Thesis for the degree

Doctor of Philosophy

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

עבודת גמר (תזה) לתואר

דוקטור לפילוסופיה

מוגשת למועצה המדעית של
מכון ויצמן למדע
רחובות, ישראל

By

Sivan Kaminski-Strauss

מאת

סיון קמינסקי-שטראוס

חקירת מנגנוני אבולוציה וההשפעות של בחירת בני זוג על אבולוציה והורשת מסוגלות בשמרים

Exploration of evolutionary strategies and the effect of mate choice on evolution and on fitness inheritance in yeast

Advisor:
Prof. Yitzhak Pilpel

מנחה:
פרופ' יצחק פלפל

December 2021

טבת תשפ"ב

1. Acknowledgments

This thesis could not have been possible without the help of many people, for whom I would like to thank.

First, to my PhD advisor and mentor Prof. Tzachi Pilpel, who shared his knowledge and expertise with me. I have learned a lot about science and much more from our discussions.

To Orna Dahan, who taught me everything I know about yeast and lab technique, and was constantly available for coffee breaks help and support.

Orna and Tzachi - you made me the scientist I am, thank you!

I thank Ruth Towers for her help in the lab, I could not have done anything in this work without her help.

I wish to thank all of the former and present Pilpel lab members, that are always willing to help, and to make hard times happier .

I would like to thank the scientific and professional staff at the Weizmann Institute life science core facilities, and specifically Ghil Jona and the Bacteriology & Genomic Repository Unit, Tomer-Meir Salame and the FACS unit, Yoav Peleg and the DNA Manipulation unit, and all of the INCPM genomic unit.

I wish to thank my PhD committee, Prof. Naama Barkai and Prof. Emmanuel Levy, for their guidance and patience. Their advice was timely, valuable and transformed the process.

Finally, everything I achieved in the recent years was always with the support of my family; my parents Issack and Ronit Kaminski, my siblings and my spouse, Yehuda Strauss who supported and was keen to provide a fresh look when needed.

The term of my PhD had taught me valuable lessons; it taught me patience and endurance, how to embrace failures and move on, and how to celebrate successes.

2. Declaration

I declare that this thesis summarizes the research projects for which I was the main contributor.
The thesis includes both published and unpublished work. Accordingly, some sections of the thesis appear in the following article: S. Kaminski Strauss *et al.*, "Evolthon: A community endeavor to evolve lab evolution," *PLOS Biol.*, vol. 17, no. 3, p. e3000182, Mar. 2019, doi: 10.1371/journal.pbio.3000182.

In addition, strains for the mate choice and fitness inheritance project were given to us from Prof. Giani Liti at IRCAN.

3. Contents:

4. Abstract

Evolution is a key process in biology since organisms must survive in constantly changing environments. In nature, organisms mainly evolve through mutations and recombination. Asexual organisms mainly use mutations and other modifications to the genetic and epigenetic material, as well as recombination through horizontal gene transfer. Sexual organisms harness in addition the evolutionary potential of mixing their genes with other individuals in their species through mating. Mutations strategies on one hand, and mating on the other, affect the genetics of the next generation and evolution of the species.

My work consists of three complementary projects. In the first project I examined diverse strategies for lab evolution. I directed the "Evolthon" project that engaged an international community of experimental evolutionary biologists in the quest to explore means to best evolve microbes in the lab towards a challenge. More than 60 participants contributed their evolved strains and I have conducted all experiments and analyses to characterize the resulting strains and evaluate their fitness. Interestingly, in yeast, that can evolve both sexually and asexually, all best performing strategies involved sexual mating. This result has naturally raised basic questions on evolution through mating such as if fitness is inherited and how, and how do organisms choose mating partners to optimize the fitness of the next generation. Thus, the second project of my PhD, aimed at exploring in unprecedented scale, yeast mating patterns and its implications on fitness and fitness inheritance as a quantitative trait. I designed and built a novel molecular genetics means that uses genomic barcoding and programmed recombination to allow the parallel recording of thousands of mating choices between hundreds of yeast stains. My results show that offspring resembles their parents' average fitness, as well as their maximal and minimal fitness, suggesting a complex manner of fitness inheritance. I found that offspring fitness correlate to the genetic distance between its parent such that fitness is high in high genetic distance. In a different trajectory of this project, I studied mate choice. I am showing here that some yeast strains mate with more strains than other, when a choice is given.

A third project studied the effects of retrotransposons on evolution. I show that cellular mRNA can be encapsulated in the Virus Like Particle formed by the yeast Ty retrotransposons, and also that genes that are depleted from it evolve slower.

The three projects presented here complement each other; in Evolthon I studied different approaches to lab evolution, in the retro transposons mediated evolution I studied a Lamarckian-like mode of evolution and in the yeast mating project I studied how mating affect evolution, and how evolution affect mating.

5. List of abbreviations

S. cerevisiae – Saccharomyces cerevisiae,

E. coli - Escherichia coli

OD – Optical Density

RT – Reverse Transcriptase

cDNA – complementary DNA

VLP – Virus Like Particle

ORF – Open Reading Frame

GD – Genetic Distance

yeGFP – yeast Green Fluorescence Protein

FACS - Fluorescence-activated cell sorting

Glu - Glucose

Gly - Glycerol

BFG – Barcode Fusion Genetics

NGS – Next Generation Sequencing

6. Introduction

It was said by Dobzhansky that "nothing in biology makes sense except in the light of evolution" [1]. Evolution is a key process in biology since organisms must survive in constantly changing environments. While evolution is the process in which organisms adapt to new environments, the capacity that enables them to do so is called evolvability [2]–[4].

Evolvability determines the potential of a species to evolve, e.g. a higher evolvable species will adapt quicker to new environments than a lowly evolvable species. Evolvability is a complex process that is determined by several mechanisms like mutation rate, population structure and dynamics, epistasis between mutations and more [2], [5]–[7].

The main theory of evolution, natural selection, that was proposed by Darwin is based on three concepts; random mutagenesis, selection of the fittest, and the inheritance of mutations; as such beneficial mutations will increase in frequency in the population [8], [9].

A different theory, by Lamarck, suggests a non-random process in adaptation of species. In this theory it is thought that the use of a particular organ would lead to its gradual functional improvement that can be inherited to the next generations [8], [10], [11].

Two main differences between Darwinian and Lamarckian theories put them on opposites edges; the randomness of mutations and their nature; while in Lamarckian evolution mutations occur preferentially in used systems in the organism (genes, organs etc.), at a given environment, and are beneficial in nature, in Darwinian evolution mutations occur at random positions in the genome, and can be beneficial, neutral or deleterious. Therefore, any deviation from total randomness into directed mutagenesis, or from beneficial-only mutations to beneficial-or-not mutations, cannot be considered as Lamarckian or Darwinian evolution. Thus, different evolutionary mechanisms can be found on the spectrum between Darwinian and Lamarckian evolution.

For example, epigenetic inheritance is being researched for many years now [12], [13] and is considered to be Lamarckian. Decades ago, structural inheritance, in which the 3D structure of a molecule dictates the 3D structure of newly formed molecules was showed in ciliates [14]. Another known and researched structural inheritance is the propagation of prions [15], [16]. Prions are a classical example of Lamarckian evolution, since the environment (i.e., the 3D structure of a prion) mediates the change of conformation in other non-prions proteins into prions. In recent years, other modes of epigenetic inheritance are being studied such as the inheritance of RNA in *Caenorhabditis elegans* [17], [18] Additionally, small RNAs area also inherited through the sperm in higher eukaryote such as in mice and even human [19]. In addition

to epigenetic mechanisms such as the inheritance of RNA, retro transposition is a process mediated by enzymes that can reverse transcribe mRNA and incorporate it into the genome [20], [21]. Since this process involves a flow from the phenotype (mRNA) to the genotype (genome), it can be considered as more Lamarckian.

On the other hand, sexual mating is a process in which two organisms mate to produce offspring. This process is Darwinian, since it is not predetermined by the environment, the combination of genetic material in the offspring is random (due to the meiosis and recombination occurring during gamete formation) and the offspring can have high, or low fitness.

To study evolution, one can use one of mainly two methods; either observe and compare organisms in nature (in recent decades, mainly by comparative genomics), and infer past conditions and species history [22]–[26], or evolve an organism in the lab and document the process [27]–[29]. Since microorganisms reproduce rapidly, it is convenient and widespread to use them in the study of evolution in the lab.

Lab evolution is widely used in the scientific community, and is being used to study the emergence of specific traits in a population, such as the emergence of resistance, response to stress etc. [30]–[32] but lab evolution technique itself remained unchanged for decades [28].

In my PhD I have studied different approaches to evolve microorganisms in the lab and its effect on strain fitness in Evolthon [33] challenging the existing paradigm in which lab evolution is conducted by a daily dilution routine of microbes in a constant environment. Nicely, among the strategies employed in Evolthon, some strategies were more Lamarckian, while other were more Darwinian. The “catching cold RNA” and “Ty-induced evolution” strategies are two of the more Lamarckian strategies, while methods involving different routines of dilutions, as well as mating with other strains can be considered as more Darwinian.

The other projects in my PhD were the study of yeast retro-transposition and the study of mate choice in yeast. In my work on retrotransposons I aimed to uncover a special mechanism of evolution that involves Lamarckian characteristics. In this project I showed for the first time a potential role for retro-elements in enhancing evolution rate by reverse transcribe cellular mRNA and integrate them into the genome. This study of retro-elements is still in progress in the lab.

As for mate choice, I focused on yeast mating, and mainly yeast mate choice to study some of its possible effects on yeast fitness and on the inheritance of fitness.

7. Evolthon: A community endeavor to evolve lab evolution

7.1 Introduction

Classical investigations in evolution are based on observing and comparing organisms in nature, and they require inference of the past conditions and species history. Though extremely insightful, this approach can be effectively complemented by "lab-evolution," a research paradigm in which organisms, typically microbes, are evolved in the lab. In this controlled setup, species can be challenged by changing environmental conditions, e.g., starvation, exposure to antibiotic drugs, high temperature, high salinity [28], [31], [32], [34], or by perturbing their genes [3], [27], [35], and then they can be followed as they evolve, inspecting a diversity of physiological and genomic means of adaptation. Therefore, rather than simply observing a snapshot, an entire evolutionary "movie" can be followed, during which the environment is not only known but can also be controlled and manipulated. The Long Term Evolutionary Experiment [28] is a famous experiment that essentially established the field, and in recent years many experiments followed [36]–[39].

Consider then the following challenge: you are given a microbe and you are asked to evolve it in the lab towards a new challenge, say to extreme temperature or to a toxic drug. What evolutionary regime will achieve "best" results? Naturally, one would expose the population to the challenge (e.g., high temperature or the drug), but open questions would include: (i) What is the optimal level of exposure to the stress? (ii) Should the stress level be constant throughout the experiment, or should it increase, decrease, oscillate, or fluctuate randomly with time? (iii) What should be the population size? Small populations feature evolutionary bottleneck and high effect of drift; (iv) If the organism can exercise sexual mating, should that be allowed? (v) Should mutation rate be manipulated, e.g., by exposing the evolving cells to a mutagen, or by working with a strain that features high mutation rate? (vi) Should cells be allowed to cycle between all stages of growth, as in serial dilution regimes [28], or should they be grown in a chemostat in constant logarithmic phase [40]?

One can be very creative in designing an evolutionary experiment, and the number of degrees of freedom is essentially unlimited. Post factum, one could ask, how did the evolutionary strategy employed affected performance? For example, it has been shown in yeast that exposure to an abruptly applied challenge, high temperature, as opposed to incremental increase in the temperature, pushed cells to evolve very

different solutions. When exposed to an abrupt increase in temperature, yeast evolved through aneuploidy, a solution that proves to be maladaptive in other stresses, and that might not endure well after short relaxation periods [34]. Therefore, an interesting possibility is that the adaptation regime applied during evolution would affect the stability and generality of the adaptation.

Although many works were done looking at individual evolution strategies, there is no larger-scale study aiming to compare the effects of different evolution strategies. In a first of its kind initiative in the evolution biology community, I conducted Evolthon, a world-wide collaboration, to study the effect of the evolutionary path on strain fitness. Along with various experimental-evolution groups worldwide, I and the lab have participated in the first Evolthon Challenge, a tournament that challenged participants to come up with creative ways to evolve microorganisms in the lab. Evolthon focused on either *Escherichia coli* or *Saccharomyces cerevisiae*, and the challenge was adaptation to low temperature. The inspiration for the Evolthon Challenge came from the successes of other community efforts to advance and generate new thought in other fields, most notably, Axelrod's Tournament in evolutionary game theory [41], the systems biology competition Dialogue for Reverse Engineering Assessments and Methods (DREAM) [42], and the International Genetically Engineered Machine (iGEM) competition in the field of synthetic biology [43]. We were eager to create a platform to enable the joint exploration of the range of possibilities in evolving a trait with the belief that such an endeavor will allow researchers and students to explore, be creative, collaborate, share knowledge and insight, to educate themselves through this process, and contribute knowledge and advance the field of lab evolution. The ultimate goal is to seed a collection of creative lab evolution strategies and generate a first-of-its-kind lab evolution strategies database, that will grow further past this initial publication, from which researchers and biotechnologists will be able to select and adapt further.

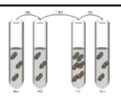
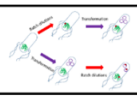
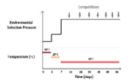
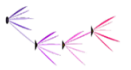
7.2 Results

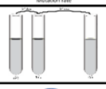
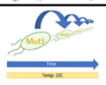
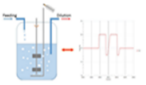
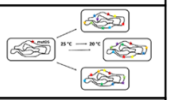
7.2.1. Participants from all over the world employed different lab evolution routines to achieve the most adapted strain

In Evolthon, 20 labs joined our community effort to study lab evolution (Figure 1). Each participant designed their own strategy for the evolution of their strain towards improved fitness under cold temperature. The different strategies that participants used were diverse and include strategies based on increasing mutagenesis, genomic engineering, population genomics, mating (in yeast) and more. Figure 1 and Table 1 summarize the strategies employed. Figure 2 localizes the various strategies on a conceptual plan that is spanned by two “principal axes”, the horizontal axis characterizes the extent of

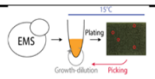
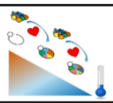
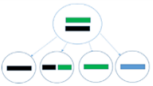
genetic manipulation, and the vertical one characterizes the environment regime employed. Strategies on the far right side employed elaborate genome engineering; those on the left side did not intervene genetically at all. High on the vertical axis are strategies that exposed cells to fluctuating temperature, and lower on this axis are those kept at a constant temperature. Some participants evolved their strains under a constant temperature, such as the announced low temperature, others applied either constantly higher or lower temperatures throughout the evolution.

(A)

#	Strategy name	logo
1	Growth advantage in stationary phase	
2	*E. coli* Manual chemostat	
3,9	Saltation-selection and vice versa	
4	Pop-Gen	
5	*E. coli* Daily dilution	
6	Survival of the fittest by means of directional selection	

#	Strategy name	logo
7,8	Variable mutation-rate selection	
10	Lazy man	
11	Accelerated Evolution	
12	Strength through diversity: the United States of *E.coli* (U.S.E)	
13	Combined chemostat with temperature fluctuation	
14	Hypermutation evolution	

(B)

#	Strategy name	logo
15	Delete and prosper	
16	Chemical mutagenesis	
17	Breeding with natural variation	
18	Simply Metabolism	
19	Adaptive lab evolution with mating	
20	*S. cerevisiae* Manual chemostat	
21	Foodie-evolution	
22	*S. cerevisiae* Daily dilution	

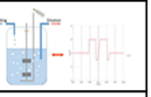
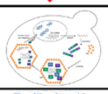
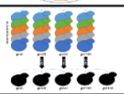
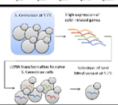
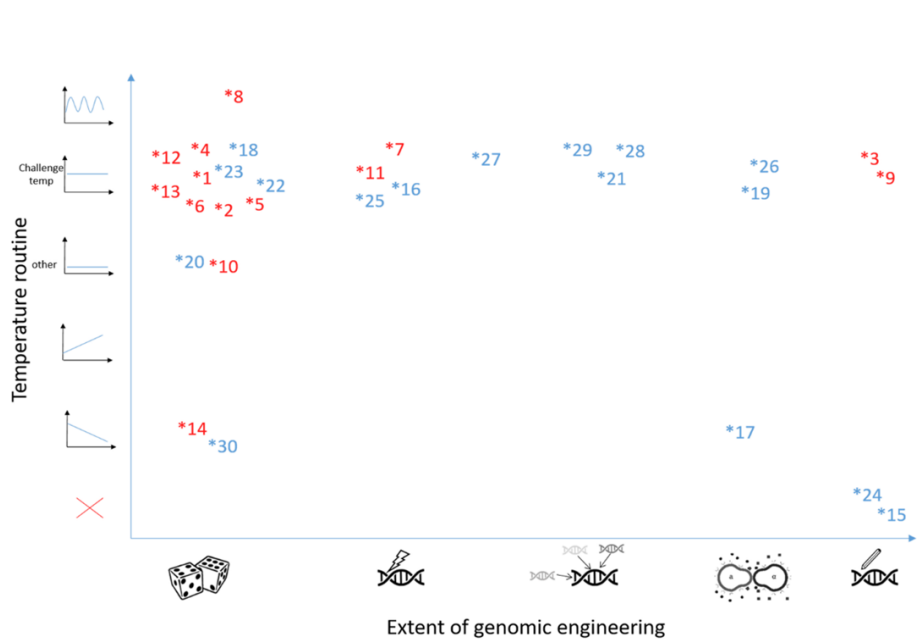
#	Strategy name	logo
23	Combined chemostat with temperature fluctuation	
24	Engineering of cold response genes using CRISPR/Cas9	
25	Cycles of random mutagenesis with selection	
26	Mating	
27	Ty-induced evolution	
28	Antarticold	
29	Catching cold RNA	
30	*S. cerevisiae* temperature gradient	

Figure 1. Summary of the strategies employed in Evolthon.
All strategies used in Evolthon are listed each strategy is characterize by identifying number, name, and logo. (A) Strategies used *for E.coli*. (B) Strategies used for *S. cerevisiae*.

