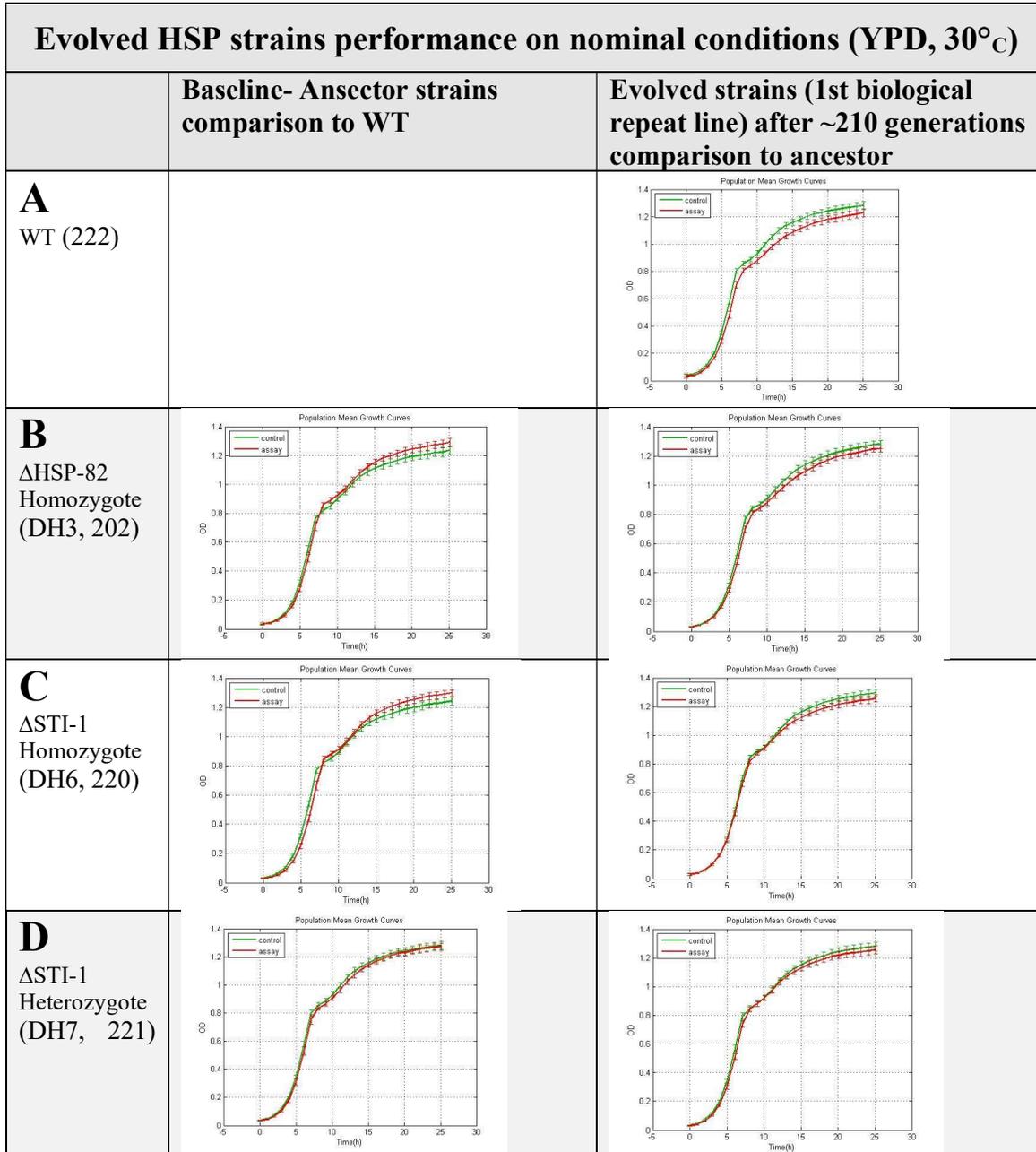


Figure 7 below for HSP90  $\Delta$ STI-1 Homozygote (DH6, 220) strain at two evolutionary points. Additional results for the other five HSP-90 strains were consistent with this result yet are not presented for brevity.



**Figure 7- HSP  $\Delta$ STI-1 Homozygote (DH6, 220) mutant evolved strain comparison to its ancestor strain for all four independent biological repeats–** The following growth curves were produced by a MATLAB analysis of comparative growth experiments that took place using the Hamilton evolution robot. Depicting OD over time, the mutant strain's four independent biological repeats were compared to their ancestor strains at two evolutionary time points- after 120 and 190 generations. The comparison took place by growing the strains on the Xylulose mixture at 30°c (evolution conditions). Results presented here are from the 28/8/14, and 31/8/14 growth tests (not to be confused with evolution dates).

Finally, as an indicator of whether their evolutionary adaptation to Xylose resulted in significant general fitness cost (specification), the evolved strains were compared to their ancestors by performing growth tests on nominal conditions (i.e. 30°C, YPD), as is demonstrated for some of the HSP-90 strains in Figure 8 below. As a baseline, the initial comparison of the ancestor mutant strains to WT before the evolution started is also presented.



**Figure 8- HSP 90 evolved strains comparison to ancestors on YPD as an indication for general loss of fitness during evolution** – The following growth curves were produced by a MATLAB analysis of comparative growth experiments that took place using the Hamilton evolution robot. Depicting OD over time, the evolved WT and some of the mutant strains, after ~210 generations of evolution on the Xylulose medium comparison to their ancestor strains while grown over YPD at 30°C (nominal conditions, the test took place on 6/9/2014), is presented in the righthand column, as an indication to their loss of generality. As a baseline, A

growth test comparing these mutants' ancestor strain to the ancestor WT while grown on nominal conditions (also from 6/9/2014)) is presented in the lefthand column, to include **(A)** WT; **(B)**  $\Delta$ HSP-82 Homozygote (DH3, 202); **(C)**  $\Delta$ STI-1 Homozygote (DH6, 220); **(D)**  $\Delta$ STI-1 Heterozygote (DH7, 221)

As can be seen from the above results, the evolved mutants' loss of generality and fitness on nominal conditions was minimal, and not significantly different from the WTs.

Thus, so far we have established-

1. Initial growth defect detected and quantified (mainly for RNAi strains, since the SUP strains had no defect, and the defect was negligible for the HSP strains to exclude  $\Delta$ HSP-82 Homozygote that showed a small defect).
2. Significant improvement was accomplished by the lab evolution for all the tested strains.
3. The above evolutionary adaptation was consistent for all the independent biological repeats.
4. No significant general fitness loss (specification) was caused due to the evolution.

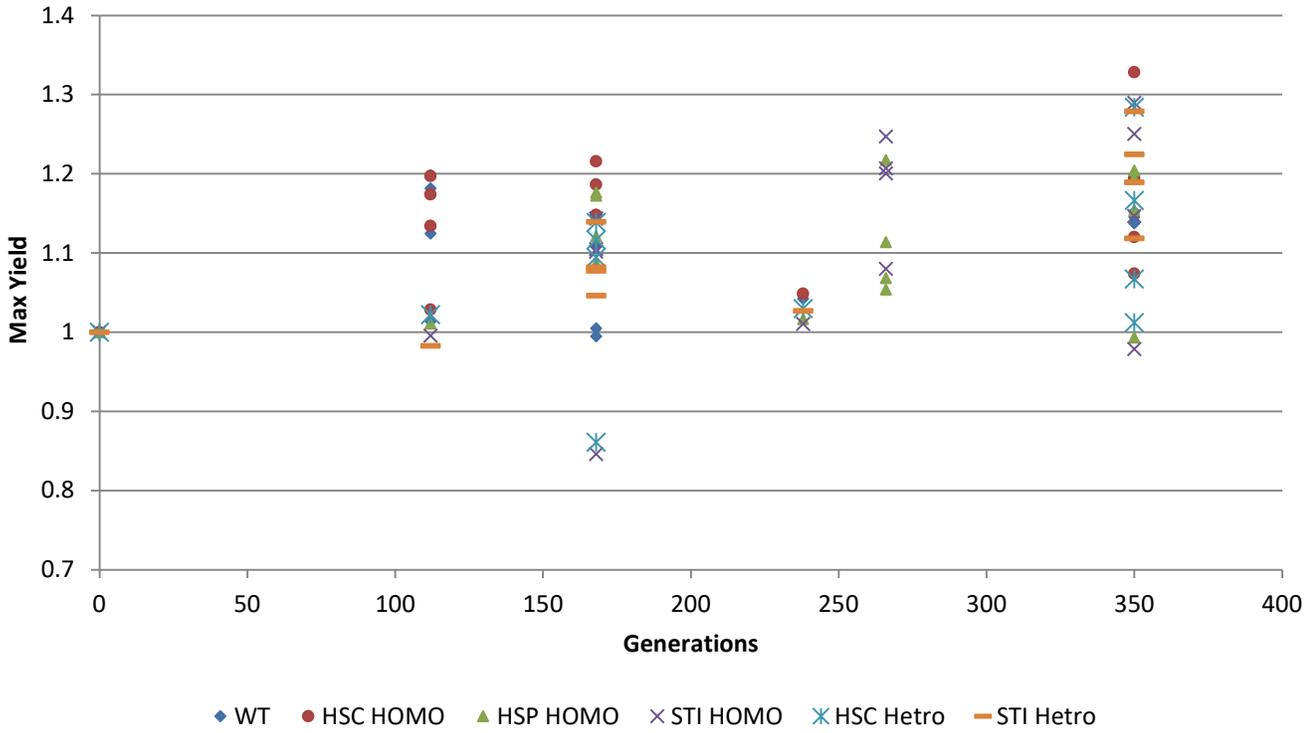
As was presented in the previous section (“Materials and Methods”), since we were interested in the evolution dynamics (i.e. fitness improvement over evolution time, and not only in the total improvement that was accomplished after it ran its course), and since visually comparing a large number of growth curves is not feasible, the curveball method [41] that was described above was used, and the strains' growth curves at different evolution time points were parameterized to enable the comparison. The following figures present the three main parameters (max growth, max yield and lag) dynamics across evolution time for each mechanism. For each parameter at each evolutionary time point its value for the evoltant is compared to its value for the ancestor strain that served as reference and the relative value is presented (for example the presented lag value =  $\text{lag}_{\text{evoltant}}/\text{lag}_{\text{ancestor}}$ ), hence equaling one at time point zero.

#### HSP-90 (protein chaperone)

Fitness parameters for the HSP-90 evolution are presented in Figure 9 below, to include a trend lines view for clarity.

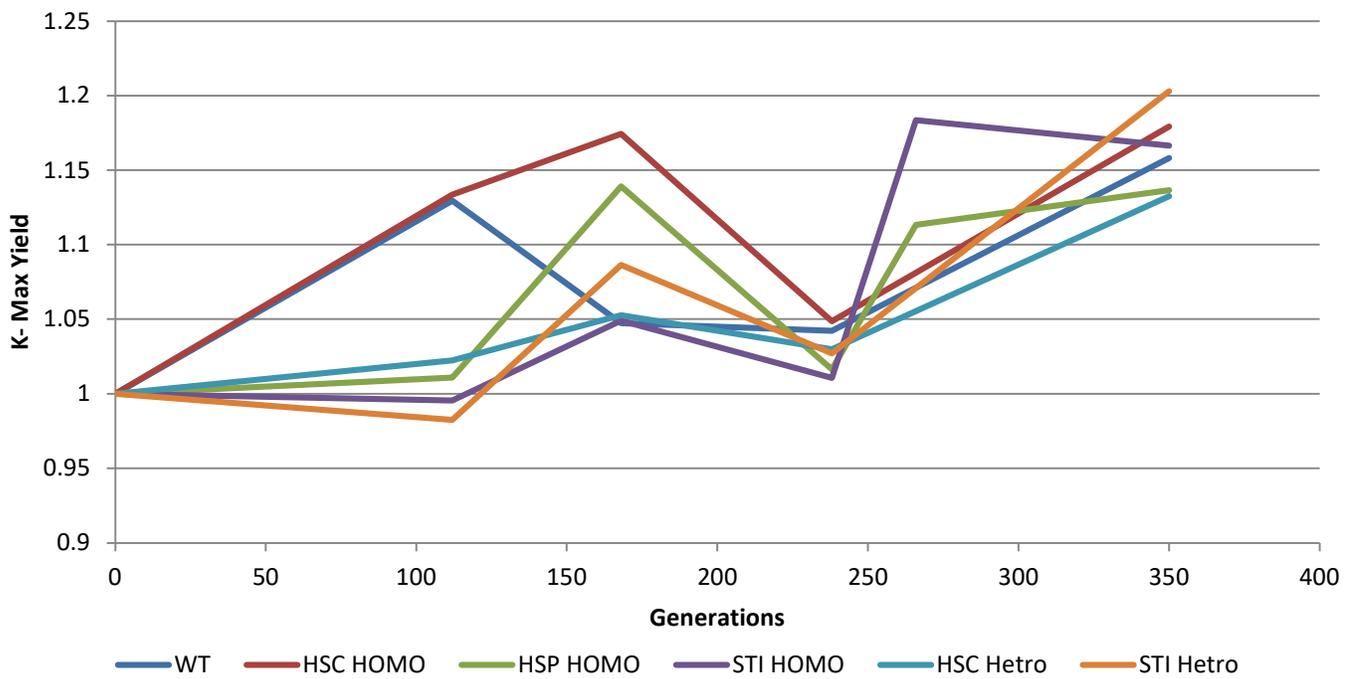
# A1

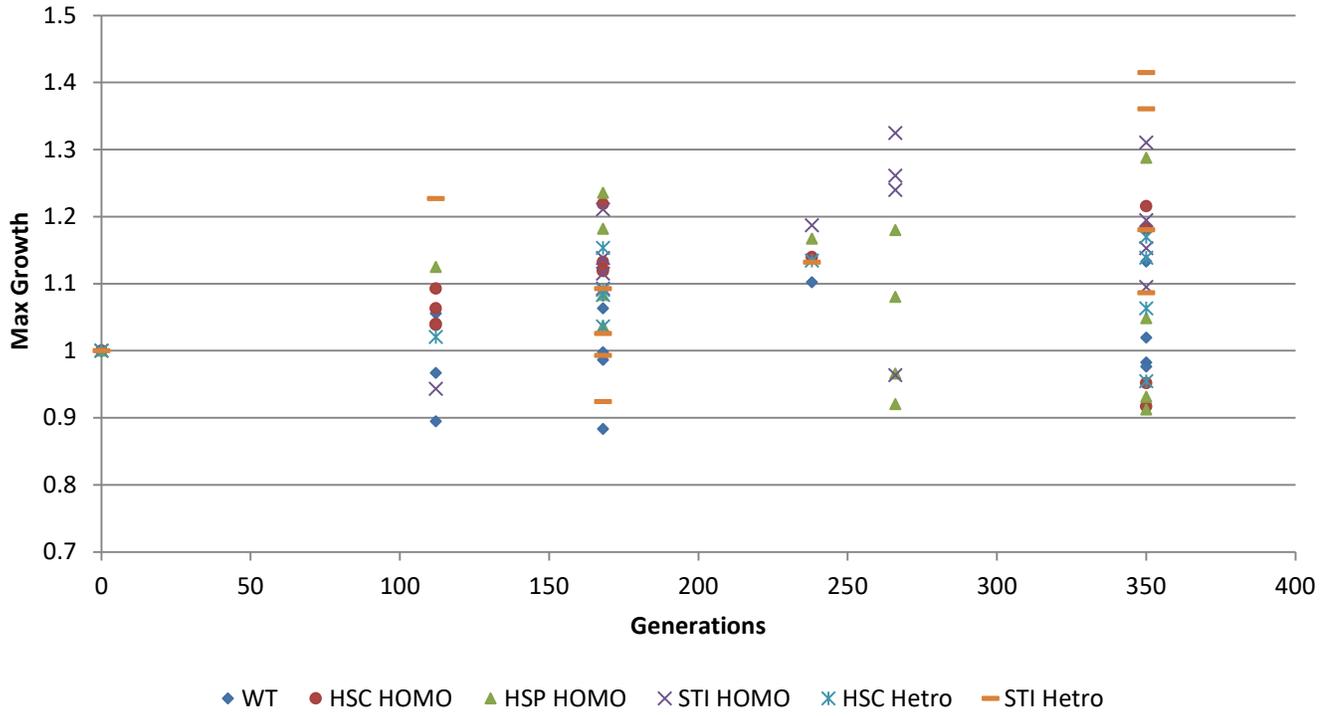
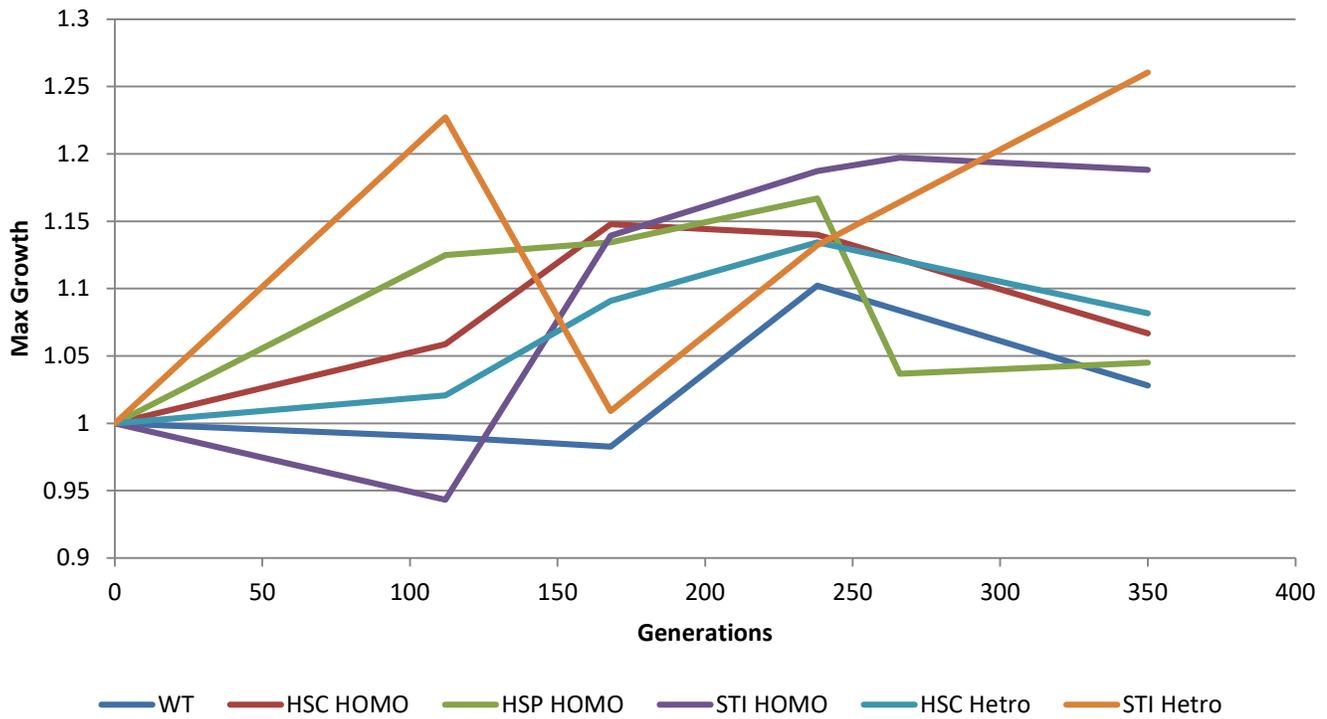
## HSP strains EVOL - Max Yield

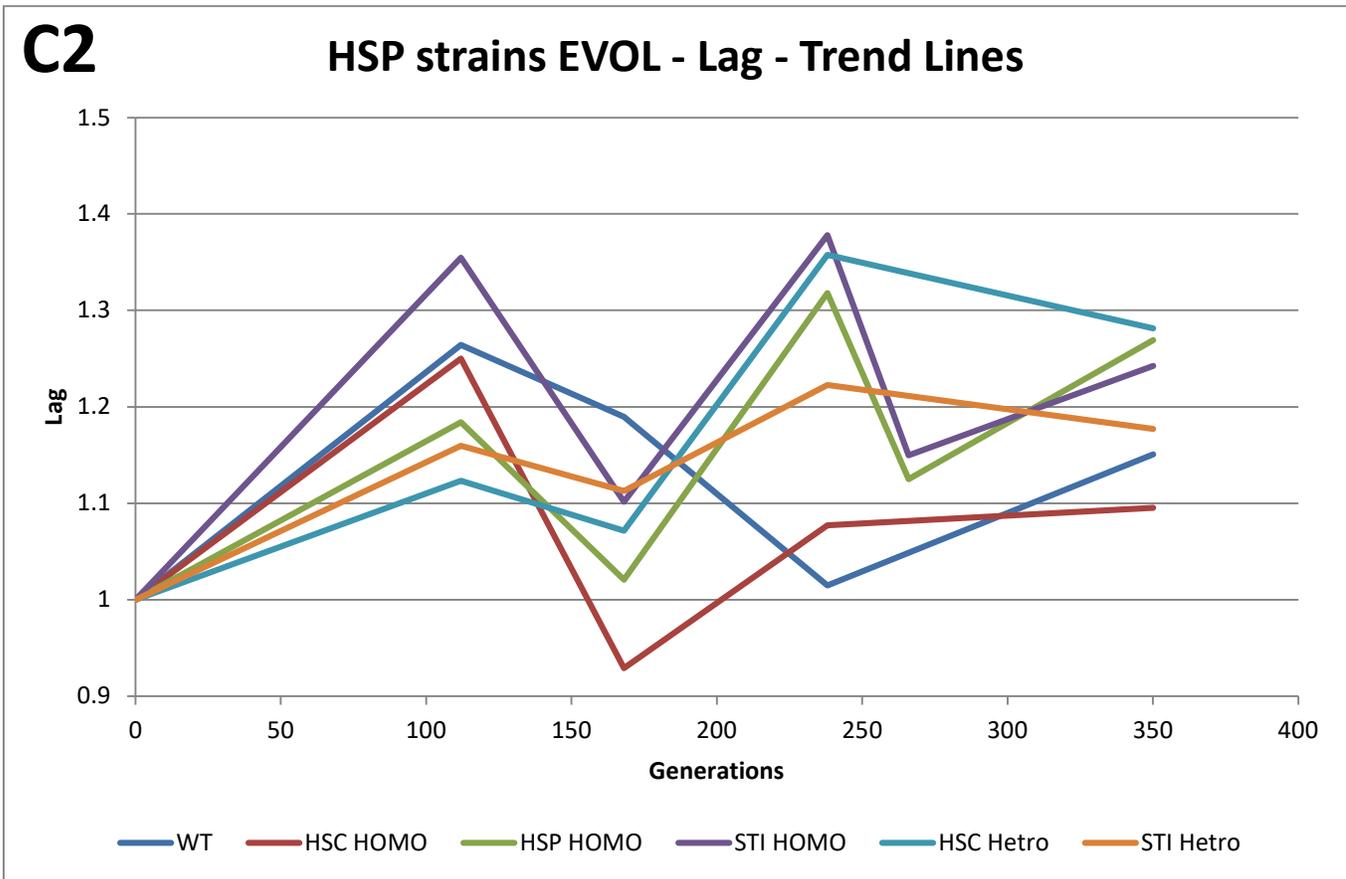
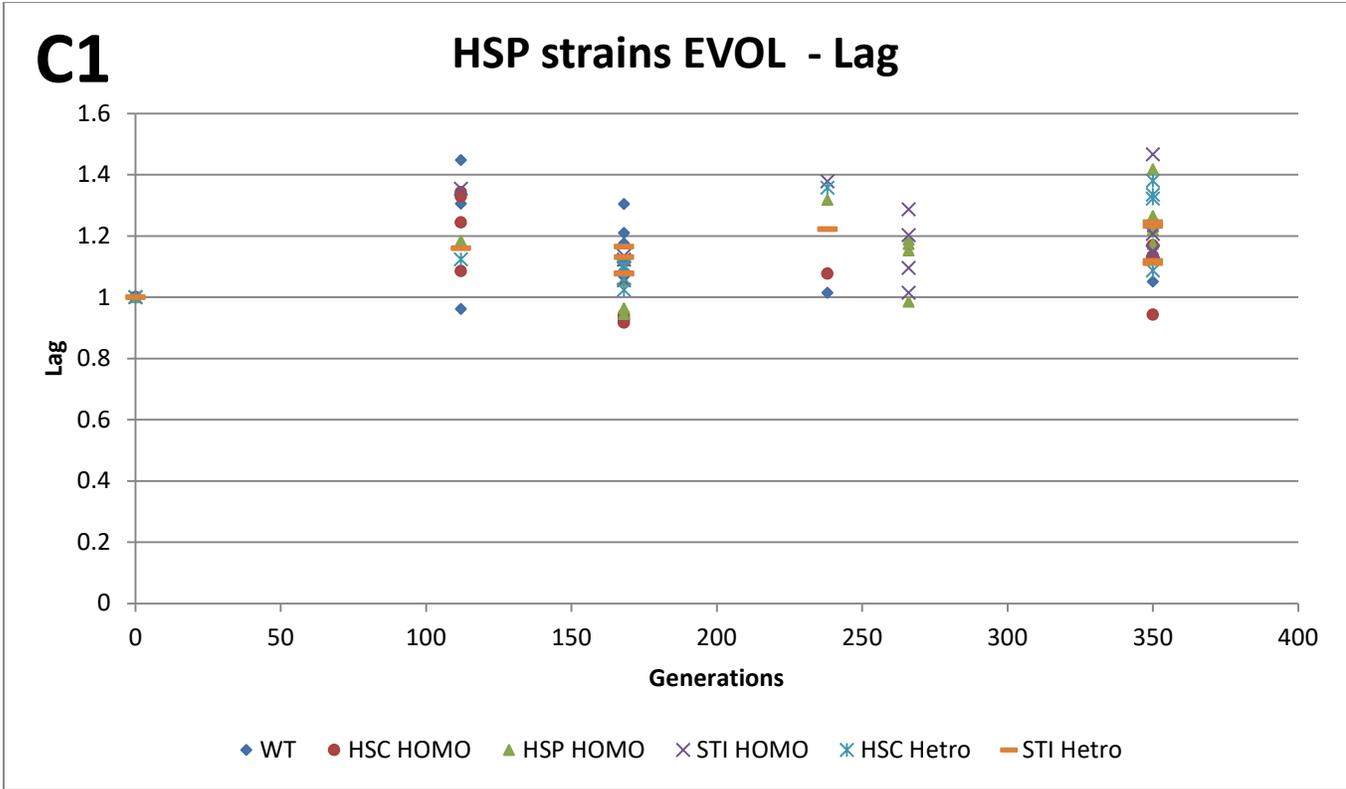


# A2

## HSP strains EVOL - Max Yield - Trend Lines



**B1****HSP strains EVOL - Max Growth****B2****HSP strains EVOL - Max Growth - Trend Lines**



**Figure 9- HSP 90 parameters across evolution time** – This figure presents HSP-90 growth parameters across evolution as calculated from the strains' growth curves by using the curveball method to include **(A)** Max Yield; **(B)** Max Growth; **(C)** Lag. For each parameter at each evolutionary time point its value for the evoltant is compared to its value for the ancestor strain that served as reference and the relative value is presented (for example lag value =  $\text{lag}_{\text{evoltant}}/\text{lag}_{\text{ancestor}}$ ) hence equaling one at time point zero. Results are presented for the WT, the  $\Delta\text{HSC-82}$  Homozygote (DH1, 022), the  $\Delta\text{HSP-82}$  Homozygote (DH3, 202), the  $\Delta\text{STI-1}$  Homozygote (DH6, 220), the  $\Delta\text{HSC-82}$  Heterozygote (DH2, 122) and the  $\Delta\text{STI-1}$  Heterozygote (DH7, 221) For Each parameter two views are presented to enable a better impression from the data- a classic view **(1)** where each biological repeat is separately represented and a trend line view **(2)** where the average parameter value for all the biological repeats of a strain at that evolutionary time point is calculated and presented as part of a trend line.

From the above initial results, it seems that the HSP-90 strains have presented different evolutionary dynamics than the WT (Parent) strain, specifically-

The max yield dynamic seemed different for the HSC homozygote (starting and remaining higher than the WT) and the STI homozygote (starting lower and ending higher, as was the STI heterozygote). The other strains behaved much like the WT, to exclude an initial rise and drop (WT only) that is based on a single time point and hence may not be that reliable.

As to the max growth dynamic, here all the mutants did better than the WT, with the STI mutants in particular (both homozygote and heterozygote).

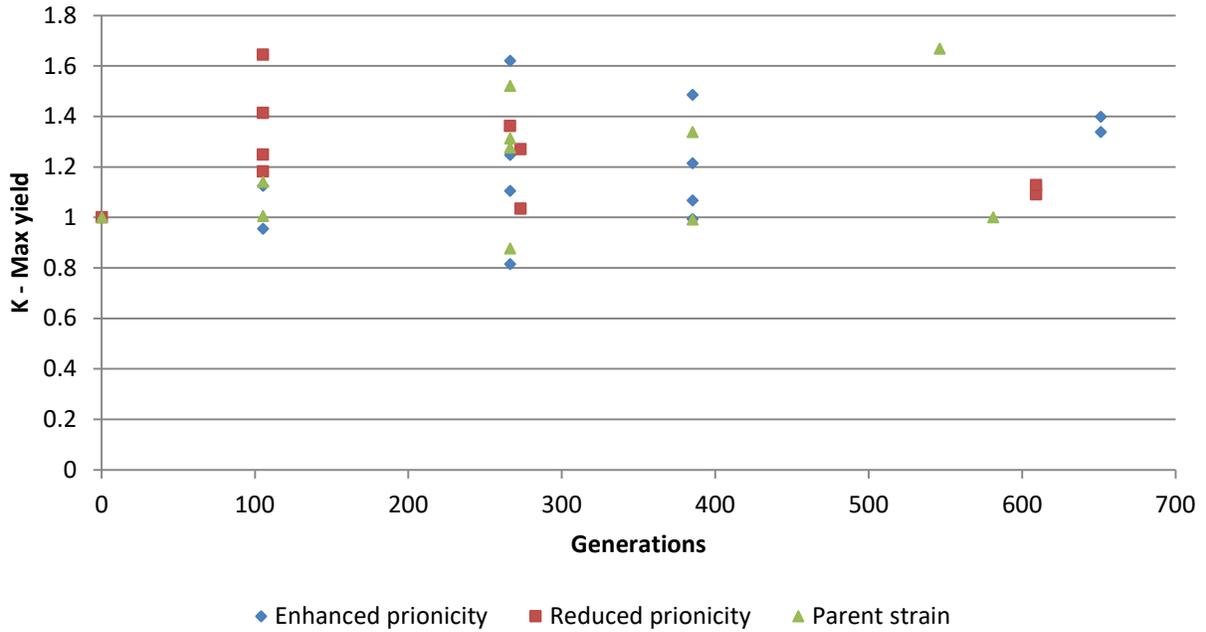
Finally, regarding the lag dynamic, here all the mutants behaved differently than the WT (except maybe the HSC homozygote), with the WT rising and then dropping to remain at approximately the original level, and the other strains rising and ending on a higher (longer lag time) level.

### SUP-35 prions

Fitness parameters for the SUP-35 evolution are presented in Figure 10 below, to include a trend lines view for clarity.