Some combinations of genetic and environmental regimes were not tried and thus left are empty regions of the plane. For example, none of the participants combined rational design with evolution under changing temperatures. Strategies also differed in other aspects of their lab evolution protocol, such as number of generations, population size and the environmental settings.



	Bacteria
1	Growth advantage in stationary phase
2	*E. coli* Manual chemostat
3	Saltation-selection and vice versa
4	Pop-Gen
5	*E. coli* Daily dilution
6	Survival of the fittest group by means of selection
7	Variable mutation-rate selection
8	Variable mutation-rate selection (with cold-shock)
9	Saltation-selection and vice versa
10	Lazy man
11	Accelerated Evolution
12	Strength through diversity: the United States of E.coli (U.S.E)
13	Combined chemostat & temperature fluctuations
14	Hypermutation evolution
	Yeast
15	Delete and prosper
16	Chemical mutagenesis
17	Breeding with natural variation
18	Simply Metabolism
19	Adaptive evolution with mating
20	*S. cerevisiae* Manual chemostat
21	Foodie-evolution
22	*S. cerevisiae* Daily dilution
23	Combined chemostat & temperature fluctuations
24	Engineering of cold response genes using CRISPR/Cas9
25	cycles of random mutagenesis with selection
26	Mating
27	Ty-induced evolution
28	Antarticold
29	Catching cold RNA
30	*S. cerevisiae* temperature gradient

Figure 2. Schematic illustration of the different evolutionary strategies location in the conceptual plan.
A conceptually and qualitatively projection of all 30 evolutionary strategies onto a plane that is spanned by two principled characteristics of many of the strategies. The x axis denotes the extent of genome engineering and mutagenesis used. The left most strategies used no engineering, the second used mutagenesis, the third used DNA transformation, the fourth used mating (in yeast), and the right most used genome engineering. The y axis denotes the temperature versus evolutionary time regimen experienced by cells during evolution, with strategies exposing cells to fluctuating temperature, constant temperature, monotonically increasing or decreasing temperature, and a strategy (marked by a red X) that involved engineering with no lab evolution. Colors represent organism, *E. coli* (blue) or *S. cerevisiae* (red).

7.2.2. Most evolved strains exhibited improved fitness in the cold environment

Comparison of the fitness of the ancestral strains with those of the evolved strains, measured by individual growth curves, revealed that all evolved *E. coli* strains significantly improved their fitness in the cold environment as compared to their ancestor (Figure 3A,B). In *S. cerevisiae,* the picture was more complex, as some strains improved their fitness whereas the fitness of others either did not change or actually declined (Figure 3A,C). To further quantify how strains adapted to the cold environment, I analyzed the individual growth curves (using curveball algorithm) and extracted three growth parameters for each strain: lag phase duration, growth rate in the exponential phase, and the yield (maximal optical density [OD]) at the stationary phase [44] (Table 1). As can be seen from Figure 3D,E, the evolved *E. coli* and *S. cerevisiae* strains (respectively) behave differently.

Table 1. Summary of strains' performance.

For each strategy and the ancestral strain the growth parameters (1/lag, growth rate, and yield) calculated from individual growth curves (see S3 Text) are shown. Fitness values are calculated using a maximum likelihood algorithm (see S3 Text) based on the pool competition. Fitness was only calculated for strains with more than 10 reads at the beginning of the competition (otherwise ND is assigned).

Abbreviations: anc, ancestor; ND, Not Determined.

Bacteria					
#	Strategy name	Growth Curves parameters			Fitness
		1/Lag	Growth rate	yield	
anc	-	0.096	0.438	0.931	
1	Growth advantage in stationary phase	0.509	0.497	1.001	−0.234
2	E. coli Manual chemostat	0.380	0.493	1.010	−0.028
3	Saltation-selection and vice versa	0.733	0.579	0.994	−0.174
4	Pop-Gen	0.697	0.620	0.974	−0.097
5	E. coli Daily dilution	0.577	0.451	1.017	−0.173
6	Survival of the fittest group by means of selection	0.819	0.586	1.015	0.013
7	Variable mutation-rate selection	0.649	0.554	1.066	0.086
8	Variable mutation-rate selection (with cold-shock)	0.813	0.587	0.985	−0.130
9	Saltation-selection and vice versa	0.808	0.525	0.983	−0.159
10	Lazy man	0.852	0.557	0.983	−0.110
11	Accelerated Evolution	0.678	0.527	0.945	−0.076
12	Strength through diversity: the United States of E.coli (U.S.E)	0.838	0.567	0.988	−0.250
13	Combined chemostat and temperature fluctuations	0.565	0.495	0.992	−0.243
14	Hypermutation evolution	0.602	0.608	0.941	−0.097
Yeast					
#	Strategy name	Growth Curves parameters			Fitness
		1/Lag	Growth rate	yield	
anc	-	0.137	0.196	1.196	
15	Delete and prosper	0.132	0.149	1.411	−0.283
16	Chemical mutagenesis	0.145	0.169	1.491	−0.187
17	Breeding with natural variation	0.162	0.224	1.963	0.281
18	Simply Metabolism	0.105	0.143	1.621	−0.437
19	Adaptive evolution with mating	0.143	0.129	1.860	−0.762
20	S. cerevisiae Manual chemostat	0.143	0.143	1.353	ND
21	Foodie-evolution	0.040	0.090	0.639	ND
22	S. cerevisiae Daily dilution	0.065	0.159	0.724	ND
23	Combined chemostat and temperature fluctuations	0.131	0.161	1.490	−0.121
24	Engineering of cold response genes using CRISPR/Cas9	0.144	0.151	1.363	−0.298
25	cycles of random mutagenesis with selection	0.069	0.154	1.027	ND
26	Mating	0.178	0.219	1.319	−0.003
27	Ty-induced evolution	0.043	0.096	0.691	−0.128
28	Antarticold	0.149	0.153	1.511	−0.351
29	Catching cold RNA	0.054	0.129	0.749	−0.318
30	S. cerevisiae temperature gradient	0.149	0.195	1.391	−0.062

All *E. coli* strains mainly evolved by significantly shortening the lag phase duration, and they also improved their growth rate to some extent, whereas their yield showed little improvement (Figure 3F). In *S. cerevisiae*, in addition to strains that improved their fitness, there were strains in which none of the parameters were improved, and even strains that performed worse than the ancestor, mainly due to increase of lag time (Figure 3G). Moreover, unlike *E. coli* that mainly improved its lag, the fittest strains in *S. cerevisiae* primarily improved their yield (Figure 3D and 3E).

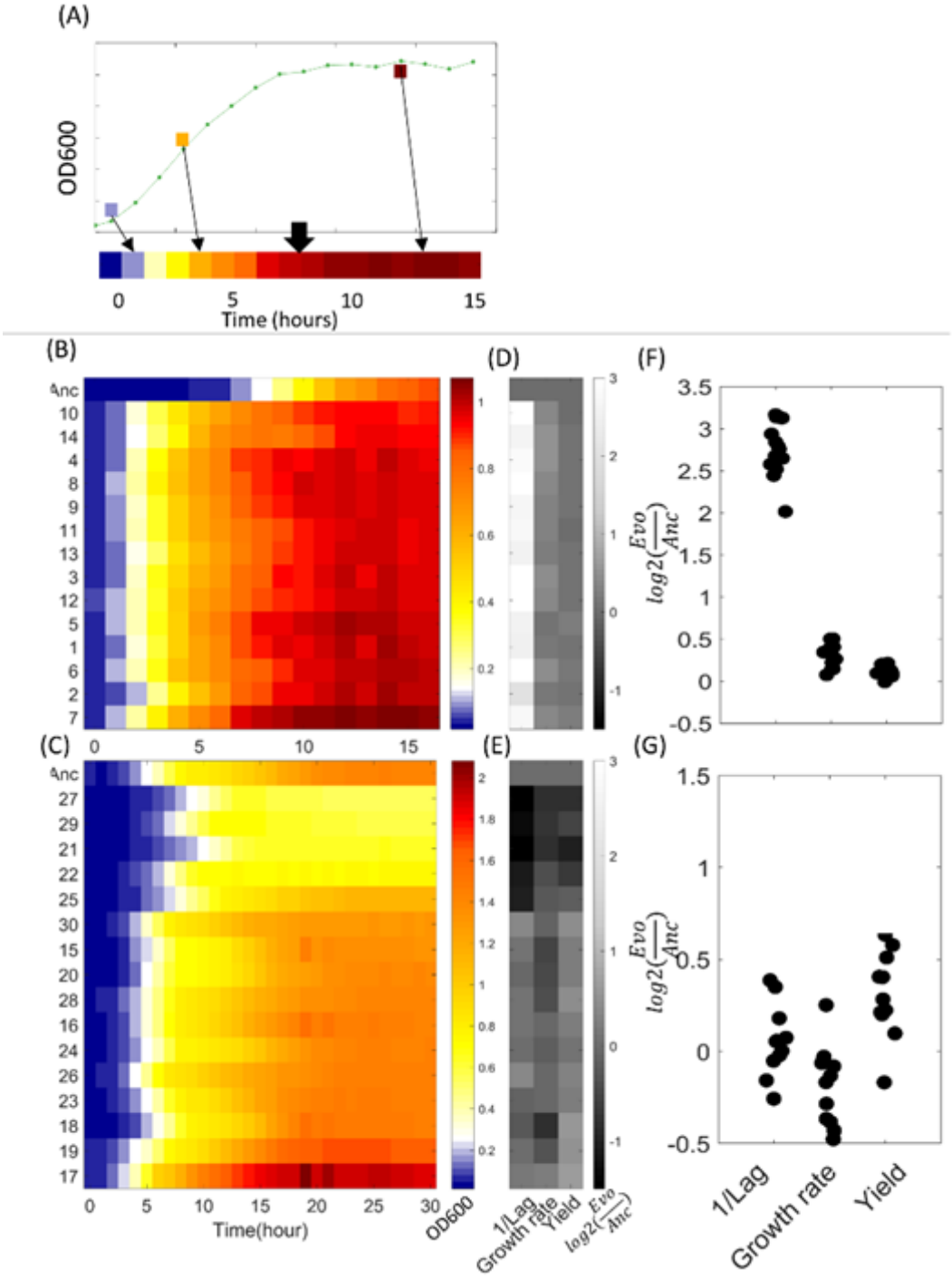


Figure 3. Growth experiments of individual strains.
All strains were grown for approximatley 30 hours in 15 °C and 20 °C (*E. coli* and *S. cerevisiae*, respectively), while measuring OD600 every approximately 1.5 hours. (A) Schematic representation of transforming growth curves into heat map figure. Each point in the growth curve is colored based on its OD600 value to obtain the heat map figure. (B-C) Growth in heat map format. Each row corresponds to a strain. Color bar represents OD600 values. Growth experiments were done in 11 replicates per strain. (B) *E. coli* (SD doesn't exceed 0.02). (C) *S. cerevisiae* (SD doesn't exceed 0.17). Strains in each species are sorted in ascending order according to final OD. (D-G) Growth parameters (lag, growth rate, and yield) were calculated based on a mathematical model for growth (for details, see S3 Text). Color bar represents log2(Evol/Anc) for each growth parameter. (D, F) *E. coli*; (E, G) *S. cerevisiae.*

Notably, the ancestral strains of *S. cerevisiae* and *E. coli* have different growth dynamics, especially a different lag phase duration (under the low temperature regimes). In *S. cerevisiae*, the lag phase is approximately 11% of the entire growth cycle (5 hours of lag phase out of 45 hours until stationary). In *E. coli*, the lag phase duration is also 5 hours, but the entire growth cycle duration is 25 hours (thus lag phase covers 20% of the cycle) (Figure 3B, C Top rows). In light of this dynamics, it seems that the benefit of shortening the lag phase is higher in *E. coli* than in *S. cerevisiae*. These results indicate that, unlike a potential naïve expectation, increase in growth rate might be less common in adapting to a new environment; in contrast shortening lag phase appears to be the immediate avenue for adaptation in *E. coli*. Shortening of lag phase was revealed as the main means of adaptation in *E.coli* population that were not exposed to such abiotic stress [45] but rather evolved to utilize non favorable carbon source. This commonality suggests that evolution through shortened lag phase in *E. coli* may be a common adaptation mean featured in different types of conditions. Because both *E. coli* and *S. cerevisiae* were evolved under the same type of stress, i.e., cold temperature, our data allowed us to compare the type of improvement featured by the two organisms. By looking on the correlation across evolving strains, in their improvements in each of the growth phases, we note a difference between the two organisms. Whereas in yeasts strategies that improved the performance of the cells in one phase typically improved performance in other phases, in *E. coli*, correlations exist only between lag phase and growth rate improvements (Figure 4).

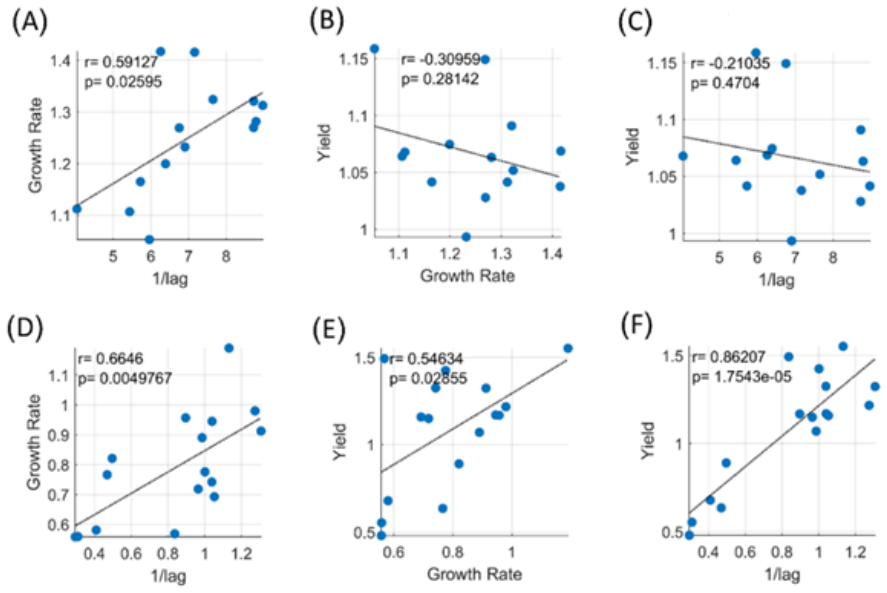


Figure 4. Growth parameters are correlated in yeast but not in bacteria.
Growth parameters (lag, growth rate, and yield) were calculated based on a mathematical model for growth [36]. Correlations between each two parameters are shown separately for *Escherichia coli* (A–C) and *Saccharomyces cerevisiae* (D–F). Correlation coefficient and statistical significance were calculated based on Pearson correlation and are presented for each plot. Data for this figure was taken from Table 1.

7.2.3. Pooled competition reveals the best performing strains of E. coli and S. cerevisiae

Different strains featured various levels of improvement in different growth phase parameters (Figure 3D-3G). Because it is not clear which of the parameters mostly affect ultimate evolutionary success, we conducted a pooled competition experiment to evaluate fitness of each strain in the presence of all others. We competed all bacteria and separately all yeast strains for up to 60 generations, employing a conventional routine of daily dilution into a fresh medium. Competition
was done in rich media at the designated low temperature (either 20˚C or 15˚C for *E. coli* and *S. cerevisiae*, respectively). We then sequenced the barcode region of the strains in generation 0, 20, 40, and 60 to follow changes in frequency over time. We estimated the fitness of each strain using a published algorithm [46], [47]. The pooled competition results show that one evolved *E. coli* variant (a strategy called "Variable mutation-rate selection") and one evolved *S. cerevisiae* variant (the "Breeding with natural variation" strategy) took over the population, hence, having the highest fitness based on the pooled competition (Figure 5A).

Because one yeast strain ("Breeding with natural variation" strategy) took over the population, we could not rank, in its presence, the rest of the strains. We thus removed this strain from the pool and repeated a competition between all other strains. The winner in this event was the "Mating" strategy, another strategy that exercised sexual mating (Data not shown). A third strategy that utilized mating, "Adaptive lab evolution with mating" strategy, obtained the second highest estimated yield in the individual growth curves analysis (Figure 3C and Table 1). Therefore, the three fittest yeast strains, coming from three independent labs, were those that utilized sexual reproduction as a means to evolve. In particular, these strategies mated the Evolthon strain with phenotypically diverse natural isolates of yeast strains. They either selected a natural isolate prior to mating, based on growth advantage in cold (the "Adaptive lab evolution with mating" and "Mating" strategies), or they mated the Evolthon strain with a library of wild isolates, selecting for growth advantage in cold after mating (the "Breeding with natural variation" strategy). The success of the mating based strategies in yeast can be rationalized because sex is very well known to improve adaptation since it allows the evolving populations to recombine beneficial mutations that would have otherwise segregated in different populations [48]. In contrast to the success of mating based strategies, several strategies that used DNA transformations of various sorts (see Table 1) did not fare very well.

In bacteria, the winning strategy was the "Variable mutation-rate selection" strategy, used a high mutation rate using an error-prone DNA polymerase, which was induced at different levels at different

repeats (Figure 5A, top panel). The best colony from the combined evolutionary repeats was chosen for submission.

Here, too, we later removed the winner strain and repeated the competition in order to reveal the second highest (Data not shown). In this case, the winner was a strategy termed "Survival of the fittest group by means of selection" that employed a more complex population genetics approach that utilizes directional selection while increasing the number of tested genotypes.

To control for possible biases originating from slight technical differences between labs and in order to examine the robustness of the competition results, we performed the pooled competition experiment under the low temperature conditions in two additional Evolthon labs (one for *E. coli* and the other for yeast) using different shakers, incubators, etc. and repeated the barcode sequencing-based fitness measurements, as described above. The results were highly correlated between these repeats, indicating that the results reflect the true ranking of the strains (Data not shown).

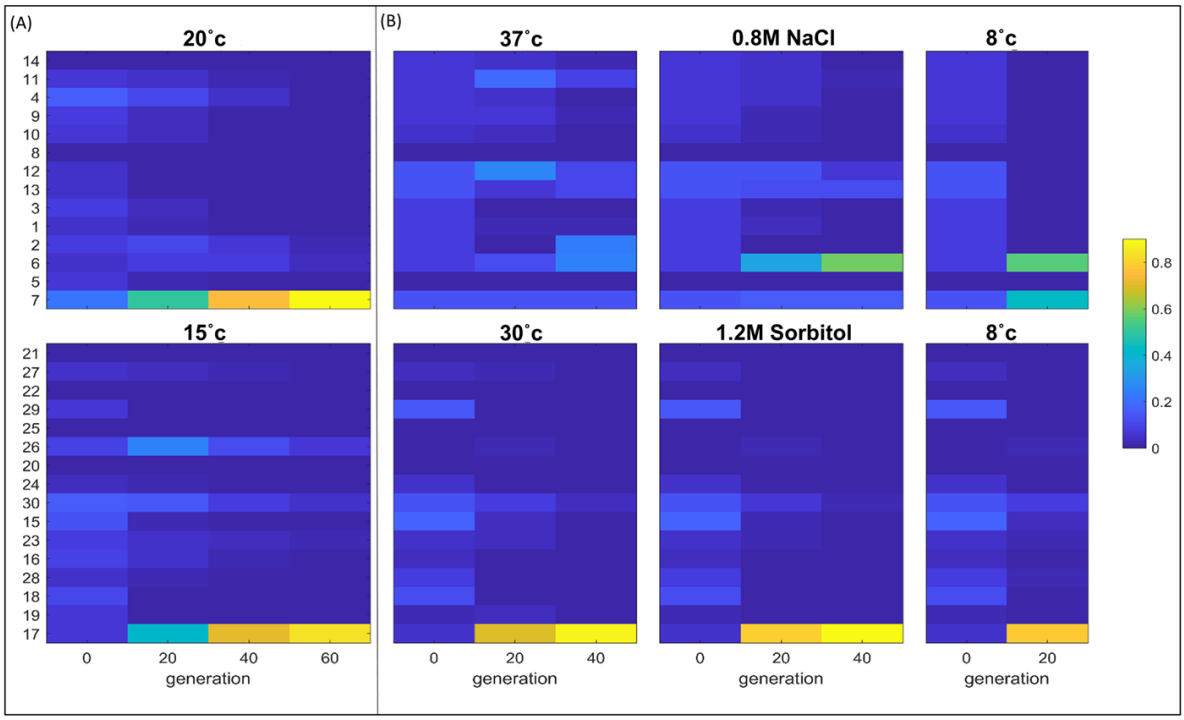


Figure 5. Pooled competition.
Strains were mixed and grown for several dozens of generations in serial dilution regimes under different growth conditions (see S3 Text for details). At different time points during the competition, barcodes were sequenced, and their frequencies are shown. (A) Challenge conditions to which strains were evolved (15 °C and 20 °C for *E. coli* and *S. cerevisiae*, respectively). Color bar represents the frequency of the strains barcode reads from total number of reads. (B) Other challenges ("evolutionary memory," 37 °C and 30 °C for *E. coli* and *S. cerevisiae*, respectively; "generalization," 0.8M NaCl and 1.2M sorbitol for *E. coli* and *S. cerevisiae*; "extremity," 8 °C for both *E. coli* and *S. cerevisiae).* Color bar represents the frequency of the strains barcode reads from total number of reads. Upper panels present *E. coli* competition results; lower panels present *S. cerevisiae* competition results

7.2.4. Revealing trade-offs, memory and generalization upon adaptation

Evolution often trades off between competing tasks. For example, when improving fitness towards a certain challenge under selection, organisms might compromise their fitness in another environment, in particular the original environment to which they were already adapted. Are there evolutionary strategies that intensify or weaken such trade-offs compared to others? We utilized our set of evolved strains to examine trade-offs by competing the strains in different conditions that were not revealed to the various participating labs when the challenge was announced. The three conditions that we chose were (1) performance at extreme temperature conditions ("extremity")—here we sought to assess how well each strain performs at an even lower temperature of 8˚C; (2) trade-off between evolutionary change and previous legacy ("evolutionary memory")—Here we were interested in assessing whether evolution toward low temperature compromised the fitness at the original "comfort-zone" temperature (37˚C and 30˚C for *E. coli* and yeast, respectively); (3) performance under a different stressor ("generalization")—here, we wanted to test whether strains that evolved toward one stress, low temperature, have also gained adaptation, perhaps as an evolutionary by-product, to another stress, an osmotic stress, using NaCl for *E. coli* and sorbitol for yeast. The results in yeast were very clear: the sexually reproducing strain ("Breeding with natural variation" strategy) outperformed all others under each of the three additional conditions (Figure 5B lower panel). In bacteria, the situation was more complex (Figure 5B upper panel). The winner in the osmotic stress condition was the "Survival of the fittest group by means of selection" strategy, but in extremity conditions, both the "Survival of the fittest group by means of selection" and the "Variable mutation-rate selections" strategies, were the best strains (Figure 5B upper panel). The behavior of the "Variable mutation-rate selection" strategy across the conditions was interesting. Although this strategy performed the best under the designated low temperature conditions, it did worse in the other, unforeseen challenges. This behavior might indicate that a mutagen can be beneficial in finding a good genetic solution to a particular environment under selection but might compromise other parts of the genome that are presently not under selection but that might prove crucial in the future. In contrast sexual mating appears to preserve and even expand on organisms' qualities, presumably as it mixes alleles and mutations that were proven useful in diverse conditions in the natural history of the species.

7.3 Discussion

Evolthon was the first community challenge in lab evolution. It was successful in engaging many labs, mainly through the independent work of students that were very creative, though often employing "backyard biology" in the lab. The joint work of many labs brought two essential assets. First, the strategies chosen were very diverse, highly creative, and they open many new possibilities for new developments. As can be seen in Figure 2, many potential combinations of strategies were not explored (here) so far. Many additional degrees of freedom may still be utilized. Second, in terms of number and heterogeneity of approaches experimented here, such a community effort can much exceed the scale that is typically achievable by individual researchers and students.

It is also important to note a central limitation of Evolthon and community challenges of this sort. Due to the very nature of this mode of science making, it cannot, and probably should not, attempt to cover and examine systematically all possible parameters and degrees of freedom in the space of strategies. For example, if a conventional research was aimed at finding the concentration of a mutagen that maximizes evolutionary adaptation, typically a single researcher in one lab, they would have carried out an orderly experiment with appropriate controls in which a whole range of concentrations were examined. However, natural evolution actually works the "Evolthon way" in the sense that genomes never evolve by systematically varying their parameters over a range of potential values (say, expression level of a gene or affinity of an enzyme to a substrate). Instead, evolution tries out sporadic solutions and continues with the fittest. In that respect, we might say that here we apply the nature of the evolutionary process to the study of evolution itself.

The conceptual directions revealed here could be important for other fields of biology. For example, in biotechnology, optimal evolutionary strategies are important. It is a common practice to use lab evolution to evolve strains with desired applied properties, such as degradation of biological products [49], [50], production of products [51], etc. The search for optimal strategies can lead towards efficient means to screen the parameter space of evolutionary strategies.

In clinical applications, such as in infectious diseases and cancer, it is crucial that the cells will not evolve resistance. The regimen of application of drugs could enhance, or perhaps suppress, evolution of resistance. Can efforts of the type conducted here reveal anti-evolution regimes, e.g., for drug application, that would allow on one hand effective treatment and, on the other, would limit the capacity of the attacked cells to evolve resistance? Perhaps the least efficient strategies tried here could be most useful in this opposite challenge.

7.4 Materials and Methods

7.4.1. Strains and growth conditions

Yeast

Yeast strains were based on *S. cerevisiae* BY4741 strain. A 20-bp barcode marked with Hygromycin B resistance gene (HygR) was introduced into the HO locus to create a collection of 35 strains, each with a unique barcode (MATa, his3Δ1, leu2Δ0, met15Δ0, ura3Δ0, HO::barcode-HygR).

Plasmid pAG32 [52] (Addgene plasmid #35122) was used as a template to amplify the HygR resistance cassette. The HygR was amplified with primers that include 40 bp homology to the HO locus, 20bp unique barcode and homology to the HygR cassata. The following primers were used

F:CTCATAAGCAGCAATCAATTCTATCTATACTTTAAAATGCTTTCTGNNNNNNNNNNNNNNNNNNNccttgacag tcttgacgt

R: CTCAAGATACAAAAAGCGTTACCGGCACTGATTTGTTTCAACCAGggcgttagtatcgaatcg

(Underlined regions are the HO homology tails, N are the 20-bp barcode, and the lower case letters are the primers for the HygR amplification from pAG32).

To introduce the barcode and the HygR cassette into the HO locus, yeast cells BY4741 were transformed with a standard LiAc protocol [53]. In short, cells in the logarithmic phase were harvested and washed twice in 1M TE+LiAc. Cells were incubated with 1M TE+LiAc, 40% PEG4000, 100mg/ml salmon sperm and 45ul of PCR product for 40 minutes in 30˚C. Cells were then incubated for 40 minutes at 42˚C and incubated over-night in YPD at 30˚C. On the following day cells were plated on YPD + Hygromycin B and grown until colonies appeared to select for transformants. Insertion of barcodes was verified by amplification of the HO region using the following primers followed by Sanger sequencing. (F: ATTGTATTCAATTCCTATTC, R: ATTGTATTCAATTCCTATTC)

Unless mentioned otherwise yeast cells were grown on YPD (10g/L yeast extract, 20g/L peptone, 20g/L glucose). YPD-hyg media is YPD with 300ug/ml Hygromycin B (Roche). YPD-Sor is YPD with 1.2M of sorbitol.

Bacteria

E. coli strains were based on MG1655 (K-12 F^{-} λ^{-} *ilvG*$^{-}$ *rfb-50 rph-1*). I have introduced a 20-bp barcode marked with Kanamycin resistance gene (KanR) into the *LacZ* locus.

KanR was amplified from a strain with genomic KanR cassette [54] (kindly given from Ron Milo's lab). The KanR was amplified with primers that include 40 bp homology to the *LacZ* locus, 20bp unique barcode and homology to the KanR cassata. The following primers were used

F:AGCGGTGCCGGAAAGCTGGCTGGAGTGCGATCTTCCTGAGGCCGATACTGNNNNNNNNNNNNNNNNNNNN ggaacccctatttgttt

R:CCGCTTGCCAGCGGCTTACCATCCAGCGCCACCATCCAGTGCAGGAGCTCggaccgaaccccgcgttta

(Underlined regions are the LacZ homology region, N represents the 20-bp barcode, and lower case letters are the primers for the KanR cassette)

Integration of barcodes into the bacterial genome was based on homologous recombination using the lambda Red recombinase system [55] using electroporation. Briefly, *E. coli* MG1655 cells harboring the pSLTS plasmid were grown over-night on LB-ampicillin media at 30˚C. 1ml of the culture was inoculated into 100ml of LB-Amp media and grown for an hour in 30˚C. L-Arabinose was added to a final concentration of 1mM and cells were allowed to grow until reaching an OD of 0.7. Cells were harvested by centrifugation (4500g for 10min) and washed twice with ice-cold 10% glycerol. Pellet was resuspended in ice-cold 10% glycerol and dispensed into 50ul aliquots.

To introduce the KanR cassette, 50-100ng of PCR products containing the cassette, barcodes and selectable marker were incubated on ice with 50ul of electrocompenet cells carrying the pSLTS plasmid. Cells were electroporated and immediately suspended in 1ml LB and incubated for 3 hours at 30˚C. Cells were then plated on LB-Kanamycin plates and grown until colonies appear to select for transformants cells. Insertion of barcodes was verified by amplification of the *LacZ* region using the following primers followed by Sanger sequencing. (F: ATGACCATGATTACGGATT, R: TTATTTTTGACACCAGACCA).

Unless mentioned otherwise *E. coli* cells were grown in LB (5g/L yeast extract, 10g/L tryptone, 10g/L NaCl). LB-Amp is LB with 100 µg/mL final concentration of Ampicilin. LB-Kan is LB with 50 µg/mL final concentration of Kanamycin. LB-NaCl is LB with 0.8M final concentration of NaCl.

7.4.2. Fitness assessment using individual growth experiments

Strains were inoculated from plates containing Hygromycin B or Kanamycin (*S. cerevisiae* or *E. coli* respectively) into YPD or LB (*S. cerevisiae* or *E. coli* respectively) and grown for two days in cold temperature (*S. cerevisiae* in 15˚C, *E. coli* in 20˚C) until reaching stationary phase.

Strains were diluted 1:50 in 96-well plate for a final volume of 150ul per well (four strains in 8 repetitions each and the ancestor in 48 repetitions in a single plate, in a checker-board format. Strains were grown under shaking conditions for ~50h at the appropriate temperature (*S. cerevisiae* in 15˚C, *E. coli* in 20˚C). OD600 was measured every 1.5 hours for ~50 hours by a plate reader (infinite 200, Tecan). All measurements were done automatically using a Hamilton robotic system.

Growth parameters (lag phase duration, growth rate at exponential phase and yield) were extracted from the obtained growth curves using the “curveball” software [36].

7.4.3. Pooled competition

Strains were inoculated from plates containing Hygromycin B or Kanamycin (*S. cerevisiae* or *E. coli* respectively) into YPD or LB (*S. cerevisiae* or *E. coli* respectively) and grown for two days in cold temperature (*S. cerevisiae* in 15˚C, *E. coli* in 20˚C).

After two days, OD values were measured to each strain and strains were mixed accordingly to reach equal cell representation. The mixture was diluted 1:120 into relevant media (Yeast were competed on the following conditions: (i) YPD-Hyg at 15˚C, (II) YPD-hyg at 30˚C, (III) YPD-hyg at 8˚C and (IV) YPD-Sorbitol at 15˚C. *E. coli* were competed on the following conditions: (I) LB-kan at 20˚C, (II) LB-Kan at 37˚C, (III) LB-Kan at 8˚C and (IV) LB-NaCl. Cells were grown in 1.2ml liquid media under shaking of 800RPM, in a 24-well plate. Every ~1-2 days, when culture reached the stationary phase, cells were diluted by a factor of 1:120 and re-grown under the same conditions. Competitions were carried out for 40-80 generations. Cells were frozen in 30% Glycerol and kept in -80˚c every 4 dilutions. All competitions were done in 5 replicates.

7.4.4. Barcode sequencing for pooled competition

At the end of the competition (~60 generations) DNA was extracted from 2, 3 or 4 time points from three repetitions (yeast: MasterPure Yeast DNA Purification Kit by epicenter, *E. coli*: Wizard Genomic DNA Purification Kit by promega). Libraries for sequencing the barcode region were constructed by designing PCR primers targeting the barcode region with tails that match Illumina adapters (yeast: F: ACGACGCTCTTCCGATCTacgtcaagactgtcaagg, R: AGACGTGTGCTCTTCCGATCTttgtattcaattcctattctaaatggc, *E. coli*: F: ACGACGCTCTTCCGATCTaaaaccctggcgttaccc, R: AGACGTGTGCTCTTCCGATCTgatccttggcggcaag. Capital letters correspond to Illumina adaptors, while lower case letters correspond to homology to the genome). A second PCR (F: AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTCCGATCT,
R: CAAGCAGAAGACGGCATACGAGAT***NNNNNNNN***GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT. ***N*** corresponds to Illumina index for library multiplexing) was carried out to attach the adapters for the Illumina run. Barcodes were sequenced using 75-nt single-end reads, on the NextSeq platform (Illumina).

7.4.5. Fitness estimation based on pooled competition

Fitness was derived by employing a Maximum-Likelihood (ML) algorithm on all frequency measurements along the competition experiment per variant (fitness was calculated only for strains with more than 10 reads in the beginning of the competition). Briefly, first, each variant fitness is estimated by using a simple loglinear regression over the first three time points. Based on these estimations, the initial relative

frequencies of each variant, and a noise model that accounts for experimental errors [47], expected trajectory of each variant is estimated and compared to the measured trajectory. Next, small changes are made to our fitness estimates, comparison is repeated, fitness is updated if they better fit the data (higher likelihood). This procedure is performed iteratively until fitness estimates are stable (maximized likelihood).

8. Reverse transcription as a potential role for Lamarckian evolution

8.1 Introduction

As described previously, different genetic elements can contribute to evolution.

One such agent is retro elements which was used as one of the methods to evolve yeast in Evolthon [33].

Reverse transcription is a process where a reverse transcriptase (RT) enzyme reverse transcribes RNA into DNA; The complementary DNA, cDNA, that is made by the RT can be incorporated into the genome via one of two mechanisms; Integration into the genome by the Integrase (Intp), creating an extra copy of a gene or by homologous recombination, replacing an existing copy [20], [21].