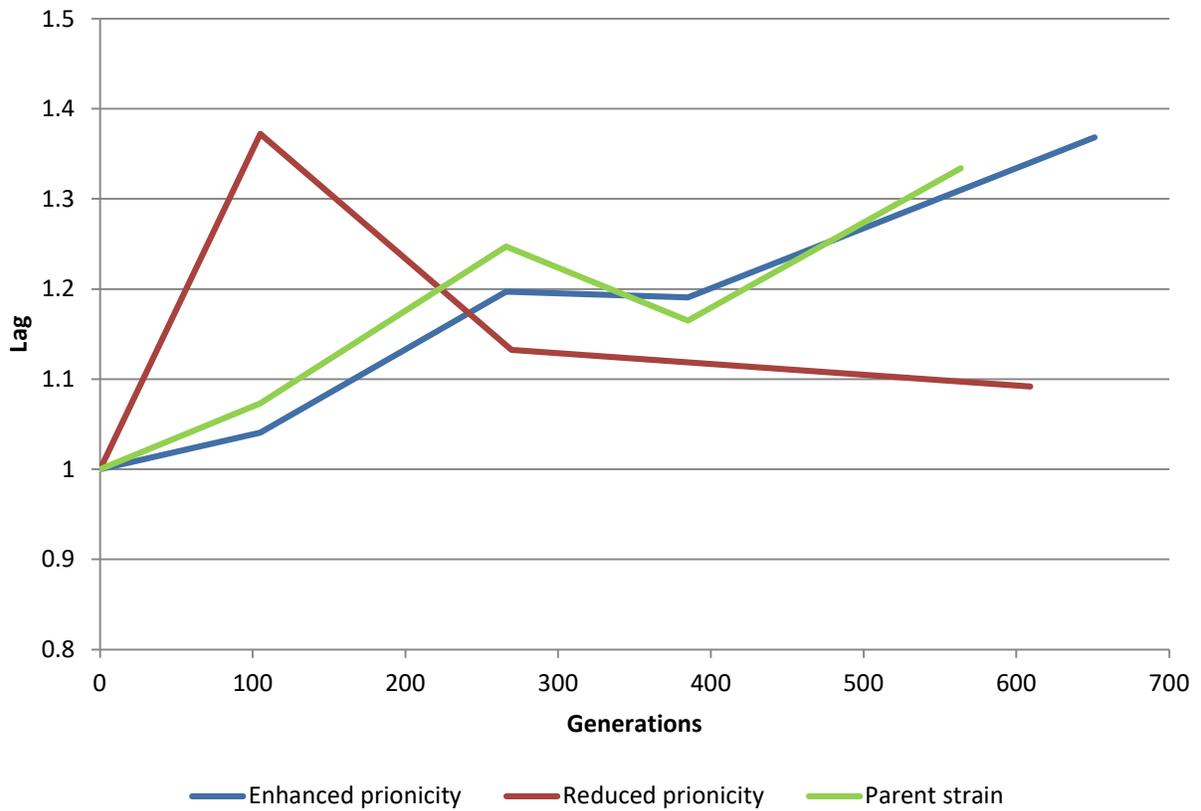
# A1

## SUP strains EVOL- Max Yield



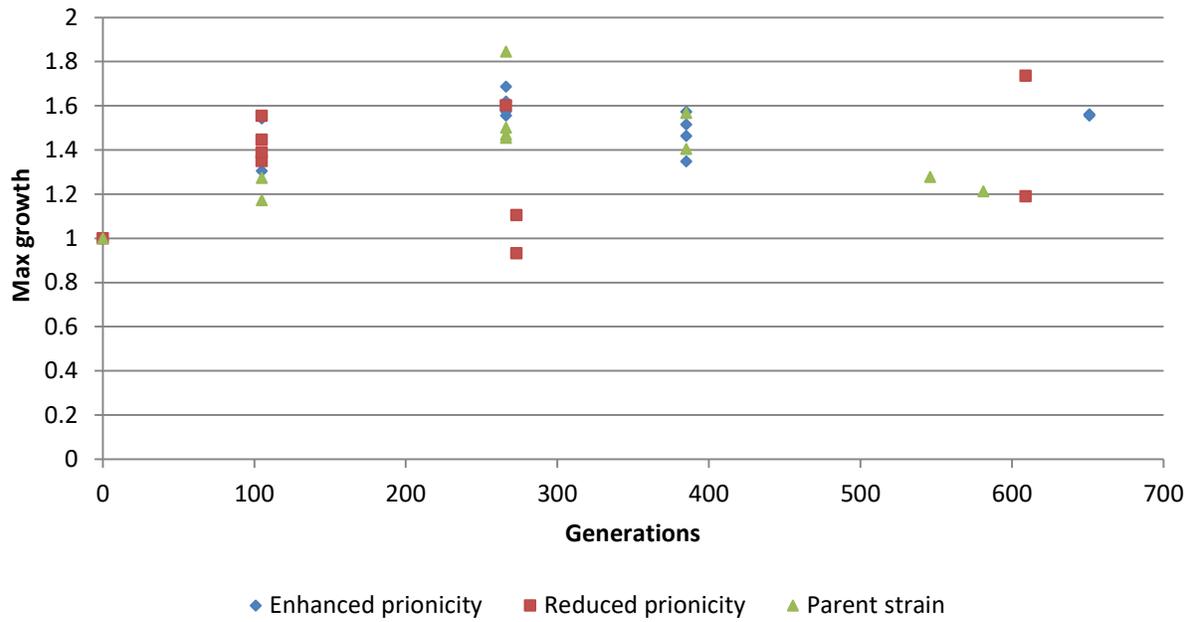
# A2

## SUP strains EVOL - Max Yield - Trend Lines



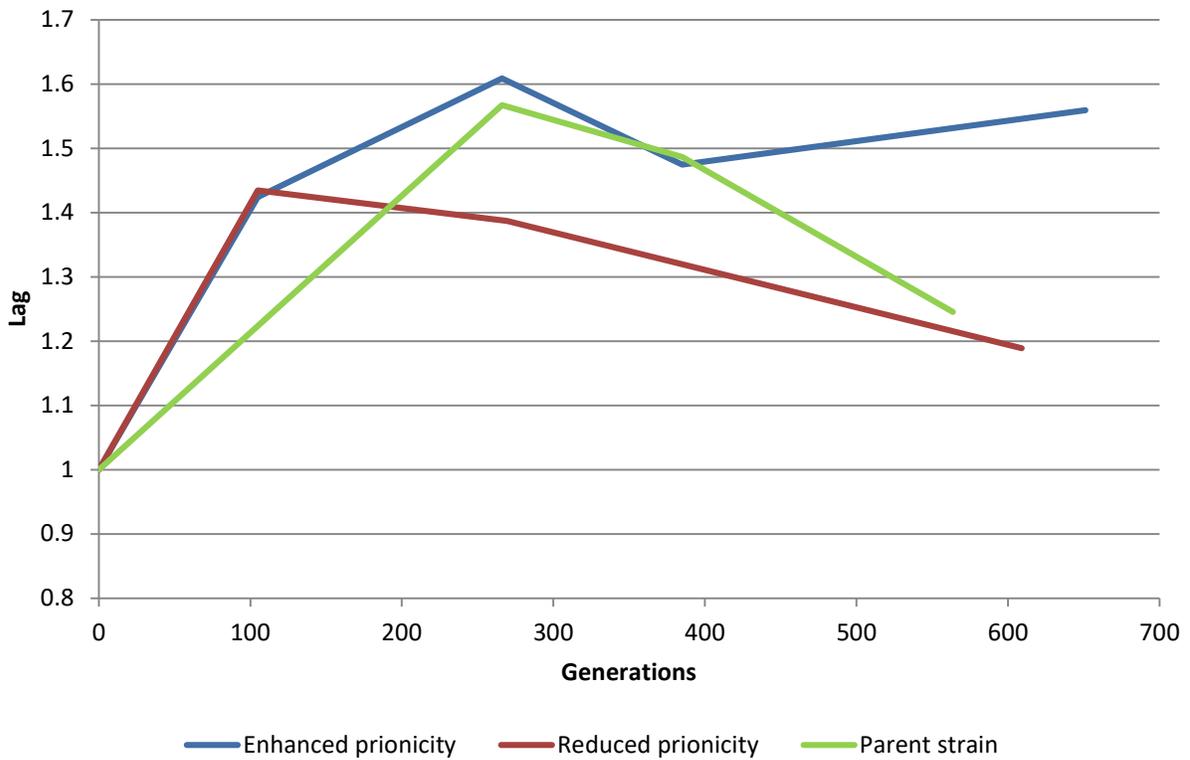
# B1

## SUP strains EVOL- Max growth



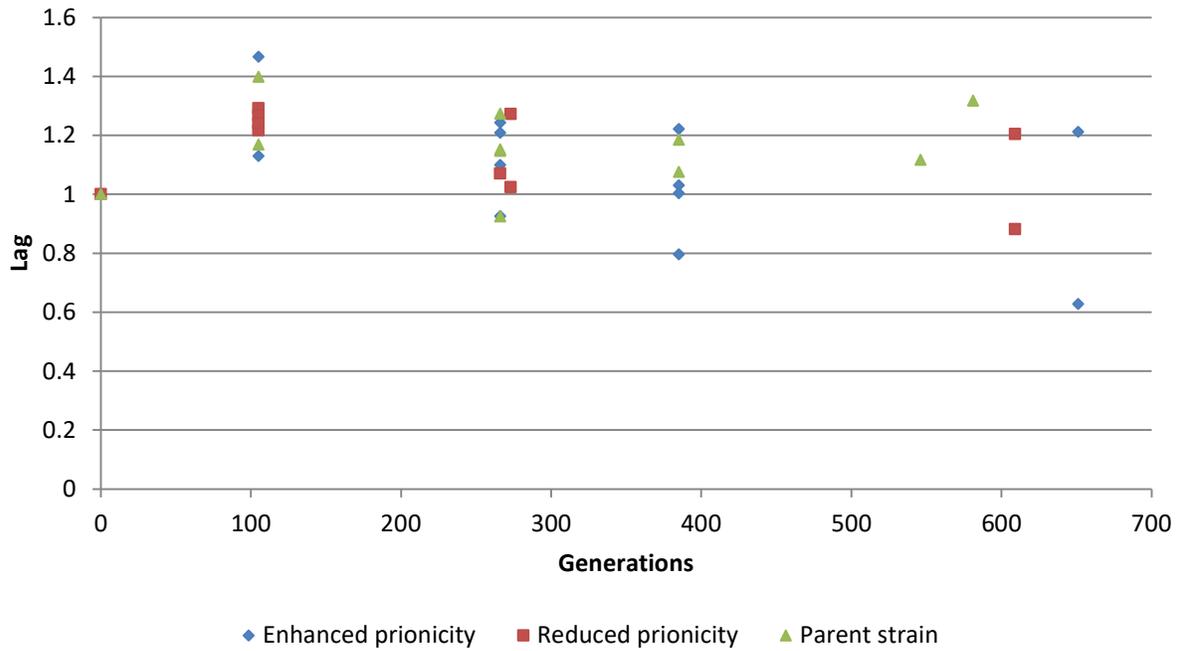
# B2

## SUP strains EVOL - Max growth - Trend Lines



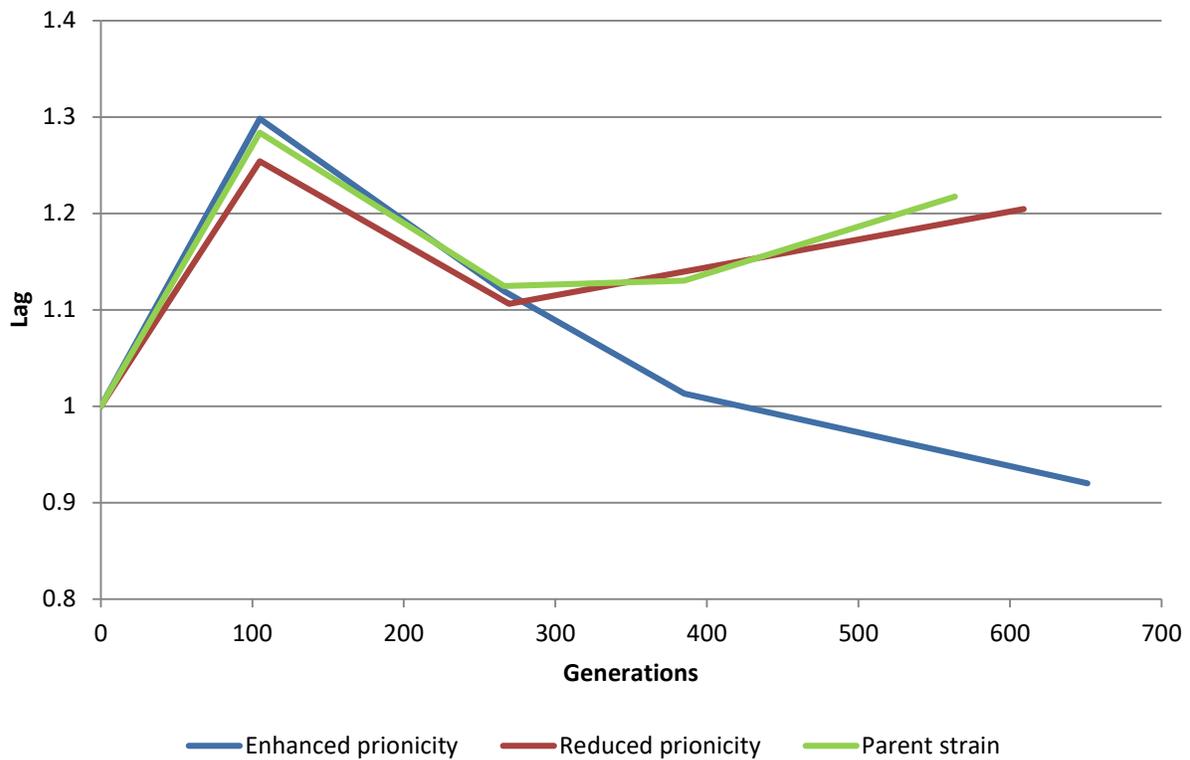
# C1

## SUP strains EVOL- Lag



# C2

## SUP strains EVOL - Lag - Trend Lines



**Figure 10- SUP-35 parameters across evolution time** – This figure presents SUP growth parameters across evolution as calculated from the strains' growth curves by using the curveball method to include **(A)** Max Yield; **(B)** Max Growth; **(C)** Lag. For each parameter at each evolutionary time point its value for the evoltant is compared to its value for the ancestor strain that served as reference and the relative value is presented (for example lag value =  $\text{lag}_{\text{evoltant}}/\text{lag}_{\text{ancestor}}$ ) hence equaling one at time point zero. Results are presented for the 406 (LQ8) enhanced prionicity strain, the 407 (LQ9) reduced prionicity strain and the 408 (LQ10) parent strain. For Each parameter, two views are presented to enable a better impression of the data- a classic view **(1)** where each biological repeat is separately represented and a trend line view **(2)** where the average parameter value for all the biological repeats of a strain at that evolutionary time point is calculated and presented as part of a trend line. At the trend line view, day 546 and 581 parent strain data and day 266 and 273 reduced prionicity strain data were averaged together to compensate for a relatively high scatter for some of the parameters, as can be seen in the classic view charts.

As can be seen above, the SUP-35 initial results were ambiguous, to say the least, With maybe a different max yield dynamics for the reduced prionicity strain (with the parent and enhanced prionicity strains practically identical), and a different max growth and lag dynamics for the enhanced prionicity strain (with the parent and reduced prionicity strains practically identical).

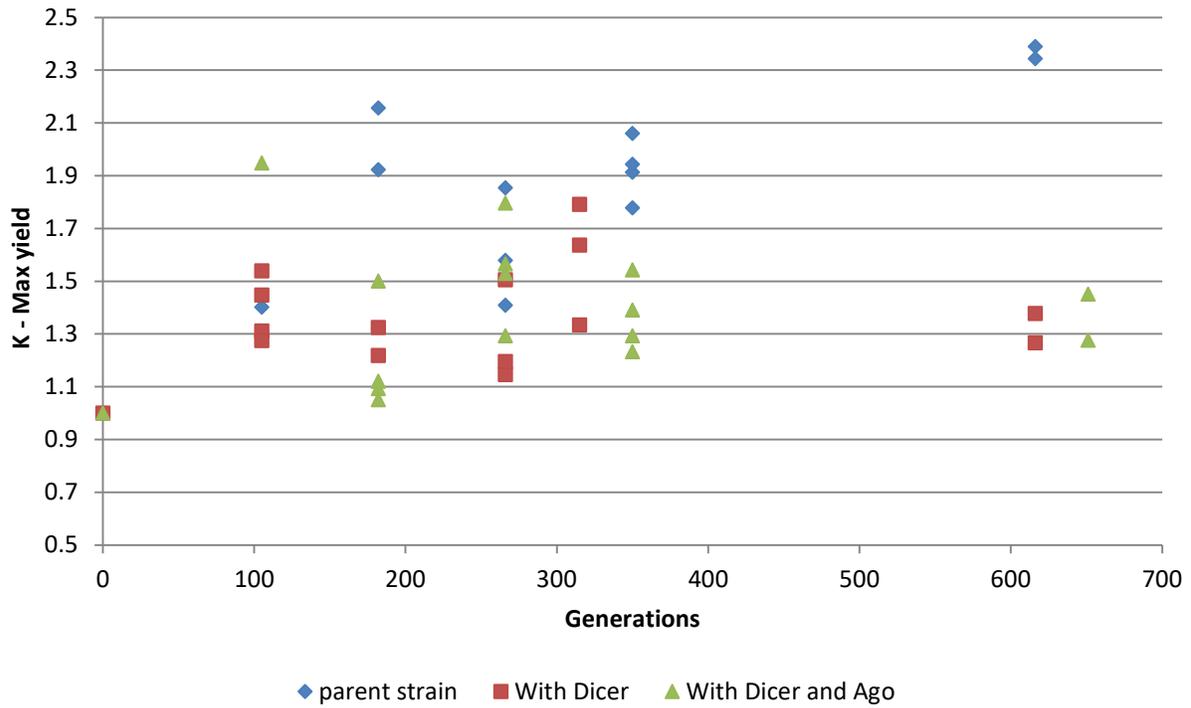
However shortly after this study was completed, other studies at the lab raised some doubts regarding these strains' prion expression conditions, and hence this assay was considered the lowest priority for further investigation and was not pursued.

### RNAi

Fitness parameters for the RNAi evolution are presented in Figure 11 below, to include a trend lines view for clarity.

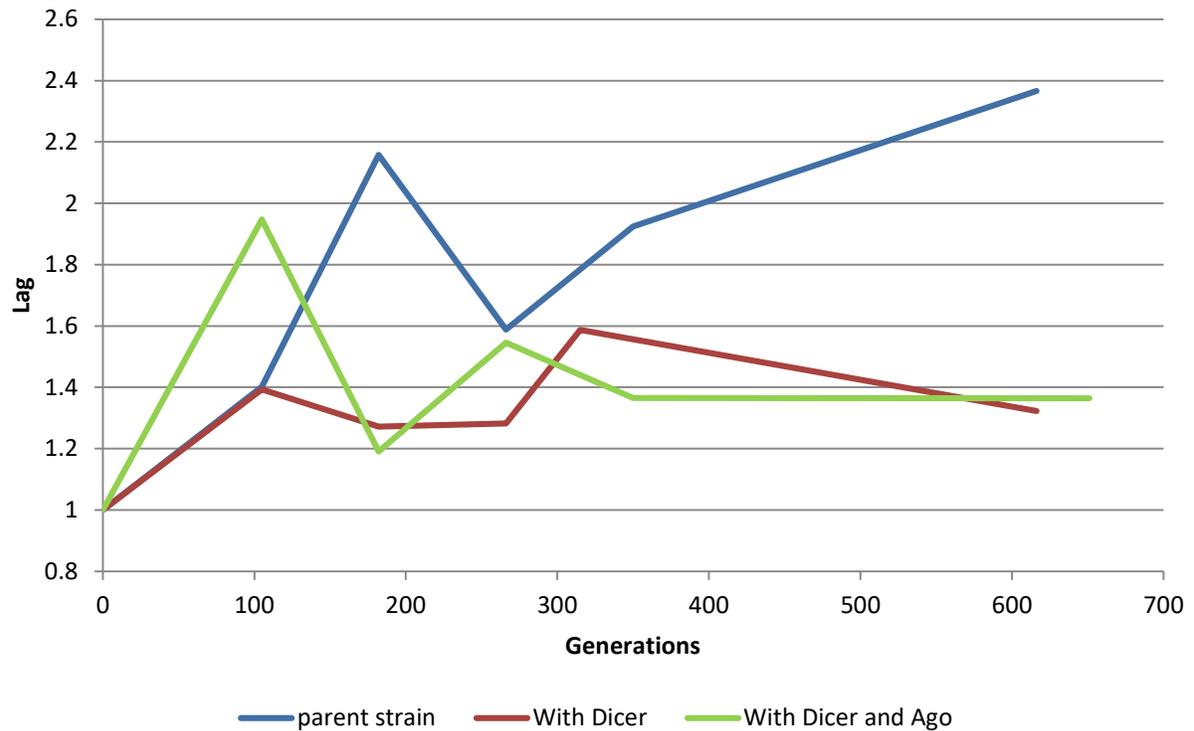
# A1

## RNAi strains EVOL- Max yield



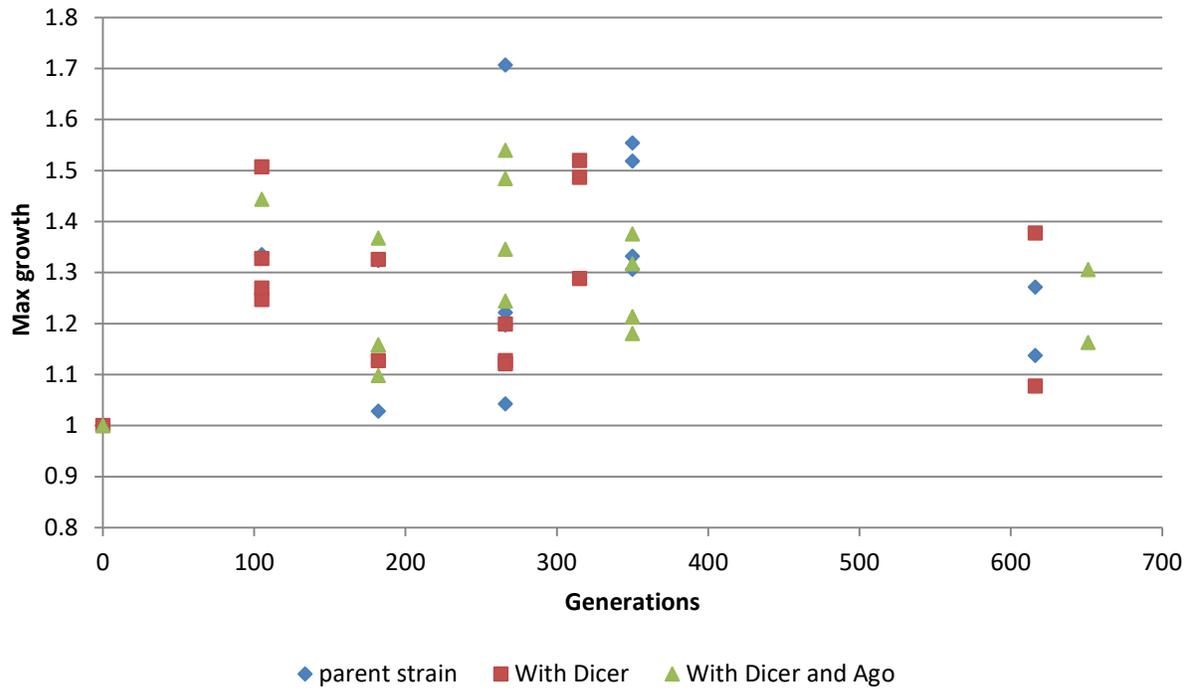
# A2

## RNAi strains EVOL - Max Yield - Trend Lines



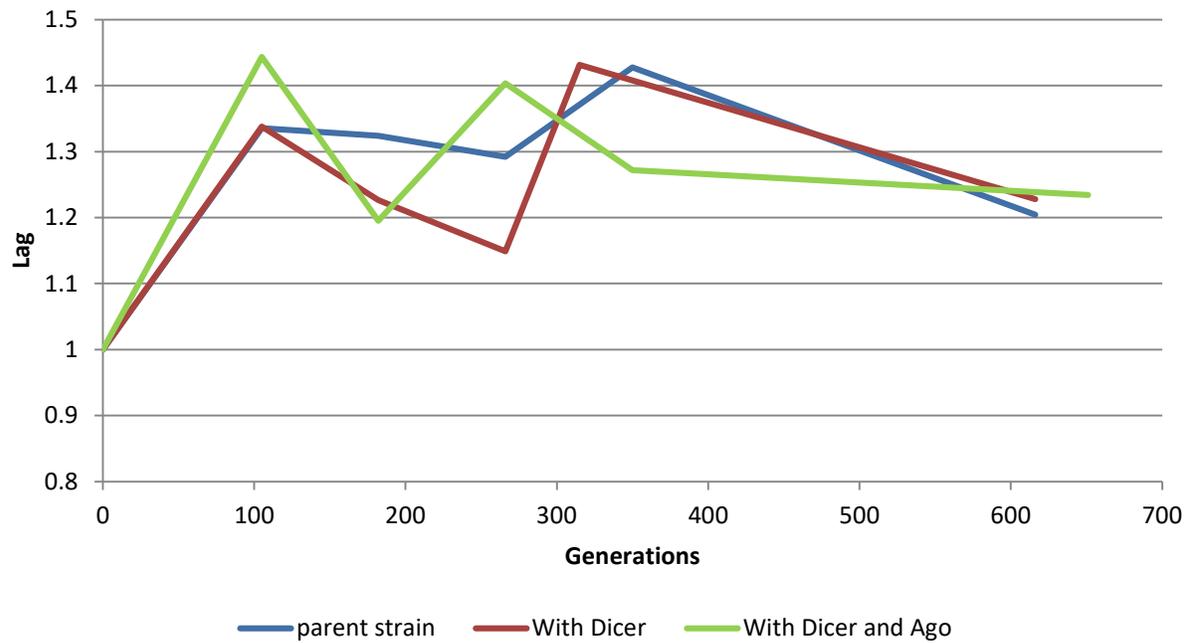
# B1

## RNAi strains EVOL- Max growth



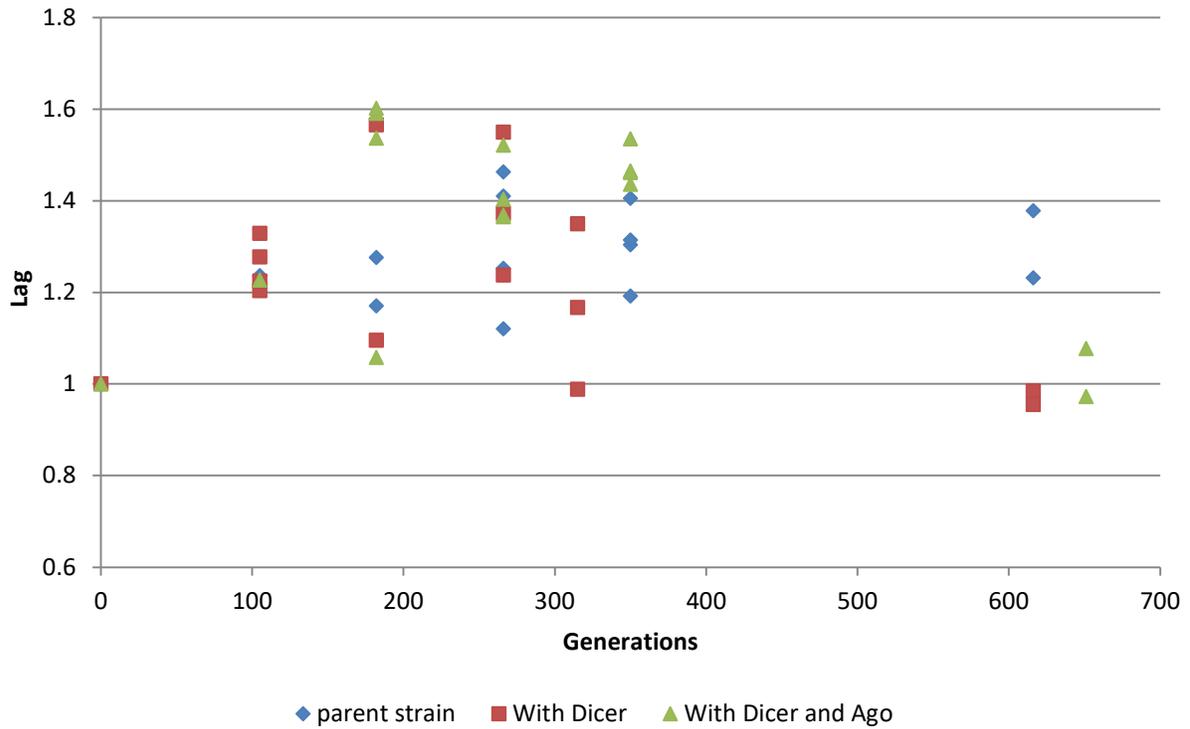
# B2

## RNAi strains EVOL - Max growth - Trend Lines



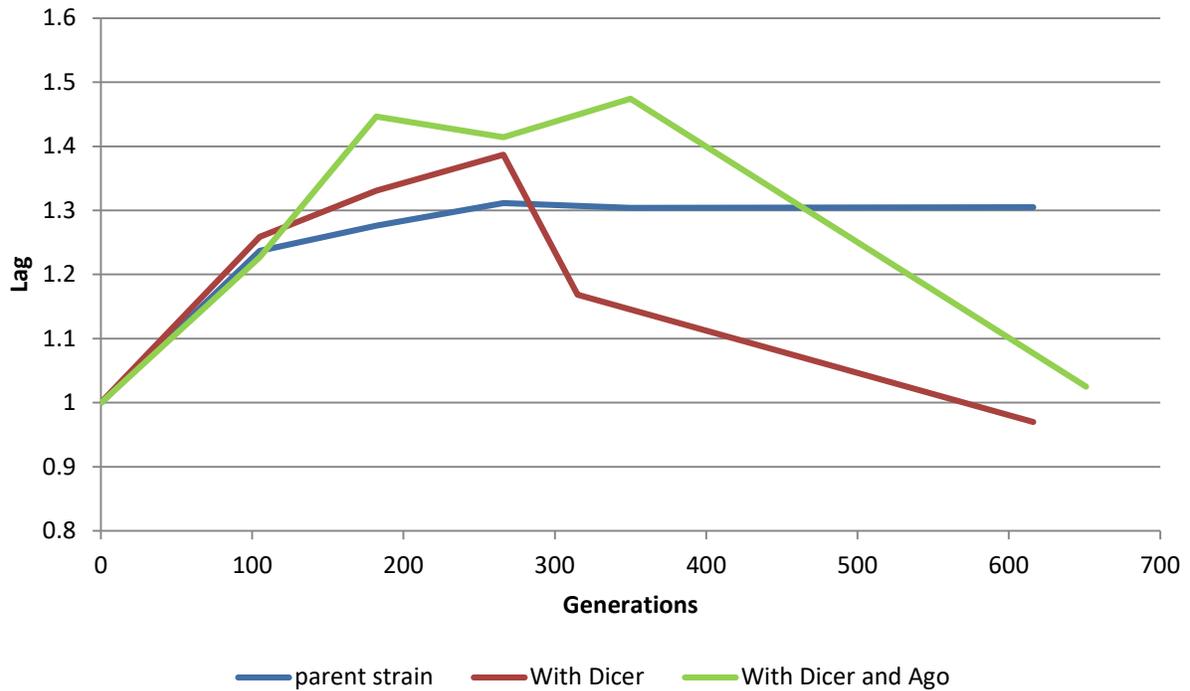
# C1

## RNAi strains EVOL- Lag



# C2

## RNAi strains EVOL - Lag - Trend Lines



**Figure 11- RNAi parameters across evolution time** – This figure presents RNAi growth parameters across evolution as calculated from the strains' growth curves by using the curveball method to include **(A)** Max Yield; **(B)** Max Growth; **(C)** Lag. For each parameter at each evolutionary time point its value for the evoltant is compared to its value for the ancestor strain that served as reference and the relative value is presented (for example lag value =  $\text{lag}_{\text{evoltant}}/\text{lag}_{\text{ancestor}}$ ) hence equaling one at time point zero. Results are presented for strains 446 (the parent strain with dicer added - partial RNAi restoration), 448 (the parent strain with dicer and argonaute added -full RNAi restoration) and 444 (the parent strain from which the other two were constructed). For Each parameter, two views are presented to enable a better impression of the data- a classic view **(1)** where each biological repeat is separately represented and a trend line view **(2)** where the average parameter value for all the biological repeats of a strain at that evolutionary time point is calculated.

As can be seen from the above results, the strains with the restored RNAi mechanism exhibited different evolutionary dynamics than the parent strain. These results are less comprehensive than desired, yet it is evident that the final max yield was significantly lower for the restored RNAi strains (even for the strain with only a partial restoration, i.e. dicer only), and so were the final lag values (shorter). For both parameters, the restored RNAi strains dynamics were similar to each other and different from the parent strain. As for the max growth parameter- no conclusive results emerge, yet no significant difference between the three strains is apparent.

### 3.4 Chapter Summary and Discussion

In this chapter, strains with tempered/enhanced epigenetic mechanisms were evolved in the lab under a carefully selected metabolic challenge, in order to gain insights into whether tempering with the epigenetic fidelity of a genome (i.e. - the genome's capability of maintaining its epigenetic landscape) can enhance evolution and evolvability under certain circumstances.

Specifically, three epigenetic mechanisms were studied including HSP-90 (protein chaperone), SUP-35/RNQ prions and RNAi, all of them with a potential capacity to accommodate rapid environmental adaptations via epigenetic mechanisms. For each of the mechanisms various strains in which they were either enhanced or tempered were evolved, alongside the WT.

Initial calibration quantified the initial growth defect (if any) and showed that no significant general fitness loss (specification) was caused due to the evolution. The evolution itself accomplished significant improvements for all the tested strains that were consistent across biological repeats.

Having accomplished the above baseline, the evolutionary dynamics of the various strains for each mechanism were compared by using the curveball method to extract representative growth curve parameters, as an indication of fitness.

From my results, it seems that the HSP-90 mutant strains have presented different evolutionary dynamics than the WT (Parent) strain. Although relevant for all the mutants, this effect was most evident for the HSC homozygote and the STI mutants (the homozygote and to a lesser extent – the heterozygote).

The HSC homozygote max yield started and remained higher than the WT, while also presenting a higher max growth all through the evolution. From all the mutant strains- it presented the lag dynamic that most resembled the WTs, ending with the shortest lag time of all.

The STI mutant strains' max yield started the lowest and ended the highest. Their max growth also ended the highest (with the homozygote starting the lowest and the heterozygote exhibiting a jump at the first time point followed by a drop at the second), and their lag dynamics were different than the WT and HSC homozygote, ending at a longer lag time.

This difference in dynamics is most intriguing, since it seems almost systematic, bearing in mind that HSC-82 is responsible for the constitutive expression of the Hsp82p (hence its deletion will cause significantly lower HSP-90 levels), while the Sti1p regulates Hsp82p/Hsc82p activity through the inhibition of ATP hydrolysis, and therefore its partial or complete depletion will increase Hsp82p activity.

Hence, these initial results indicate that HSP-90 does influence evolvability, yet as to the nature and extent of that influence additional growth tests should be held and analyzed. Additionally, sequencing and comparing the final evoltants to their ancestors for the different mutants may yield some interesting insights regarding the underlying mechanism.

As for the SUP-35 strains, the initial results were quite ambiguous as was detailed in the results section above, and shortly after this study was completed, other studies at the Pilpel lab raised some doubts

regarding these strains' prion expression conditions. Therefore, no further investigation was neither pursued nor recommended until resolving the expression issue.

Finally, for the RNAi strains, it is evident even from the initial results that re-introducing the RNAi mechanism (even partially) resulted in different evolution dynamics than the parent strain. Both restored strains exhibited similar evolutionary dynamics, with a significantly lower final max yield and lag time (no significant differences in the max growth parameter dynamics).

These initial results imply that re-introducing the RNAi mechanism (even partially) could indeed influence evolvability, yet as to the nature and extent of that influence, additional growth tests should be held and analyzed. Additionally, comparing the final evoltants to their ancestors by sequencing both DNA and RNA for the three strains may yield some interesting insights regarding the underlying mechanism.

## 4 Chapter 2- A Whole-genome evolvability screen

### 4.1 Introduction

After completing the epigenetic study described in the previous chapter (and specifically- the HSP-90 mutant strains investigation), a motivation arose to investigate further in order to identify additional genes that play a role in the evolutionary process.

It is well known that mutation in some genes such as DNA repair genes (“Mutators”), genes that code for proliferation proteins (“Oncogenes”), and cell cycle regulator genes (“tumor suppressors”) may promote different evolutionary dynamics [50, 51] by, for example, increasing mutation rate. However, putting those “immediate suspects” aside, does additional, evolution-enabling genetic-based mechanisms exist?

In the spirit of the long-lasting yeast community tradition, I choose to conduct a whole-genome evolvability screen, serving to identify genes that play a role in the evolutionary process, by competitively evolving all the known non-essential yeast deletion strains, to find those that will exhibit non-trivial evolutionary dynamics.

This evolution-competition assay was constructed to expose “Anti evolution” or “Evolution decelerator” genes, that their deletion will improve the yeast adaptation capability, rather than “Pro evolution” or “Evolution accelerator” genes that substantiate the process. Ideally, I would have gone for the second type, yet it requires an evolution of the over-expression mutants’ collection that is much more technically challenging, and I’ve reasoned that significant insight regarding the underlying mechanism can be also obtained from the first type. The pitfalls of using this competitive approach and a pool of gene-deleted strains as a starting evolutionary point are evident from the get-go, with the main two being (1) The significant level of environmental complexity- due to possible ecological interactions emerging between the competing strains in the pool that creates a very complex environment that is also very unstable as it constantly changes with changes in the population composition. Such an environment undoubtedly influences evolution in a manner that is very hard to model or predict. (2) The competition between strains, together with the fitness differences and the initial imbalance that will be discussed may (and actually did) eliminate a large fraction of them from the population well before de novo mutations had a chance to occur and affect the evolutionary dynamics and outcome, resulting in potentially massive loss of relevant candidates. Regardless, we argued that due to the methods’ large scale and high throughput, some insights will be there to be gained despite these reservations.

Based on the solid infrastructure of the yeast deletion project [42,43], this novel approach combined three powerful methodologies- yeast whole-genome screens [52], lab evolution and high throughput sequencing [53] to produce an overall evolvability-impact genome mapping.

The screen pinpointed some interesting candidates, that were thoroughly studied by conducting separate lab evolutions under various conditions, and comparative sequencing of the ancestors and the evoltants to gain insight into the underlying mechanisms.

## 4.2 Materials and Methods

### The pool- imbalance and lack of WT

As was briefly mentioned above, this screen took place by using the yeast deletion project [42,43] haploid collection (courtesy of the Schuldiner lab), and creating a pooled deletion library for all non-essential deletion mutants. The pool was created by printing out the strains from the Revco on fresh plates using a 284-pin robot, incubating them overnight at 30°C, and then scraping the resulting colonies off into a 5ML medium. This method had a few shortcomings (in comparison to purchasing a commercial library), the main ones being missing strains, lack of WT and most importantly- an inherent imbalance. Since colony sizes were affected by the deletion strains' fitness and geographic location on the plates, they differed in size and their resulting representation in the pool was different and biased toward the stronger strains. This imbalance was documented to be addressed analytically. WT addition was attempted by printing colonies in the same method that was used for the pool preparation and adding them to the mixture, yet sequencing revealed that the added quantity was miscalculated and insufficient and therefore ineffective, resulting in the WT decimation in the pool at the very beginning.

### Evolution- conditions, repeats and days

The above pool was then evolved manually at the lab for 80 “Days” (~560 generations). During this evolution, the strains were grown in 24-well plates, containing 1.2ml of evolution media until entering the stationary phase (=“Day”/ ~7 generations), and then diluted at 1:120 ratio by transferring 10µl of the old culture into a well filled with fresh media.

This evolution took place under five conditions-

1. **YPD**- YPD, 30°C- yeast-rich medium, nominal conditions.
2. **pH** - YPD, pH 7.6- High pH medium, an osmotic stress
3. **XYL**- Xylulose ( $C_5H_{10}O_5$ ) - a mild metabolic challenge (borderline "opportunity", as was explained above)
4. **SD**- SD Complete- Poor medium, a metabolic stress.
5. **HS**- YPD, 39°C- Temperature stress.

YPD (Yeast Extract Peptone Dextrose) [(1 L) 10 g yeast extract, 20 g peptone, 20 g dextrose, and DDW (Double distilled water)] was used as a rich medium and SD (Synthetic Defined) [(1 L) 6.7 g yeast nitrogen base, 20 g dextrose, 1.5 g amino acids mixed powder, and DDW] was used as a minimal medium. The XYL medium was identical to the one used (and described) for the epigenetic study in the previous chapter- a YP medium (Yeast Extract Peptone- [(1 L) 10 g yeast extract, 20 g peptone and DDW (Double distilled water)]) supplemented with the xylose/xylulose mixture as described above, to a final concentration of 2% xylulose.

The high pH medium was prepared by mixing 9/10 YPD with 1/10 1M Tris –HCL buffer [(1 L) 121.14g Tris, 800ml DDW, ~60ml concentrated HCL, titrated for pH 7.6].

To all the media, 30µg/ml doxycycline was added.

For each tested condition, three independent lines (biological repeats) were evolved. Evolution samples for each repeat were frozen every ~6 days (~42 generations).

### Library construction and sequencing- getting the barcodes, getting the fitness

As was mentioned above, the deletion strains used were produced by the yeast deletion project. This project used a PCR-based gene deletion strategy in order to generate a start-to-stop codon deletion of each of the ORFs in the yeast genome. As part of the deletion process, each gene disruption was replaced with a KanMX (antibiotic resistance) module and uniquely tagged with one or two 20mer sequence(s), i.e. “Barcodes”. The barcode sequences were flanked with uniform up and down sequences as depicted in Figure 12. This modular construct has enabled me to extract via PCR only 150 base pair segments that contained the barcodes, in order to map each mutant frequency in the pool. This mapping was obtained by ligation library construction and high throughput sequencing using Nextseq.

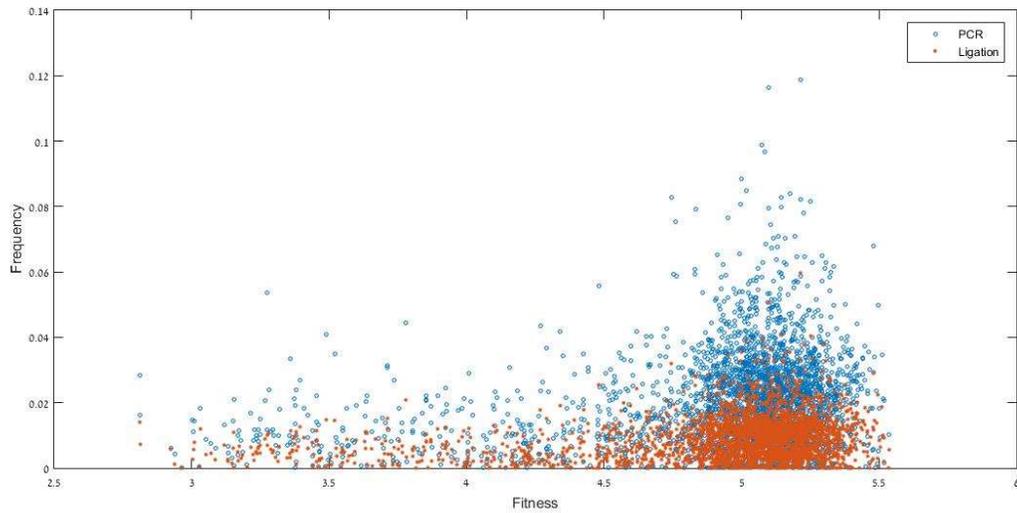


**Figure 12- The Yeast deletion project construct** – This figure, taken from the yeast deletion project website [43] presents the standard deletion cassette structure that was used. As part of the deletion process, each gene disruption was replaced with a KanMX (antibiotic resistance) module and uniquely tagged with one or two 20mer sequence(s), i.e. “Barcodes”. The barcode sequences were flanked with uniform up and down sequences.

This was done for all 5 conditions X 3 biological repeats at day 0 (ancestral pool under the relevant condition), day 40 (280 generations) and day 80 (560 generations).

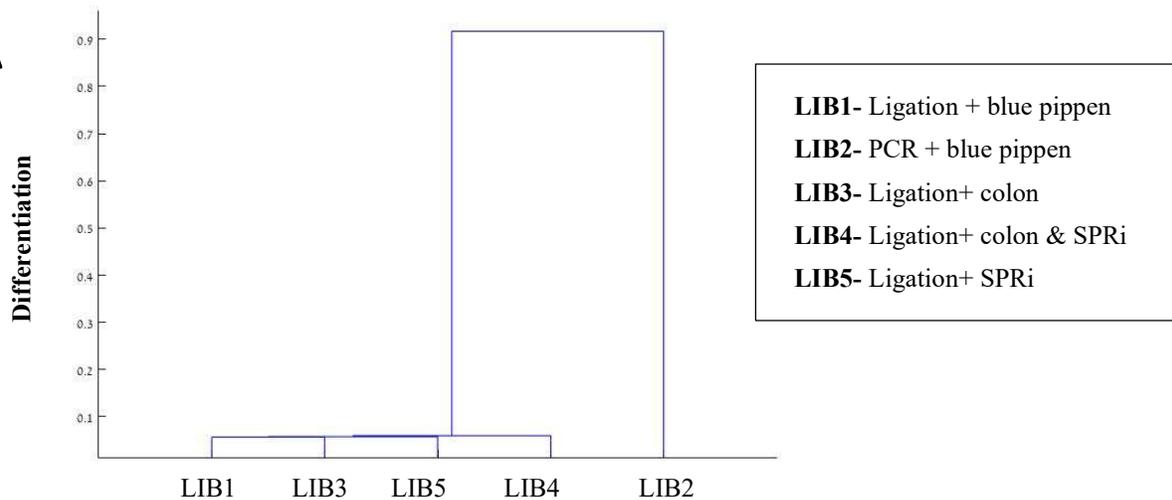
However, in order to deduce fitness, the relative mutant frequency at those remote evolution times was insufficient, since I was actually interested in the derivative. Therefore, I have performed Ad Hoc growth assays for all the above points. At those assays, starters were grown in 5ml tubes from the relevant frozen samples, over a sufficient amount of time to reach the stationary phase. Approximately  $2 \times 10^8$  cells were inoculated into 100ml fresh medium and grown under the relevant conditions. These cultures were sampled after 2 hours and after the equivalent of one evolution “day” (either 24 or 48 hours), and libraries were constructed for both time points. The ratio between each mutant’s normalized frequencies at those time points was used as a measure of its fitness at that evolution point.

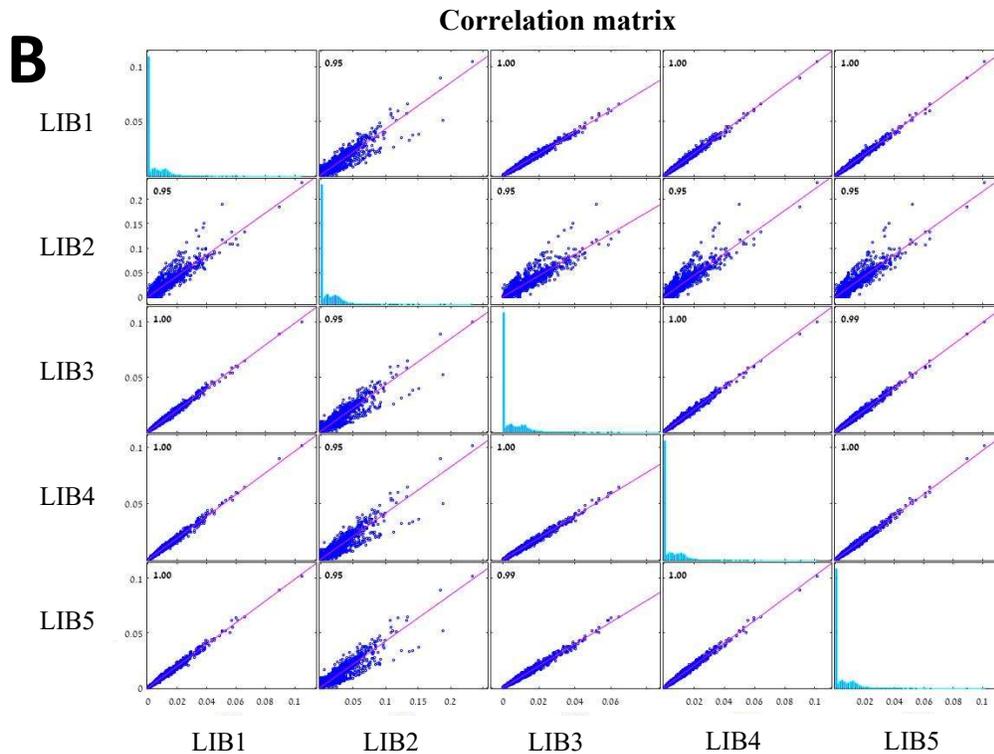
In order to reduce bias to the outmost minima, ligation library contraction for each sample was divided into 8 separate reactions with 100ng template each, which underwent a minimal number of rounds and colon cleaning. This procedure, although pricy and labor-intensive was selected after careful calibration, to include comparative sequencing of libraries prepared and cleaned using various techniques. As it turned out- constructing the libraries by using PCR created a bias toward higher frequency mutants (as is demonstrated in Figure 13 below), while the cleaning method had no tangible effect on the sequencing results (Figure 14).



**Figure 13- PCR vs. Ligation libraries construction comparison** – The data in this figure was analyzed and visualized with Dr. Hila Gingolds’ kind assistance. The figure presents the results of a calibration test, to include the relative frequency vs. the relative fitness of the different mutants that were identified at two separate libraries. These two libraries were prepared from the same stock and extraction- one by using PCR (20 cycles) and one by using the ligation method (both cleaned by the same method- blue Pippen). A considerable bias toward higher-frequency mutants is evident in the PCR library.

**A**

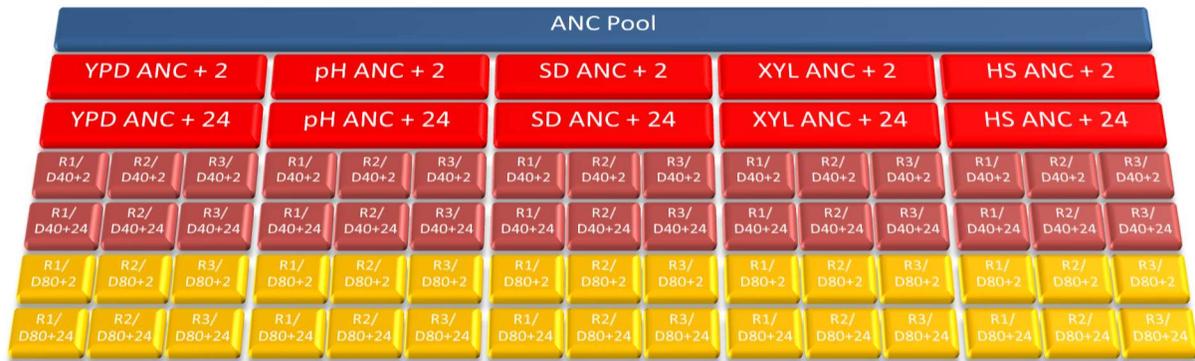




**Figure 14- Libraries cleaning methods comparison** – The data in this figure was analyzed and visualized with Dr. Hila Gingolds’ kind assistance. The figure presents an additional result of the calibration test- **(A)** A dendrogram and **(B)** A correlation comparison matrix of the mutants’ relative frequencies as were mapped for five separate libraries. These five libraries were prepared from the same stock and extraction, by using different preparation and cleaning methods, including- (1) A ligation library cleaned by blue Phippen; (2) A 20 cycles PCR library cleaned by blue Phippen; (3) A ligation library that was colon cleaned (small fragments only); (4) A ligation library that was cleaned using both colon and SPRi beads; (5) A ligation library that was double sided SPRi beads cleaned; As can be seen from the figure, the cleaning method has no significant impact on the sequencing results, unlike the preparation method that does.

The final pool sequencing took place by using the WIS/INCPM Illumina NextSeq machine and performing a high-output run (150 cycles).

After an elaborate data reduction, the frequencies and fitness of each deletion mutant were available for each of the conditions. The full data set structure is depicted in Figure 15.



**Figure 15- The data set structure** – This figure depicts the data set structure. Overall 71 pool libraries were sequenced, to include (1) the original ancestral pool- one library marked in blue (2) the ancestral pool grown in each of the five conditions (YPD, pH, SD, XYL and HS)- after 2 hours of growth and after a days' worth of growth (24/48 hours according to growth rate, marked +24 for clarity) to allow for fitness calculation, to a total of 30 libraries marked in red (3) Day 40 evlotatnts – two time-points for each of the 3 biological repeats of the 5 conditions, to a total of 30 libraries marked in dark red (4) Day 80 evlotatnts – two time-points for each of the 3 biological repeats of the 5 conditions, to a total of 30 libraries marked in yellow.

### Sequencing data analysis

The sequencing data reduction consisted of the following stages-

- A) File merge - to construct the original libraries from the sequencing lanes files
- B) Initial quality control by using [Fastqc](#) – used to assess the run quality that turned out to be high (especially considering the non-classic low complexity data)
- C) Removing the 3’ and 5’ uniform parts of the amplicon (**Figure 16**) to extract the barcode, using a CutAdapt script. Poli-N reads were also removed.

Conceptual-

5 Random nucleotides PrimerkanKANanalogyD2\_19nucleotidesBarcode\_20nucleotidesD1\_17nucleotidesRC5 Random nucleotides

Actual-

NNNNCAATCGTATGTGAATGCTGTCGCTATACTGCTGTCGATTTCGATACTAACGCCGCCATCCAGTGTGCAAAAAGAGCTCGAATTCATCGATXXXXXXXXXXXXXXXXXXXXCTA  
CGAGACCGACACCGNNNN

NNNNCGGTGTCGGTCTCTAGXXXXXXXXXXXXXXXXXXATCGATGAATTCGAGCTCTTTTCGACACTGGATGGCGGCTTAGTATCGAATCGACAGCAGTATAGCGACAGCA  
TTCACATACGATGNNNN

**Figure 16- The 4K amplicon construct** – This figure depicts the amplicon construct (conceptual and with the actual nucleotide sequences, for both upstream and downstream barcodes). The unique barcode (marked yellow) is flanked by constant adaptors, primers and analogy sequences. The standard structure and sequence enable us to pinpoint and extract the barcode.

- D) Poly-G reads cleanup- When no definitive read is available (especially for “T”s), NextSeq automatically assigns “G” for that position. Therefore reads with more than 20 “G”s were discarded.
- E) Size selection- From the remaining reads only reads with 19-21 nucleotides length were selected (As can be recalled- barcodes are 20 nucleotides long).
- F) Barcode comparison- the reads were compared to the Yeast deletion project barcodes list. To compensate for alignment errors without losing reads- a single **side** addition/deletion was tolerated, and those 19/21 nucleotide long reads were mapped with the matching 20 nucleotide barcode reads.
- G) Mismatch identification- Once more than 10 reads contained the same 19-21 nucleotide sequence that did not match any known barcode, its edit distance to the entire known barcode list was checked.

Out of 35 such significantly represented sequences (some of them reverse complement of others) with Edit distance=1 (not to include side addition/deletions), eleven were recognized as valid gene barcodes and marked GeneName-1MM (for one mismatch) if their reverse complement sequence was symmetrically represented or if that genes' original barcode was missing.

After all the above manipulations, the sequencing run and library quality were re-evaluated to assess validity, including read count vs./ allocation, remaining read percentage after applying the filters and identified barcode percentage. Overall, all were considered satisfactory as can be seen in the results section (following).

### **The selection process**

The resulting data were then visualized and analyzed to verify reasonability and internal consistency, as again will be described in the following results section.

After establishing sufficient confidence in the data, the quest for the above-mentioned infamous evolution decelerator genes has begun. The optimal candidate would present a significant fitness increase (“acceleration”) over evolution time, and preferably- over various conditions. In addition, genes with high representation (frequencies) were preferred, since the fitness data validity for them was higher.

Therefore, after the frequency of the read was determined for each condition, fitness was calculated for all the relevant reads by comparing each barcode frequency after 24/48 hours (=evolution day) to the one after 2 hours at the same condition/biological repetition. Acceleration was then calculated by comparing fitness values between day 40/day 0 and between day 80/day 40.

Since some barcode frequencies, though significant at some of the conditions/ times/ days dropped below the detection threshold at others, those preventing acceleration or even fitness calculation for that strain and rendering the data set incomplete, relatively permissive criteria were selected for the initial screening process to include-

- >10rpm (=reads per million) @at least one point (=a condition/biological repetition/evolution time combination, to a total of 35)

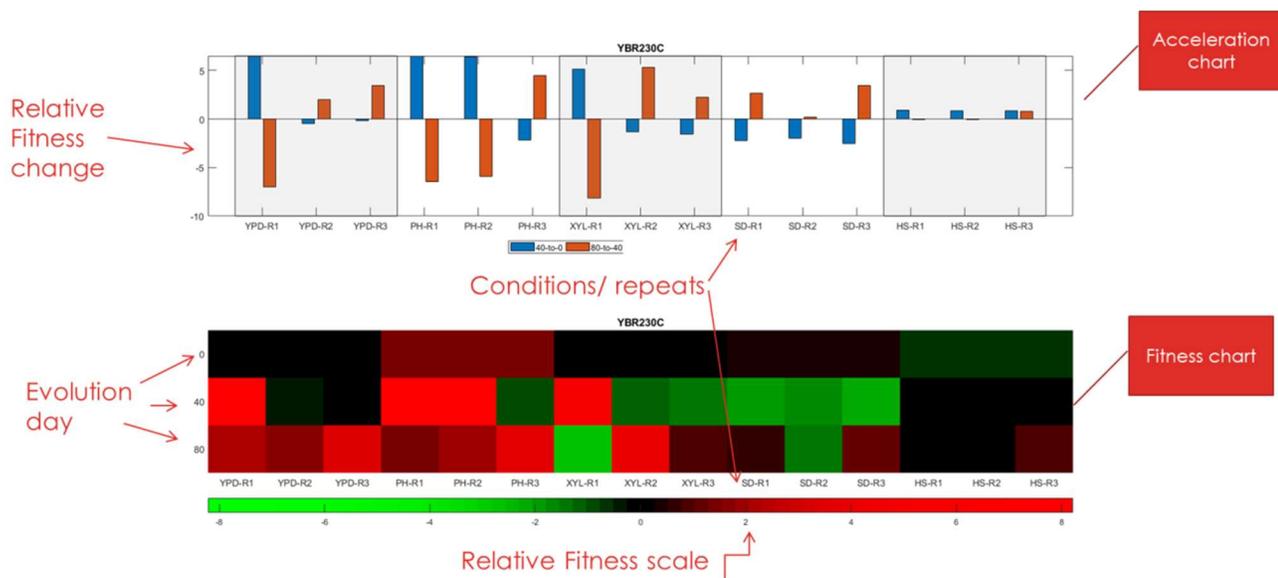
- At least one ancestor with a finite fitness value
- At least one finite day 80/day 40 acceleration value

For each point, the candidates' relative fitness was ranked, and for each candidate- the number of points at which it was represented (out of the possible 35), its average ranked fitness across them and its maximal measured rpm were calculated.

Based on this calculation of fitness across conditions, a total of 172 candidates were selected for the final batch. These candidates presented high average ranked fitness over a large number of points and specifically were either-

- Represented at >25 points (=condition/time/repeat points)
- Represented at >15 points with an average normalized fitness rank > 0.6
- Represented at >8 points with an average normalized fitness rank > 0.65

For this final batch, a dedicated selection tool was designed, to visually present for review the attributes of each candidate separately, and specifically- its fitness and acceleration across conditions, repeats and evolution days, as is depicted in **Figure 17**.



**Figure 17- The selection tool** – The data in this figure was analyzed and visualized with Dr. Hila Gingolds' kind assistance. This figure depicts the selection tool (“dashboard”) that was designed to automatically generate a visual representation of each of the 172 final candidate attributes to enable their efficient final review and ranking for further investigation. The lower panel of the tool presents a fitness chart- the candidates' relative fitness for each of the five conditions (YPD, pH, XYL, SD and HS) three biological repeats (columns) and evolutions days – day 0, day 40 and day 80 (rows). It uses a green (low) to red (high) relative color scale that is presented below. For easier interpretation, the upper panel of the tool presents an acceleration chart- the relative fitness change between day 40 and day 0 (blue) and day 80 to day 40 (red), again – for each of the five conditions three biological repeats (columns). The header of each panel displays the relevant candidates' systematic gene names.

## The candidates

Using the selection tool, all 172 final candidates were manually reviewed, to select the ones that presented the most interesting behavior of fitness increase dynamics during the evolution. The most prominent ones were classified into three priorities (19 genes were marked “priority 1”, 7 genes were marked “priority 2” and 7 more were marked “priority 3”, see results) and reviewed in great detail, to include:

- Gene Description (SGD) and relevant literature
- Selection Tool Chart
- Relative location on the Fitness across conditions chart (Fish)
- Relative location on the Frequency/fitness chart for each condition
- Relative location on the Fitness histogram for each condition
- Relative location on the acceleration chart for each condition

Conservation and expression profile data were also taken into account.

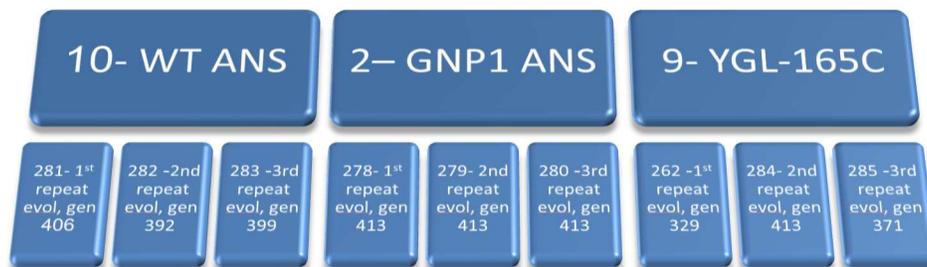
## Candidates Evolutions

After the above process, separate evolutions for the most promising candidates have begun. A total of nine strains (plus WT controls) were evolved separately under three conditions- YPD, high pH and SD. Each strain was evolved in three independent biological repeats. Since time was of the essence- strain verification (via PCR, Sanger sequencing and FACS) took place in parallel, resulting in losing three of the evolutions due to negative verification results.

These evolution results were assessed by performing growth tests. SD evolved strains' growth was evaluated by using SD medium, and pH evolved strains were evaluated by using several high pH media, including pH 7.6 (the evolution pH), pH 8.6, pH 9.0 and a YPD reference.

Two of the strains- YDR508C (**GNP1**) and YGL165C performed significantly well at the separate evolution growth essays (on high pH medium) and were therefore sequenced to better understand the underlying mechanism.

This sequencing took place by performing a 300-cycle MiniSeq run at TAU. Libraries were constructed using The Nextera DNA Flex Library Prep kit and protocol and covered the ancestor strains of both WT and the two mutants, and the final results of the three biological repeats of their evolutions. The full data set structure is depicted in Figure 18.



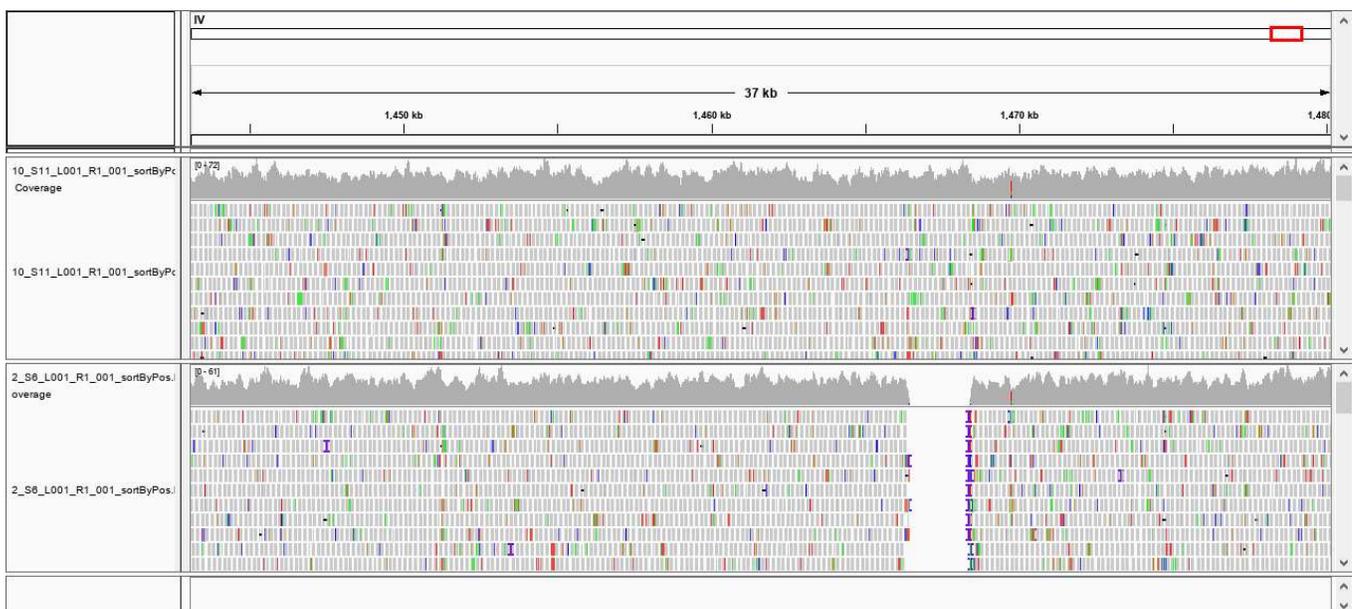
**Figure 18- MiniSeq dataset structure** – This figure depicts the data set structure. Overall, 12 libraries were sequenced, to include (1) the three ancestral strains- the WT (strain #10), YDR508C (GNP1- strain #2) and YGL165C (strain #9); (2) The final evolution product of each strain’s 3 independent biological repeats (ranging between 329-413 generations as is noted), to a total of 9 libraries.