RT can potentially affect the evolvability of cells by a local effective increase in mutation rate of expressed genes. RNA polymerase has orders of magnitude higher mutation rate than the DNA polymerase [56]; therefor, by reverse transcribing mRNA and integrating it back into the genome, RT can locally increase effective mutation rate of expressed genes while allowing constant and low mutation rate in the rest of the genome. In addition, since the cDNA can be integrated into the genome, it can alter the other genes in the genome by a knock-out (in case the cDNA replaces the ORF or abolishes regulatory sequences), reduced or enhanced activity (by interfering with regulatory sequences).

Saccharomyces cerevisiae has a reverse transcription capability in the form of a retro element called Ty. There are 5 families of Ty elements in yeast (called Ty1-Ty5) each of them is found in multiple copies in the genome and all together comprise about 3% of its genome [57]. The most active and researched Ty is Ty1.

Ty1 has ~30 copies in *S. cerevisiae* genome; it is composed of two open reading frames (ORF), the TyA ORF, also known as the Gag ORF and TyB ORF found in the +1 frame of TyA and is known as the Pol ORF. The TyA ORF contains a structural protein (gag) that is used to form a capsid. TyB is the catalytic ORF containing 3 proteins: RT, integrase (Intp) and protease. The protease is auto-cleaved from a single TyB peptide, and catalyzes the cleavage of the Intp and the RT. The Ty1 life cycle begins with transcription and translation of both ORFs. The gag proteins form a Virus Like Particle (VLP) that encapsulates the Ty1's mRNA and all Ty1's proteins as well as a tRNA Met that is used as a primer for reverse transcription. In the VLP, a reverse transcription process of the Ty1 mRNA takes place followed by the import of the cDNA into the nucleus and its integration into the genome by the Intp into an ectopic location, or by homologous recombination replacing an existing TY1 copy [57]–[60].

In this part of my PhD I wanted to explore the possibility that reverse transcription can enhance evolution by reverse transcribing cellular mRNAs (perhaps some that also contain mutations due to the high error rate of RNA polymerase) and integrating them into the genome.

To affect evolution via high mutation rate of expressed genes, cellular mRNAs should be encapsulated in the VLPs and be reverse transcribed. It is therefore important to know if cellular mRNAs can be encapsulated in the VLPs and if they can be reverse transcribed. Currently, there are a few anecdotal evidences supporting the notion that non-Ty1 mRNA can be encapsulated in the VLPs, reverse transcribed and integrated into the genome. Curcio and Garfinkel showed that the *HIS3*'s mRNA is being reverse transcribed by the Ty element [61]. In addition, Maxwell et al. showed that the VLPs contain other mRNA such as the Y' element and that these Y` element transcripts are reversed transcribed and integrated at sub-telomeric region [62]. However, a more systematic analysis of the scope and type of mRNAs that can utilize this system is still missing. Thus in this part of my PhD I aim to address this question in a systematic manner.

8.2 Results

8.2.1. Extraction and sequencing of RNA in VLPs

In order to characterize the set of RNAs encapsulated within VLP I started by setting an experimental system for isolating VLPs from yeast cells (this project originated during my M.Sc degree and was followed up in my PhD). I adapted Eichinger and Boeke's protocol for the isolation of VLPs using a 20%, 30% and 70% sucrose step-gradient [59]. I performed VLP isolation from two strains (either a strain that contains a plasmid with Ty under the strong inducible Gal promoter (pTy strain) or a strain with an empty vector (control strain). The experiment was done in three and two biological replicates, respectively. By using multiple molecular and biochemical methods, e.g. qRT-PCR on the Ty gene, western blotting against the Gag protein that constitute the VLP envelop, and RT activity assay I identified the VLP-containing fractions in the pTy strain, while no VLPs found in the control strain (examples for the gradients are in Figure 6). The corresponding fractions from the control strain serve as a control.

I then extracted RNA from the chosen fractions of both strains as well as total RNA from both strains and subjected them to NGS sequencing. My library preparation and sequencing method was custom made (Based on Zheng et al. paper [63]) and allowed me to detect all RNA including small RNAs.

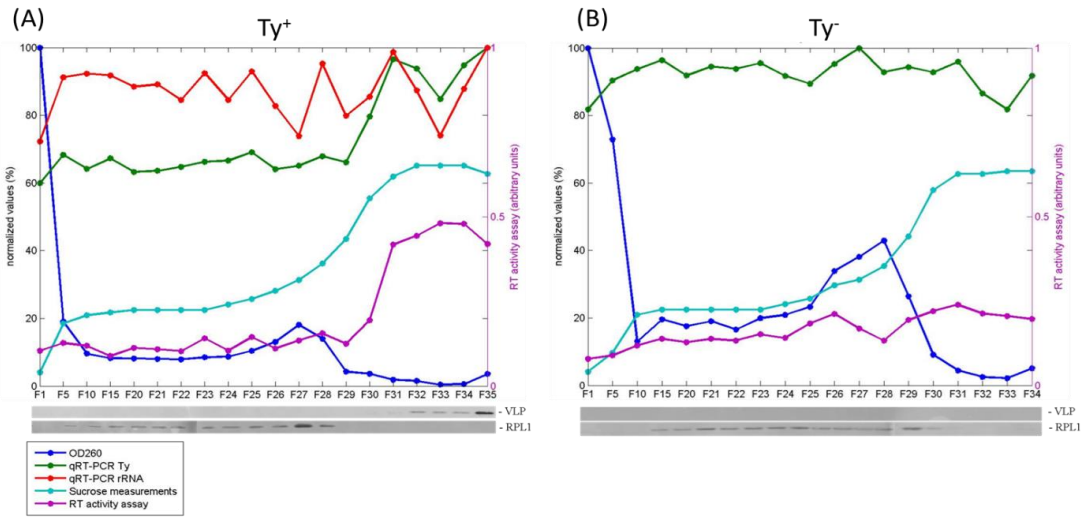


Figure 6. VLPs are found in ~60% sucrose fractions of the pGal-Ty strain

A sucrose gradient was made to isolate VLPs. The gradient was manually fractionated to (A) 34 fractions or (B) 38 fractions. Multiple assays were done on each of the fractions to identify VLPs containing fractions, as well as ribosomes containing fractions. OD 260 measurements (blue line) were done on fractions; qRT-PCR analysis using both Ty primers and ribosomes primers (green and red lines, respectively) were done on total RNA extracted from the fractions; each of those assays results were normalized. RT-activity assay was also done on fractions (purple line). Western analysis using both anti-VLPs and anti-rbosomes were done on fractions (bottom panel). The sucrose percentage in the fractions was determined using a refractometer (cyan line).
(A) Ty strain, (B) control strain

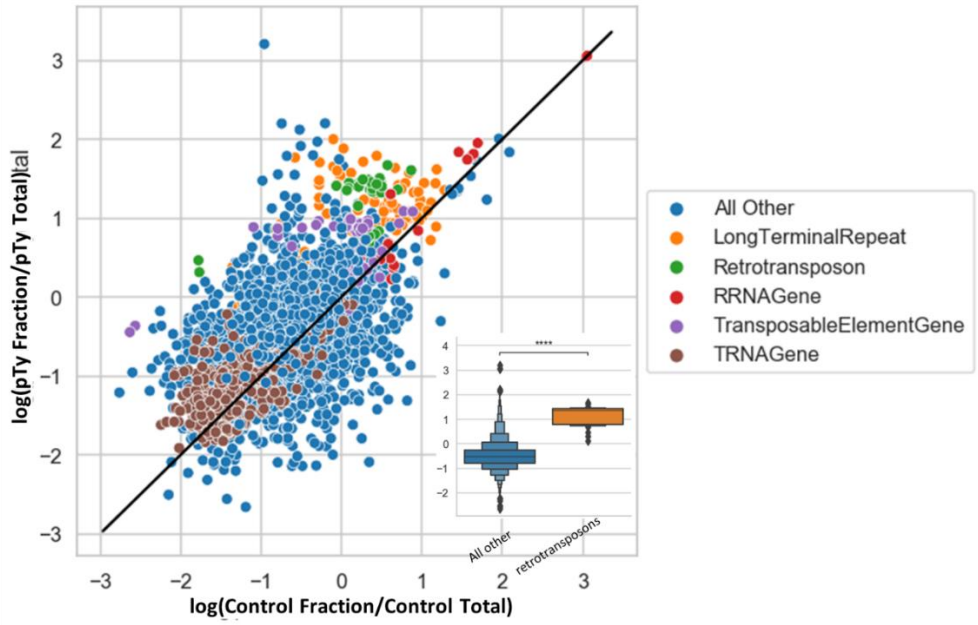


Figure 7. Some RNAs tend to be encapsulated in the VLPs

Each dot represents a transcript (mRNA, or other kinds of RNAs). The x axis shows the expression level of each gene in the cytoplasm of a Ty-containing strain. The y axis shows the fraction of the gene in the extracted VLPs. In green are the Ty genes, in red are all genes in the genome. The encapsulation rate of Ty is significantly higher than the genome.
Inset – the Ty tend to be encapsulated in the VLPs, X axis is retrotransposons (right, orange) or all other (left, blue), Y axis is the normalized pTy fraction to total RNA divided by the normalized control fraction to total RNA.

8.2.2. Different RNAs are encapsulated to different extent in the VLPs

Figure 7 presents a dot plot showing for each transcript its RNA level within VLPs normalized to its levels in the total RNA in the pTy strain against its levels from the control gradient normalized to total RNA from the control strain. The figure shows that most RNAs are found equally in the VLP fraction of the pTy strain and in the equivalent fraction from the control strain, meaning that they are not encapsulated nor depleted from the VLPs. However, some mRNAs are higher in the pTy fraction than in the control. Reassuringly, retrotransposons were found to be among the highly encapsulated genes (green dots) as well as long terminal repeats sequences (orange dots). As the Ty are known to be encapsulated in the VLPs, they serve as a positive control for the extraction method, as well as the strain used in this assay, reassuring that assay is valid. rRNA and tRNA were also colored (red and brown dots, respectively) and show no tendency to be depleted nor encapsulated in the VLPs.

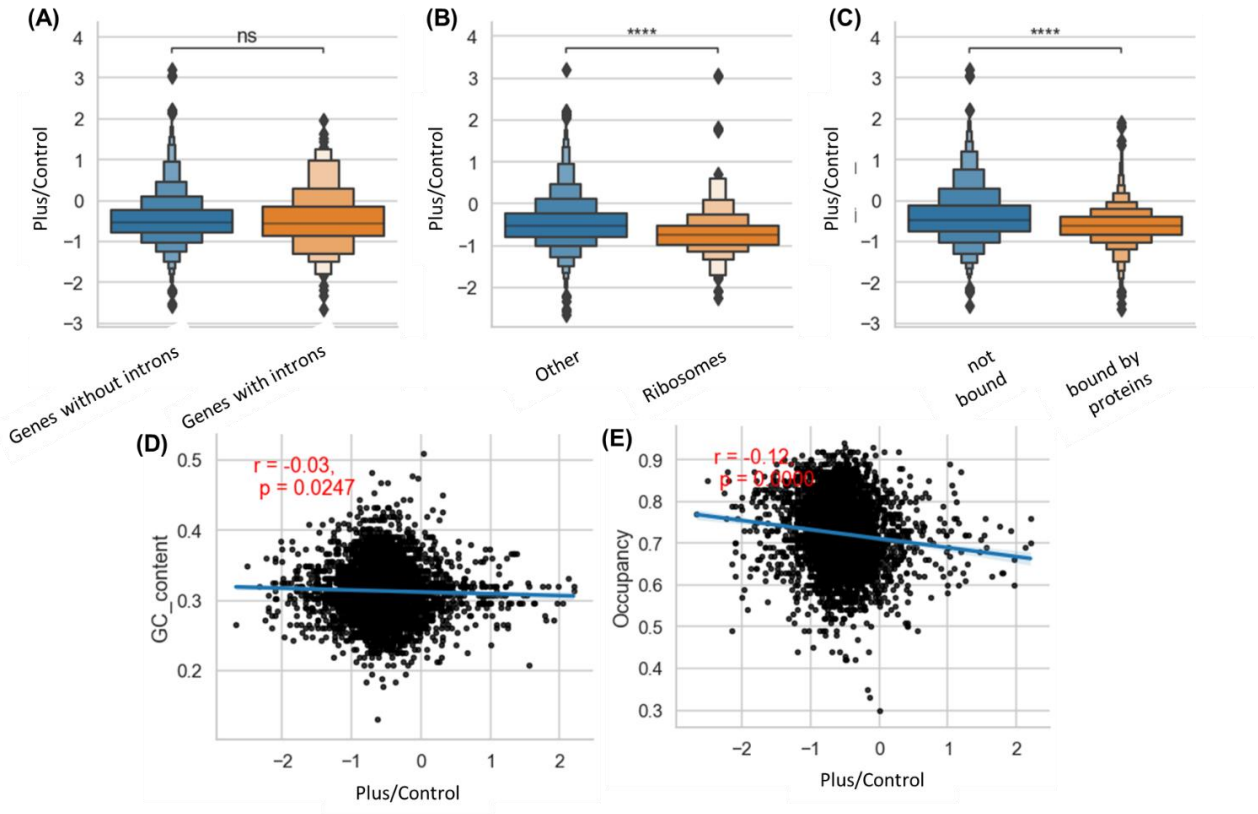


Figure 8. mRNA that encodes for ribosomes, and mRNAs that are bound by RNA binding proteins tend to be depleted from the VLPs.
(A) boxplot of intron containing genes (blue - without introns, orange - intron containing genes).
(B) boxplot of mRNA encoding for ribosomes (blue - not ribosomal, orange - ribosomal genes), pval < 0.001
(C) boxplot of genes that are bound by RNA binding proteins (blue - mRNA that are not bound, orange - mRNA that are bound), pval < 0.001
(D) GC content as a function of normalized pTy Fraction to Control, (E) ribosome occupancy as a function of normalized pTy Fraction to Control

8.2.3. Properties of encapsulated RNAs

In order to understand if some properties of RNA can dictate which will be encapsulated or depleted, I analyzed the results with respect to different parameters and traits of the genes; some that migh influence encapsulation, and some that might be affected by encapsulation (Figure 8). For example, if an intron containing gene is reverse transcribe and integrated into the genome, it will create an intron-less version of the gene (splicing occur early, before encapsulation). Thus, I hypothesized that the presence or absence of introns could be different between genes that are encapsulated or not. Nonetheless, there is no difference in encapsulation rate between genes that contain introns to genes that don't (Figure 8A). Other tested traits such as the GC content and ribosomes occupancy [64], which I hypothesized could influence the encapsulation, are slightly negatively correlated to encapsulation rate (Figure 8D, E). On the other hand, mRNAs that encodes for ribosomal proteins are depleted from the VLPs (Figure 8B). Another interesting characteristic of an mRNA is its tendency to be bound to RNA-binding proteins. If an mRNA is bound by a protein, it might not be free to be encapsulated. Indeed, Figure 8C shows that mRNA that are bound by RNA binding proteins [65] tend to be depleted from the VLPs.

8.2.4. Low evolution rate for non-encapsulated genes

A main hypothesis in this research is that genes that are being reverse transcribed can evolve faster. To test it, I used data on evolutionary rate [66]. In this paper, *S. cerevisiae* genes were divided into 5 groups – from low evolution rate to high evolution rate. Each group contains the same number of genes. Interestingly, even when using only 5 groups to describe evolution rate, low evolutionary rate genes tend to be depleted from VLPs (Figure 9). There is a significant difference between the low evolution rate group (in blue) and all other groups (except for the med-low group) in their tendency to be encapsulated. In contrast, the high evolution rate group is not significantly enriched in the VLPs. These results are consistent with the hypothesis that encapsulation of and RNA of a gene from the genome into the VLP might be followed in some cases by its reserve transcription and re-integration into the genome, thus accelerating its evolution through accumulation of mutations in transcription and in reverse transcription.

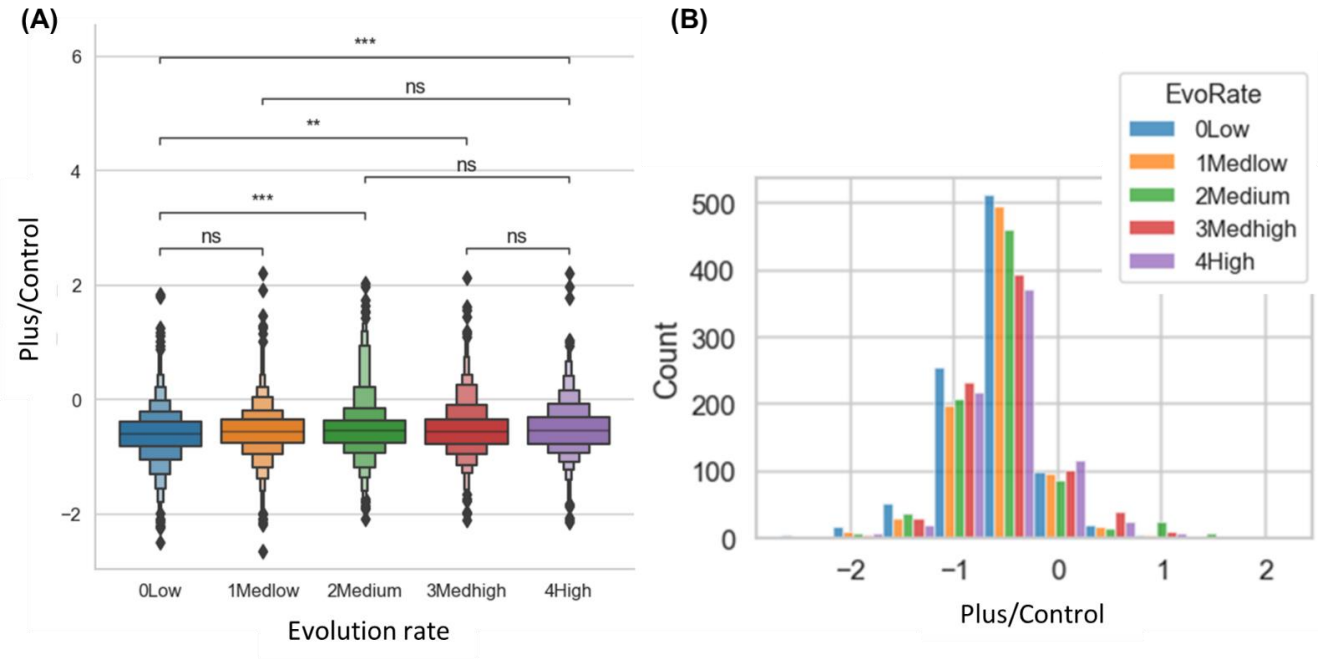


Figure 9. encapsulated genes tend to have higher evolution rate than depleted genes
(A) boxplots showing encapsulation rate (calculated by normalized pTy Fraction to control, y axis) for each evolution (x axis), pairwise t-test with multiple hypothesis correction (FDR) is indicated by star notation.
(B) A side-by-side histograms describing the encapsulation rate of genes with different evolution rate (different colors). a one-way ANOVA test shows a significant change between some of the groups.