### Candidates’ evolutions sequencing data reduction

Miniseq data were aligned by using the bowtie 2 aligner software (<https://bowtie-bio.sourceforge.net>) and compared to the SGD reference genomes-

- GNP1 - <https://www.yeastgenome.org/locus/S000002916>
- YGL-165C - <https://www.yeastgenome.org/locus/S000003133>
- WT- <https://www.yeastgenome.org/strain/BY4741>

This was done by using the IGV genome browser (<https://software.broadinstitute.org/software/igv/>), and initial confidence in the data was gained by verifying that all libraries had proper coverage of the genome, to include sanity checks in the form of the ORF deletions, as is depicted in Figure 19 for example.



**Figure 19- IGV showing GNP1 ORF deletion from the data** – An example of sequencing results’ sanity check using the IGV genome browser. WT (#10) and GNP1 (#2) sequences are comparatively aligned, showing proper reads coverage and the GNP1 ORF deletion (chromosome IV, 1466453-1468444).

Additional verification of sequencing coverage took place by comparing each final evoltant (for all three biological repeats) to the ancestor strain- chromosome by chromosome by using a dedicated script to produce visual coverage maps such as the one demonstrated in Figure 20. This was done to enable the detection of big deletions/additions, significant under covered/overcovered areas, ploidy changes, or significant unaligned areas. Anomalies that were detected visually were analyzed by using the SGD [JBrowse](#) function.



A comparison between the two variant calling methods showed significant differences both in the amount of SNPs received, and in the identity of the ORFs in which they were found, yet some ORFs emerged using both as is detailed in the following results section. Since GATK is considered to be the current standard, its results were used as the basis for continuing the analysis. Regardless, some of the most interesting results have emerged by both methods, a fact that increased the confidence in their validity.

Venn diagrams (produced by [Venny 2.1](#)) and correlation matrixes were used to visualize and analyze the SNPs repertoire for each mutant strain and biological repeat.

Once initial confidence in the data set was obtained, an in-depth analysis took place, focusing on SNPs within known ORFs. These SNPs distribution between the different strains was analyzed for possible enrichments using [GOrilla](#), and then manually analyzed for additional insights.

## 4.3 Results

### The pool- imbalance and lack of WT

As was mentioned in the method section above, this study began by evolving the entire yeast haploid non-essential deletion mutants as a pool, under five conditions (YPD, pH, XYL, SD and HS) for ~80 “evolution days” (approx. 560 generations).

Sequencing then took place for all 5 conditions X 3 biological repeats at day 0 (ancestral pool under the relevant condition), day 40 (280 generations) and day 80 (560 generations). At each point, Ad Hoc growth assays were performed, and the pool was sampled after 2 hours and after the equivalent of one evolution “day” (either 24 or 48 hours) to enable fitness estimation. The data set structure is depicted in Figure 15.

### Sequencing data analysis

After sequencing, an elaborate data reduction took place to include file merge, initial quality control ([Fastqc](#)), 3’ and 5’ uniform amplicon sequences removal, poly-G reads cleanup, size selection (19-21 nucleotides were allowed) and barcode comparison.

Finally, a mismatch identification process took place. As was mentioned above, once more than 10 reads contained the same 19-21 nucleotide sequence that did not match any known barcode, its edit distance to the entire known barcode list was checked.

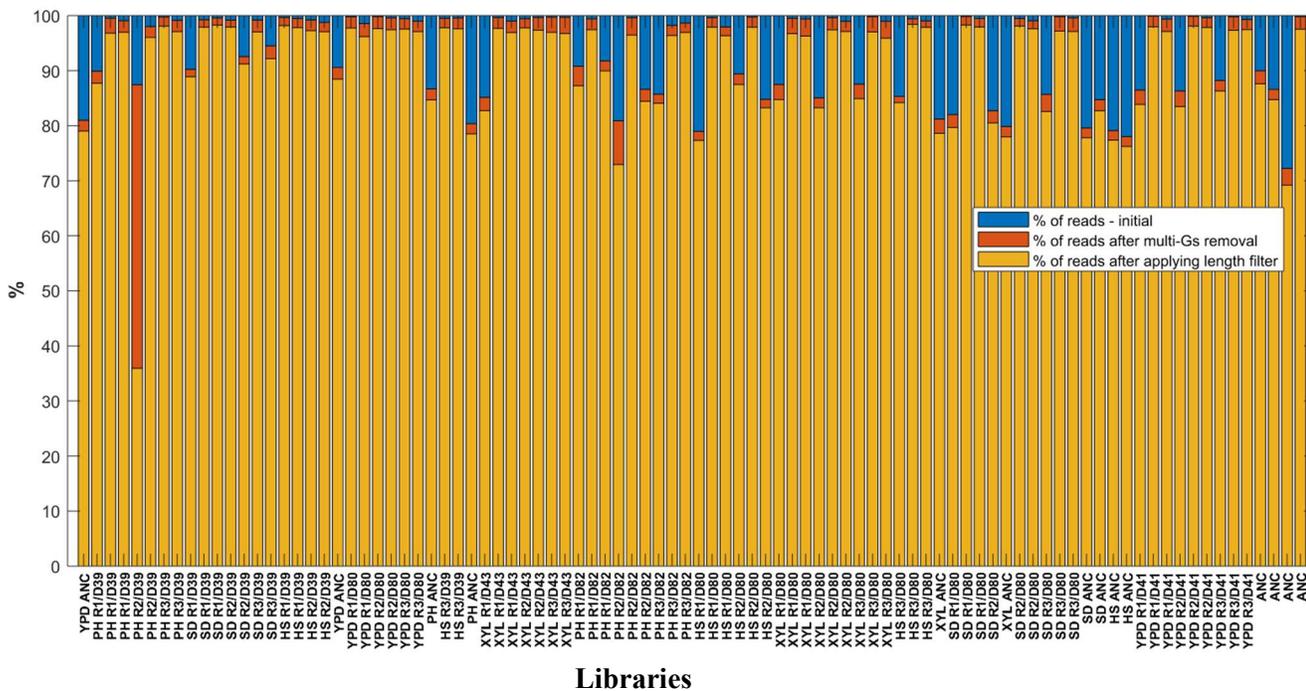
Out of 35 such significantly represented sequences (some of them reverse complement of others) with Edit distance=1 (not to include side addition/deletions), eleven were recognized as valid gene barcodes and marked GeneName-1MM (for one mismatch) if their reverse complement sequence was symmetrically represented or if that genes' original barcode was missing. They are detailed in Table 6.

Original index	Gene name	Sequence in fastq	Significant reverse complement detected?	Addition at the edge?	Original Barcode detected?
ACCGATGCGTCAT <b>T</b> TCGTCT	YBR141C-1MM	ACCGATGCGTCAT <b>C</b> TCGTCT	Yes		
TTAGAT <b>A</b> GTTTCGCTGCCGG	YBR230C-1MM	TTAGATGTTTCGCTGCCGG			
GCGT <b>C</b> ICTTAAACCTTCTGA	YDL194W-1MM	GCGTCCTTAAACCTTCT <b>I</b> GA			
GTTGAGCGTGTGT <b>T</b> AAATACC	YDR001C-1MM	GTTGAGCGTGTGT <b>A</b> AATACC	Yes		Yes
CCGGTTATCCA <b>A</b> G <b>T</b> GTACTA	YDR508C-1MM	CCGGTTATCCAATGTACTA			
<b>C</b> AAACATGGACCTCCGTAGG	YGL002W-1MM	<b>G</b> CAAACATGGACCTCCGTAGG		Yes	Yes
GACCTAATTCGAC <b>C</b> AGTTA	YGL165C-1MM	GACCTAATTCG <b>A</b> CCAGTTA	Yes		Yes- very low
<b>G</b> CCAGAAATGTT <b>C</b> AGCACTC	YHL017W-1MM	<b>A</b> CCAGAAATGTT <b>C</b> AGCACTC		Yes	Yes
ACATTGCACTGAAC <b>G</b> T <b>C</b> ATC	YOR101W-1MM	ACATTGCACTGAACGT <b>C</b> A <b>C</b>			
<b>A</b> TGCGTGCTCCG <b>T</b> CCAGATA	YOR293C-A-1MM	<b>I</b> AAGCGTGCTCCG <b>T</b> CCAGATA		Yes	
TAGC <b>G</b> <b>I</b> TCAGCACCTTGTGG	YPL006W-1MM	TAGCGTCAGCACCTTGTGG			

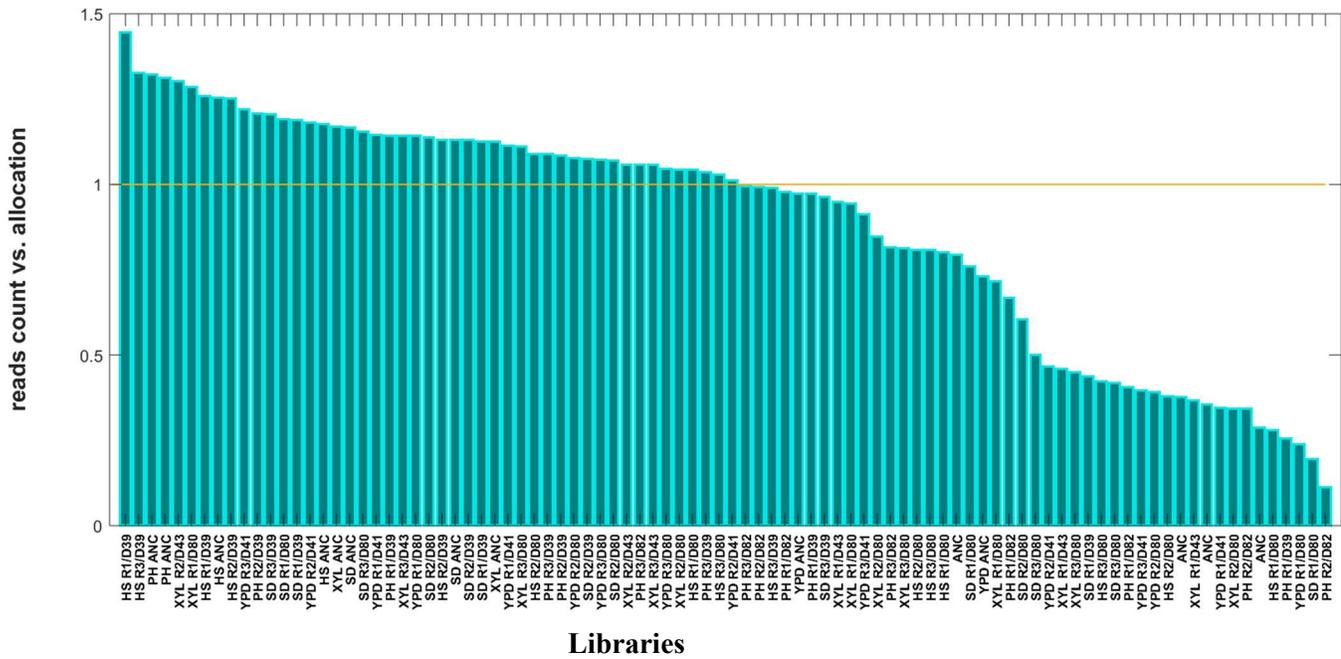
**Table 6 – Mismatched barcode list**– This table presents the eleven significantly represented sequences with Edit distance=1 that were recognized as valid gene barcodes and marked GeneName-1MM (for one mismatch). They were selected out of 35 such sequences based on reverse complement representation, original barcode presence and the location of the mismatched nucleotide.

After the above initial processing, the sequencing run and library quality were re-evaluated to assess validity, including read count vs/ allocation, remaining read percentage after applying the filters and identified barcode percentage analysis. Overall, all were considered satisfactory as can be seen in Figure 21.

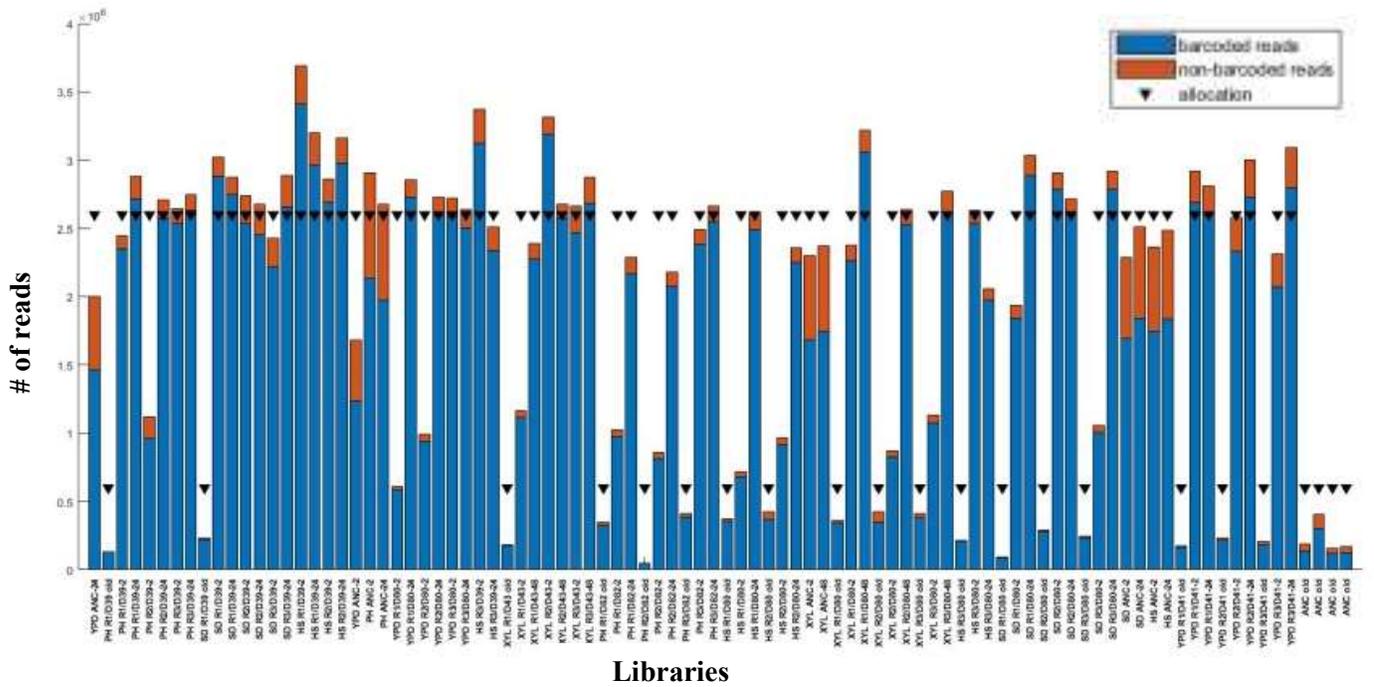
## A Reads % after applying length and poly-G filters



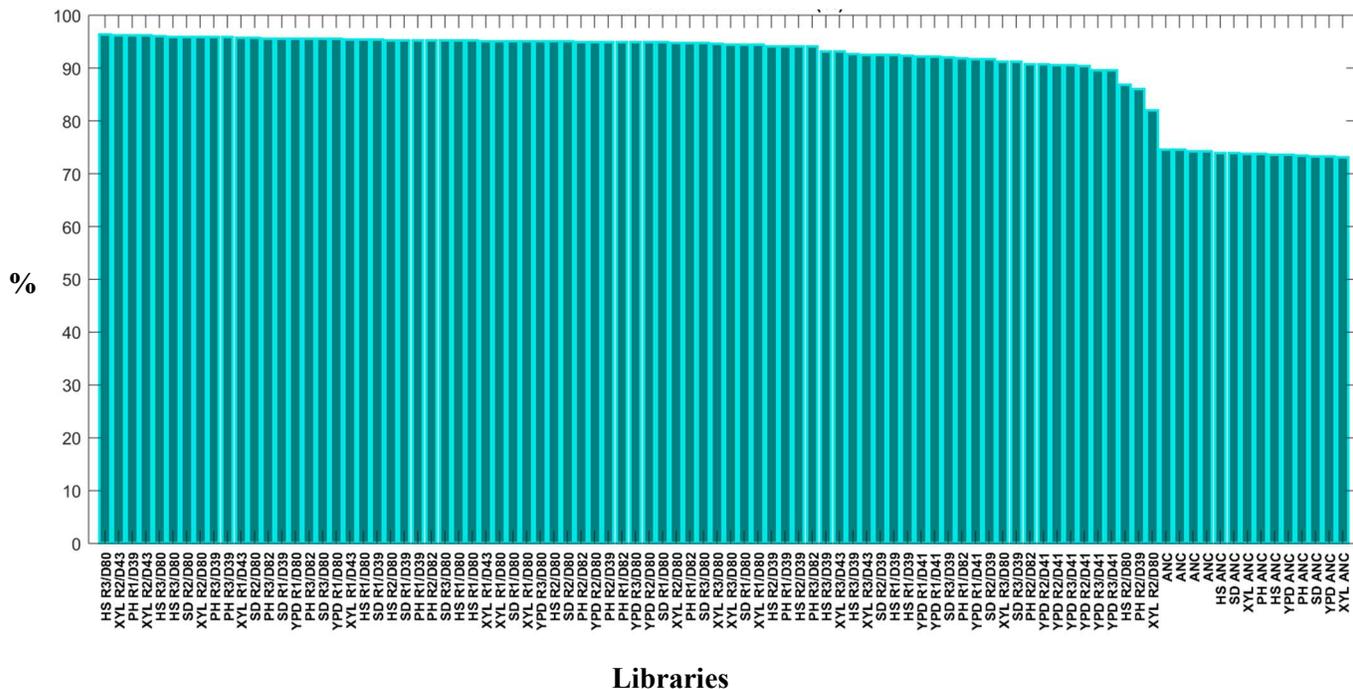
## B Read count vs. Allocation



## C Reads per library- barcodes and allocation



## D Barcodes identification %

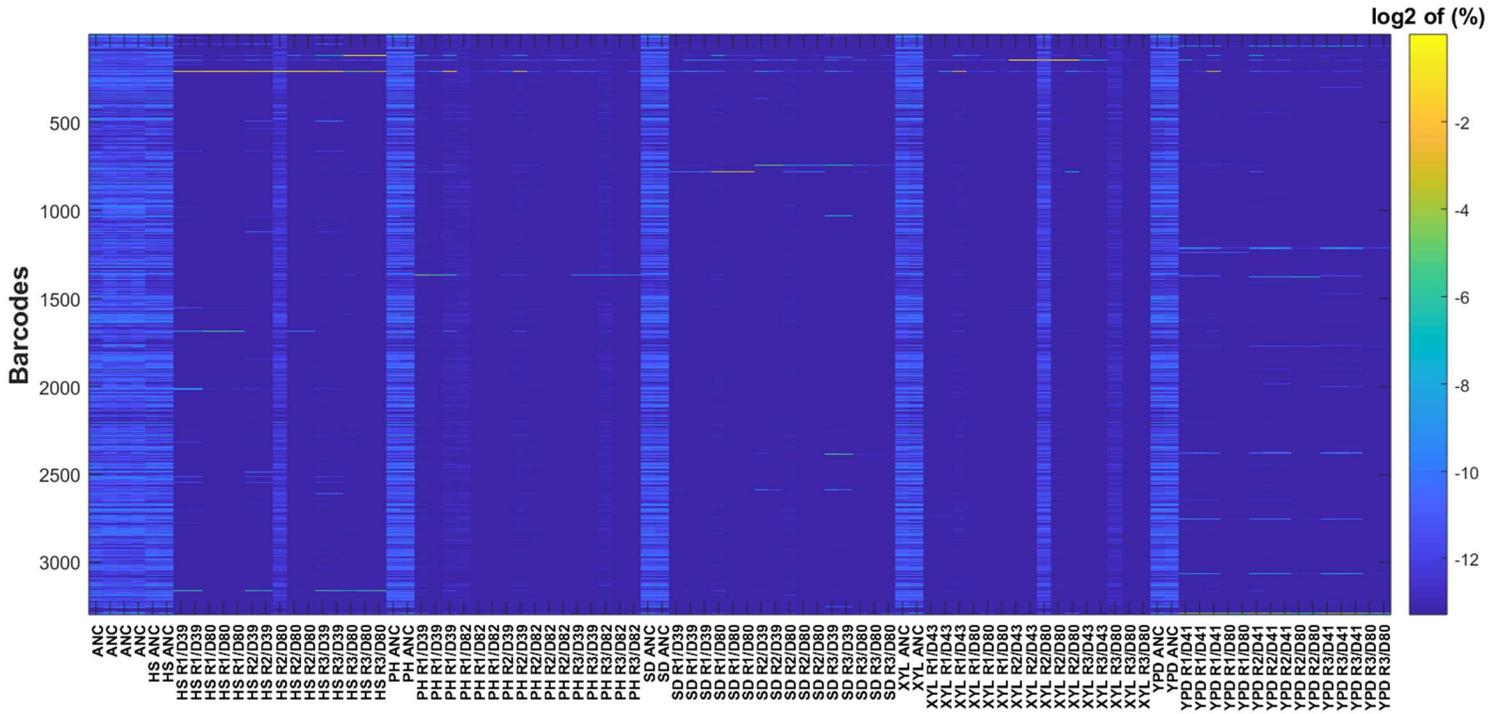


**Figure 21- Run quality and processing sanity check analysis** – The data in this figure was analyzed and visualized with Dr. Hila Gingolds’ kind assistance. The figure presents the run quality and analysis sanity check for all the libraries. The 92 libraries are presented on the x-axis, including the condition, evolution day and repeat nomenclature (two points for each- one at +2hrs and one after one evolution day. For brevity and legibility, no differentiation between these points is supplied except for chart C) and the Y axis presents (A) Reads % remaining for each library after applying the length and poly-G filters (given in yellow). The red segment shows the read % lost due to length criteria and the blue- due to poly-G criteria; (B) Read count vs. Allocation, where accurately meeting the allocation will be marked as 1; (C) Reads per library- barcodes and allocation- the triangle marks the allocation, barcoded reads are marked blue and un-barcoded are marked red; (D) Barcodes identification %.

As can be seen from the data, 22 of the libraries received a lower read allocation. These are old libraries that were sequenced only for reference in addition to the final 70 libraries. Run quality was overall good, as can be seen from the different metrics- most libraries met their reads allocation (and the ones that didn’t show a sufficient amount of reads rather than decimated), to exclude one library, the loss of reads due to length and poly-G was minor (and in the one where it wasn’t- the amount of the remaining reads was still sufficient), and the barcode identification percentage was high, hence the run technical quality was considered sufficient.

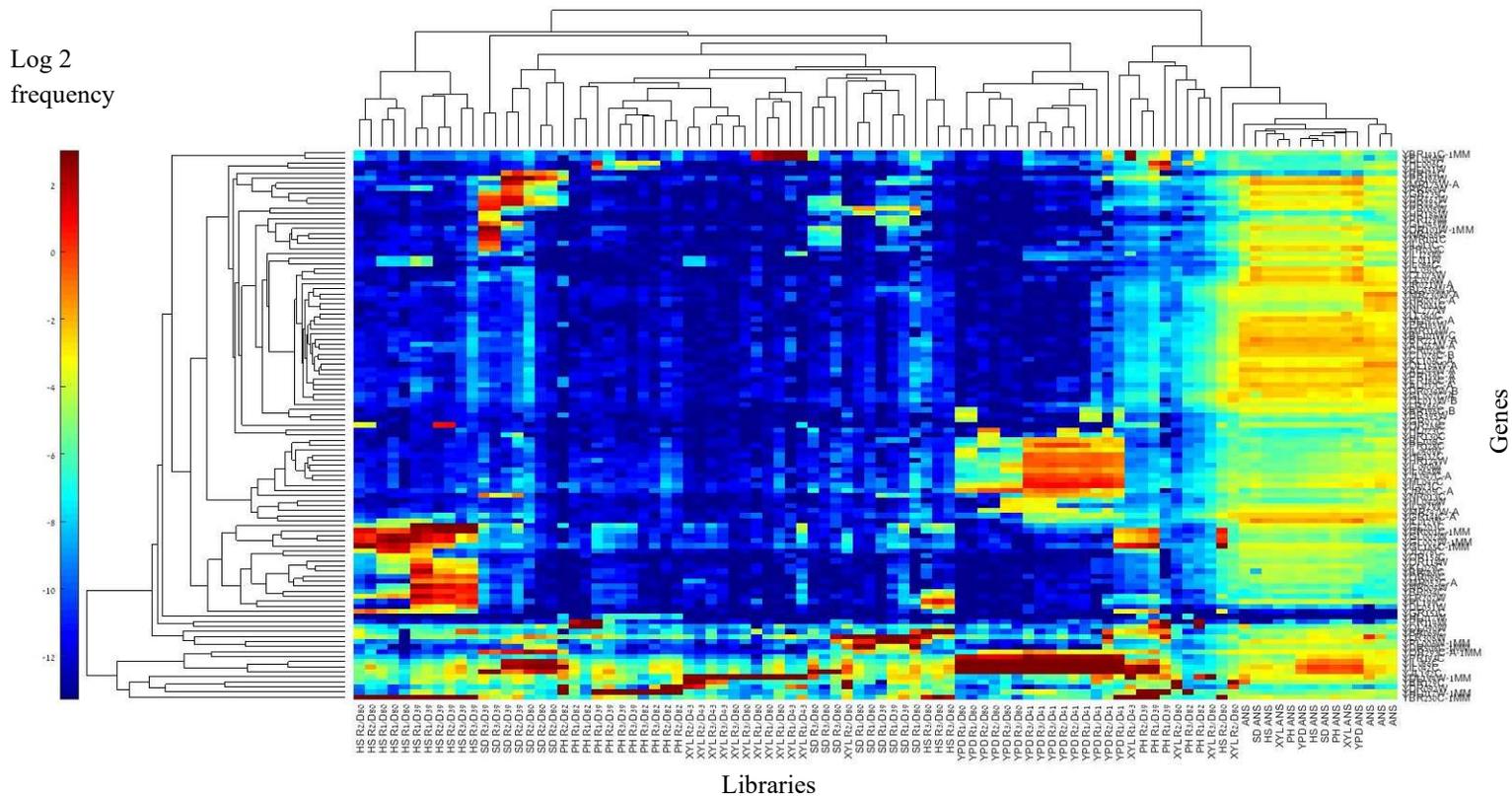
## Gaining confidence in the data validity given a decimated number of strains, the selection process

After establishing the sufficiency of the technical quality, a brief look at the data showed a great loss of heterogeneity that was to be expected due to the unbalanced nature of the competition, resulting in many strain frequencies dropping below the detection threshold at various stages and making the set incomplete (Figure 22).

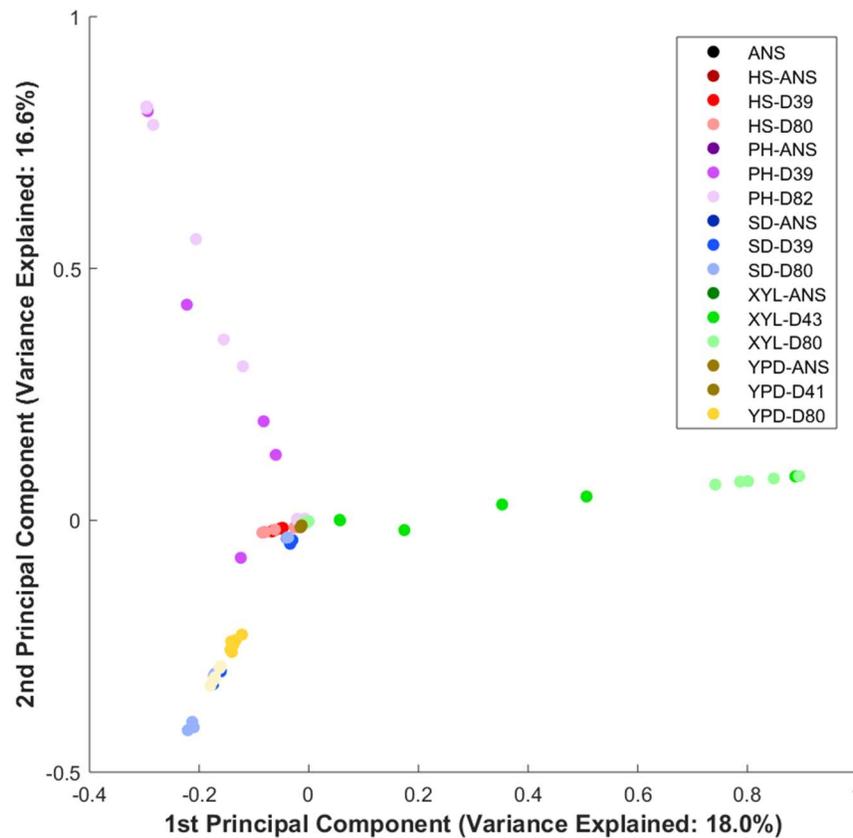


**Figure 22- Pool loss of heterogeneity** – The data in this figure was analyzed and visualized with Dr. Hila Gingolds’ kind assistance. The figure presents a visualization of the relative representation of each of the 4,852 barcodes at each of the libraries that were sequenced. The 92 libraries are presented on the x-axis, including the condition, evolution day and repeat nomenclature (two points for each- one at +2hrs and one after one evolution day. For brevity and legibility, no differentiation between these points is supplied) and the Y axis presents a color-coded line for each barcode, where high relative representation is marked yellow, and low/none is marked dark blue as can be seen at the right-hand scale.

As was mentioned above- this reduction in heterogeneity at day 40, and even more at day 80 is to be expected by the nature of the assay, and in addition, a preliminary analysis that took place to gain some confidence in the data set showed great internal consistency while focusing on the top represented strains. Examples are depicted in Figure 23 and Figure 24. As can be seen, at each condition a unique set of genes has prevailed, and the divergence of those repertoires has increased over evolution time.



**Figure 23- Top genes for each condition clustergram**– The data in this figure was analyzed and visualized with Dr. Hila Gingolds’ kind assistance. The figure presents the 109 top-represented genes relative representation at each condition. The 92 libraries are presented on the x-axis, to include the condition and evolution day nomenclature (six points for each- one at +2hrs and one after one evolution day for each biological repeat. For brevity and legibility, no differentiation between these points is supplied) and the Y axis presents a color-coded line for each barcode, where high relative representation is marked red and low/none is marked dark blue as can be seen at the left-hand scale. This figure also presents a calculated taxonomy of the relative similarity between libraries and genes, charted above and to the left of the chart.



**Figure 24- PCA for gene frequencies at each condition, and each evolution time** – The data in this figure was analyzed and visualized with Dr. Hila Gingolds’ kind assistance. The figure presents a primary component analysis of the data (mutant frequencies at the different libraries). Each color represents a condition, and the evolution time is represented by the hues- going from the dark (ancestor) to light as evolution time progresses. No differentiation between the same condition/ same evolution day and different biological repeats is provided.

As can be seen from Figure 24 above, originating from the same point, the Xylose and high pH evolutions diverged significantly over time in two completely different directions, while the YPD and HS evolutions took a third, similar direction of their own. Both the similarity and the differences are evident in Figure 23, with significant common prominent genes repertoire (at the lower part of the chart) and an additional unique repertoire for each condition at the higher parts. The Heat shock evolution showed the smallest amount of diversion over time, suggesting rapid fixation at an early stage, as is also reflected in Figure 23.