8.3 Discussion

Prior works have looked on reverse transcription role in evolution by mediating genomic aberrations due to its large, and changeable copy number in the genome [58]. But in this work I examined for the first time the potential role of reverse transcription as an evolvability agent that if expressed in cells might serve as an agent for Lamarckian evolution. As described earlier, the two most distinctive characteristic of Lamarckian evolution is the inheritance of the phenotype, and the effect of the environment on the phenotype [8]. mRNA can be considered as part of the phenotype for two main reasons; first, it is not part of the DNA that is normally inherited, and second, more importantly, it varies in quantity and in quality based on the environment. Epigenetic inheritance, including the inheritance of RNA, is considered Lamarckian inheritance but one should note that it fades over time (and/or generations) [12], [67]. Thus, the RT is a special mode of Lamarckian evolution; mRNA is differentially expressed in different environments, but rather than being inherited as RNA and fade, RT re-write it into the genome to allow a more stable mode of Lamarckian inheritance and evolution.

For that, I have created a platform to study VLP encapsulation in our lab, including the creation of sucrose gradients, performing all biochemistry assays, and extraction of RNA.

As mentioned before, the RT could cause high mutagenesis on expressed genes only if other mRNAs (but the Ty mRNA) are being reverse transcribed in the VLP. It is known that some cytoplasmic mRNAs are being reverse transcribed [62] but the extent of the phenomena is not clear.

Maxwell *et al* [62] have extracted VLPs from *S. cerevisiae* cells and found using micro-arrays that the VLP contains mRNAs resulting from ~1500 genes. I have further analyzed Maxwell *et al.* microarray data and have noticed that the transcriptome consists of two distinct sets of genes (figure 10): one of which shows preferential enrichment within the VLP. Surprisingly, the rRNAs are found in the enriched subset; we hypothesized that it means that ribosomes and VLPs are found in the same fraction of the sucrose step gradient, and that translated mRNAs can be found in those fractions even though they are not encapsulated in the VLPs. The VLP extraction protocol used here avoids ribosomes in the relevant fractions, as shown in Figure 6.

I have repeated the same type of analysis reported by Maxwell et al. [62] yet making sure to reduce ribosomal contaminations as well significantly increase the results quality by using Illumina platform to sequence the mRNA found in the VLP (as well as the cytosol). Indeed, as shown in Figure 7, the ribosomes are not found to be enriched in my data, suggesting that no ribosome contamination exists.

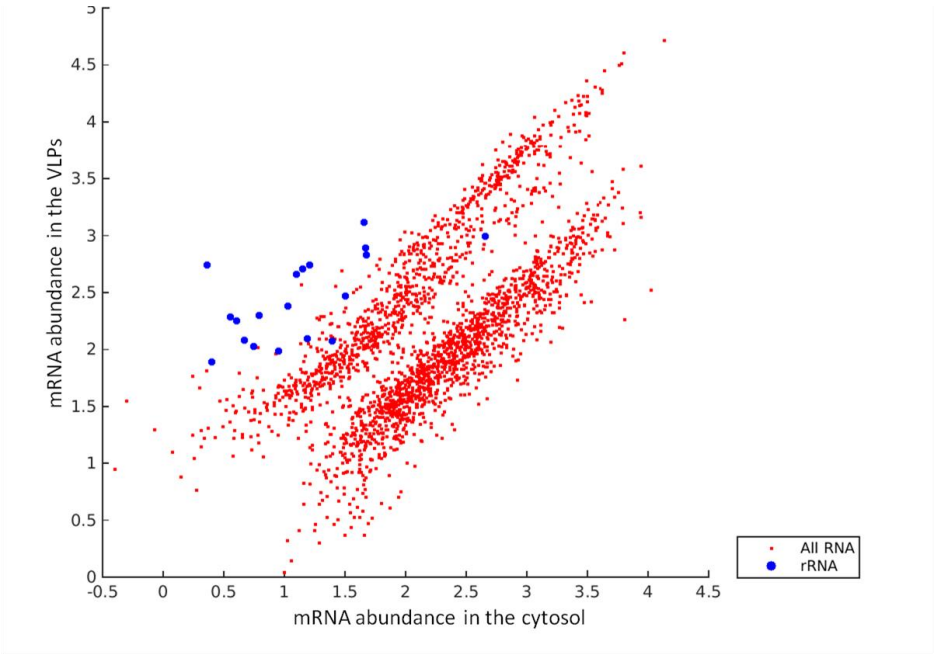


Figure 10. Analysis of microarray data shows two clouds
The analysis was done on data adapted from [62]. The upper cloud represents mRNAs that are encapsulated in the VLPs while the bottom cloud represents mRNAs that are depleted from the VLPs. In blue are rRNAs that are found mainly in the upper cloud.

A strong support to our hypothesis that RT can increase evolution is shown in figure 9 by showing that genes with low evolution rate are depleted from the VLPs. This work gives the first support for RT-mediated evolution of cellular mRNA.

In addition, I found that mRNA that are bound by at least one out of a collection of 35 RNA binding proteins (taken from [65]) are depleted from the VLPs (Figure 8C). This result is in line with what is known from the literature on encapsulation of mRNA in virion particles, and in the case of VLPs, that the encapsulation of mRNAs into the VLPs is mediated by their binding to the Gag protein [68]–[71]. Thus, it is reasonable to assume that mRNA that are already bound by other RNA proteins are less likely to be encapsulated in the VLP. Moreover, the fact that mRNA encoding for ribosomal proteins (Figure 8B) are depleted from the VLPs indicates that specific gene families can actively be avoided from the VLP, perhaps to avoid high mutation rate, or copy number increase in the genome.

This project has continued in the lab since I established this system and the above results were obtained by myself. In recent years this project was led by Yonat Gurvich, a postdoc in the lab, who conducted sucrose gradients and sequencing of both mRNA and cDNA from relevant fractions. The sequencing of cDNA from VLPs was done by Yonat for the first time in the lab. In addition, Yonat's used different conditions to study how the environment affects retrotransposition activity. More recently a M.Sc student, Ran Ashkenazi joined the project for bioinformatics analysis. Interestingly, their results show that some of the VLP encapsulated mRNAs are indeed converted into cDNA within the VLP. Further, they see an enrichment of telomeres' helicases in the RNA and cDNA extracted from VLPs. Nicely, telomeres' helicases gene resides in the Y' element region that was shown to be enriched at Maxwell *et a*l.

Yonat's data also show a distinction between RNA and cDNA encapsulated in the VLP for the first time. Mainly, Yonat's shows that mitochondria genes encoded in the nucleus are found in the VLPs as mRNA based on their basal expression rate, but are depleted from the cDNA data. This may suggest an active mechanism that reverse-transcribes specific genes inside the VLP (data not shown).

This project is currently continuing in the lab. We are now working on post-analysis of the data, examining characteristics that can promote encapsulation (using the same type of analysis used here to show that being bound by proteins reduces encapsulation), and also characteristics that can be influenced by reverse transcription (such as evolution rate of genes).

8.4 Materials and Methods

8.4.1. Strains, Plasmids, Primers and Media

Yeast strain RM11-a (*MATa leu2Δ0 ura3-Δ0 HO::kanMX*) which is a Ty-less strain was kindly sent to us by Pscale lesage's lab.

The plasmids that were used in this project;

pGal-Ty – a ~15kb plasmid that was kindly sent to us by Pascale Lesage's lab, the plasmid is described in [53]. The plasmid is a 2µ plasmid and it contains a URA3 marker and a Ty element under the regulation of a Gal promoter. The plasmid also contains a *HIS3* gene, with an artificial introns enabling quantification of transposition events (not used here).

pGal-control – based on the pGal-Ty plasmid. Using restriction free (RF) cloning the entire Ty element was deleted, resulting in a ~9kb plasmid. The primers used for the RF cloning are:

F-CCTGGCCCCACAAACCTTCAAATGAGAGCAATCCCGCAGTCTTCAGT;

R-ACTGAAGACTGCGGGATTGCTCTCATTTGAAGGTTTGTGGGGCCAGG.

The deletion was verified with both Sanger sequencing and Ty expression (data not shown).

Media used in this project: **SC-Ura –** media composed of nitrogen base, amino acid and 2% Glucose (according to [72]). **SC-Ura, Galactose/Raffinose –** media composed of nitrogen base and amino acid (according to [72]). Galactose (for the induction of Ty) or Raffinose (for inactivation of Ty) were added after autoclaving for a final concentration of 2%.

8.4.2. VLP extraction

Single colonies of RM11-a harboring the pGal-Ty or the pGal-control plasmid were picked from SC-ura glucose plates (approximately 10^8 cells), inoculated into 500 ml of liquid SC-ura+2% raffinose media, and shaken overnight at 30°c, Galactose was then added to a final concentration of 2% and the cultures were shaken for 72hr at 22°c until cell density reached ~$1*10^7$. Cells were harvested by centrifugation, washed with 10 ml of distilled sterile water, and re-suspended in 5ml of cold buffer B/Mg (10mM HEPES-KOH (pH 7.8) 15mM KCl, 3mM DTT, 10µg/ml aprotinin , 5mM $MgCl_2$). All subsequent steps were carried out on ice or at 4°C. Cells were lysed by adding 8g of cold, nitric acid-washed glass beads and vortexed at 4°C for 5 min intervals, alternating with 1 min incubations on ice repeated 4 times. Cell lysis was monitored by phase-contrast microscopy. Glass beads were separated from the lysate by puncturing the tubes and collecting the lysate into new tubes. The lysate was centrifuged at 4000rpm, 4°C, for 10min. The supernatant (approximately 8 ml) was layered onto a sucrose step gradient composed of 5ml of 70% sucrose in buffer B and containing 10 mM EDTA (buffer B/EDTA), 5ml of 30% sucrose in buffer B/EDTA and 20 ml of 20% sucrose in buffer B/EDTA in a Beckman SW28 polyallomer tube. The gradients were centrifuged for 3hr at 25,000 rpm, 4°C, and manually fractionated from the top (~1ml per fraction). OD260 was measured manually using Nanodrop on each of the fractions. Fractions were kept at -80 until further use. Additional assays (qRT-PCR, western analysis and RT-activity assay and RNA extraction and sequencing) were done on fractions 1,5,10,15, and 20 to end to identify the VLP-containing fractions.

8.4.3. RNA extraction and qRT-PCR

Total RNA was extracted from each of the fractions mentioned using Bio-Tri RNA reagent according to manufacture protocol and used as a template for quantitative RT–PCR using light cycler 480 SYBR I master (Biosystems)(LightCycler 480 system) according to the manufacture instructions. The absence of genomic DNA in RNA samples was checked by real-time PCR by using the RNA in the qRT-PCR. A blank (Nno template control) was also incorporated in each assay. The qRT-PCR was done using two sets of primers to identify levels of Ty's mRNA and rRNA. The Ty primers' sequences are: Ty-F-CGCTACACACGTCATCGACAT; Ty-R-GCGAGAATCATTCTTCTCATCACT; the rRNA primers are against the 18S subunit of the ribosomes are their sequences are: rRNA-F-TGGCGAACCAGGACTTTTAC; rRNA-R-CCGACCGTCCCTATTAATCAT.

8.4.4. Western analysis

20µl from each of the fractions were mixed with 60µl, 4X sample buffer and boiled for 10 minutes. 20µl of the boiled fraction was loaded on a 10% SDS-gel (lower gel: 4ml 30% acrylamide mix, 4.5 ml 1 M Tris pH=8.8, 120ul 10% SDS, 120ul 10% APS, 5ul TEMED, 3.2 ml DDW, upper gel: 1ml 30% acrylamide mix, 750 ul 1 M Tris pH=8.8, 120ul 10% SDS, 120ul 10% APS, 5ul TEMED, 4.1 ml DDW). Proteins were transferred onto nitrocellulose membrane using a semi-dry (BioRad) protocol. Membrane was then blocked for 1hr while shaking at room temperature in PBS-5% milk. First antibody (anti-VLP ab) was kindly sent to us by Jef Boeke's lab. Membrane was incubated in 1% milk-PBS + first antibody (1:10,000) in 4˚C overnight while shaking. 1hr incubation in room temperature was done for the secondary antibody, anti-rabbit-HRP (1:20000). Membrane was then stripped using DDW and NaCl (100mM). The stripped membranes were used again, for anti-ribosome antibody (anti-RPL1, dilution 1:2500) given to us by Jef Gerst's lab starting from the blocking step and continuing regularly the second antibody was the anti-rabbit-HRP (1:20000).

8.4.5. RT activity assay

A Retro-Sys kit by innovagen was used to quantify RT-activity of the fractions. 20µl from each of the fractions was used in the kit. Manufacture's protocol was followed with the following exceptions; incubation with the Alkaline phosphatase enzyme was done 3 times, every two hours, the 2-hour measurement was used in figure 6.

8.4.6. RNA sequencing and read analysis

tRNA sequencing protocol was adapted from Zheng et al., with minor modification.

3 VLP containing fractions, 2 control fractions and total RNA from either the pTy strain or the control strain were used for library preparation. Depletion of rRNA was done using RiboZero kit (Illumina, cat MRZY1324) by the manufacturer instructions. The entire volume after RiboZero (20ul) was used to continue. tRNA was uncharged using 100mM Tris-HCl, pH = 9 for 30 minutes in 37C, HCl was neutralized using 100mM NaAc (pH=4.8) followed by Ethanol precipitation. Zinc fragmentation (ife Technologies; 37002D) and Silane cleanup (Dynabeads® MyOne™ Silane, cat 370-02D) were done based on manufacturer instructions. Reverse transcription was done using TGIRT™-III Enzyme (InGex, LLC) with primers for library preparation. Primers are DNA-RNA hybrids. After Silane beads cleanup a 3' adaptor was ligated to the cDNA using T4 ligase (NEB; M0202S). The library was amplified using NEBNext PCR mix and cleaned using SPRI-beads. Samples were pooled and sequenced using a 75bp single read output run on MiniSeq high output reagent kit.

Read were trimmed using homerTool12. Alignment to the genome and mature tRNAs gene sequence was done using Bowtie2 with parameters --very-sensitive-local. Reads aligned with equal alignment score to the S288C genome. Read count was done using BedTools-coverage count. Read counts were normalized to library size and used for the analysis.

9. Mate choice and fitness inheritance

9.1 Introduction

in Evolthon, The strategy that made the fittest yeast strain, was mating with natural isolates [33]. Also, the 2nd and 3rd ranking strategies in yeast involved mating. Thus, In the following part of my PhD I decided to focus on Yeast mating and examined the effect of mate choice and parental fitness on the fitness of the next generation.

Sexual reproduction and is wide spread in nature [73]. Even among prokaryote, that do not reproduce sexually, one can find other method of genetic exchange (also known as horizontal gene transfer) between individuals, such as the uptake of free DNA in *Bacillus subtilis*, conjugation, and transduction [74]–[79]. Although many species of eukaryotes can reproduce asexually (especially in plants and fungi), most eukaryotes (mainly multicellular eukaryotes) reproduce sexually [80]–[85].

Although sexual reproduction has started hundreds of millions of years ago [82], the evolutionary forces that made it so wide spread are not entirely understood. Many works have studied the benefits of sex, mostly considering clonal interference (i.e., combining beneficial mutation from different genomes together, and rescuing of beneficial mutation from a disruptive background) [86]–[89].

In sexual reproduction, offspring fitness might depend on both the fitness of their parents and on the genetic distance (GD) between them. However, the exact genes that determine fitness and the mode of inheritance of fitness as a quantitative trait, are not clear, as is the case of many other multi-loci traits [90], [91]. Fitness is likely to correlate with parents' fitness, as fitter parents will have fitter offspring, and *vice versa*. In addition to the fitness of the parents, the genetic distance contributes to fitness via either heterosis or incompatibility [92]–[101]. In heterosis, parents' genes complement each other in the offspring, and might even produce a hybrid that is fitter than any of its two parents (a phenomenon also called "hybrid vigor"), probably via additive or synergistic effects [96]–[101]. Genetic incompatibility happens when parents' genes have a destructive effect [92]–[95]. For example, Wei and Zhang showed that parents with an intermediate level GD had the offspring with the highest fitness in yeast (*Saccharomyces cerevisiae*), plants (*Arabidopsis thaliana*), and in animals (*Mus musculu)* [102].