### The selection process

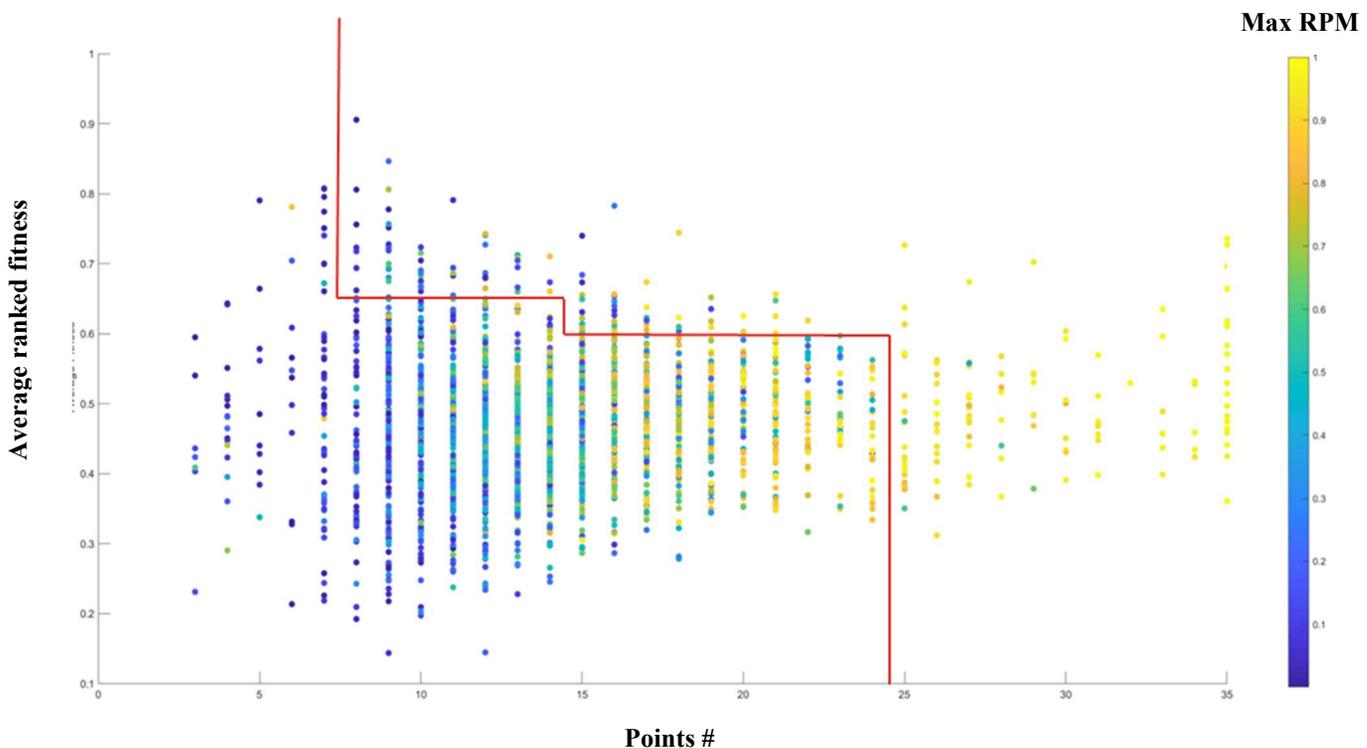
Once this initial confidence in the data was gained, the quest for the above-mentioned evolution decelerator genes began. As was mentioned in the methods section above, the optimal candidate would present a significant fitness increase (“acceleration”) over evolution time, and preferably- over various conditions. In

addition, genes with high representation (frequencies) are preferred, since the fitness data validity for them will be higher.

Therefore, after the read frequency was determined for each condition, relative fitness and acceleration values were calculated as described. This caused an inherent loss of heterogeneity since some barcode frequencies, though significant at some of the conditions/ times/ days dropped below the detection threshold at others, thus preventing acceleration or even fitness calculation for that strain and rendering the data set incomplete.

In order not to lose many interesting candidates, relatively permissive criteria were used. For each point (=a condition/biological repetition/evolution time combination, to a total of 35), the candidates' relative fitness was ranked, and for each candidate- the number of points at which it was represented (out of the possible 35), its average ranked fitness across them and its maximal measured rpm were calculated. This is presented in Figure 25.

Using Figure 25 visualization of fitness across conditions, a total of 172 candidates that presented high average ranked fitness over a large number of points were selected for the final batch (these are the candidates above and to the right of the red threshold line).



**Figure 25- Fitness across conditions and the Final candidates batch selection logic** – The data in this figure was analyzed and visualized with Dr. Hila Gingolds' kind assistance. The figure presents all the candidate mutants that (1) Had >10rpm (=reads per million) @at least one point (=a condition/biological repetition/evolution time combination, to a total of 35) (2) Had at least one ancestor with a finite fitness value and (3) Had at least one finite day 80/day 40 acceleration value.

The candidates are presented according to the number of points at which they were represented (X-axis, out of the

possible 35), their average ranked fitness across them (Y-axis) and their maximal measured rpm (color scale where yellow is high and blue is low, as is presented at the left-hand side of the chart. This served as a measure of significance).

This visualization was also used for the final batch selection, which included candidates with high average ranked fitness over a large number of points and specifically were either (1) Represented at >25 points or (2) Represented at >15 points with average normalized fitness rank > 0.6 or (3) Represented at >8 points with average normalized fitness rank > 0.65. These 172 candidates are the ones above and to the left of the red line.

## The candidate genes

Using the dedicated selection tool that was described in the Method section above, all the final batch candidates were manually reviewed, to select the ones that presented the most interesting behavior of fitness increase dynamics during the evolution. The most prominent ones (based on signal quality assessment only, as was reflected in the selection tool) were classified into three priorities as presented in Table 7.

Priority 1	Priority 2	Priority 3
YBR230C/ <b>OM14</b>	YBR129C / <b>OPY1</b>	YLR207W/ <b>HRD3</b>
YDL194W/ <b>SNF3</b>	YBR141C/ <b>BMT2</b>	YLR361C/ <b>DCR2</b>
YDR092W/ <b>UBC13</b>	YDR001C / <b>NTH1</b>	YOL013W-B
DR508C/ <b>GNP1</b>	YGR204C-A	YPR114W
YGL002W/ <b>ERP6</b>	YIL012W	YPR159C-A
YGL165C	YIL086C	YPR196W
YHL017W	YJL078C / <b>PRY3</b>	YOR293C-A
YIL042C / <b>PKP1</b>		
YJL192C/ <b>SOP4</b>		
YJR005C-A/ <b>LSO1</b>		
YKL029C / <b>MAE1</b>		
YLR308W/ <b>CDA2</b>		
YMR230W-A		
YOL090W / <b>MSH2</b>		
YOR293C-A		
YPL038W_1MM / <b>MET31</b>		
YPL199C		
YPR008W / <b>HAA1</b>		
YPR201W / <b>ARR3</b>		

**Table 7 - The final candidates**—This table presents three gene lists, reflecting their priority for further study. Each gene is represented by its systematic name and standard name (if exists). The genes within a list are not ranked and share the same priority. These genes were selected from the final candidates batch by using the selection tool that was developed for the study.

As can be seen from Table 7, 19 Genes were classified as priority 1, 7 Genes as priority 2 and 7 Genes as priority 3. These priority 1 final candidates were reviewed and discussed in great detail, including:

- Gene Description (SGD) and relevant literature
- Selection Tool Chart
- Relative location on the Fitness across conditions chart (Fish)

- Relative location on the Frequency/fitness chart for each condition
- Relative location on the Fitness histogram for each condition
- Relative location on the acceleration chart for each condition

Conservation and expression profile data were also taken into account.

## Candidates Evolutions

After the above process, separate evolutions for the most promising candidates have begun. Due to incubator capacity, a total of nine strains (plus WT controls) were evolved separately under three conditions- YPD, high pH and SD. Each strain was evolved in three independent biological repeats. Since time was of the essence- strain verification (via PCR, Sanger sequencing and FACS) took place in parallel, resulting in losing part of the evolutions due to negative verification results as is detailed in Table 8 below.

plate #	Gene name	Gene #	Condition	Incubator	Evolution details
1	WT	10	YPD	1	Since Strain #4 was not verified- This control YPD evolution was frozen at day 11 (generation #77)
2	YBR230C (OM14)	4	YPD	1	This strain was not verified (turned out to be a diploid). Evolution was frozen on day 11 (generation #77). A haploid strain was ordered yet arrived too late to allow ample evolution time.
3	WT	10	pH	1	Overall 56-58 evolution days (392-406 generations) Dilution every 2>1 days (regime updated due to fitness increase).
4	YDR508C (GNP1)	2	pH	1	Overall 59 evolution days (413 generations) Dilution every 2>1 days (regime updated due to fitness increase).
5	YGL165C	9	pH	1	Overall 53-58 evolution days (371-406 generations) Dilution every 2>1 days (regime updated due to fitness increase).
6	YHL017W	1	pH	2	This strain was verified as YBR258C (SHG1) due to original pool mislabeling. Evolution was frozen at day 4 (generation #28). The correct strain was ordered yet arrived too late to allow ample evolution time.
7	YJR005C-A (LSO1)	5	pH	1	This strain was not verified at first via PCR and evolution was frozen at day 4 (generation #28). Once it was finally verified evolution was retried but to no avail (insufficient growth under high pH)
8	WT	10	SD	2	Overall 58 evolution days (490 generations)
9	YDL194W (SNF3)	7	SD	2	Overall 58 evolution days (490 generations)
10	YDR092W (UBC13)	8	SD	2	Overall 58 evolution days (490 generations)
11	YOL090W (MSH2)	3	SD	2	Overall 56 evolution days (476 generations)
12	YPL199C	6	SD	2	Overall 60 evolution days (504 generations)

**Table 8 - Leading Candidates' Separate Evolutions**– This table details the prominent candidates' separate evolutions. Overall nine strains (numbered 1-9) were evolved over three conditions (one was evolved on YPD, 4 were evolved on SD and 4 were evolved on high pH), along with a WT control at each condition (numbered 10). Four of the strains were not verified, rendering their evolutions irrelevant, and the other five were evolved for 371-504 generations as is presented in the table.

These evolution results were assessed by performing growth tests. SD evolved strains' growth was evaluated by using SD medium, and pH evolved strains were evaluated by using several high pH media, including pH 7.6 (the evolution pH), pH 8.6, pH 9.0 and a YPD reference.

The results are presented in Figure 26 (for SD) and Figure 27 (for high pH). Please note that the Ancestor strain is always A, depicted in light blue, and the other three strains are the evoltant's three biological repeats. In addition- please note that both X and Y scales vary between charts, due to different yields for different conditions (Y), and tests failing before completion (X).

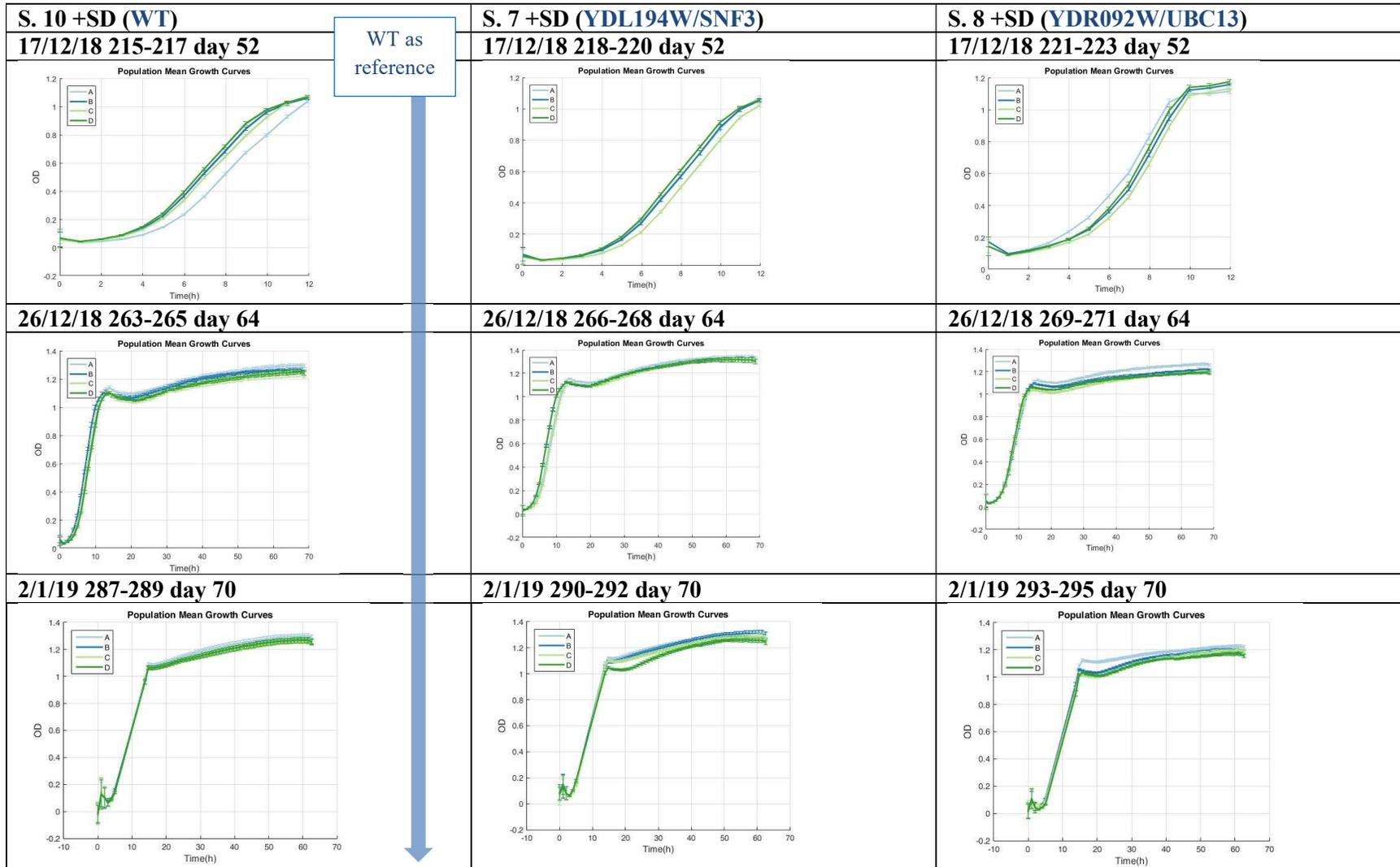
**Figure 26- SD growth test results**- This figure presents the results for growth tests that were performed on the four mutant strains and the WT control that were evolved on SD. Each column represents a strain (the WT strain results are repeated twice, at the left-hand column of each page, to provide reference). For each strain, growth tests were performed at different stages of the evolution. Their results are presented from the top (=the earliest stage) to the bottom (=the final stage) of each column (i.e. at the rows of the table presents stages in the evolution). Each growth curve presents the OD vs. time for the ancestor strain (marked as strain "A" and depicted in light blue) and the three evoltants' independent biological repeats (strains "B", "C" and "D"). Please note that both X and Y scales vary between charts, due to different yields for different conditions (Y), and tests failing before completion (X). **Please mind the different time scales used in this figure- some of the experiments were considerably shorter due to robot malfunctions.**

A = Ancestor strain  
 B }  
 C } = Mutants' evolution biological repeats  
 D }

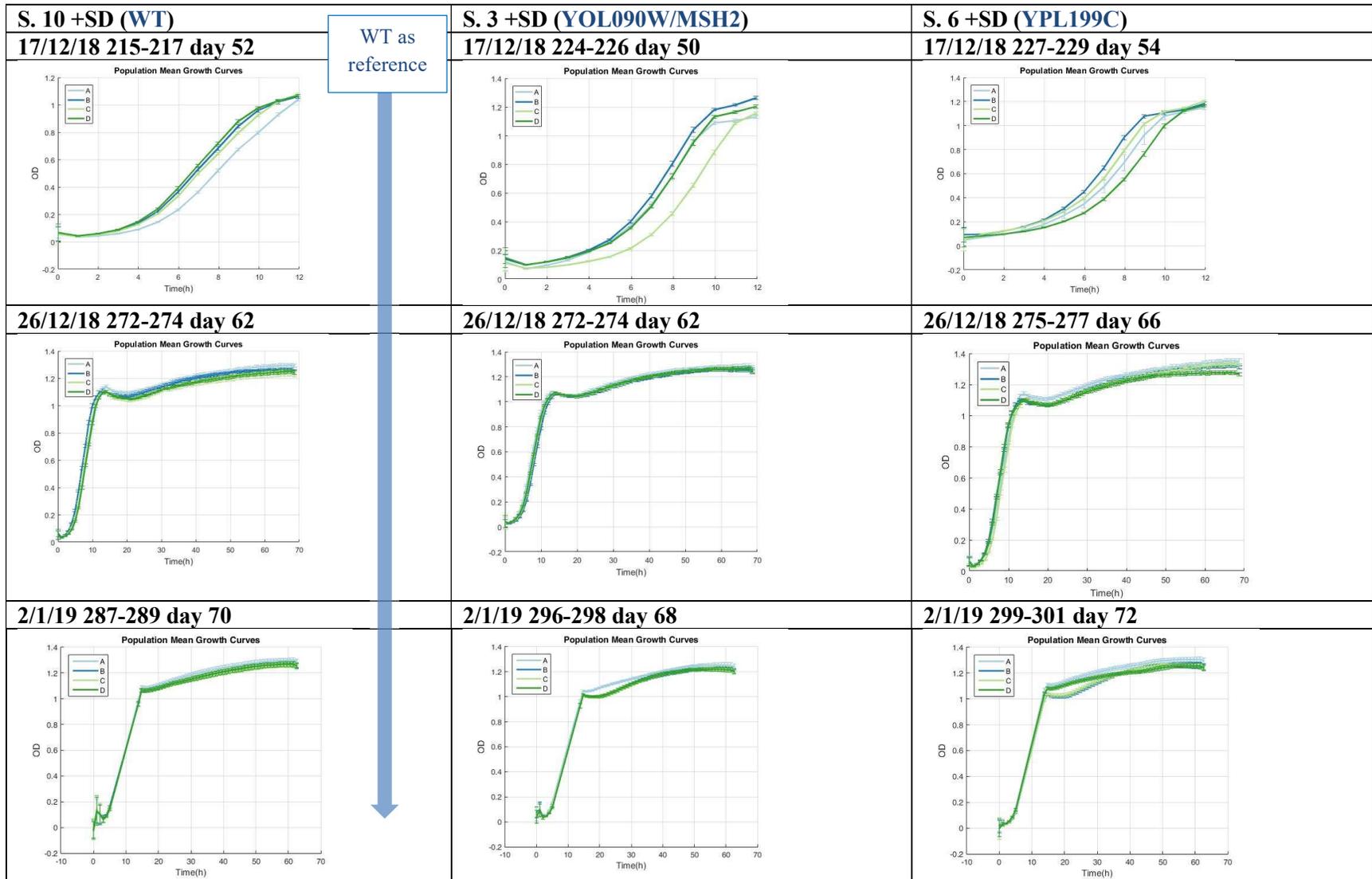
Evolution Time



WT as reference

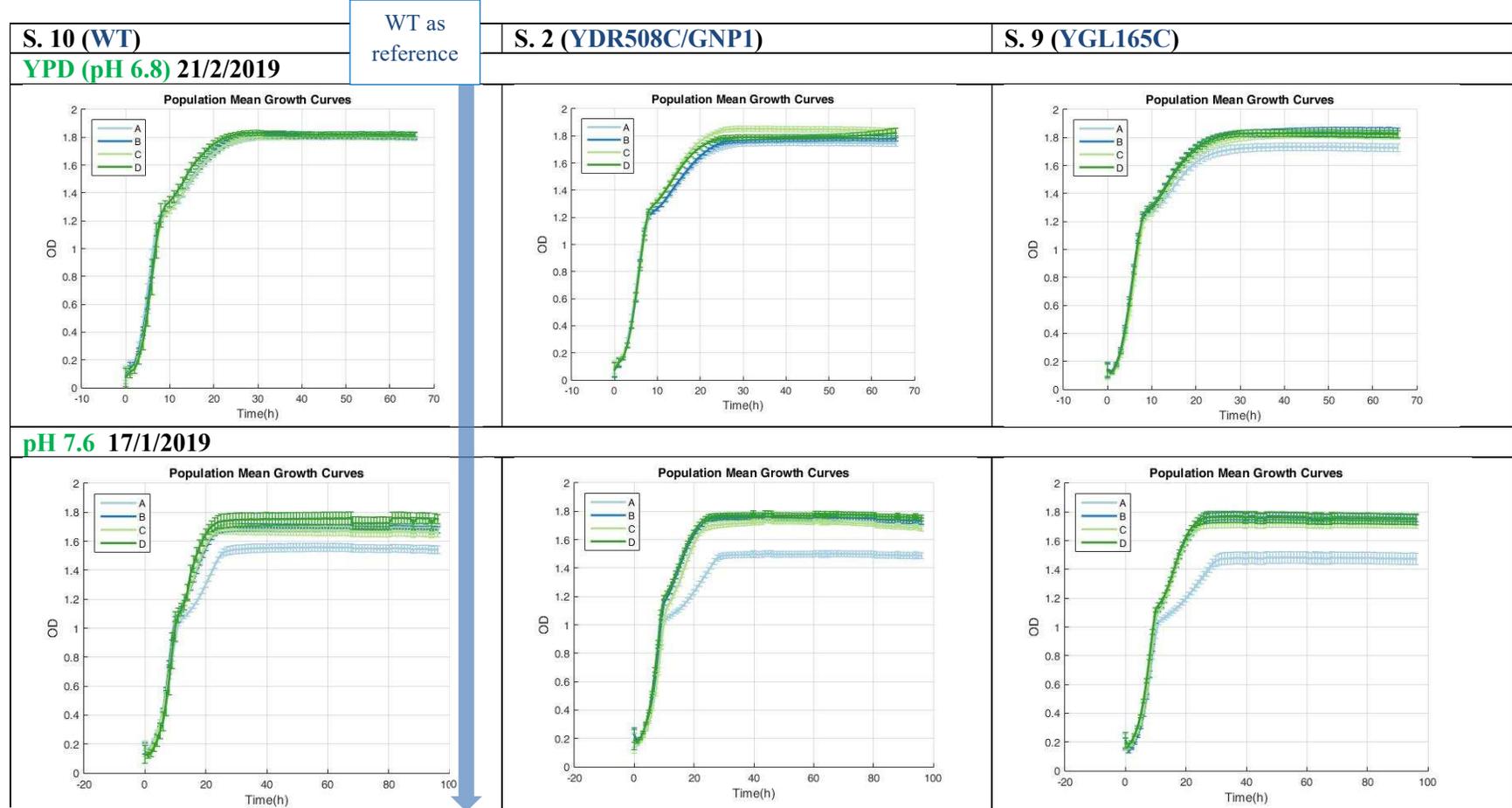
A = Ancestor strain  
 B } = Mutants' evolution biological repeats  
 C }  
 D }



Evolution Time

A = Ancestor strain  
 B } = Mutants' evolution biological repeats  
 C }  
 D }

**Figure 27- pH growth test results (final evoltants)** - This figure presents the results for growth tests that were performed on the two mutant strains and the WT control that evolved on the high pH medium (pH 7.6). Each column represents a strain (the WT strain is presented in the left column). For each strain's final evoltants, growth tests were performed on media with different pH values. Their results are presented from the top (YPD, pH 6.8) to the bottom (pH 9) of each column (i.e. at the rows of the table). Each growth curve presents the OD vs. time for the ancestor strain (marked as strain "A" and depicted in light blue) and the three evoltants' independent biological repeats (strains "B", "C" and "D"). Please note that both X and Y scales vary between charts, due to different yields for different conditions (Y), and tests failing before completion (X). **Please mind the different OD scales used in this figure.**



pH value

More pH values next page

A = Ancestor strain  
 B }  
 C } = Mutants' evolution biological repeats  
 D }

pH growth test results (final evoltants) continued-

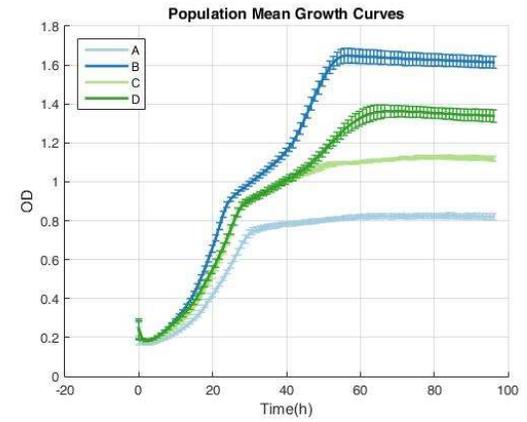
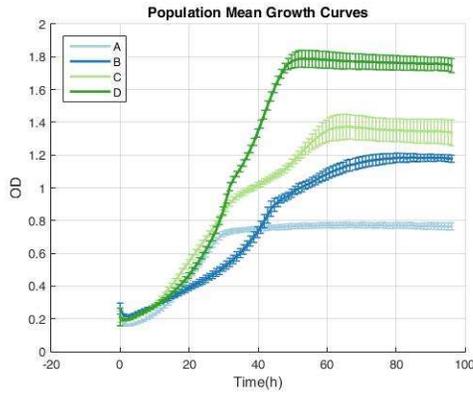
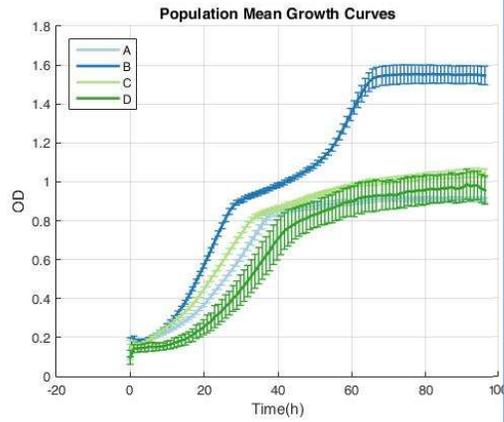
WT as  
 reference

S. 10 (WT)

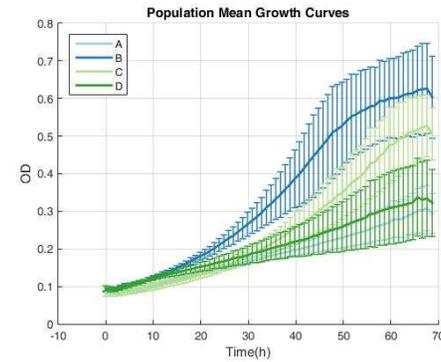
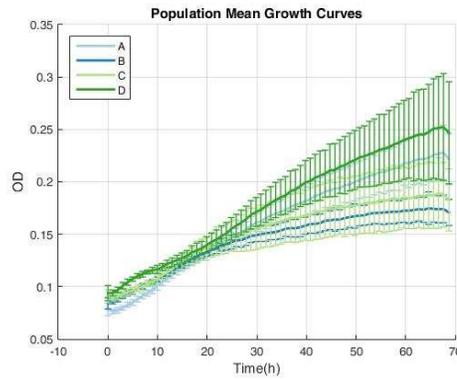
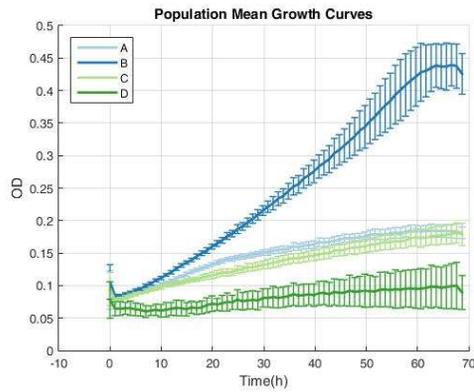
S. 2 (YDR508C/GNP1)

S. 9 (YGL165C)

pH 8.6 7/3/2019



pH 9 7/3/2019



pH value

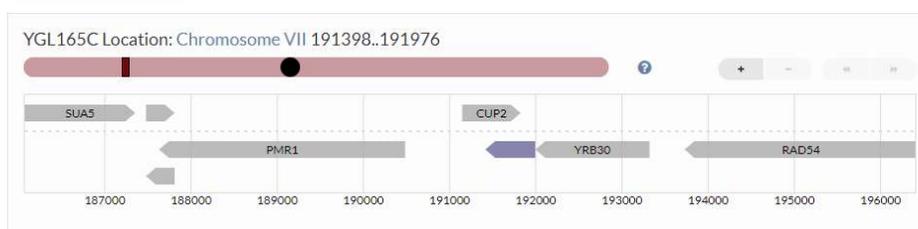


As can be seen from Figure 26, no significant improvement was accomplished for the SD evolutions, and the strains' dynamics over time were similar to the WT's'.

However, for the high pH evolutions, significant improvement in pH tolerance was achieved during the evolution for both the mutant strains and the WT. As can be seen in Figure 27, although starting from a slightly lower fitness level than the WT, the YDR508C (GNP1) and YGL165C strains were able to improve to similar or higher fitness levels, a phenomenon more evident for higher pH levels, hence implying a change in evolutionary pattern. An attempt was also made to map the evolutionary dynamics (growth tests at different evolutionary time points, as was done for SD), yet low equipment technical reliability plagued these tests too and prevented the collection.

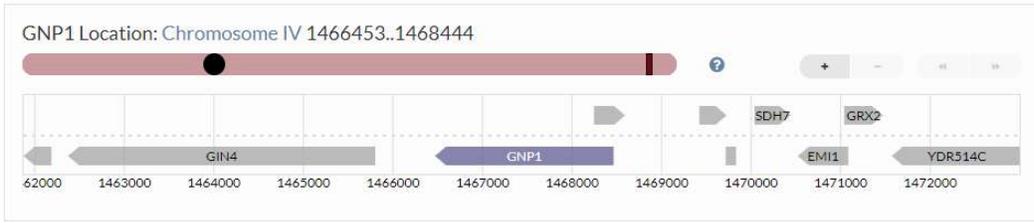
These differences in the extent and nature of the improvement between the WT and candidates (especially at a higher pH than the one that was used for the original evolution) are interesting, especially in view of the fact that those genes were not previously known to have any effect on evolvability.

Specifically, YDR508C (GNP1) codes for a High-affinity glutamine permease and also transports Leu, Ser, Thr, Cys, Met and Asn, and YGL165C is considered a dubious open reading frame that based on available experimental and comparative sequence data is considered unlikely to encode a functional protein. However, it partially overlaps the verified ORF of CUP2/YGL166W (Figure 28), which is a Copper-binding transcription factor. CUP2 activates transcription of the metallothionein genes CUP1-1 and CUP1-2 in response to elevated copper concentrations and is required for regulation of copper genes in response to DNA-damaging reagents, but those mutants did not emerge in our screen in their own right.



**Figure 28- YGL165C location**– This figure, taken from the SGD website [54] presents the chromosome map of the YGL165C ORF location. Located on chromosome VII at the anti-sense strand, it is adjacent to the RAD54 ORF and partially overlaps the CUP2 ORF on the sense strand.

As can be seen in **Figure 29**, YDR508C (GNP1) Also partially overlaps a small ORF, named YDR509W. Described as a dubious open reading frame that is unlikely to encode a functional protein, this ORF too did not emerge in our screen in its own right, same as its other adjacent neighbors that are not known to have any effect on evolvability.

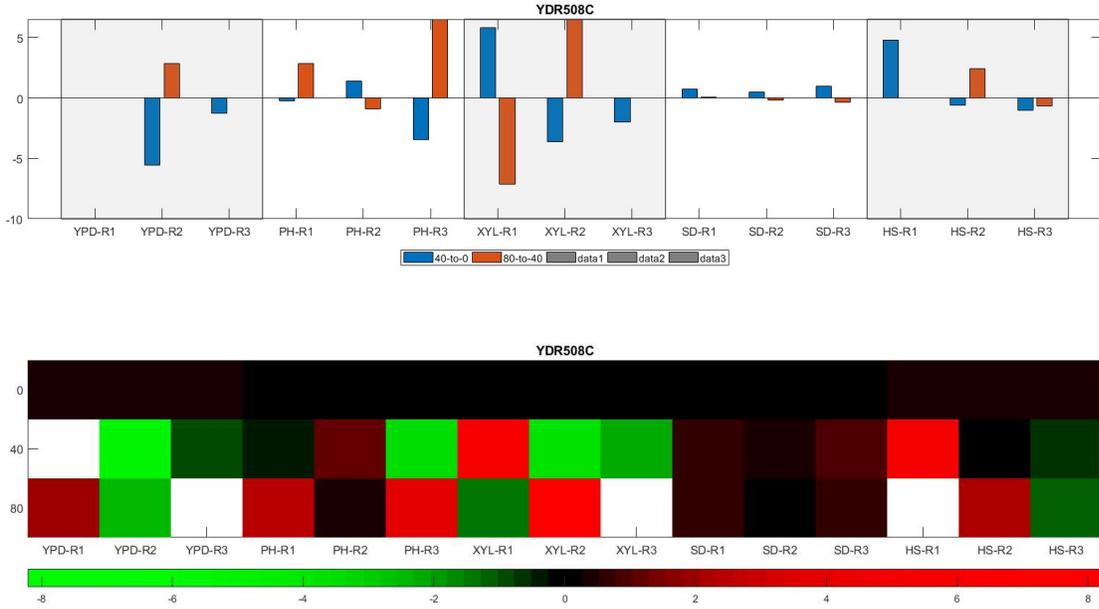


**Figure 29- YDR508C/GNP1 location**– This figure, taken from the SGD website [54] presents the chromosome map of the YDR508C/GNP ORF location. Located on chromosome IV at the anti-sense strand, it partially overlaps the YDR509W ORF on the sense strand.

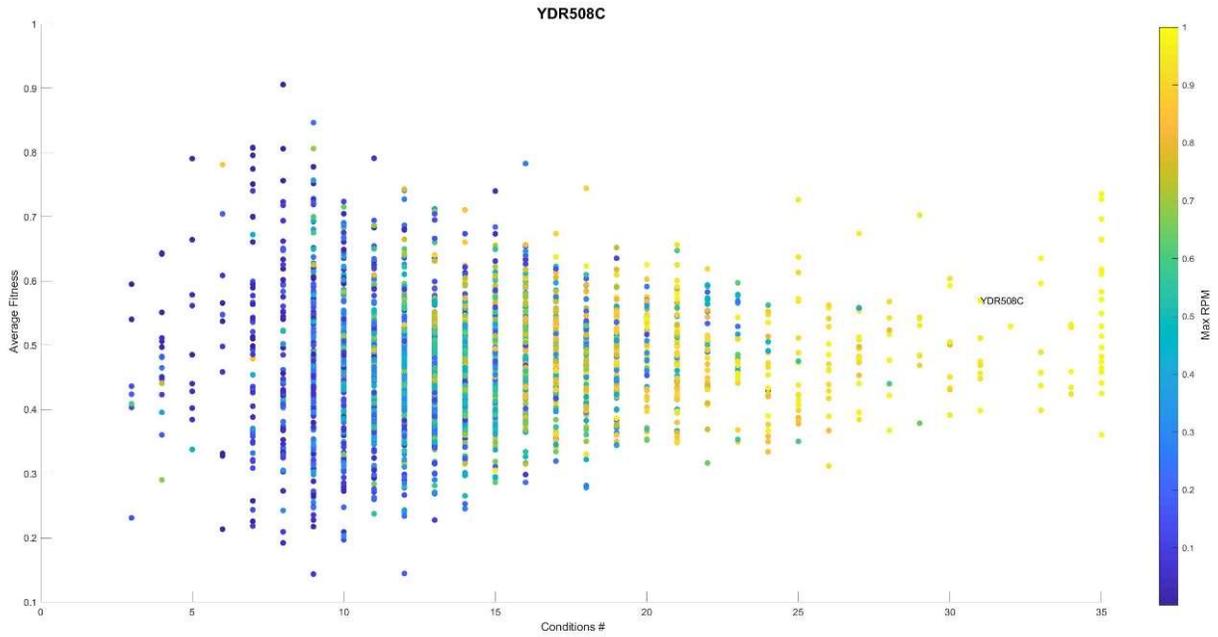
Both YDR508C (GNP1) and YGL165C performed significantly well at the pooled evolution, as can be seen in Figure 30 and Figure 31 (and hence selected for separate evolutions). Therefore, after verifying the evolvability effect via the separate evolutions and growth essays, sequencing took place in order to better understand the underlying mechanism.

**Figure 30- YDR508C (GNP1) pool evolution data**– The data in this figure was analyzed and visualized with Dr. Hila Gingolds’ kind assistance. The figure presents YDR508C/GNP1 results in the preliminary analysis that took place in order to select candidates for separate evolutions. These results consist of (A) Its Selection Tool Chart; (B) Its relative location on the Fitness across conditions chart (Fish); (C) Its relative location on the Frequency/fitness chart for each condition; (D) Its relative location on the Fitness histogram for each condition; (E) Its relative location on the acceleration chart for each condition;

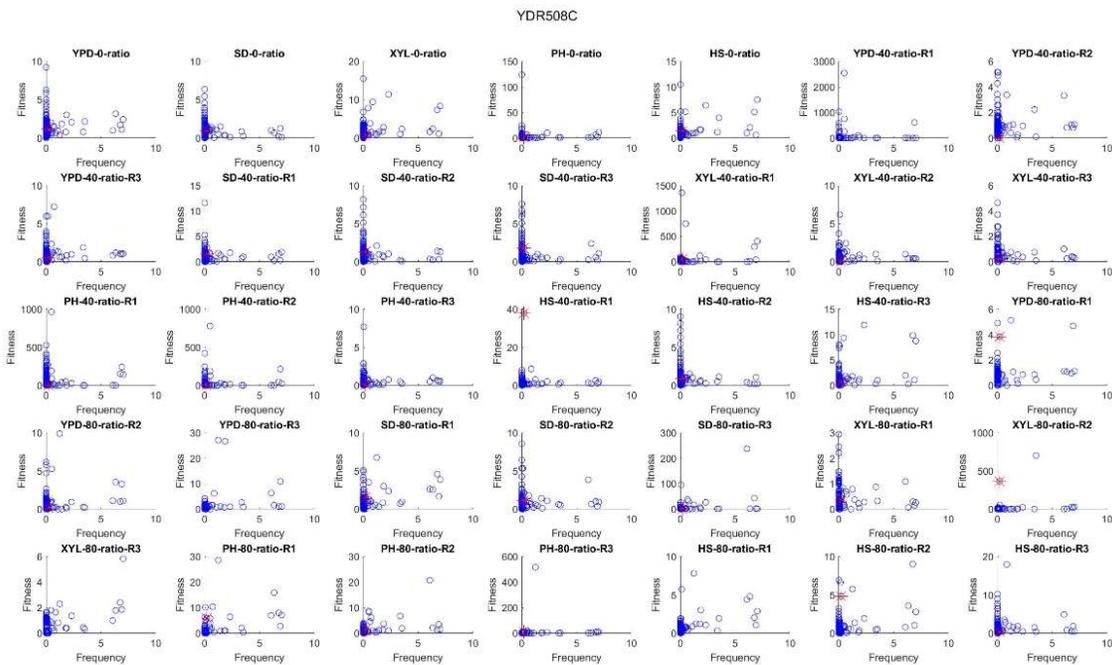
# A Selection Tool Chart



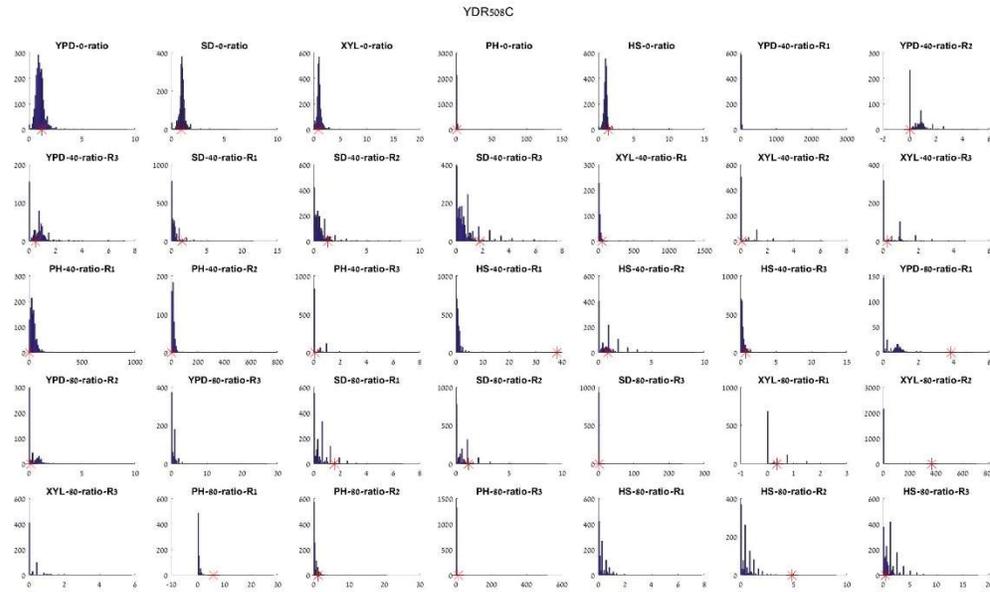
## B Relative location on the Fitness across conditions chart (Fish)



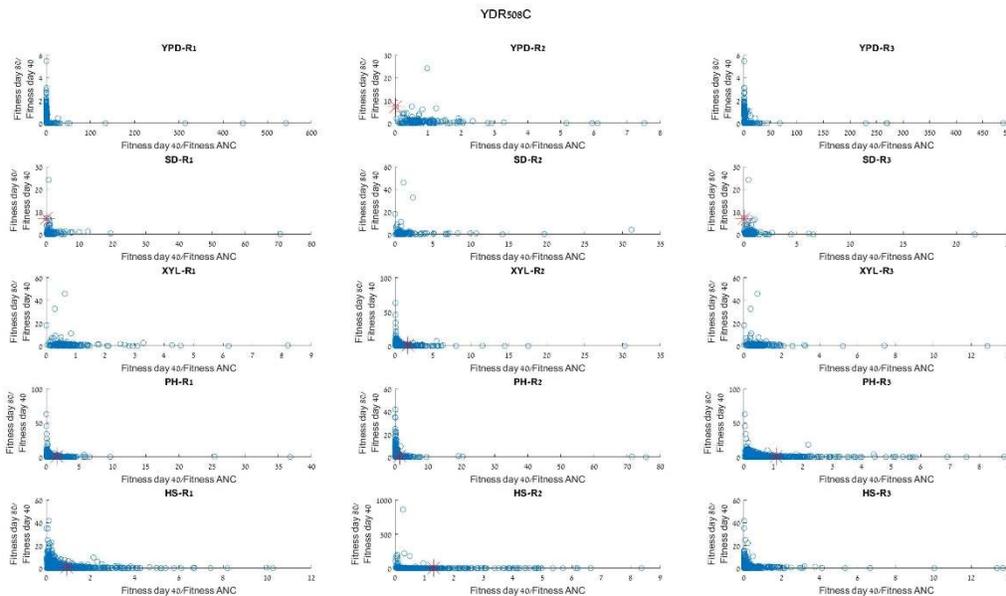
## C Relative location on the Frequency/fitness chart for each condition



## D Relative location on the Fitness histogram for each condition

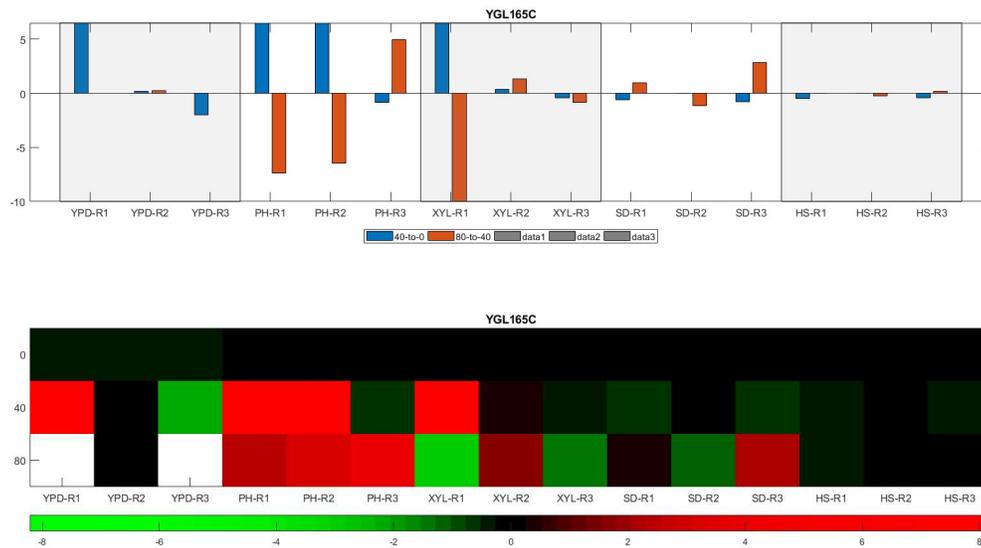


## E Relative location on the acceleration chart for each condition

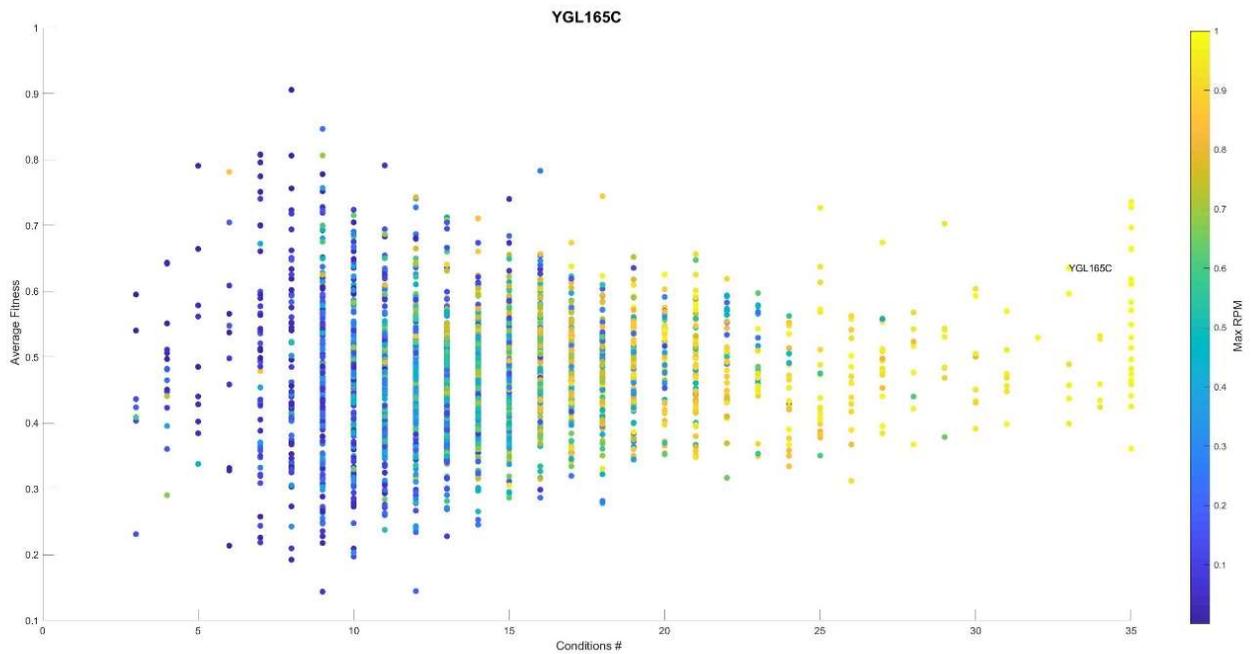


**Figure 31- YGL165C pool evolution data** -The data in this figure was analyzed and visualized with Dr. Hila Gingolds' kind assistance. The figure presents YGL165C results in the preliminary analysis that took place in order to select candidates for separate evolutions. These results consist of (A) Its Selection Tool Chart; (B) Its relative location on the Fitness across conditions chart (Fish); (C) Its relative location on the Frequency/fitness chart for each condition; (D) Its relative location on the Fitness histogram for each condition; (E) Its relative location on the acceleration chart for each condition;

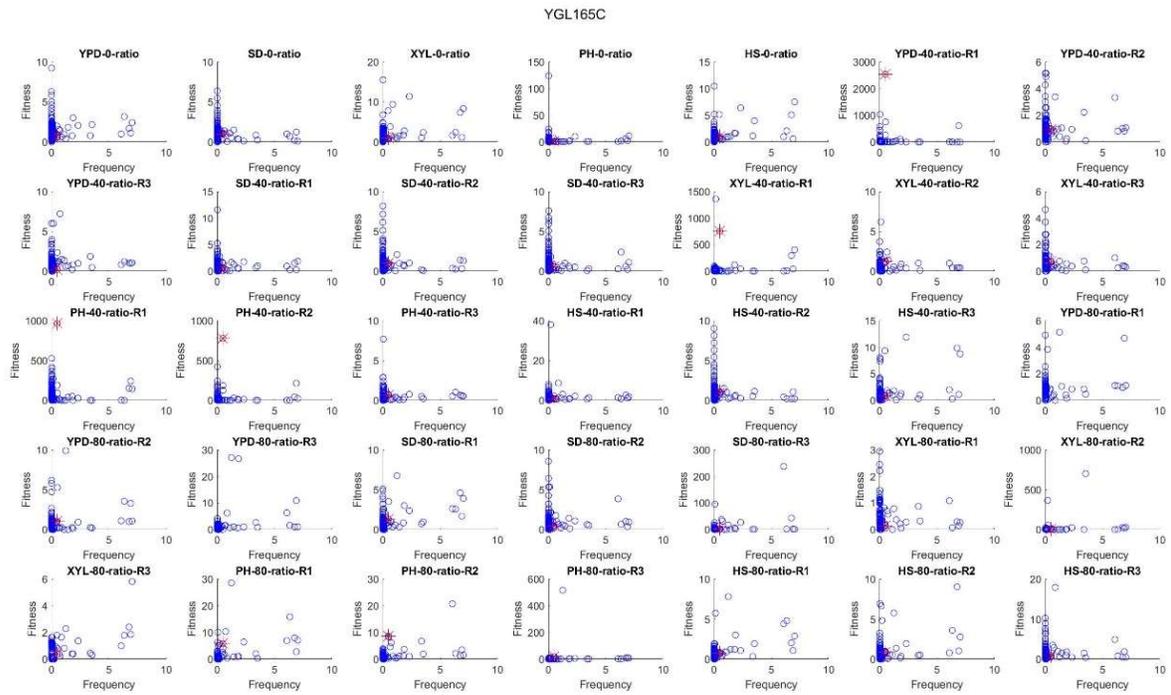
# A Selection Tool Chart



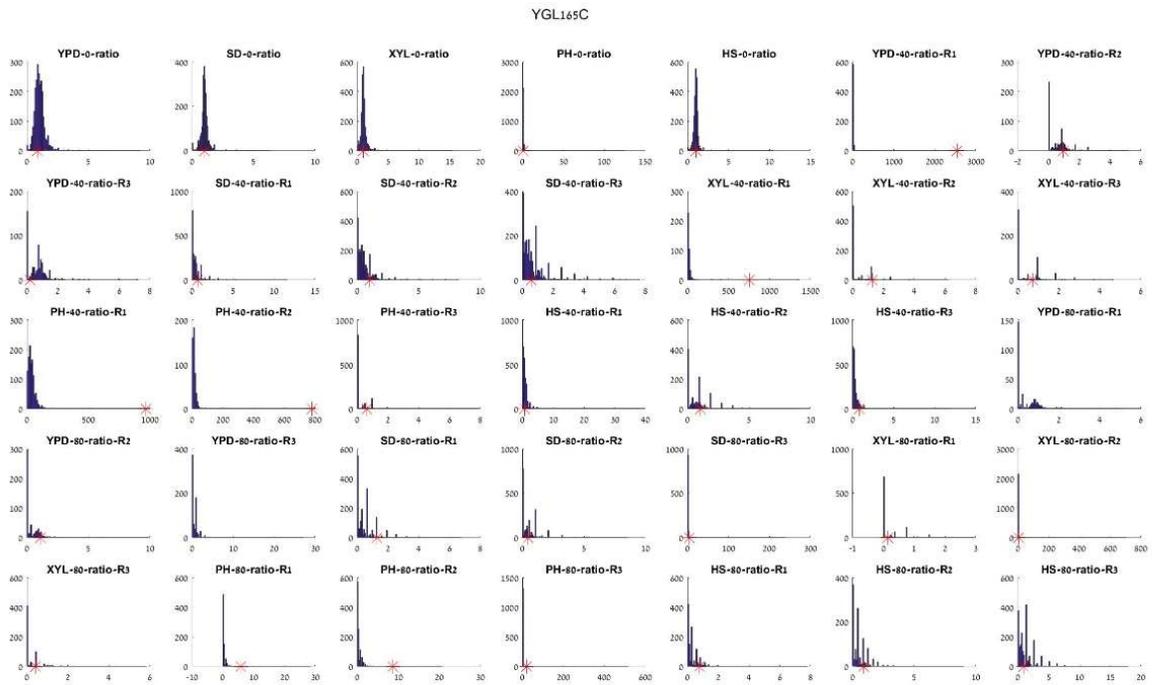
# B Relative location on the Fitness across conditions chart (Fish)



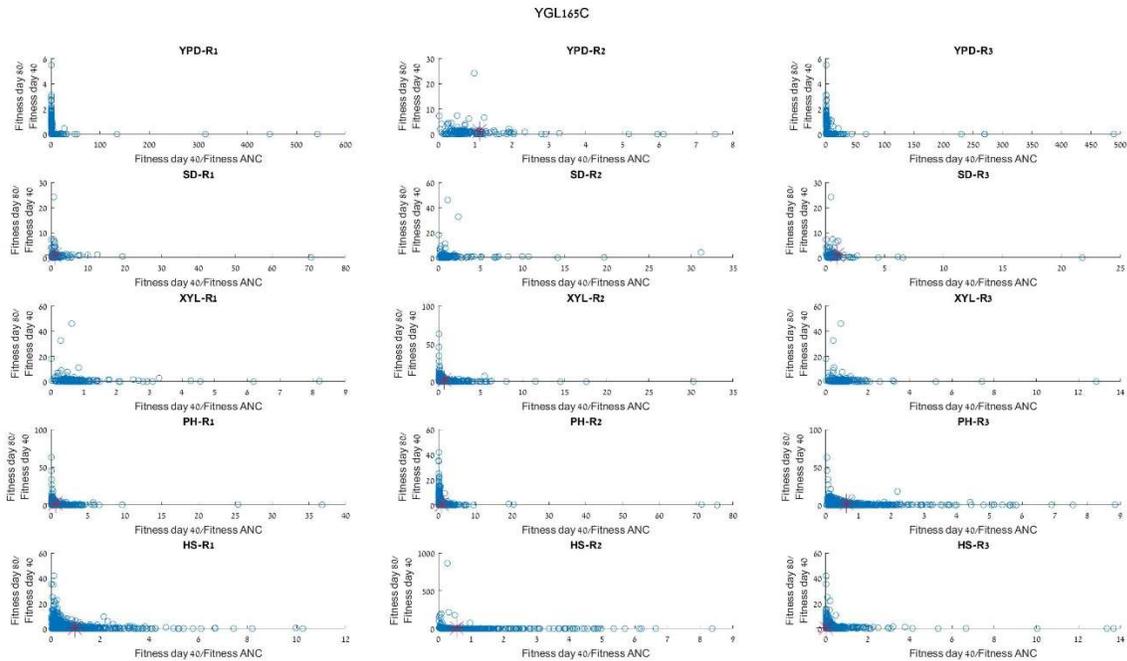
# C Relative location on the Frequency/fitness chart for each condition



# D Relative location on the Fitness histogram for each condition



## E Relative location on the acceleration chart for each condition



### Candidates' evolutions sequencing data reduction results

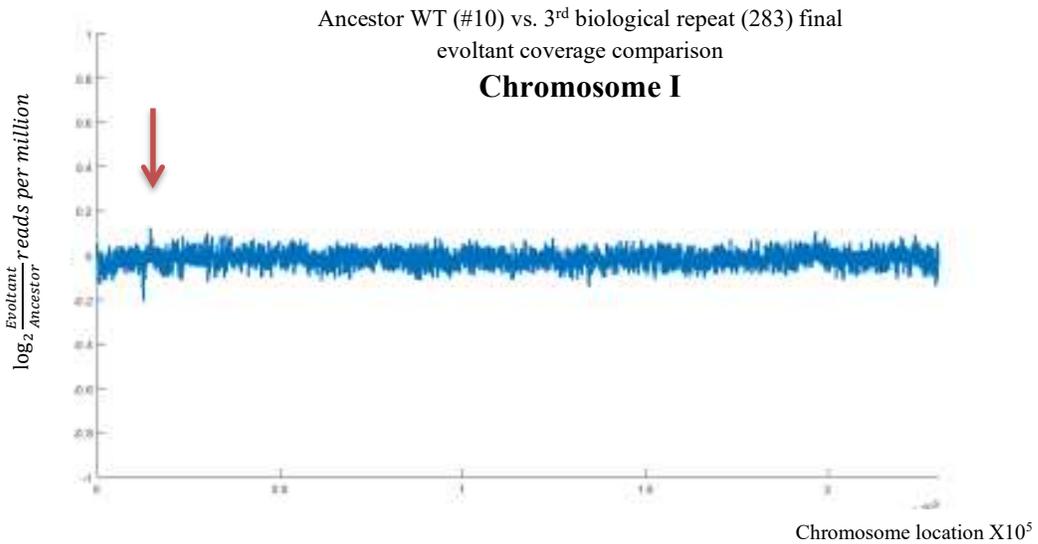
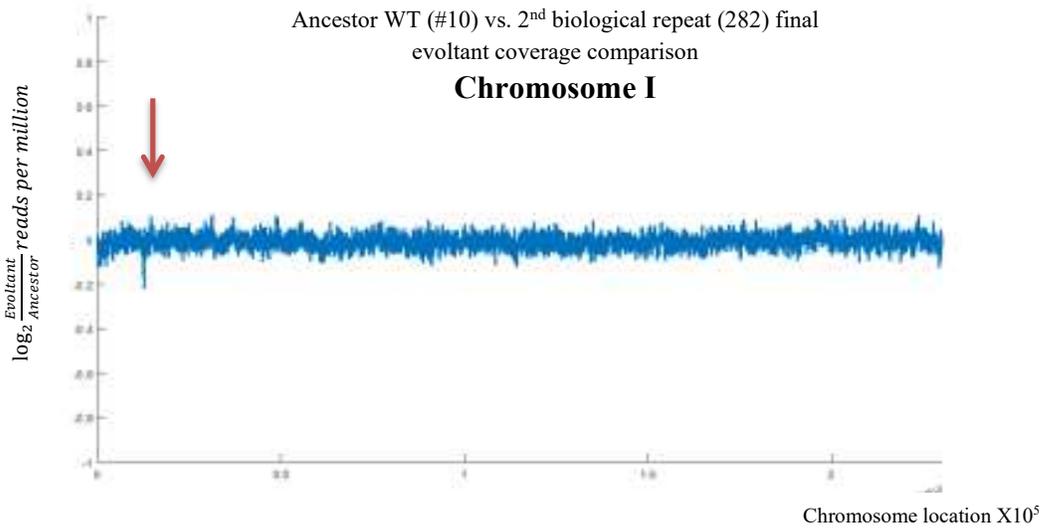
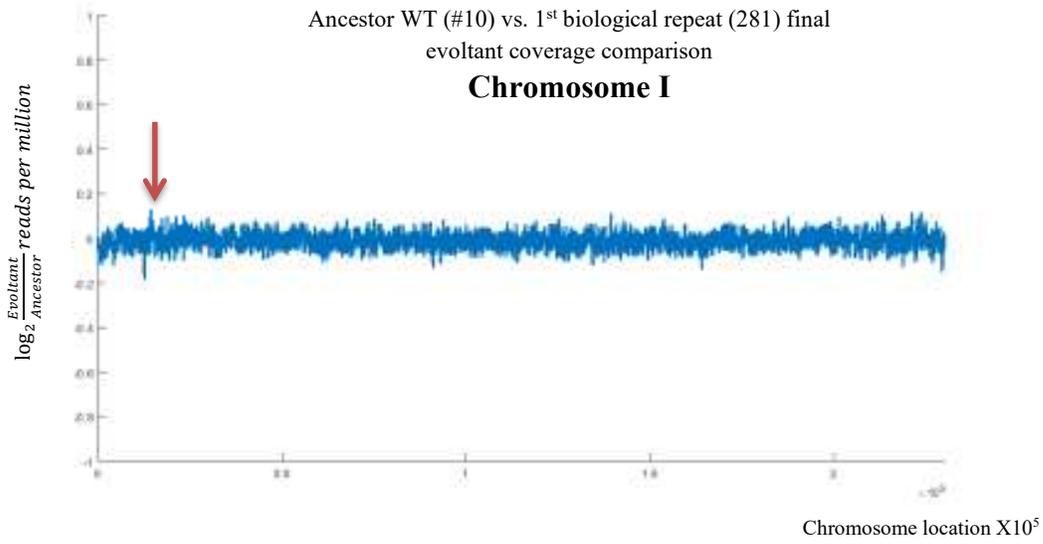
As was mentioned above, YDR508C (**GNP1**) and YGL165C were sequenced to better understand the underlying evolutionary mechanisms. The data set structure was depicted in Figure 18 in the method section and consisted of the three strains (the two mutants + the WT) ancestors and the final evoltants of the three independent evolutions of each to a total of 12 libraries.

As was described in the method section, sequencing data was aligned, and initial confidence in the data was gained by verifying that all libraries had proper coverage of the genome, including sanity checks in the form of the ORF deletions.

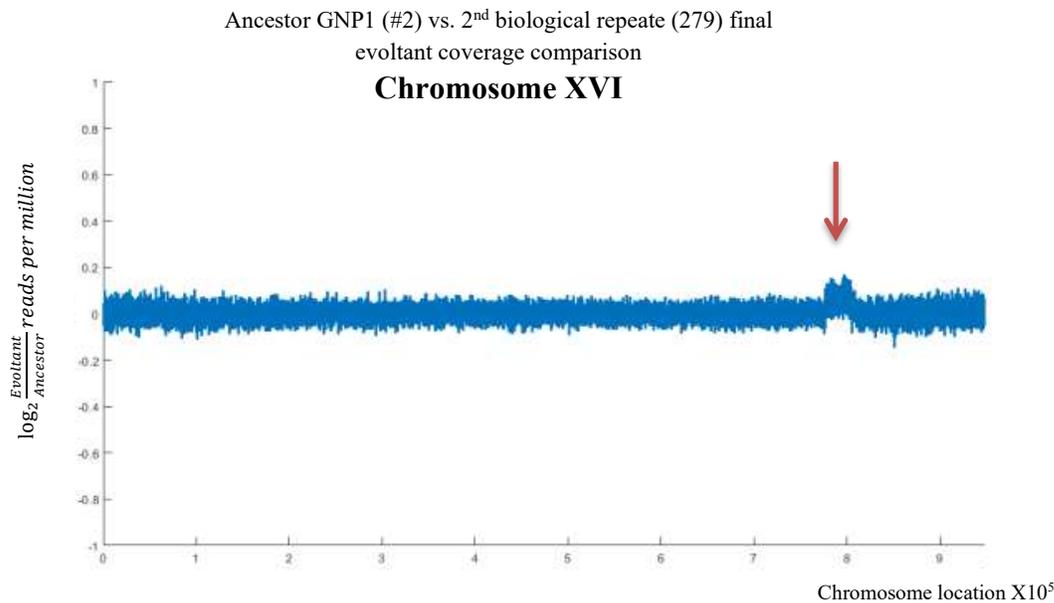
Then, an additional verification of the sequencing coverage took place by comparing each evoltant to the ancestor strain- chromosome by chromosome by using a dedicated script to produce coverage maps and reviewing under/over-sampled areas.

This analysis revealed six anomalies -

- A small area of no coverage at Chromosome III – around 150,000 – repeated for all the libraries.
- A small area of irregular (over/under) coverage at Chromosome VIII – around 220,000 – repeated for all the libraries.
- An area of irregular (over/under) coverage at Chromosome XII – around 460,000 – repeated for all the libraries.
- Irregular coverage of the mitochondria - repeated for all the libraries.
- A very small area of under coverage at Chromosome I – around 15,000- repeated only for the WT libraries (see Figure 32)
- An area of significant over-coverage at Chromosome XVI – around 780,000 – was discovered only for one GNP1 evolution biological repetition library (279)- see Figure 33.



**Figure 32- WT Chromosome I under-coverage** - The data in this figure was analyzed and visualized with Dr. Hila Gingolds' kind assistance. It presents a comparative chromosome coverage maps of each of the three final WT evoltants. Reads coverage was compared to the relevant ancestor sequencing results to detect big deletions/additions, significant under covered/overcovered areas, ploidity changes, or significant unaligned areas. This figure presents chromosomes' I coverage comparison of the three WT biological repeats (281, 282 and 283) final evoltant to their (#10) ancestor. The X-axis presents chromosome location, and the Y-axis presents the normalized relative coverage. As can be seen, the same under-coverage area is present at all the repeats, at around 15,000bp (marked by a red arrow).



**Figure 33- GNP1 Chromosome XVI over-coverage (repetition 279)** -The data in this figure was analyzed and visualized with Dr. Hila Gingolds' kind assistance. It presents a comparative chromosome coverage maps of one of the final GNP1 evoltants. Reads coverage was compared to the relevant ancestor sequencing results to detect big deletions/additions, significant under covered/overcovered areas, ploidity changes, or significant unaligned areas. This figure presents chromosome XVI coverage comparison of the 2<sup>nd</sup> GNP1 biological repeats (279) final evoltant to its (#2) ancestor. The X-axis presents chromosome location, and the Y-axis presents the normalized relative coverage. As can be seen, a significant over-coverage area is present at around 780,000 bp (marked by a red arrow).