As parents influence vastly on offspring's fitness, many organisms developed mechanisms to choose their mating partner [103]–[106]. Evidence suggests that on top of choosing partners based on their fitness, organisms assess GD between them and partners and select accordingly, perhaps avoiding too close or too far mates [107]–[111]. Though genetically based mate choice was shown in several organisms,

including plants [103], mice [112] and humans [110], [113], a systematic dissection at significantly bigger numbers of parental strains and mating option combinations is very much needed.

The yeast *S. cerevisiae* is a great model organism for the study of mating. *S.* cerevisiae has both sexual and asexual life cycles. In the asexual life style, a diploid or a haploid cell mitotically divide into two identical cells. In its sexual reproduction process, two haploids cells from different mating types (a and α) mate through a well-studied process mediated mainly by pheromones and their binding to a specific receptor on the surface of the opposite mating type cell [114]. In addition, yeast can be easily manipulated with many genetic tools, and large collections of natural strains exist [115], [116].

We initiated a collaboration with Prof. Gianni Liti from IRCAN, a world leader in yeast population genomics and phylogeny who collected and characterized numerous wild isolates [115]. Prof. Liti shared with us his recent collection of 1011 strains of *S. cerevisiae* collected from different geographic locations, different niches, wild or domesticated etc. in addition, each strain is characterized by a set of genetic (ploidy, aneuploidy, zygosity, etc.) and phenotypic traits (fitness in different conditions, sporulation efficiency).

By using this collection of strains, I conducted a study with thousands of variants, more than was ever done in the field of mating choice and fitness inheritance [97], [102].

In this project I aimed to study different properties of mate choice in yeast as well as fitness inheritance. More specifically, I asked if different strains of *S. cerevisiae* mate in the same efficiency, and which characteristics contribute to it. In addition, I wanted to study if yeast can choose mating partners, meaning, that if a yeast strain has multiple strains surrounding it, will it mate with each strain equally, or will mate more with specific strains. If mating is not performed uniformly in this scenario, which properties influence it? Lastly, I asked how offspring relate to their parents; are they similar to them? better or worse? Are they more similar to one of the parents?

9.2 Results

9.2.1. An experimental platform to study mating choice and fitness inheritance in yeast on a massive scale

Strains choice

Strains for this project were chosen from a collection of ~1000 *S.cerevisiae* strains isolated from diverse ecological niches all over the world (described in Peter *et al*. [115]). From those, I chose 200 strains to perform genetic engineering. Strains were chosen based on the following parameters; Euploid diploids, intact HO and efficiency sporulation rate were all chosen to allow easier manipulation of strains during

the genetic engineering steps. Euploid diploids were chosen to reduce difficulties of mating efficiency, an intact HO was chosen to allow transformation into the HO in the genetic engineering part. Efficient sporulation is needed to allow sporulation after transformation. In addition to the technical criteria above, three parameters were optimized to allow enough statistical power to address the roles of GD and parents' fitness on offspring as well as on mate choice. Yeast usually have two opposing metabolic preferences for energy extraction, fermentation or respiration. While yeast that live on Glucose ferments, different carbon sources such as Glycerol or Ethanol cause yeast to respirate aerobically. Using Peter *et al* data, we observed a negative correlation between strains fitness in Glucose (YPD) and Ethanol (YPEthanol) (shown in Figure 11A). Since one hypothesis is that cells choose mating partners based on fitness, using the two carbon sources and revealing if mate choice changes accordingly can shed light on the relationship between fitness and mate choice. Strains were chosen to represent the following groups, either high fitness on glucose and low on Ethanol, high fitness on Ethanol and low on glucose or intermediate fitness on both. A few strains with low fitness on both media were added. A second criterion for strain choice was low heterozygosity of the strains. The genetic distance between strains was determined by the data of Peter *et al*, and as strains had to be sporulated in the construction process high heterozygosity strains will produce haploid cells that differ from the original strains and thus deflect the genetic distance matrix I obtained. In addition, strains were chosen to represent the original distribution of GD as in the entire collection. As explained in the introduction, organisms tend to choose mate partners to avoid too far or close partners. Thus, a main hypothesis was that yeast choose mating partners based on GD. To that end, it was important to have as many genetic distances as possible between pairs of strains. Furthermore, strains can be binned in three bins: low, intermediate and high GD to allow different statistical analysis methods (Figure 11, left panel).

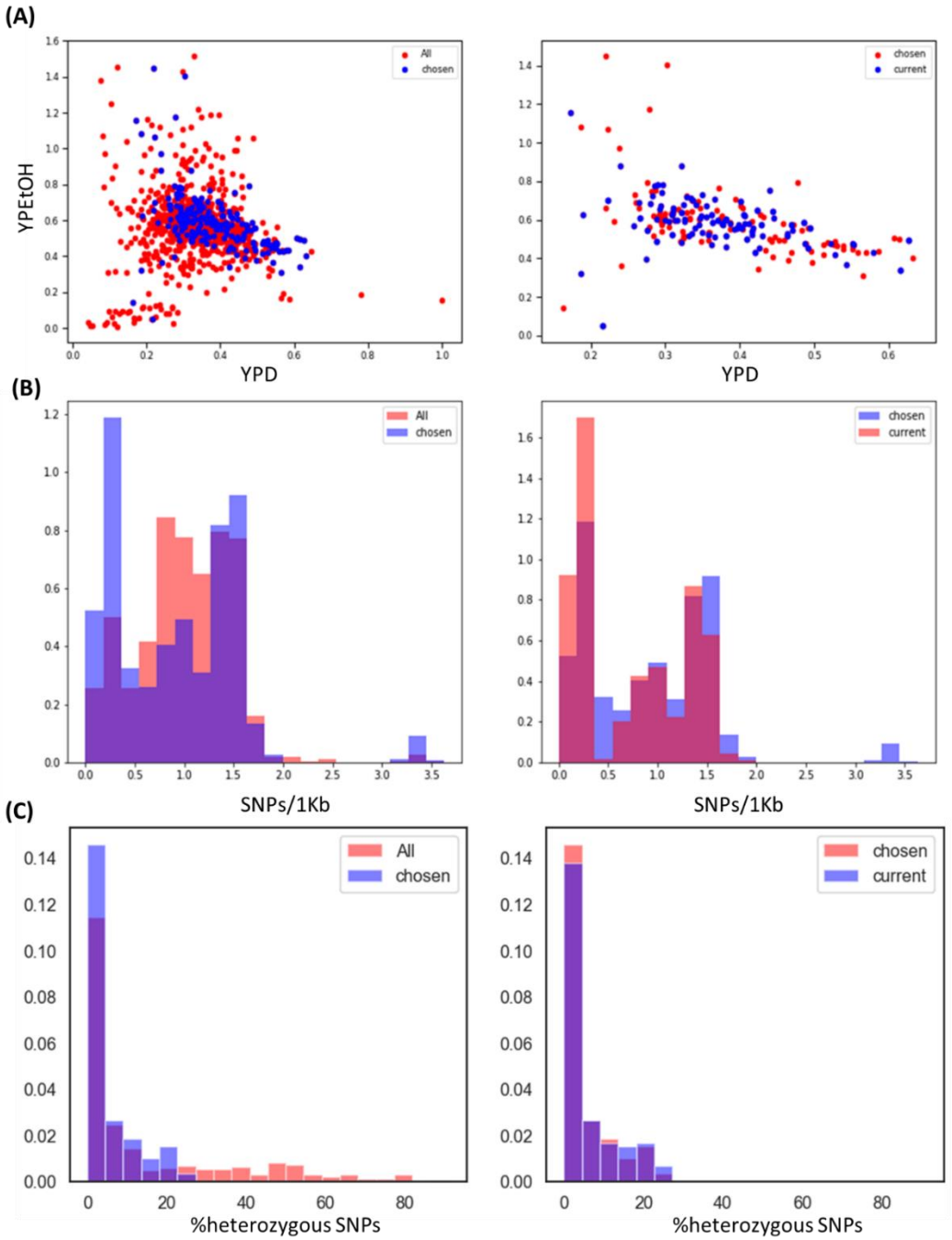


Figure 11 . Engineered strains vs all 1000 strains properties

(A) Fitness of strains in YPD (x-axis) and YPEthanol (y-axis) based on liti 2018 (3) paper. Left - Red dots represent the entire collection, while the blue dots represent the strain chosen for engineering strains, right - Red dots represent the chosen for engineering strains, while the blue dots represent the verified strains after cloning

(B) SNPs distribution in the collection. Left - Red represent the entire collection, while the blue represent the chosen for engineering strains, right - Red represent the chosen for engineering strains, while the blue represent the verified strains after cloning

(C) Heterozygosity level distribution in the collection. Left - Red represent the entire collection, while the blue represent the chosen for engineering strains, right - Red represent the chosen for engineering strains, while the blue represent the verified strains after cloning

Strains' construction

Massive genetic engineering had to take place to be able to work with natural isolates in this complex and high throughput experiments. As original strains were diploids, I had to sporulate them and disabled mating type switching [117] by knocking out the HO. In addition, I added constitutive markers to allow selection as well as detection in FACS analyzer of the different mating types, and the offspring. "Magic Markers" (adapted from [118]) were added as well, enabling the selection of a specific mating type after sporulation. Lastly a sophisticated system enabling fusion of barcodes originating from different chromosomes was added as well to allow identification of offspring after *en masse* mating. Below is an explanation of each of the above requirements. The full construct is shown in Figure 12 and it includes the following regions: (i) genome homology region: 500bp homology to the HO locus on both ends of the construct (Figure 12A). The specific sequences of homology were chosen to have the least amount of SNPs between most strains, thus enabling favorable conditions for genome integration. Integration to the HO and by knocking it out is essential to achieve stable haploids, as needed for further steps in the project (ii) Barcode Fusion Genetic (BFG): (Figure 12B, Figure 13). This region was kindly given to us by the lab of Fredrick Roth [119]. This region enables the identification of the two parents of the hybrids. Each parental strain is labeled with two barcodes, flanked by a loxP/lox2272 sites. Activating Cre recombinase in the offspring diploid (with tetracycline antibiotic), fuses the parental barcodes, resulting in one linear fragments containing one barcode originating from parent MatA and one barcode of parent Matα that could be identified using Next Generation Sequencing (NGS) (Also see Figure 13) (iii) Constitutive markers and fluorescence proteins (Figure 12C). Constitutive markers enable the selection of diploids after transformation. Diploids that received the MatA design are selected using Hygromycin (Hyg) and further validated by FACS for the presence of GFP. Diploids that received the Matα design are selected using Nourseothricin (Nat) and further validated FACS for the presence of mCherry. (iv) Haploid selecting markers (Figure 12D). After transformation the strains are sporulated, and haploids that contain either MatA or Matα designs are selected by Zeocin or Geneticin (G418), respectively. Since resistance cassettes are under the control of mating type specific promoters; Ste2 promoter is active in MatA cells only, while Ste3 promoter is active in Matα cells only, this part of the construct allows the selection of the desired haploids.

To achieve final strains, strains were transformed, selected on antibiotic and verified by FACS, and then sporulated. After sporulation, haploid strains were selected and eventually verified by Sanger sequencing. At the end of this process from 200 initial diploid strains chosen as described in the section "strain choice" I ended up with 102 MatA engineered strains, and 55 Matα engineered strains. The verified final strain

set has similar properties to the chosen strain set, without any dramatic effect on the collection fitness, GD or heterozygosity (Figure 11, right panel, Table S1).

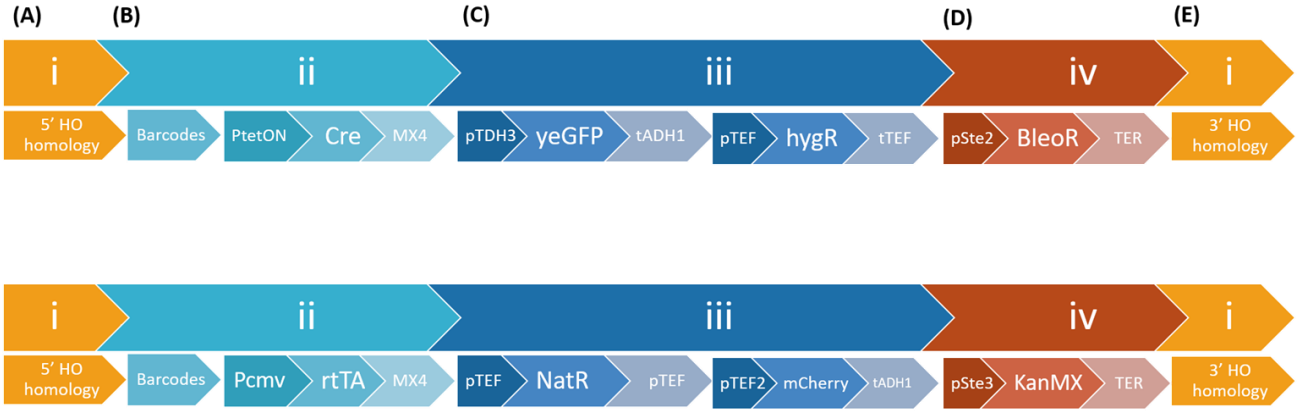


Figure 12 . Design maps that used for strains' engineering

Top panel – construct design for creating matA strains, bottom panel – construct design for creating matα strains

(A) HO homology region that was used for homologous recombination to the HO locus after transformation

(B) The Barcode Fusion Genetics system. Composed of the barcodes and *lox* sequences (see maps in Figure 14) and the Tet-ON system. matA strains contains the Cre enzyme that is regulated by the rtTA inducer that is found on the matα design.

(C) Constitutive markers. matA contains yeGFP and Hyg resistance cassettes, while matα contains mCherry and NAT resistance cassette

(D) "magic marker" haploids mating type specific markers. matA cells induce the BleoR resistant markers and thus can grow on Zeocin, while matα strains induce the KanMX resistance marker thus can grow on Kanamycin (G418)

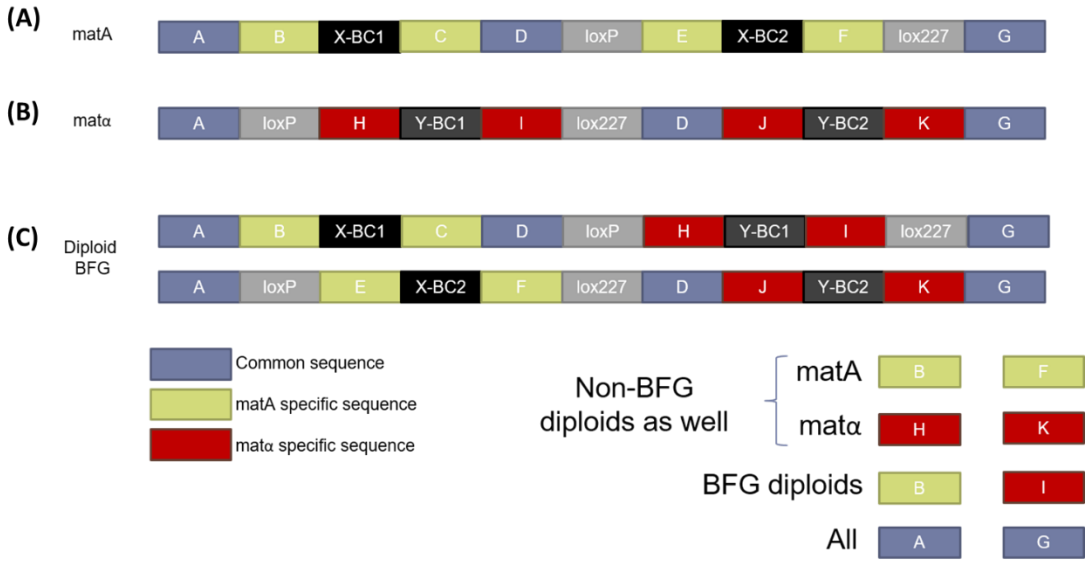


Figure 13 . Zoom in on barcodes region from the design

In this map, each letter corresponds to ~25nt length sequence. Same letter in different panels means that the same sequence is present in both regions. (A) matA design, (B) matα design, (C) the two fragments created in the offspring after fusion of the barcodes. Different combinations of sequences can be used to amplify specific regions only for NGS.

9.2.2. Outline of the project

This project addressed a few key questions in mating choice and fitness inheritance, summarized in Figure 14.