As intriguing as the above anomalies may be, their consistency over samples, correlation to highly repetitive areas and relative scarcity make them insignificant in our view, and the sequencing coverage was therefore considered satisfactory.

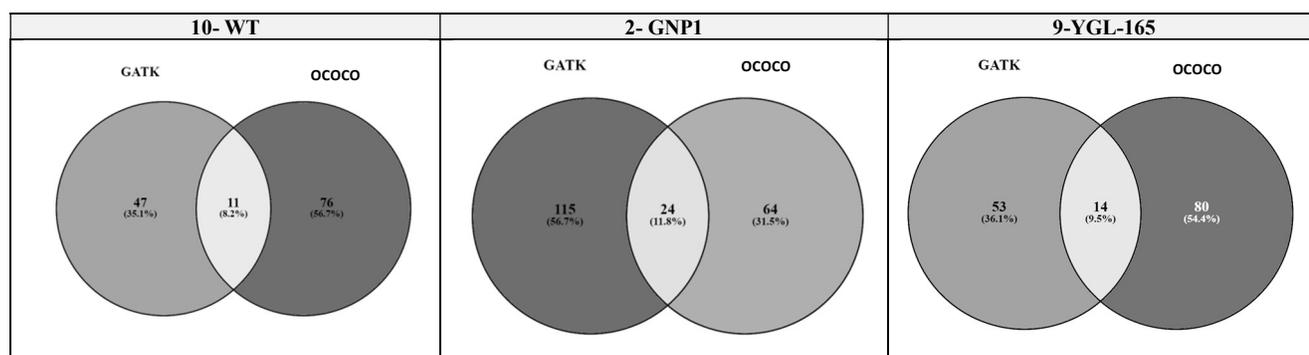
Therefore, since overall no big deletions/additions, significant under covered/overcovered areas, ploidity changes, or significant unaligned areas were found, the analysis continued by identifying SNPs and mapping point mutations to known ORFs by using two variant calling tools- OCOCO and GATK as was described in the method section above.

The total number of resulting SNPs (where the evoltant is different from the appropriate ANS) and more for both variant calling methods are presented in Table 9 below.

Method ↻	Total # of SNPs at all the biological repetitions			Total # of SNPs that fell within known ORFs at all the biological repetitions			Total # of ORFs containing SNPs at all the biological repetitions		
	Strain ↻	10	2- GNP1	9-YGL	10	2- GNP1	9-YGL	10	2- GNP1
OCOCO	636	545	554	213	192	207	87	88	94
GATK	338	848	363	76	168	76	58	139	67

**Table 9 - SNP analysis (GATK vs OCOCO)**– This table presents the results of the analysis performed on the sequencing results to identify SNPs and mapping point mutations to known ORFs. The analysis was performed by using two alternative methods- OCOCO and GATK (presented in the rows) on the WT (10), GNP1 (2) and YGL-165 (9) results, to include all three biological repeats for each. The total # of SNPs at all the biological repetitions, the total # of SNPs that fell within known ORFs at all the biological repetitions and the total # of ORFs containing SNPs at all the biological repetitions are presented in the columns for each strain.

An additional comparison between the results of using the two variation calling methods is presented in Figure 34 below-

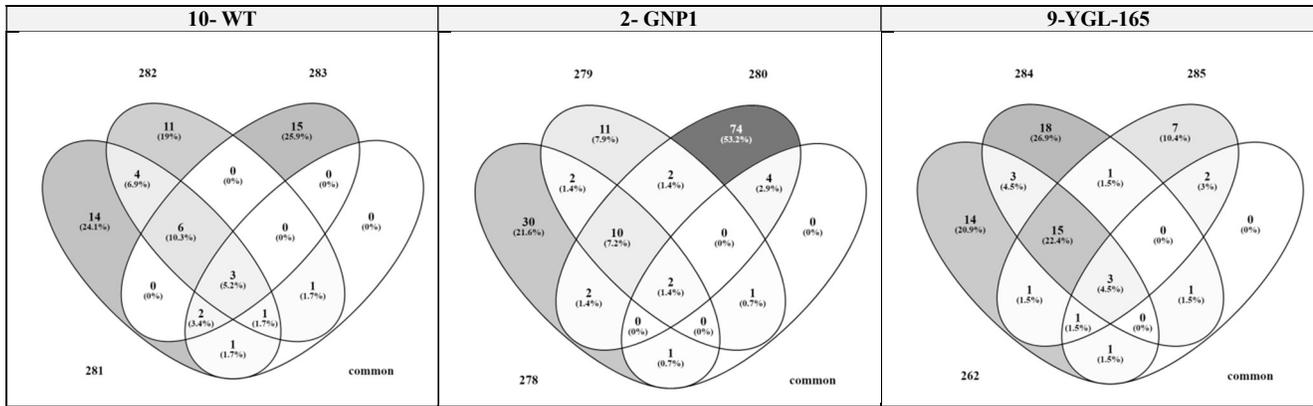


**Figure 34- Variation calling methods comparison- ORFs containing SNPs** – These Venn diagrams present for each strain (to include the #10- WT, #2- GNP1 and #9- YGL-165) the number of known ORFs to which SNPs were mapped by each of the variants calling methods (OCOCO and GATK).

As can be seen from the above table and figure- significant differences were observed between the two methods, yet some ORFs emerged using both. Since GATK is considered to be the current standard, its results were used as a basis for continuing the analysis, and the following discussion is based on them. Regardless, some of the most interesting results have emerged by both methods (will be noted when applicable), a fact that increases our confidence in their validity.

As can be seen from Table 9 above, both WT and the YGL-165 (strain #9) mutant have an approximately similar number of SNPs, while the GNP1 (strain #2) mutant exhibits a higher one. From past experience, this amount of SNPs seems logical (albeit a bit on the high side).

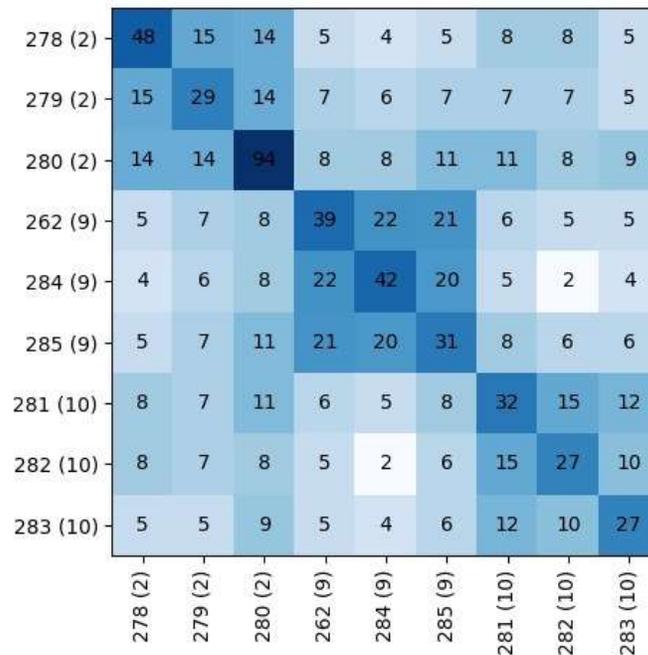
A significant amount of the GNP1 (strain #2) mutant SNPs was detected at one specific biological repetition (#280) as is shown below. Comparing the number of SNPs at the other two repeats gives a much similar number to WT and the YGL-165 (strain #9) mutant results. The “Known ORFs containing SNPs” distribution between biological repeats for each of the strains is presented in Figure 35.



**Figure 35- ORFs containing SNPs for different biological repeats** – These Venn diagrams present for each strain (to include the #10- WT, #2- GNP1 and #9- YGL-165) the number of known ORFs to which SNPs were mapped for each of the three independent biological repeats by using the GATK variants calling method.

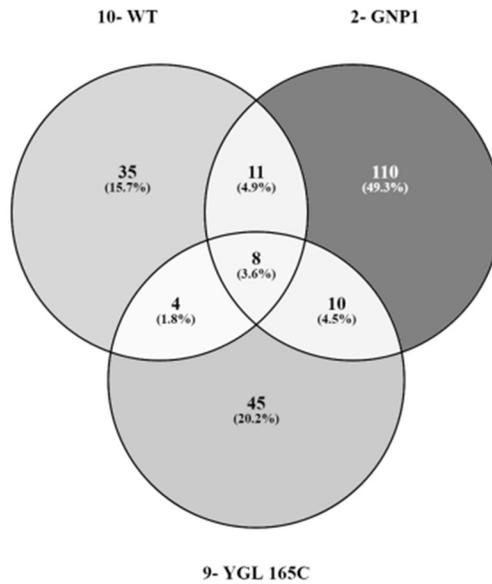
A numeric view of the SNPs for the various strains and biological repeats analysis is presented in Figure 36. As can be seen, a significant correlation was observed between libraries belonging to the same strains' biological repeats.

Overall- this logical pattern of SNPs increases the confidence in the data set.



**Figure 36- SNP correlation between libraries** – This diagram shows for all the nine libraries (that consists of three biological repeats for each of the three strains) given in both the rows and the columns, the number of known ORFs to which SNPs were mapped into, for the two libraries that define that square.

Following the above ORF number analysis, a complementary, in-depth analysis took place, focusing on the number and nature of the SNPs within known ORFs. The SNPs distribution between the different strains is presented in Figure 37.



**Figure 37- Mutated Genes for the three strains** – This Venn diagram presents for each strain (to include the #10- WT, #2- GNP1 and #9- YGL-165) the number of known ORFs to which SNPs were mapped, by using the GATK variants calling method.

From the above figure, it is evident that each strain exhibited a unique repertoire of mutated genes of its own (although a small part of the repertoire was common for all three strains).

Using [GORilla](#), enrichment for the SET 3 complex components was detected on the mutated genes that were mapped for YGL-165, as is presented in Table 10. Additional enrichments presented a FDR q-value that didn't meet the significance threshold (lower than 0.05), or referred to highly repetitive elements (FLO9 and FLO1) in the genome and are therefore considered artifacts.

**Table 10 - Enrichment analysis** – This table presents GO categories analysis of the number of known ORFs to which SNPs were mapped for each of the three strains (to include the #10- WT, #2- GNP1 and #9- YGL-165) and their combinations (ORFs that were mapped for more than on strain). This analysis was performed using GOrilla. The table presents the group, the nature of the enrichment (process/function/component), the relevant GO-Term and its description, the P-value and the FDR q-value (the percentage of significant tests that are expected to result in false positives), the enrichment =  $b/n/B/N$  (where N=5520, i.e. the functional yeast genes number, B- the number of genes belonging to that GO category, n- the size of the group selected and b- the number of genes belonging to that category in the group- that are also detailed in the next column). Rows with FDR q-value below 0.05 are highlighted.

Group	Enrichment	GO Term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
WT	Process	GO:0070534	protein K63-linked ubiquitination	4.21E-04	1.00E+00	58.72 (5520,4,47,2)	MMS2 - e2 ubiquitin-conjugating protein mms2; RAD6 - e2 ubiquitin-conjugating protein rad6
WT	Process	GO:0000128	flocculation	4.21E-04	1.00E+00	58.72 (5520,4,47,2)	FLO9 - flo9p; FLO1 - flo1p
WT	Function	GO:0003676	nucleic acid binding	1.29E-04	3.24E-01	2.20 (5520,1120,47,21)	NOG2 - nog2p; MPT5 - mpt5p; SPO11 - spo11p; DBP2 - dbp2p; BRR2 - brr2p; NPL3 - npl3p; REF2 - ref2p; SWI1 - swi1p; EST2 - est2p; CCR4 - ccr4p; PUF3 - puf3p; XBP1 - xbp1p; POL12 - pol12p; MAC1 - mac1p; DBP9 - dbp9p; NOP2 - nop2p; HMO1 - hmo1p; RAD6 - e2 ubiquitin-conjugating protein rad6; PRP40 - prp40p; NUP157 - nup157p; RPL22A - ribosomal 60s subunit protein l22a
WT	Function	GO:0003723	RNA binding	2.65E-04	3.33E-01	2.78 (5520,591,47,14)	NOG2 - nog2p; MPT5 - mpt5p; DBP2 - dbp2p; BRR2 - brr2p; NPL3 - npl3p; REF2 - ref2p; EST2 - est2p; CCR4 - ccr4p; PUF3 - puf3p; DBP9 - dbp9p; NOP2 - nop2p; PRP40 - prp40p; NUP157 - nup157p; RPL22A - ribosomal 60s subunit protein l22a
WT	Function	GO:0140097	catalytic activity, acting on DNA	8.33E-04	7.00E-01	5.30 (5520,133,47,6)	EST2 - est2p; SPO11 - spo11p; POL12 - pol12p; DBP9 - dbp9p; RAD6 - e2 ubiquitin-conjugating protein rad6; SWI1 - swi1p
WT	Component	GO:0031371	ubiquitin-conjugating enzyme complex	2.12E-04	2.24E-01	78.30 (5520,3,47,2)	MMS2 - e2 ubiquitin-conjugating protein mms2; RAD6 - e2 ubiquitin-conjugating protein rad6
WT	Component	GO:0044428	nuclear part	2.32E-04	1.23E-01	2.19 (5520,1075,47,20)	SRC1 - src1p; NOG2 - nog2p; SEN54 - sen54p; SPO11 - spo11p; BRR2 - brr2p; SET3 - set3p; NPL3 - npl3p; REF2 - ref2p; SWI1 - swi1p; EST2 - est2p; CCR4 - ccr4p; POL12 - pol12p; DBP9 - dbp9p; NOP2 - nop2p; HMO1 - hmo1p; CKA2 - cka2p; PRP40 - prp40p; NUP157 - nup157p; RRT14 - rrt14p; CDC14 - cdc14p
GNP1 (2)	Process	<a href="#">GO:0051302</a>	regulation of cell division	5.07E-04	1.00E+00	7.38 (5524,35,107,5)	FIR1 - fir1p; SLI15 - sli15p; GLC7 - glc7p; ACE2 - ace2p; CDC14 - cdc14p
GNP1 (2)	Process	<a href="#">GO:0033554</a>	Cellular response to stress	5.87E-04	1.00E+00	1.96 (5524,657,107,25)	ATG7 - atg7p; SNF5 - snf5p; MSH2 - mismatch repair atpase msh2; HAC1 - hac1p; RTT103 - rtt103p; MMS1 - mms1p; RIM15 - rim15p; EAF1 - eaf1p; ACT1 - actin; MMS2 - e2 ubiquitin-conjugating protein mms2; HSP12 - hsp12p; UBC13 - e2 ubiquitin-conjugating protein ubc13; RPO21 - rpo21p; HRD1 - e3 ubiquitin-protein ligase hrd1; CDC14 - cdc14p; HSC82 - hsp90 family

Group	Enrichment	GO Term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
							chaperone hsc82; PBS2 - pbs2p; MLP1 - mlp1p; SYG1 - syg1p; MMS22 - mms22p; PKH1 - pkh1p; GLC7 - glc7p; FLO1 - flo1p; YBP1 - ybp1p; XRS2 - xrs2p
GNP1 (2)	Process	<a href="#">GO:0051716</a>	cellular response to stimulus	6.01E-04	1.00E+00	1.89 (5524,736,107,27)	ATG7 - atg7p; SNF5 - snf5p; HAC1 - hac1p; MSH2 - mismatch repair atpase msh2; RTT103 - rtt103p; MMS1 - mms1p; RIM15 - rim15p; MLP2 - mlp2p; EAF1 - eaf1p; ACT1 - actin; MMS2 - e2 ubiquitin-conjugating protein mms2; HSP12 - hsp12p; UBC13 - e2 ubiquitin-conjugating protein ubc13; RPO21 - rpo21p; HRD1 - e3 ubiquitin-protein ligase hrd1; CDC14 - cdc14p; HSC82 - hsp90 family chaperone hsc82; PBS2 - pbs2p; MLP1 - mlp1p; DNF3 - aminophospholipid-translocating p4-type atpase dnf3; SYG1 - syg1p; MMS22 - mms22p; PKH1 - pkh1p; GLC7 - glc7p; FLO1 - flo1p; YBP1 - ybp1p; XRS2 - xrs2p
GNP1 (2)	Component	<a href="#">GO:1990234</a>	transferase complex	9.00E-04	9.50E-01	2.31 (5524,380,107,17)	CLB1 - clb1p; CLG1 - clg1p; RTT103 - rtt103p; MMS1 - mms1p; EAF1 - eaf1p; ACT1 - actin; EST2 - est2p; MMS2 - e2 ubiquitin-conjugating protein mms2; MMS22 - mms22p; FAB1 - 1-phosphatidylinositol-3-phosphate 5-kinase; POL12 - pol12p; CLB5 - clb5p; UBC13 - e2 ubiquitin-conjugating protein ubc13; RPO21 - rpo21p; HRD1 - e3 ubiquitin-protein ligase hrd1; FKS1 - fks1p; SUS1 - sus1p
YGL-165 (9)	Process	<a href="#">GO:0010564</a>	regulation of cell cycle process	9.28E-05	4.79E-01	4.25 (5522,220,59,10)	VRP1 - vrp1p; REC8 - rec8p; SET3 - set3p; WHI2 - whi2p; GLC7 - glc7p; SIF2 - sif2p; ZDS1 - zds1p; CDC27 - anaphase promoting complex subunit cdc27; ASK1 - ask1p; CDC14 - cdc14p
YGL-165 (9)	Process	<a href="#">GO:0090068</a>	positive regulation of cell cycle process	2.14E-04	5.53E-01	6.85 (5522,82,59,6)	GLC7 - glc7p; WHI2 - whi2p; ZDS1 - zds1p; CDC27 - anaphase promoting complex subunit cdc27; ASK1 - ask1p; CDC14 - cdc14p
YGL-165 (9)	Process	<a href="#">GO:0032874</a>	positive regulation of stress-activated MAPK cascade	3.34E-04	5.75E-01	62.40 (5522,3,59,2)	SNT1 - snt1p; SIF2 - sif2p
YGL-165 (9)	Process	<a href="#">GO:0032872</a>	regulation of stress-activated MAPK cascade	3.34E-04	4.31E-01	62.40 (5522,3,59,2)	SNT1 - snt1p; SIF2 - sif2p
YGL-165 (9)	Process	<a href="#">GO:0070302</a>	regulation of stress-activated protein kinase signaling cascade	3.34E-04	3.45E-01	62.40 (5522,3,59,2)	SNT1 - snt1p; SIF2 - sif2p
YGL-165 (9)	Process	<a href="#">GO:0070304</a>	positive regulation of stress-activated protein kinase signaling cascade	3.34E-04	2.87E-01	62.40 (5522,3,59,2)	SNT1 - snt1p; SIF2 - sif2p
YGL-165 (9)	Process	<a href="#">GO:0051495</a>	positive regulation of cytoskeleton organization	3.78E-04	2.78E-01	7.93 (5522,59,59,5)	VRP1 - vrp1p; BNR1 - bnr1p; GLC7 - glc7p; ASK1 - ask1p; CDC14 - cdc14p

Group	Enrichment	GO Term	Description	P-value	FDR q-value	Enrichment (N, B, n, b)	Genes
YGL-165 (9)	Process	<a href="#">GO:0045787</a>	positive regulation of cell cycle	4.02E-04	2.59E-01	6.10 (5522,92,59,6)	GLC7 - glc7p; WHI2 - whi2p; ZDS1 - zds1p; CDC27 - anaphase promoting complex subunit cdc27; ASK1 - ask1p; CDC14 - cdc14p
YGL-165 (9)	Process	<a href="#">GO:0007059</a>	chromosome segregation	8.57E-04	4.91E-01	5.30 (5522,106,59,6)	REC8 - rec8p; TOP2 - dna topoisomerase 2; GLC7 - glc7p; TUB1 - tub1p; ASK1 - ask1p; CDC14 - cdc14p
YGL-165 (9)	Component	<a href="#">GO:0034967</a>	Set3 complex	3.93E-05	4.15E-02	40.11 (5522,7,59,3)	SNT1 - snt1p; SET3 - set3p; SIF2 - sif2p
YGL-165 (9)	Component	<a href="#">GO:0070210</a>	Rpd3L-Expanded complex	9.94E-04	5.25E-01	14.78 (5522,19,59,3)	SNT1 - snt1p; SET3 - set3p; SIF2 - sif2p
GNP1 (2) and YGL-165 (9) cross-section	Process	<a href="#">GO:0006986</a>	response to unfolded protein	2.92E-04	1.00E+00	73.59 (5519,25,6,2)	HAC1 - hac1p; GLC7 - glc7p
GNP1 (2) and YGL-165 (9) cross-section	Process	<a href="#">GO:0035966</a>	response to topologically incorrect protein	4.23E-04	1.00E+00	61.32 (5519,30,6,2)	HAC1 - hac1p; GLC7 - glc7p
GNP1 (2) and WT cross-section	Function	<a href="#">GO:0005537</a>	mannose binding	1.47E-05	3.71E-02	306.67 (5520,6,6,2)	FLO9 - flo9p; FLO1 - flo1p
GNP1 (2) and WT cross-section	Function	<a href="#">GO:0048029</a>	monosaccharide binding	1.50E-04	1.88E-01	102.22 (5520,18,6,2)	FLO9 - flo9p; FLO1 - flo1p
GNP1 (2) and WT cross-section	Function	<a href="#">GO:0030246</a>	carbohydrate binding	4.81E-04	4.04E-01	57.50 (5520,32,6,2)	FLO9 - flo9p; FLO1 - flo1p
GNP1 (2) and WT cross-section	Process	<a href="#">GO:0000128</a>	flocculation	5.90E-06	3.04E-02	460.00 (5520,4,6,2)	FLO9 - flo9p; FLO1 - flo1p
GNP1 (2) and WT cross-section	Process	<a href="#">GO:0098743</a>	cell aggregation	2.75E-05	7.08E-02	230.00 (5520,8,6,2)	FLO9 - flo9p; FLO1 - flo1p
GNP1 (2) and WT cross-section	Process	<a href="#">GO:0098630</a>	aggregation of unicellular organisms	2.75E-05	4.72E-02	230.00 (5520,8,6,2)	FLO9 - flo9p; FLO1 - flo1p
GNP1 (2) and WT cross-section	Process	<a href="#">GO:0051704</a>	multi-organism process	6.10E-04	7.86E-01	51.11 (5520,36,6,2)	FLO9 - flo9p; FLO1 - flo1p

Following that analysis and zooming into the specific genes mutated for each strain, an interesting picture arose. Table 11 depicts ORFs of interest with non-synonymous SNPs for the different strains and biological repeats. The SNPs depicted are all de-novo SNPs, that were not present at the ancestor strain.

As can be seen from the data, each biological repeat of GNP1 deficient strains exhibited a non-synonymous SNP at a gene belonging to the Error-free replication repair pathway (each one on a different gene), and each biological repeat of YGL-165C deficient strains exhibited a non-synonymous SNP at a gene belonging to the histone deacetylase pathway (same).

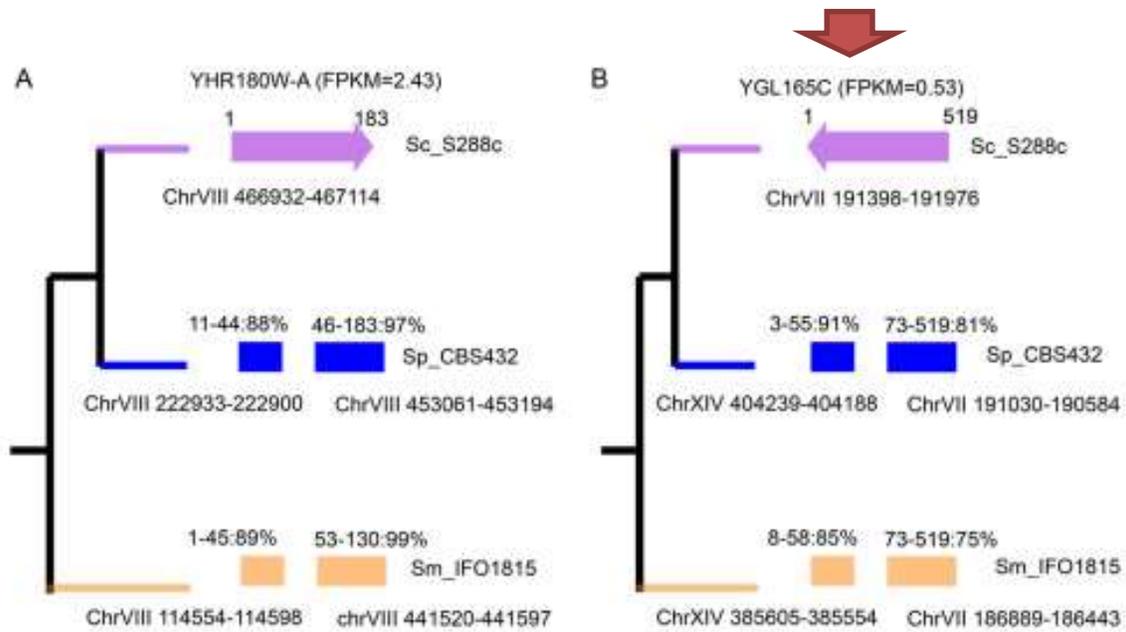
	10- WT			2- GNP1			9- YGL-165C		
	281	282	283	278	279	280	262	284	285
<b>DDR48</b>						X			
<b>MMS1</b>					X				
<b>MMS2</b>	X		X			X		X	
<b>UBC13</b>				X					
<b>HOS2</b>					X				
<b>SIF2</b>							X		
<b>SNT1</b>								X	
<b>SET3</b>			X						X

**Table 11 - SNP analysis** – This table presents ORFs at which non-synonymous de-novo (i.e.- that were not present at the ancestor strain) SNPs were detected for some of the libraries (=strain/biological repeat combinations). The rows represent the ORFs and the columns- the libraries.

This result is fascinating- it seems that each original (“anti-evolvability”) gene mutation changes the evolution landscape in a manner that dictates an alternative optimal evolutionary course, resulting in reaching higher local maxima.

A literature survey revealed no connection between GNP1 (that codes for a High-affinity glutamine permease) and the Error-free replication repair pathway or YGL-165C (described as a dubious open reading frame and considered unlikely to encode a functional protein) and the histone deacetylase pathway.

However- one intriguing result has emerged from the literature- a 2018 article by Wu and Knudson [44] identified by comparative evolutionary analyses 84 genes originating de novo from previously noncoding regions in *S. cerevisiae* S288C since the divergence with their sister groups. Specifically, YGL-165C was named as one of those genes, and as an example of de-novo gene creation via a DNA shuffling mechanism as is depicted in **Figure 38**



**FIG 4** DNA shuffling shaping the birth of *de novo* genes (A) YHR180W-A and (B) YGL165C. The identity and corresponding region between *de novo* genes and noncoding DNA are labeled above the colored boxes. The genomic locations are labeled under the colored boxes. *S. cerevisiae* S288C is shown in pink, *S. paradoxus* CBS432 in blue, and *S. mikatae* IFO 1815 in cream.

**Figure 38- YGL-165C as an example of de-novo gene creation via DNA shuffling (Wu and Knudson 2018)**– This figure shows that YGL165C is a recent addition to the yeast genome, created *de novo* at *S. cerevisiae* via DNA shuffling of elements that exist separately in the ancestral specie (and its parallels along the evolution tree- *S. paradoxus* and *S. mikatae*).

From this and my above data, it seems that YGL-165C may function as an “evolutionary cork”- a means to stop accelerated evolution once a good-enough fitness is achieved (“anti-evolution gene”), hence its removal induces a more vigorous adaptation.

## 4.4 Chapter Summary and Discussion

The study described in this chapter aimed to identify genes that play a role in the yeast evolutionary process by conducting a whole-genome evolvability screen. This novel approach consisted of competitively evolving all the non-essential yeast deletion strains under various conditions, to identify genes that affect yeast adaptation capability in a non-trivial manner, and specifically, to uncover "Evolution deceleration" genes that, when deleted, would enhance yeast adaptation capabilities, hence playing a role in shaping new evolutionary trajectories.

The research consisted of evolving the pooled deletion library in the lab under various conditions to include nominal (YPD), metabolic, osmotic and temperature stresses (SD, pH and HS) and a metabolic "opportunity" (XYL), and then performing high-throughput sequencing at several evolutionary time points to analyze their evolutionary dynamics. The focus was on identifying deletion strains that exhibited significant fitness and fitness rate improvement over time and across conditions.

Despite a significant loss of pool heterogeneity over evolution time due to the unbalanced nature of the competition and the original pool, and the increased environmental complexity and instability induced by the inherent changes in the population composition, a thorough data analysis proved the assays' validity by showing great internal consistency among the top-represented strains, with unique gene repertoires emerging to prominence under each condition. Notably, distinct evolutionary patterns were observed based on conditions, suggesting different adaptive routes. Based on this analysis, a meticulous selection process took place, and several promising candidate strains were chosen for separate lab evolutions. Assessing the resulting growth characteristic, two genes, YDR508C (GNP1) and YGL165C, were found to exhibit significant improvement in pH tolerance during their separate evolutions. Interestingly, these genes were not previously associated with evolvability.

Comparative sequencing and data analysis revealed the presence of single nucleotide polymorphisms (SNPs) in the evolved strains compared to their ancestors. From this analysis, it seems that each strain exhibited a unique repertoire of mutated genes (that was different also from the WT control), suggesting a strain-specific impact on evolutionary trajectories. Enrichment analysis revealed an association between the YGL165C mutant and the SET 3 complex, potentially influencing epigenetic regulation [55].

Furthermore, each biological repeat of the two mutant strains exhibited a non-synonymous SNP at a gene belonging to the same pathway (for each strain a different pathway and for each repeat a different gene) - the histone deacetylase pathway for the YGL-165C deficient repeats, and the Error-free replication repair pathway for the GNP1 deficient repeats, hence suggesting that the mutation of original "anti-evolution" genes alters the evolutionary landscape, leading to unique alternative optimal paths.

In a 2018 paper by Wu and Knudson [44], YGL-165C was identified as a de novo addition to the *S. cerevisiae* S288C genome, emerging from DNA shuffling of previously noncoding regions. This may suggest it functions as an "evolutionary cork"- a means to stop accelerated evolution once a good enough fitness is achieved, i.e., an "evolution deceleration" gene), hence its removal induces a more vigorous adaptation.

Overall, this study shed some light on the intricate interplay between genetic factors and environmental conditions in shaping evolutionary outcomes. The emergence of "evolution deceleration" genes like YGL165C and GNP1 adds complexity to the understanding of evolutionary dynamics and challenges the notion of genes solely associated with slowing evolution.

Future plans should consist of further investigations into the molecular mechanisms underlying these genes' roles and their potential effects on adaptation. Such investigations may include additional evolutions of these strains under additional conditions, together with comparative genomics analyses, functional validation (epistasis for example) and overexpression evolutions.

Finally, as part of the study, a novel approach and framework were developed and validated. Based on yeast whole-genome screens, lab evolution, high-throughput sequencing and a solid analytic tool kit, it is available and has the potential to enable the identification of additional genes that may affect yeast evolvability. Such future studies may include separate investigations (each consisting of separate evolutions on additional conditions and sequencing) of more of the 33 final candidates (or even the 172 final batch) under various conditions, current data enrichment by sequencing an additional evolutionary time point (sequencing libraries for day 20 were prepared as part of this study but not sequenced) and even repeating the study while using a commercial, balanced yeast deletion pool as a starting point. This has the potential to identify additional evolvability-related genes and patterns.

## 5 Summary

This thesis dwells on the intricate mechanisms and effectors influencing the evolutionary process, by attempting two complementary and possibly synergetic approaches, both based on lab evolution of the yeast *Saccharomyces cerevisiae* as a model organism. The first, focused approach, was aimed at understanding how epigenetic factors influence evolvability, and the second spanned a wider net by developing a framework and performing a whole-genome high-throughput evolvability screen, thus in order to explore both epigenetic effects and genome-wide adaptations in the context of yeast evolution.

The two chapters of this thesis examine these aspects from different angles by using the above approaches, contributing to our understanding of how evolution unfolds and the factors that shape it.

### Chapter 1: Epigenetic Effects in Evolution

The first chapter presents the focused approach. It revolves around whether manipulating the epigenetic fidelity of a genome, i.e., its capacity to maintain its epigenetic landscape, can potentially enhance evolution and evolvability under specific circumstances. Three distinct epigenetic mechanisms were examined: HSP-90 (a protein chaperone), SUP-35 prions, and RNA interference (RNAi). These mechanisms were selected to provide both diversity and relative technical assay maturity. The relevant strains were evolved at the lab under a mild metabolic challenge (“opportunity”) and their evolutionary dynamics were assessed by observing their growth characteristics at several evolutionary time points.