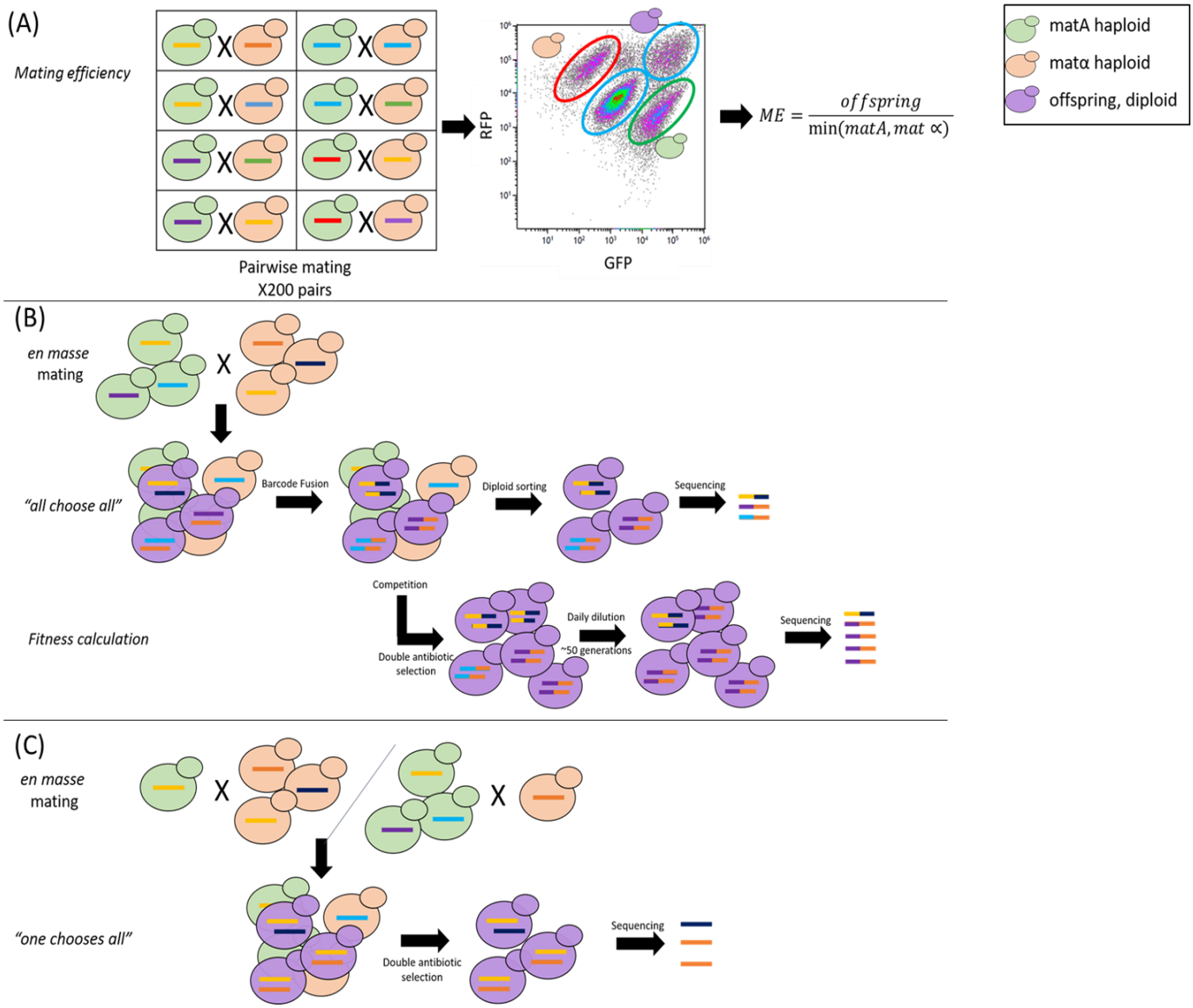


Figure 14 . Schematic representation of research questions

(A) All strains from both mating types (matA and matα) were mixed together and allowed to mate. After mating, culture were either sorted for offspring (diploid) cells to measure preference of the strains toward each other (top), or culture was diluted daily in a diploid selecting media to allow competition of the offspring (bottom). At the end of the experiments offspring identity was determined by NGS

(B) A different approach to measure mate preference of a single strain toward all other. In that case, one cell was inoculated with all other cells of the opposite mating type. After mating, culture was continued to grow on double antibiotic to select for offspring. after ~15 generations mate choice was determined by NGS

(C) Measurement of mating efficiency. In this experiment cells were mated in pairs, such that the frequency of offspring (as can be determined by FACS analysis) represents the mating efficiency of the pair. Give short explanation to the FACS image

(1) Are all strains able to mate with each other? Do some pairs mate in higher efficiency than others? Measuring the offspring fraction after pairwise mating reveals the mating efficiency of each pair of strains (Figure 14A). To explore this question I mated pairs of strains and measured the offspring fraction using FACS analyzer taking advantage of the fact that each of the two haploid parents (mat A and mat alpha) were marked with different markers.

(2) How is offspring fitness related to parental fitness and to other parental characteristics? Does offspring fitness depend on genetic distance (GD) between the two parents? To answer those questions, I performed *en masse* mating, followed by a pooled competition of all offspring to measure their fitness (Figure 14B).

(3) Do yeast choose their mating partners? Are some strains more attractive than others? What modulates mating choice in yeast, if any? This question further divides into two hypothetical questions which were addressed in two different experiments; (i) "one chooses all", one strain (of matA or mat α) is given a choice of many other strains of the reciprocal mating type to mate with (Figure 14C) and (ii) "all choose all", where all strains from both mating types are inoculated together to allow mutual choices of both mating types (Figure 14B).

As explained before, yeast lifestyle depends on the carbon source, while they ferment on Glucose, and respirate on Glycerol. Figure 11A shows a negative correlation between fitness on Glucose and on Glycerol. To understand if fitness inheritance and mate choice are dependent upon carbon source, and thus metabolic preference, I chose to perform most experiments on both Glucose and Glycerol.

9.2.3. Diploid fitness correlates with parents' fitness

Fitness is the ultimate evolutionary trait that integrates over many traits that an organism has, of which some are quantitative and others are not. As fitness is measured here, growth rate of a strain at in a competition with other strains, it is a quantitative trait whose inheritance can be complex. An interesting question in biology is how fitness is inherit. In particular, given the fitness of two parents at a given condition, can we predict the fitness of their offspring? The inheritance of other quantitative traits in biology reveals several models. For example, the inheritance of height in human show a simple trend in which offsprings' height correlate with the average height of parents [120], [121]. Other modes of inheritance can be that offspring will correlate with the minimal value of a trait among its two parents, or the maximal of the two. More complicated models can be envisaged too.

To reveal the nature of inheritance of fitness as a quantitative trait in yeast I have measured fitness of each parental haploid strain, and of each offspring. To measure offspring fitness, all parental haploid strains from both mating types were mixed and mated *en masse*. After mating, cells continued growing

in a daily dilution manner on double antibiotics (Hyg + NAT) to eliminate haploid parents. This allowed selecting for diploid offspring only. Offspring were competed for 50 generations, to allow their relative fitness estimation. Offspring frequency in the population changed over time according to their relative fitness. In addition to the *en masse* mating and offspring competition, haploids competition was also conducted (separately for matA and matα). As mentioned, mating and competition experiments were done on media containing one of two carbon sources Glucose or Glycerol. Fitness was assigned to each variant (offspring from diploids competition with double antibiotics and parents from haploids competitions) based on pooled competition results [46], [47]. Fitness is derived from the following equation: $f(\mathrm{t}) = f(\mathrm{anc}) \cdot (1+s)^t \approx f(\mathrm{anc}) \cdot e^{st}$,which define that the frequency of a strain dependent upon its initial frequency and its relative fitness (f = frequency of a strain at a given time point, t = time, anc = strain in time point zero, s = selection coefficient). To derive fitness, I used the algorithm described in Levy *et a*l. [47] using maximum likelihood.

With fitness measurements of parents and offspring I could ask if offspring's' fitness correlates with a simple function of parents' fitness. In order to explore that, the correlation between offspring fitness and that of their haploid parents is shown in figure 15. A weak but significant correlation is observed between diploids fitness and the average, maximal and minimal fitness of the parents in the two media examined (Figure 15).

These correlations suggest that offspring fitness is not a simple function of its parents' fitness, but much more complex than that.

9.2.4. Fitness of offspring is maximized when parents are genetically distant from one another

I next asked if fitness inheritance is dependent upon GD between parents. I binned all offspring based on the GD of their parents into 4 bins and looked at the offspring fitness value distributions in each GD bin. Figure 16 shows that offspring fitness is higher in high GD, in both glucose and glycerol, yet more so in Glycerol media. These results show the advantage of outbreeding in this yeast collection. One potential reason for the lower fitness of offspring whose parents are close genetically is the homozygosity of recessive traits that might be avoided when parents are less similar genetically. Offspring fitness resembles the minimum, maximum and average fitness of its parents in all bins of GD as in the entire data (data not shown), indicating that offspring takes after their parents in all GD.

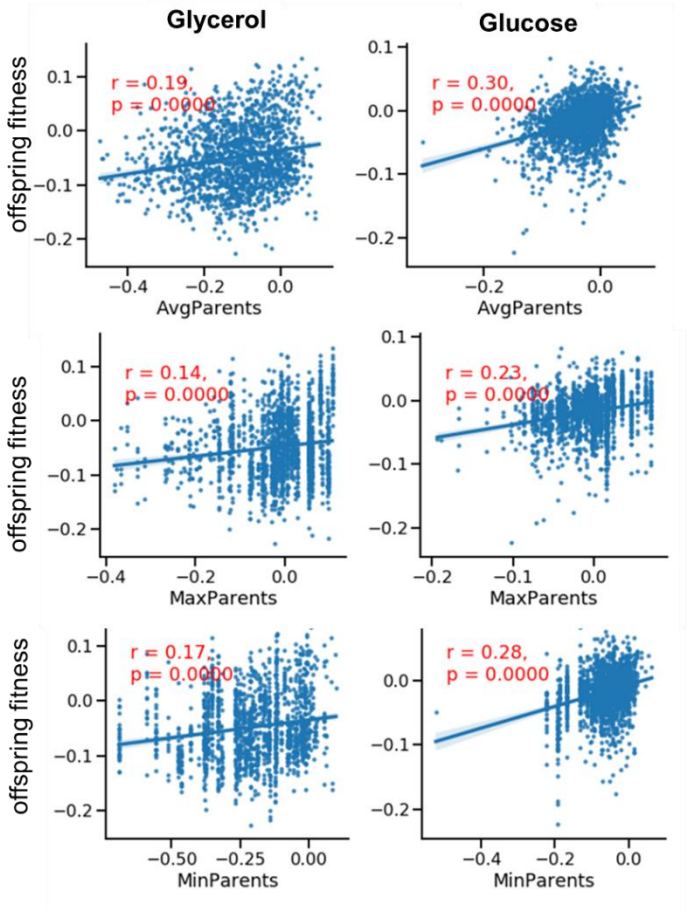


Figure 15. offspring fitness is correlated to parents fitness
Left – experiments done on Glycerol media, right – experiments done on Glucose media. Each doe represent an offspring. The y-axis is the average fitness of the offspring fitness across 6 repetitions. The y-axis is (A) parents average fitness, (B) parents maximum fitness (C) parents minimum fitness.

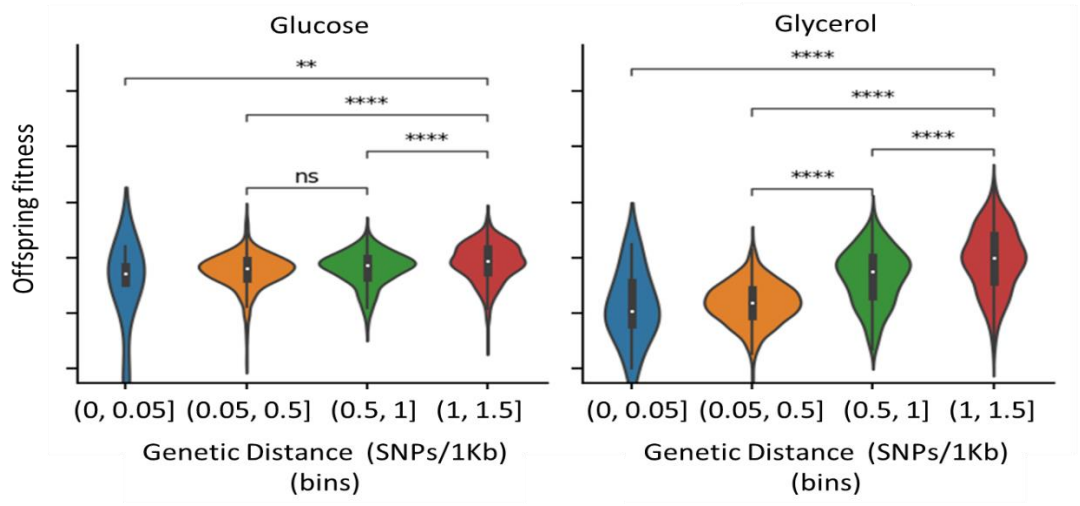


Figure 16 . offspring fitness is higher in high GD
Data was binned into 4 bins based on GD between parents – 0<GD≤0.05, 0.05<GD≤0.5, 0.5<GD≤1, 1<GD. Violin plot of offspring fitness values in each bin of GD are presented.
ns: P > 0.05, *: P ≤ 0.05, **: P ≤ 0.01,***: P ≤ 0.001, **** :P ≤ 0.0001 (For the last two choices only)

9.2.5. Mating efficiency vary between different pairs

Can any two strains mate, and is mating efficiency similar between different pairs? To study mating efficiency, I have conducted a pairwise mating experiment, in which each pair of strains is allowed to mate in isolation from other strains. In these experiments, mating efficiency between pairs of strains was measured by FACS analysis of fluorescent markers (Figure 14A). Mating efficiency is defined by the tendency of each pair to mate without having other potential partners in the environment. As described earlier, all haploids matA are labeled with yeGFP while all matα strains are labeled with mCherry, thus offspring are labeled with both yeGFP and mCherry, and can be detected in FACS analyzer by double positive fluorescence signal. Mating efficiency was calculated based on this equation: $\#offspring/min(\#matA, \#mat\alpha)$. Table S2 shows all pairs that were measured. Figure 17A shows that mating efficiency varies; while some pairs of strains almost do not mate (~0.2% mating efficiency), others mate well (close to 100% mating efficiency). I could thus ask if mating efficiency depends on the GD between the parents. Interesting, I found that mating efficiency does depend upon GD of the mated strains. In particular, I found that intermediate GD maximize mating efficiency (Figure 17B).

Mating in yeast depends upon the secretion and the recognition of pheromones from one cell to its opposing mating type partner [114]. If pheromones are incompatible, maybe due to high GD between partners, cells will mate poorly, or might not mate at all. Figure 17B might suggest yeast are very sensitive to high GD and thus mating efficiency is reduced in the high GD (>1 SNPs/1kb) bin.

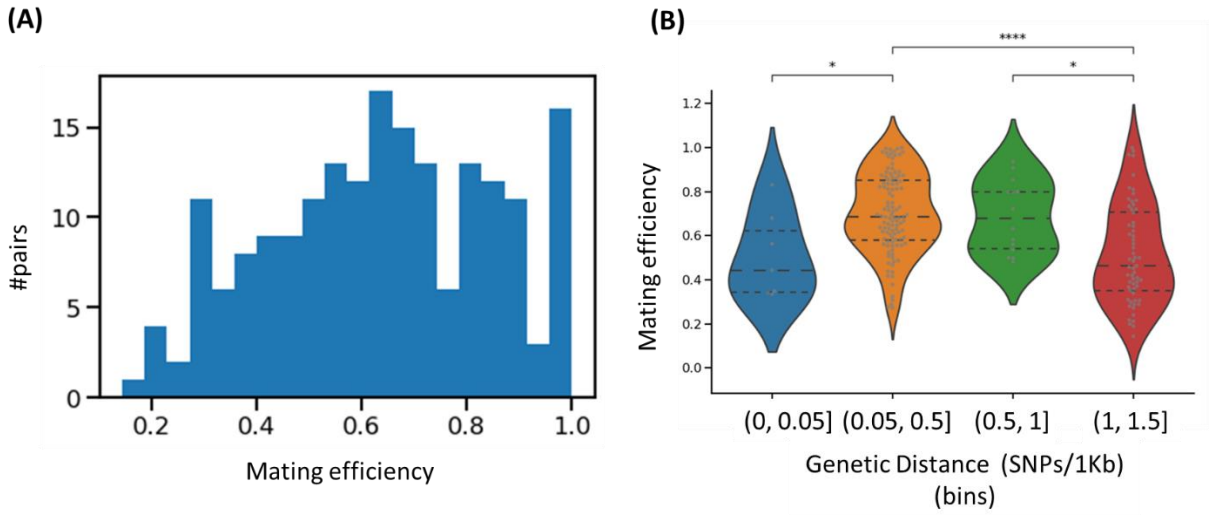


Figure 17 . Mating efficiency is high in intermediate genetic distances

(A) Mating efficiency distribution across all 200 pairs that were tested

(B) Violin plots of mating efficiency values in each Genetic distance bin: 0<GD≤0.05, 0.05<GD≤0.5, 0.5<GD≤1, 1<GD. ns: P > 0.05, *: P ≤ 0.05, **: P ≤ 0.01,***: P ≤ 0.001, **** :P ≤ 0.0001

9.2.6. BFG efficiency vary between strains, but can be predicted per pair by the parents' average BFG efficiency

In order to detect each offspring, yeast were transformed with a sophisticated system enabling the fusion of barcodes initially found on different chromosomes (in this case, the two parental chromosomes). Thus, the barcode fusion (BFG) efficiency is an essential part of detecting mate choice. To calculate BFG efficiency, 16 matA strains and 12 matα strains were chosen (matA: AKG, BLV, BSD, CDQ, BKM, BQC, BTH, CFE, BHC, BMA, CCG, SACE-YAM, BMH, BPM, CDF, matα: AKP, AIM, BHB, BQG, BNK, BPQ, BQI, BRD, BMA). Pairwise mating between all of them (192 pairs) was done. After pairwise mating, offspring were selected by double antibiotics and sequenced to find BFG efficiency (see "Materials and Methods", BFG efficiency part for more details). BFG efficiency values varied across different pairs of strains (Figure 18A). In addition, the average BFG efficiency of strains also varies (Figure 18B).