**HSP-90-** HSP-90, a ubiquitous heat shock protein, plays a critical role in buffering genetic variance and releasing it in response to environmental stress [12]. It has been proposed to affect evolution through various mechanisms, including transposon mobilization [14,23], the induction of aneuploidy [13], and interactions with DNA repair pathways. This study focused on evolving several HSP-90 related genes deletion strains that were constructed and found that they seemed to influence the evolutionary dynamics in what appeared to be a systematic manner, particularly the HSC homozygote and STI mutant strains (HSC-82 deficiency reduces HSP-90 levels and STI-1 deficiency increases it), hence suggesting that HSP-90 indeed affects evolvability. Yet as to the nature and extent of that influence, additional growth tests should be held and analyzed. Additionally, sequencing and comparing the final evoltants to their ancestors for the different mutant strains may yield some interesting insights regarding the underlying mechanism.

**SUP-35 Prions:** [PSI<sup>+</sup>], the prion state of SUP-35p, can be seen as an epigenetic switch that enhances phenotypic plasticity and evolvability by unleashing Cryptic Genetic Variation (CGV) [15]. It causes a loss-of-function mutation (stop-codon read-through) that reveals phenotypic variation which in turn facilitates faster evolution to new adaptive peaks [21]. Several strains with enhanced/reduced prionicity were obtained (Curtesy of the Lindquist lab) and evolved, yet preliminary ambiguous results together with concerns that arose regarding the prion expression levels of these SUP-35 strains, prompted a pause in further investigation until these issues are resolved.

**RNAi:** RNA interference (RNAi) mechanisms were reintroduced into *Saccharomyces cerevisiae* by Bartel *et al.* [29]. In this study, strains with fully/partially restored RNAi mechanism lab evolution resulted in altered evolutionary dynamics in comparison to the parent strain. Both the fully and partially restored RNAi strains exhibited differences in max yield and lag time, indicating that reintroducing RNAi could influence

evolvability. Sequencing analysis of evolved strains may provide deeper insights into the underlying mechanisms.

In conclusion, this research suggests that manipulating epigenetic mechanisms such as HSP-90 and RNAi can indeed impact the evolvability of yeast strains. However, the specific nature and extent of this influence require further investigation through additional growth tests and comprehensive sequencing analysis of evolved strains. Additionally, resolving expression issues related to SUP-35 prions is necessary to explore their role in evolvability further. These findings contribute to our understanding of how epigenetic mechanisms can shape the evolutionary trajectories of organisms, providing valuable insights into the field of evolutionary biology.

Initially designed to investigate epigenetic effects in evolution, due to technical challenges the research pivoted, leading to a comprehensive exploration of high-throughput evolvability screening offering a complementary, synergetic approach, as is described in the next chapter.

## **Chapter 2: A Whole-Genome Evolvability Screen**

The second chapter of this thesis presents the second approach and embarks on a journey to unravel the genetic determinants of yeast evolutionary adaptation. As an aftermath of the previous chapter's epigenetic study results (the HSP-90 mutant strains in particular), It spanned a wider net by developing a framework and performing a whole-genome high-throughput evolvability screen.

As mentioned, the central theme revolved around identifying additional genes that influence the evolutionary process in yeast. While well-known "immediate suspects" like DNA repair genes, oncogenes, and tumor suppressors are already recognized for their roles in promoting different evolutionary dynamics [50, 51], this research sought to uncover additional relevant genetic-based mechanisms, and specifically, to identify "evolution-deceleration" genes that, when deleted, enhanced yeast adaptation capabilities, shaping unique evolutionary paths in the process.

The research encompassed lab evolutions of the pooled deletion library under different conditions, including nominal, metabolic, osmotic and temperature stresses, and a metabolic "opportunity" (similar to the one used in the previous chapter). High-throughput sequencing at multiple time points enabled the analysis of evolutionary dynamics, with a primary emphasis on identifying deletion strains exhibiting substantial improvements in fitness and fitness rate across conditions.

Despite challenges related to the starting point, the competition dynamics, and the resulting loss of pool heterogeneity, a rigorous data analysis confirmed the validity of the approach. Notably, unique gene repertoires emerged to prominence under each condition, indicating distinct adaptive routes.

Candidate strains that exhibited interesting evolutionary dynamics within the pool were selected for separate lab evolutions, resulting in the discovery of two genes, YDR508C (GNP1) and YGL165C, which exhibited significant improvement in pH tolerance. Intriguingly, these genes were not previously associated with evolvability (Furthermore- YGL165C was described as a dubious ORF, not associated with any known function). Comparative sequencing revealed single nucleotide polymorphisms (SNPs) in

evolved strains, suggesting a strain-specific impact on evolutionary trajectories. Enrichment analysis linked the YGL165C mutant to the SET 3 complex, potentially influencing epigenetic regulation.

Moreover, each biological repeat of the mutant strains exhibited non-synonymous SNPs in genes belonging to specific pathways (different genes for each repeat, same pathway for each strain to include WT), such as the histone deacetylase pathway for YGL-165C deficient strains and the Error-free replication repair pathway for GNP1 deficient strains. These findings suggest that the mutation of original "evolution deceleration" genes alters the evolutionary landscape, leading to unique alternative optimal paths.

Future plans should consist of further investigations into the molecular mechanisms underlying YGL-165C and GNP1 genes' roles and their potential effects on adaptation, such as more evolutions under additional conditions, comparative genomics analyses, functional validation and overexpression evolutions.

However, based on yeast whole-genome screens, lab evolution, high-throughput sequencing and a solid analytic tool kit, the novel approach and framework were developed and validated in this part of the study are available and have the potential to enable the identification of additional genes that may affect yeast evolvability. Such future studies may vary in scope to include -

Re-doing the study altogether while using a better starting point in the form of a commercial, fully balanced yeast deletion pool that includes WT (hence trying to avoid or reduce the loss of heterogeneity);

Enriching the existing database by sequencing an additional evolutionary time point (sequencing libraries for day 20 were prepared as part of this study but not sequenced)

Investigating more promising candidates that were already identified in the current data set by performing more separate evolutions under various conditions.

Hopefully, if attempted with a better starting point as described, this chapters' novel approach and framework may be used to further investigate the HSP-90 mutants from the previous chapter, that were decimated in this study due to the unbalanced nature of the initial pool.

## **Conclusion**

The 2018 discovery of YGL-165C as a de novo addition to the yeast genome, emerging from DNA shuffling of previously noncoding regions [44], implies its role as an "evolutionary cork." The RNAi removal effect (though it requires additional validation) seems to be similar in nature, i.e. suggesting that we are facing a gene that if removed and a mechanism that if added- induces more vigorous adaptations, hence they are added or removed accordingly by the yeast once a sufficient fitness level is achieved.

The same may be argued (although again- requires additional verification) regarding the HSP-90 levels, as induced by the different genetic backgrounds, influence on the evolutionary rate.

This thesis contributes to the understanding of evolution by investigating both epigenetic influences and genome-wide adaptations. Through rigorous experimentation and analysis, it sheds light on the complex

interplay between genetic factors and environmental conditions in shaping evolutionary outcomes. These findings, and specifically the novel approach and framework that were developed and the "evolution deceleration" genes that were identified invite further research into the molecular mechanisms underlying these effects and the potential for identifying additional evolvability-related genes.

As a whole, this work expands our comprehension of the mechanisms driving evolution and challenges existing paradigms.

## 6 References

1. Jean-Baptiste Lamarck Zoological Philosophy trans. Hugh Elliot, 1914, p.113
2. Chen B, Zhong D, Monteiro A (2006). "Comparative genomics and evolution of the HSP90 family of genes across all kingdoms of organisms". *BMC Genomics*7: 156. doi:10.1186/1471-2164-7-156. PMC 1525184. PMID 16780600
3. Csermely P, Schnaider T, Soti C, Prohászka Z, Nardai G (August 1998). "The 90-kDa molecular chaperone family: structure, function, and clinical applications. A comprehensive review". *Pharmacol. Ther.* 79 (2): 129–68. doi:10.1016/S0163-7258(98)00013-8. PMID 9749880.
4. Nathan DF, et al. (1997) In vivo functions of the *Saccharomyces cerevisiae* Hsp90 chaperone. *Proc Natl Acad Sci U S A* 94(24):12949-56
5. Thomas JG, Baneyx F (October 1998). "Roles of the Escherichia coli Small Heat Shock Proteins IbpA and IbpB in Thermal Stress Management: Comparison with ClpA, ClpB, and HtpG In Vivo". *J. Bacteriol.* 180 (19): 5165–72. PMC 107554. PMID 9748451
6. Borkovich KA, et al. (1989) hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol* 9(9):3919-30
7. Burnie JP, et al. (2006) Fungal heat-shock proteins in human disease. *FEMS Microbiol Rev* 30(1):53-88
8. Picard D (2002) Heat-shock protein 90, a chaperone for folding and regulation. *Cell Mol Life Sci* 59(10):1640-8
9. Prodromou C and Pearl LH (2003) Structure and functional relationships of Hsp90. *Curr Cancer Drug Targets* 3(5):301-23
10. Borkovich KA, et al. (1989) hsp82 is an essential protein that is required in higher concentrations for growth of cells at higher temperatures. *Mol Cell Biol* 9(9):3919-30
11. Richter K, et al. (2001) Coordinated ATP hydrolysis by the Hsp90 dimer. *J Biol Chem* 276(36):33689-96

12. Rohner N. et al (2013), Cryptic Variation in Morphological Evolution: HSP90 as a Capacitor for Loss of Eyes in Cavefish, *Science* 13 December 2013: 342 (6164), 1372-1375. [DOI:10.1126/science.1240276]
13. Chen G. et al (2012), Hsp90 stress potentiates rapid cellular adaptation through induction of aneuploidy, *Nature* 482, 246–250 (09 February 2012) doi:10.1038/nature10795
14. Taipale, M., Jarosz, D. F. & Lindquist, S. HSP90 at the hub of protein homeostasis: emerging mechanistic insights. *Nature Rev. Mol. Cell Biol.* 11,515–528 (2010).
15. J. Shorter, S. Lindquist (2005), Prions as adaptive conduits of memory and inheritance, *Nature Reviews Genetics* 6, 435-450 (June 2005) | doi:10.1038/nrg1616
16. Doel SM, et al. (1994) The dominant PNM2- mutation which eliminates the psi factor of *Saccharomyces cerevisiae* is the result of a missense mutation in the SUP35 gene, *Genetics* 137(3):659-70
17. Lindquist S, et al. (2001) Investigating protein conformation-based inheritance and disease in yeast. *Philos Trans R Soc Lond B Biol Sci* 356(1406):169-76
18. Derkatch IL, et al. (2004) Effects of Q/N-rich, polyQ, and non-polyQ amyloids on the de novo formation of the [PSI<sup>+</sup>] prion in yeast and aggregation of Sup35 in vitro. *Proc Natl Acad Sci U S A* 101(35):12934-9
19. Funakoshi Y, et al. (2007) Mechanism of mRNA deadenylation: evidence for a molecular interplay between translation termination factor eRF3 and mRNA deadenylases. *Genes Dev* 21(23):3135-48
20. Salnikova AB, et al. (2005) Nonsense suppression in yeast cells overproducing Sup35 (eRF3) is caused by its non-heritable amyloids. *J Biol Chem* 280(10):8808-12
21. Bergman, A. & Siegal, M. L. Evolutionary capacitance as a general feature of complex gene networks. *Nature* 424, 549–552 (2003).
22. Sangster, T. A., Lindquist, S. & Queitsch, C. Under cover: causes, effects and implications of Hsp90- mediated genetic capacitance. *Bioessays* 26, 348–362 (2004).
23. Specchia, V. et al. Hsp90 prevents phenotypic variation by suppressing the mutagenic activity of transposons. *Nature* 463, 662–665 (2010).
24. Tomari Y, Zamore PD. *Genes Dev.* 2005; 19

25. Malone CD, Hannon GJ. *Cell*. 2009; 136
26. Farazi TA, Juraneck SA, Tuschl T. *Development*. 2008; 135
27. Nakayashiki H, Kadotani N, Mayama S. *J Mol Evol*. 2006; 63
28. Laurie JD, Linning R, Bakkeren G. *Curr Genet*. 2008; 53
29. Drinnenberg I, Weinberg D, Xie K, Mower J, Wolfe K, Fink G, Bartel D (2009), RNAi in Budding Yeast, *Science* 23 October 2009: 326 (5952), 544-550 [DOI:10.1126/science.1176945]
30. Romano GH, Gurvich Y, Lavi O, Ulitsky I, Shamir R, Kupiec M. (2010), Different sets of QTLs influence fitness variation in yeast, *Mol Syst Biol*. 2010;6:346. doi: 10.1038/msb.2010.1. Epub 2010 Feb 16.
31. Tirosh I, Barkai N (2008), Two strategies for gene regulation by promoter nucleosomes, *Genome Res*. 2008 Jul;18(7):1084-91. doi: 10.1101/gr.076059.108. Epub 2008 Apr 30.
32. Gerstein AC, Otto SP (2009), Ploidy and the causes of genomic evolution, *J Hered*. 2009 Sep-Oct;100(5):571-81. doi: 10.1093/jhered/esp057.
33. Mitchell A, Romano GH, Groisman B, Yona A, Dekel E, Kupiec M, Dahan O, Pilpel Y (2009), Adaptive prediction of environmental changes by microorganisms, *Nature* 460, 220-224 (9 July 2009) | doi:10.1038/nature08112
34. Springer M, Weissman JS, Kirschner MW (2010), A general lack of compensation for gene dosage in yeast, *Mol Syst Biol*. 2010 May 11;6:368. doi: 10.1038/msb.2010.19, PMID: 20461075
35. Brinkmann U, Mueller RH, Babel W (1990), The growth rate-limiting reaction in methanol-assimilating yeasts, *FEMS Microbiol Rev*. 1990 Dec;7(3-4):261-5. Review. PMID: 2094286
36. Halfmann R., Jarosz D., Jones S., Chang A., Lancaster A., Lindquist S., (2012) Prions are a common mechanism for phenotypic inheritance in wild yeasts, *Nature* 482, 363–368 (16 February 2012) doi:10.1038/nature10875
37. Rechavi O, Houry-Zeevi L, Anava S, Sho W Goh, Kerk SY, Hannon GJ, Hobert O (2014), Starvation-Induced Transgenerational Inheritance of Small RNAs in *C. elegans*, *Cell*, Volume 158, Issue 2, 17 July 2014, Pages 277-287

38. Koonin EV (2014), Calorie Restriction à Lamarck, *Cell*, Volume 158, Issue 2, 17 July 2014, Pages 237–238
39. Grishok, A. (2013), Biology and Mechanisms of Short RNAs in *Caenorhabditis elegans*, *Adv. Genet.* 83, 1–69.
40. Rechavi, O. (2014), Guest list or black list: heritable small RNAs as immunogenic memories, *Trends Cell Biol.* 24, 212–220.
41. Ram Y., Dellus-Gur E., Bibi M., Obolski U., Berman J., Hadany L. (2017), Predicting microbial relative growth in a mixed culture from growth curve data, bioRxiv 022640
42. Giaever, G; Chu, AM; Ni, L; et al. (July 2002). "Functional profiling of the *Saccharomyces cerevisiae* genome". *Nature.* 418: 387–91.
43. Yeast deletion project web page-  
[http://www-sequence.stanford.edu/group/yeast\\_deletion\\_project/deletions3.html](http://www-sequence.stanford.edu/group/yeast_deletion_project/deletions3.html)
44. Wu, B; Knudson, A. (July 2018), "Tracing the De Novo Origin of Protein-Coding Genes in Yeast". *mBio - ASM Journals.* Vol. 9, No. 4.
45. Goldberg, A; David Allis, C; Bernstein, E (2007), Epigenetics: A Landscape Takes Shape, *Cell*, Volume 128, Issue 4, 2007, Pages 635-638, ISSN 0092-8674,  
<https://doi.org/10.1016/j.cell.2007.02.006>.
46. Heard, E; Martienssen, R. (2014), Transgenerational Epigenetic Inheritance: Myths and Mechanisms, *Cell*, Volume 157, Issue 1, 2014, Pages 95-109, ISSN 0092-8674,  
<https://doi.org/10.1016/j.cell.2014.02.045>,
47. Waddington C.H. (1957), *The Strategy of the Genes; a Discussion of Some Aspects of Theoretical Biology*, Allen & Unwin, London
48. Tamari Z, Rosin D, Voichek Y, Barkai N (2014) Coordination of Gene Expression and Growth-Rate in Natural Populations of Budding Yeast. *PLoS ONE* 9(2): e88801.  
<https://doi.org/10.1371/journal.pone.0088801>
49. Olsson L, Linden T, Hahn-Hägerdahl B (1994) A rapid chromatographic method for the production of preparative amounts of xylulose. *Enzyme Microb Technol* 16: 388–394.  
[https://doi.org/10.1016/0141-0229\(94\)90153-8](https://doi.org/10.1016/0141-0229(94)90153-8)

50. Thompson, J, Woodruff, R (1978) Mutator genes—pacemakers of evolution. *Nature* 274, 317–321. <https://doi.org/10.1038/274317a0>
51. Casás-Selves M, Degregori J (2011) How cancer shapes evolution, and how evolution shapes cancer. *Evolution (N Y)*. 2011 Dec;4(4):624-634. doi: 10.1007/s12052-011-0373-y. PMID: 23705033; PMCID: PMC3660034.
52. Stephen G (1997) From gene to screen with yeast, *current opinion in genetics & development*, Volume 7, Issue 3, pages 405-409, ISSN 0959-437X, [https://doi.org/10.1016/S0959-437X\(97\)80156-6](https://doi.org/10.1016/S0959-437X(97)80156-6).
53. Cooper V (2018) Experimental evolution as a high-throughput screen for genetic adaptations, *ASM Journals, mSphere*, Vol. 3, No. 3, <https://doi.org/10.1128/msphere.00121-18>
54. *Saccha08romyces* genome database website- <https://www.yeastgenome.org/>
55. Pijnappel W et al. (2001), The *S. cerevisiae* SET3 complex includes two histone deacetylases, Hos2 and Hst1, and is a meiotic-specific repressor of the sporulation gene program, *Genes Dev*. 2001 Nov 15;15(22):2991-3004. doi: 10.1101/gad.207401. PMID: 11711434; PMCID: PMC312828.
56. Liebman S, Chernoff Y (2012), Prions in Yeast, *Genetics*, Volume 191, Issue 4, 1 August 2012, Pages 1041–1072, <https://doi.org/10.1534/genetics.111.137760>
57. Rutherford, S., Lindquist, S. Hsp90 as a capacitor for morphological evolution. *Nature* 396, 336–342 (1998). <https://doi.org/10.1038/24550>
58. Queitsch C, Sangster TA, Lindquist S. Hsp90 as a capacitor of phenotypic variation. *Nature*. 2002 Jun 6;417(6889):618-24. doi: 10.1038/nature749. Epub 2002 May 12. PMID: 12050657.
59. Sollars, V., Lu, X., Xiao, L. et al. Evidence for an epigenetic mechanism by which Hsp90 acts as a capacitor for morphological evolution. *Nat Genet* 33, 70–74 (2003). <https://doi.org/10.1038/ng1067>
60. Earl, D. J. & Deem, M. W. (2004) Evolvability is a selectable trait. *PNAS* August 2, 2004, 101 (32) 11531-11536. <https://doi.org/10.1073/pnas.0404656101>

61. True, H., Lindquist, S. A yeast prion provides a mechanism for genetic variation and phenotypic diversity. *Nature* 407, 477–483 (2000). <https://doi.org/10.1038/35035005>

בנוסף, במחקר משנת 2018 התגלה אחד הגנים - YGL165C כתוספת חדשה יחסית לגנום השמרי, מה שעשוי אולי לרמז על תפקידו כ"פקק אבולוציוני". האפקט של היעדר מנגנון RNAi בשמר האפיה נראה (כתלות בניסויי אימות נוספים) גם הוא דומה באופיו, קרי גן שאם מוחקים אותו או מנגנון שאם מוסיפים אותו גורמים לתהליכי הסתגלות מואצים, ועל כן הם מוספים או אובדים בהתאמה ברגע שמושגת רמת כשירות מספקת בסביבה יציבה. כתלות בניסויי אימות נוספים, נראה גם שטענה דומה ניתן לטעון גם לגבי השפעת רמות ה-HSP-90, כפי שמושרות ע"י הרקעים הגנטיים השונים, על קצב האבולוציה.

ממצאים אלו, ובפרט הגישה והמתודולוגיה החדשה שפותחו והגנים "מאיטי האבולוציה" שזוהו פותחים פתח למחקרים נוספים לזיהוי גנים דומים נוספים, והמנגנונים המולקולריים העומדים בבסיסם.

## 2. סקר כלל גנומי למיפוי פוטנציאל אבולוציוני (Evolvability)

סקר זה התבסס על אוסף ההפלואידים של פרויקט מחיקות הגנום השמרי (The yeast deletion project) שקובץ ליצירת ספריית מוטנטים אחת ובה כל המחיקות של גנים לא חיוניים בשמר. ניסיון להוסיף גם את זן הבר (WT) נכשל עקב ריכוז נמוך מדי שהפך אותו ללא אפקטיבי.

לאחר בנייתה בוצעו בספרייה אבולוציות-מעבדה תחת חמישה תנאים שונים כולל תנאים נומינליים, עקות מטבוליות ואוסמוטיות, עקת טמפרטורה ומדיית קסילוז (אתגר מטבולי קל, בדומה לפרק הקודם). דגימות מזמנים אבולוציוניים שונים ושילבי גידול שונים (מניסויי אד-הוק שנערכו) רוצפו לקביעת תדירותם היחסית של המוטנטים השונים.

בעזרת אנליזות מקיפות מוצו מתוצאות הריצופים קירובים לערכי הכשירות (fitness) של כל מוטנט בכל נקודת זמן אבולוציונית על מנת לייצר תמונה של הדינמיקה האבולוציונית שלהם כפי שעולה משינוי רמת הכשירות לאורך זמן האבולוציה, להלן- התאוצה.

מתוצאות האנליזה עלו מספר מועמדים מבטיחים (שהציגו תאוצות אבולוציוניות משמעותיות), בהם בוצעו אבולוציות מעבדה נפרדות. התוצאות הראשוניות עבור שניים מהם- YDR508C (GNP1) ו-YGL165C (שלא היו ידועים בעבר כבעלי קשר כלשהו לפוטנציאל אבולוציוני) באבולוציית מעבדה עם מדיום YPD בעל רמת בסיסיות גבוהה (pH 7.6), הראו הבדלים משמעותיים במידת וטבע השיפור שהושג באבולוציות אלו בינם לבין זן הבר (WT), בפרט בערכי בסיסיות (pH) גבוהים יותר מזה שבו בוצעה האבולוציה.

תוצאות אלו אומתו בסדרת ניסויי גידול חוזרת עבור YDR508C (GNP1) ו-YGL165C, כולל נקודות זמן אבולוציוניות וערכי pH נוספים, ועל כן רוצפו התרבויות בתחילת וסיום האבולוציות הנפרדות (קרי- זני האב והאבולטנטים הסופיים של כל החזרות הביולוגיות) ונותחו בקפידה על מנת לנסות ולהבין את מנגנוני ההסתגלות.

מהניתוח לעיל התגלה כי בכל החזרות הביולוגיות של מוטנט מסוים (הן YDR508C/GNP1 והן YGL165C), קיים סט ייחודי של מוטציות נקודתיות שלהן השפעה על רצף החלבון (nonsynonymous), ואשר התמפו לגנים שונים על אותו המסלול הביולוגי. מסלול זה היה ייחודי לכל מוטנט, ונותר לרוב ללא פגע באבולוציות של המוטנט השני או של זן הבר, מה שמעלה את הסברה שהמוטציה המקורית דחפה את הדינמיקה האבולוציונית לנתיב ייחודי, ובכך גרמה לו להשיג בסיומה של האבולוציה יתרון פנוטיפי משמעותי בדרך המיוחדת רק לו.

לסיכום, מחקר זה כיוון לנסות ולשפוך מעט אור על המנגנונים והגורמים המשחקים תפקיד בתהליכים אבולוציוניים, תוך שימוש בשתי שיטות משלימות (ובמידה מסוימת- סינרגטיות), ובשמר האפייה (*Saccharomyces cerevisiae*) כבאורגניזם מודל. השיטה הראשונה כוונה להבין את השפעתם של גורמים אפיגנטיים על הפוטנציאל האבולוציוני, בפרט ע"י שימוש בזנים בעלי מנגנונים אפיגנטיים ידועים שתוגברו/הוחלשו – חלבון סוכך HSP-90, פרוינים מסוג SUP-35/RNQ ו-RNAi. תוצאות ראשוניות ממחקר זה הצביעו על הבדלים דינמיקה האבולוציונית (בהשוואה לזן הבר) בזנים עם RNAi משוחזר ועם מוטציות בגנים הקשורים ל-HSP-90, המעלים מוטיבציה לחקירה נוספת בנושא.

השיטה השנייה פרשה רשת רחבה יותר ע"י פיתוח מתודולוגיה וביצוע סקר גנומי מלא למיפוי פוטנציאל אבולוציוני, וזאת ע"י ביצוע אבולוציה במעבדה בתנאים שונים לספרייה אחודה של כל המחיקות של גנים לא חיוניים בשמר. למרבה הצער, עקב קשיים טכניים, ריכוז אותם הגנים הקשורים ל-HSP-90 שנחקרו בחלק הראשון בספרייה לא היה מספק, ועל כן לא ניתן היה למפות את האפקט שלהם במסגרת הסקר של החלק השני אך מועמדים מבטיחים אחרים עלו במקומם, ובפרט YDR508C (GNP1) ו-YGL165C שהציגו שיפור אבולוציוני משמעותי בסיבולת לחומציות הסביבה (pH). חקירה מדוקדקת חשפה מוטציות שלהן השפעה על רצף החלבון (nonsynonymous) במסלולים ביולוגיים שהיו ייחודיים לכל זן, מה שמצביע על קיומם של מסלולים אבולוציוניים שונים המובילים ליתרון פנוטיפי.

## תקציר

בעבודה זו התמקדתי בחקר התהליך האבולוציוני, ובפרט המנגנונים העומדים בבסיסו והגורמים המשפיעים עליו.

במקור, תוכנן המחקר לעסוק באפקטים אפיגנטיים באבולוציה, אך אתגרים טכניים לא צפויים ותובנות שעלו במהלך הניסויים היטו את מסלולו של המחקר עצמו לכדי סקר כלל גנומי בחיפוש אחר גנים המשפיעים על הפוטנציאל, קרי יכולתו האבולוציונית של האורגניזם להסתגל לסביבה (היינו Evolvability, פוטנציאל אבולוציוני).

המחקר התבסס על ביצוע אבולוציות במעבדה תוך שימוש בשמרי אפיה (*Saccharomyces cerevisiae*) כאורגניזם מודל, ושתי מטרותיו העיקריות היו כמפורט להלן-

### 1. חקירת אפקטים אפיגנטיים באבולוציה

חלק זה במחקר כוון למציאת מנגנונים אפיגנטיים בעלי יכולת השפעה משמעותית על הפוטנציאל האבולוציוני (קרי- היכולת לעבור תהליך אבולוציוני אפקטיבי) של השמר, כאשר המטרה העיקרית הייתה לחקור לעומק ולמדל את מנגנוני ההסתגלות שהם נוקטים בפרט, ומכך לשאוב תובנות אודות התהליך האבולוציוני בכלל.

שיטת הניסוי שנבחרה כללה ביצוע אבולוציות במעבדה למספר זנים מייצגים, בהם תוגברו או דוכאו מנגנונים אפיגנטיים מוכרים, ובפרט HSP-90 (חלבון סוכך- protein chaperone), פרוינים מסוג SUP-35/RNQ ו-RNAi- כולם בעלי פוטנציאל משמעותי לקדם תהליכי הסתגלות סביבתית מואצת באמצעים אפיגנטיים. האבולוציות בוצעו ע"י שימוש במדיית גידול מבוססת קסילוז (Xylose), המציבה בפני השמר אתגר מטבולי קל (או שמא "הזדמנות").

תוצאות ראשוניות הראו ששמרים בהם נכח, אפילו חלקית, מנגנון RNAi הציגו דינמיקה אבולוציונית שונה בהשוואה לזן המקורי (נעדר ה-RNAi). תוצאות אלו דורשות חקירה ואימות נוספים, אך בתרבויות הניסוי הריכוז הסופי (max yield) וזמן השיהוי (lag) שנמדדו היו נמוכים משמעותית בהשוואה לזן המקור אפילו עבור הזן בו בוצע שחזור חלקי בלבד של מנגנון ה-RNAi (תוספת של גן ה-Dicer בלבד, ללא גן ה-Argonaute הנדרש למנגנון מלא ופעיל), מה שמצביע על השפעה אפשרית של שחזור מנגנון ה-RNAi בשמר על הפוטנציאל האבולוציוני (Evolvability) שלו.

ניתוח ראשוני הראה גם שזנים עם מוטציות שונות הקשורות ל-HSP-90 הציגו גם הם דינמיקות אבולוציוניות שהיו שונות באופן שיטתי מזן האב ממנו נבנו (זן הבר-WT). הבדלים אלו היו משמעותיים במיוחד עבור זני המחיקה של STI, וזן המחיקה הכפולה (הומוזיגוט) של HSC (מחסור ב-HSC-82 מוריד את רמות ה-HSP-90 בשמר, ומחסור ב-STI-1 מעלה אותן), מה שמרמז שאכן קיימת השפעה של HSP-90 על הפוטנציאל האבולוציוני של השמר, אם כי נדרשים ניסויים נוספים לכימות טבעה ועוצמתה של השפעה זו.

באשר לזני הפרוינים מסוג SUP-35/RNQ, הרי שמתוצאות ראשוניות לא נראה שנוכחות מוגברת/ מועטת שלהם הביאה לאפקט אבולוציוני משמעותי יחסית לזן האב, ומחקרים נוספים במעבדה העלו ספקות לגבי רמות ביטוי הפרוינים בפועל.

נערכו מספר ניסיונות להשלמת המידע ואימות התוצאות שהתקבלו בזני ה-RNAi וחלבוני הסוכך לעיל על ידי ביצוע ניסויי גידול נוספים, אך למרבה הצער ניסיונות אלו עלו בתוהו בגין אמינות הצידוד וקשיים טכניים מתמשכים. לכן, ולאור כמות המשאבים והזמן המשמעותיות שמחייבים ניסויים אלו והרצון להתכנס למסגרת הזמן הקצובה, הוחלט לנקוט בגישה כוללת יותר המבוססת על סקר כלל גנומי, ולמקד את המאמצים בפרויקט השני המתואר להלן.

## תוכן העניינים

6	תקציר	1
9	הקדמה	2
11	פרק 1- הקירת אפקטים אפיגנטיים באבולוציה	3
11	הקדמה	3.1
14	שיטות וחומרים	3.2
21	תוצאות	3.3
47	סיכום הפרק ודיון	3.4
49	פרק 2- סקר כלל גנומי למיפוי פוטנציאל אבולוציוני	4
49	הקדמה	4.1
50	שיטות וחומרים	4.2
61	תוצאות	4.3
93	סיכום הפרק ודיון	4.4
95	סיכום	5
99	מקורות	6



TEL AVIV UNIVERSITY

GEORGE S. WISE FACULTY OF LIFE SCIENCES  
THE SMOLARZ FAMILY GRADUATE SCHOOL

אוניברסיטת תל-אביב

הפקולטה למדעי החיים ע"ש ג'ורג' ס. ויז  
המדרשה לתארים מתקדמים ע"ש משפחת סמולרש

עבודה זו נעשתה בהדרכתם של

פרופ' עודד רחבי מאונ' ת"א ופרופ' צחי פלפל ממכון וויצמן

*Ortal Rechavi*

פרופ' עודד רחבי

*Tzachi Pilpel*

פרופ' צחי פלפל



TEL AVIV UNIVERSITY

GEORGE S. WISE FACULTY OF LIFE SCIENCES  
THE SMOLARZ FAMILY GRADUATE SCHOOL

אוניברסיטת תל-אביב

הפקולטה למדעי החיים ע"ש ג'ורג' ס. וייז  
המדרשה לתארים מתקדמים ע"ש משפחת סמולרז

המחלקה לנוירוביולוגיה

# אפקטים אפיגנטיים באבולוציה

חיבור לשם קבלת התואר

"דוקטור לפילוסופיה"  
מאת

**דורית חיזי**

הוגש לסנאט אוניברסיטת תל-אביב

מרץ 2024