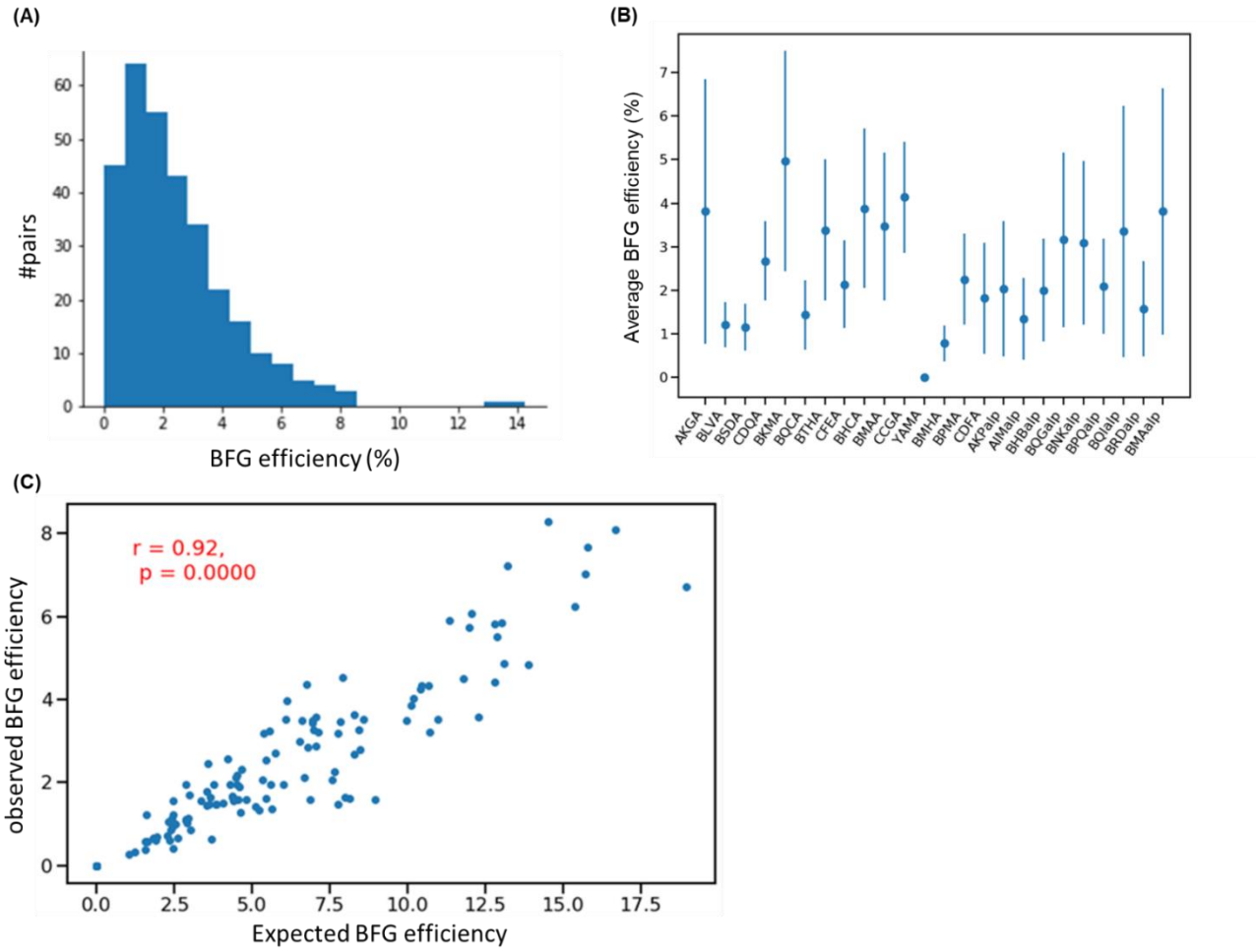


Figure 18 . BFG efficiency varies between strains

(A) Histogram of BFG efficiency across different pairs of strains

(B) Average BFG efficiency per strain (both matA and matα), error bars represent the standard deviation of a strain

(C) Correlation between the expected BFG efficiency (parents' average BFG efficiency multiplication) and observed BFG efficiency as calculated by NGS (pearson R=0.92, p-val < 0.005)

These results indicate that the ability to induce the recombinase and or complete the fusion event differ between different genetic backgrounds. Since BFG is instrumental for my ability to identify the identity of diploids I needed to check whether this efficiency can be estimated for all possible pairs. For that I did the following analysis- Based on parents' average BFG efficiency, I calculated for each pair its expected BFG efficiency, I then examined the correlation between the calculated efficiency and the observed one as measured in my experiment. The correlation between expected efficiency and calculated efficiency is very high, and significant (Figure 18C). This result suggests that BFG efficiency is not unique per pair of strains, but rather unique for each haploid strain, and that the offspring BFG efficiency is solely a simple multiplication between the parents.

9.2.7. Some strains are preferred by many strains

To assess mating preference, i.e., which strains are preferred when more than one choice is present, a couple of experiments were made. In the first experiment, termed "one chooses all", one strain (either matA or matα type) was mixed with a pool of many of the reciprocal mating type strains to allow mating. I performed many different combinations of the "one chooses all" experiments, each with a different chooser. Also, while in most of the cases, choosers were mixed with strains only from a given yeast phylogenetic clade (clade #I as defined in Peter *et al*. [115]) in some experiments I also allowed a mixture of potential mates in which partners were also available outside that clade. The focus on Clade I in this experiment was due to the fact that most strains in my collection are from clade I. Furthermore, I wanted to study if mate choice in this context rely on the fitness of the strains rather than on GD; by letting strains choose from members from Clade I, I controlled for genetic distance (since strains in this clade span smaller GD range compare to the entire set of strains).

After mating, cultures were sequenced to undercover the choices each "chooser" strain made. For each chooser strain, I calculated which is his preferred strain i.e., the strain it mated the most with. Figure 19 counts for each preferred strain, how many chooser strains mated with it the most. As seen, some strains, such as BRD1 (matα) and BLP1 (matA) were preferred by many strains (25 or 18, respectively), while some other strains were not preferred by any.

This experiment shows that yeast prefer mating with specific strains, more than they are with other strains. It also suggests that some strains of yeast are more "popular", having a lot of other strains mate with them. The underlying reasons for preference is yet unknown, since it is not explained by mating efficiency (although mating efficiency was measured for 200 pairs only) nor fitness (data not shown).

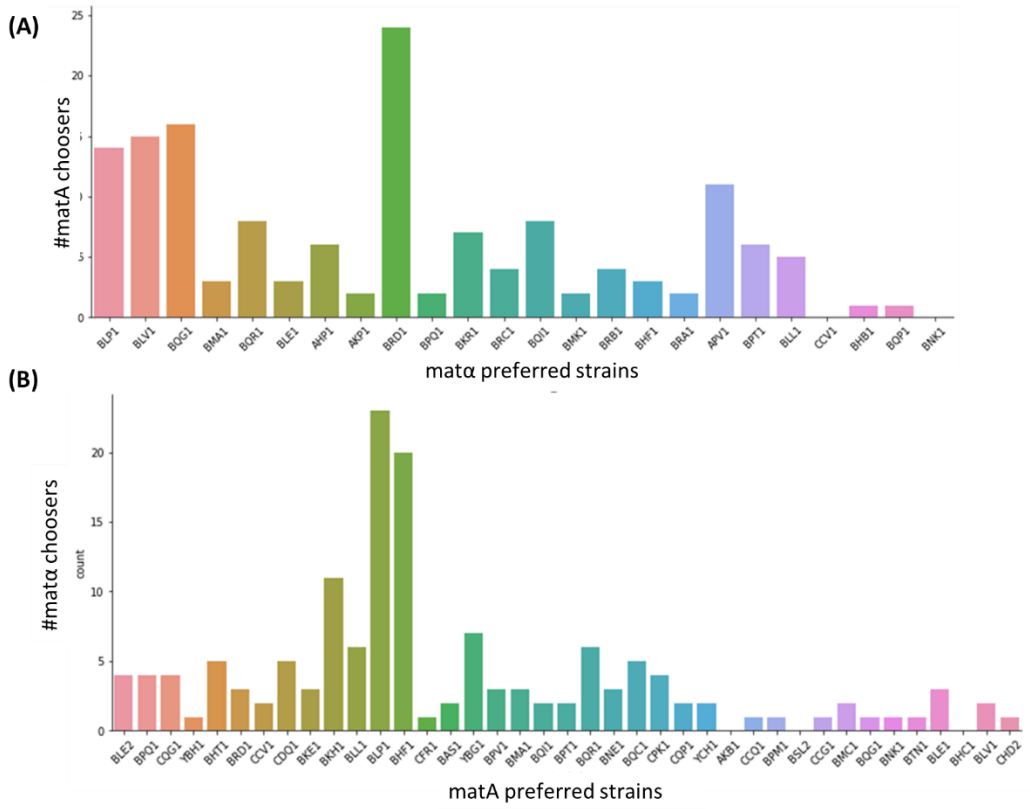


Figure 19 . preferred strains in "one-chooses-all" experiments
The X axis shows strain that were preferred by at least one strain, the Y axis shows number of chooser strains that preferred that strain as their first choice . In (A) chooser strain was matA strain, and candidates were matα strains, while in (B) chooser were matα strains and candidates were matA

9.2.8. Preferred strains vary when many choosers exist

In addition to the "one chooses all" experiment, an "all choose all" experiment was performed. In this case all of the verified strains from both matA and matα were mixed and allowed to mate *en masse.* After mating, cells were sorted based on double fluorescence (As described earlier, all haploids matA are labeled with yeGFP while all matα strains are labeled with mCherry, thus diploid offspring are labeled with both yeGFP and mCherry, and can be sorted based on double positive fluorescence signal) instead of imposing a selection period to avoid fitness affects. After sorting, barcodes were sequenced to recover which diploids have been formed. Interestingly, there is no correlation between choices in the "one chooses all" and in the "all choose all" for most strains (Data not shown). This result may indicate that strain choice itself varies when other strains are present, or that strains cannot choose its favorite partner in the "all against all" setting.

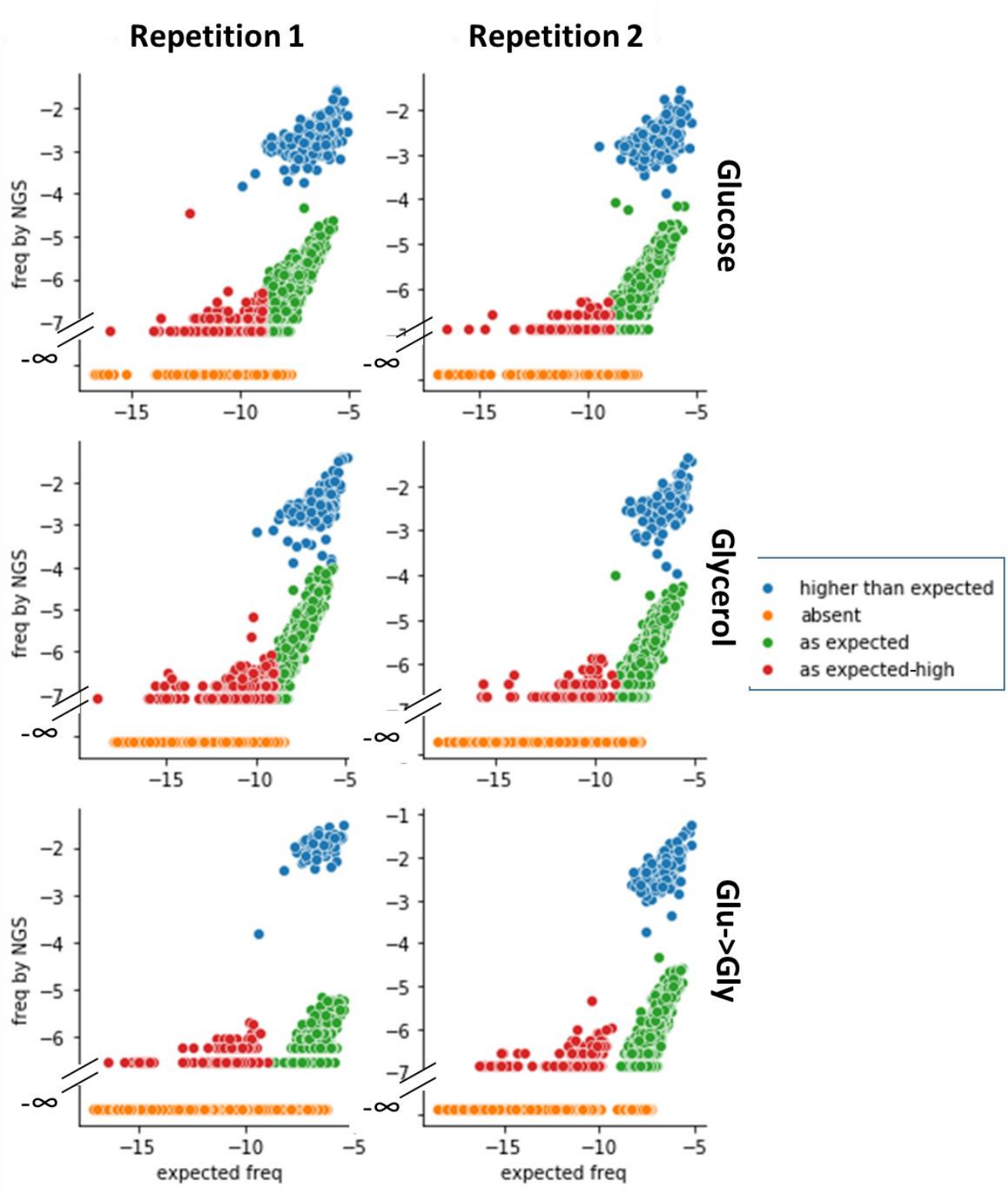


Figure 20 "All-against-all" experiment reveals different groups of offspring

Experiments were done in three conditions (media with Glucose in upper panel, media with Glycerol in intermediate panel and Glu->Gly in bottom panel) in two repetitions (left and right). The axis represents offspring frequency as measured by NGS (y-axis) and expected frequency as calculated based on the activity of each parental strain (x-axis). Most of the offspring are found on the x = y line (green cloud) - as expected while some offspring are slightly higher than expected (red cloud). In addition, some offspring are found in much higher frequency than expected (blue clouds), and some are not found in the results at all (absent strains, orange cloud which are indicated with minus infinity on the y-axis.

9.2.9. Specific offspring are being produced more than expected

The “all choose all” experiments were done in different media types. As in the competition assay aimed on measuring fitness, experiments were done on media containing Glucose or Glycerol as carbon sources. In addition, I was intrigued to see whether different mate choices on different carbon sources depends on the current carbon source (i.e., at the time of mating), or on the history of the strains (i.e., the physiological response to the media before mating). For that I performed an additional experiment in which cells were grown on Glucose as carbon source, but switched to Glycerol upon mating (see material and methods for more information). To study if mate choice occurs in yeast, mating activity was calculated for each strain (either matA strains or matα) as the sum of offspring it has in the experiment. If no mate choice occurs in yeast, the number of offspring per any two strains should only be based on the mating activity of the two partners. Figure 20 shows the results of the experiments in the different carbon sources, in two repetitions each. The expected offspring frequency, if no mate selection exists (i.e., the multiplication of parents’ initial frequency in the strains’ pool and the mating activity of the parents) is presented on the x axis and the observed offspring frequency as measured by NGS is plotted on the y axis. If no mate choice occurs in this experiment, we would expect the data to reside on the x = y line. It is clear that although the majority of the population is found on x = y line (green dots), in all experiments done, there is also a minority of offspring that are found at much higher frequency than expected (shown in blue) (will be referred to as “upper cloud”). In addition to the upper cloud, there are many offspring that were not formed at all; those are the “absent” offspring and are labeled as –infinity in Figure 20 (shown in orange). Moreover, as mentioned above, most of the data is found in an intermediate position on those axis, (referred to as “intermediate cloud”) but this cloud can be further divided into two – on the right side, offspring that are as prevalent as expected (green dots), and on the right side, offspring that are slightly enriched in comparison to expected assuming no choice (red dots).

We hypothesize that the upper cloud represents the true preference of the strains, while the rest of the data represents the background mating when no choices exist. As such, I further analyzed the offspring in the upper cloud. Interestingly, I found that the high cloud offspring are not the same in the biological repeats (Figure 21A). Encouragingly, though, the intersection between biological repeats is higher than the intersection between experiments done in different media types. For example, 99 shared offspring between the two experiments done on Glucose, only ~10-13 shared offspring between experiments done on Glucose and Glycerol. In addition, the intersection is higher in the real data than in a shuffled data set (Figure 20B, 99 shared offspring in the two repetitions of Glucose, vs 45 in the resampled data). Although the upper cloud intersection seems minor, it is higher than expected by random; in addition, the fact that

the intersection is higher between experiments done on the same media type suggests that mate choice varies in different conditions.

Further analysis showed that absent strains are enriched for offspring whose parents are of high GD (in Glucose and in one repetition of Glu->Gly) (Figure 22A). In addition, the clouds of absent offspring are enriched for offspring that are not found in the competition assays, and thus have no fitness calculated for. Lastly, Figure 22B also shows that the upper cloud offspring tend to be depleted of offspring that were not observed in the competition assays. Taken together, this results open the possibility that yeast choose mating partners to avoid low fitness offspring (as they are depleted in the upper cloud, and enriched in the absent cloud). Although, this result may be trivial, as offspring that are not made are also not participating in the competition, one should think of the massive cells numbers that participated and the portion that was sequenced. Strains in the absent cloud have read count of zero, but only 10^7 cells were sequenced. If those strains were made, even in very low numbers, but yet, were not seen in the competition at all, then indeed they are not absent from the competition but yet has low fitness. Thus, the fact that those "absent" fitness strains are depleted from the upper cloud is an important point in understanding mate choice in yeast.

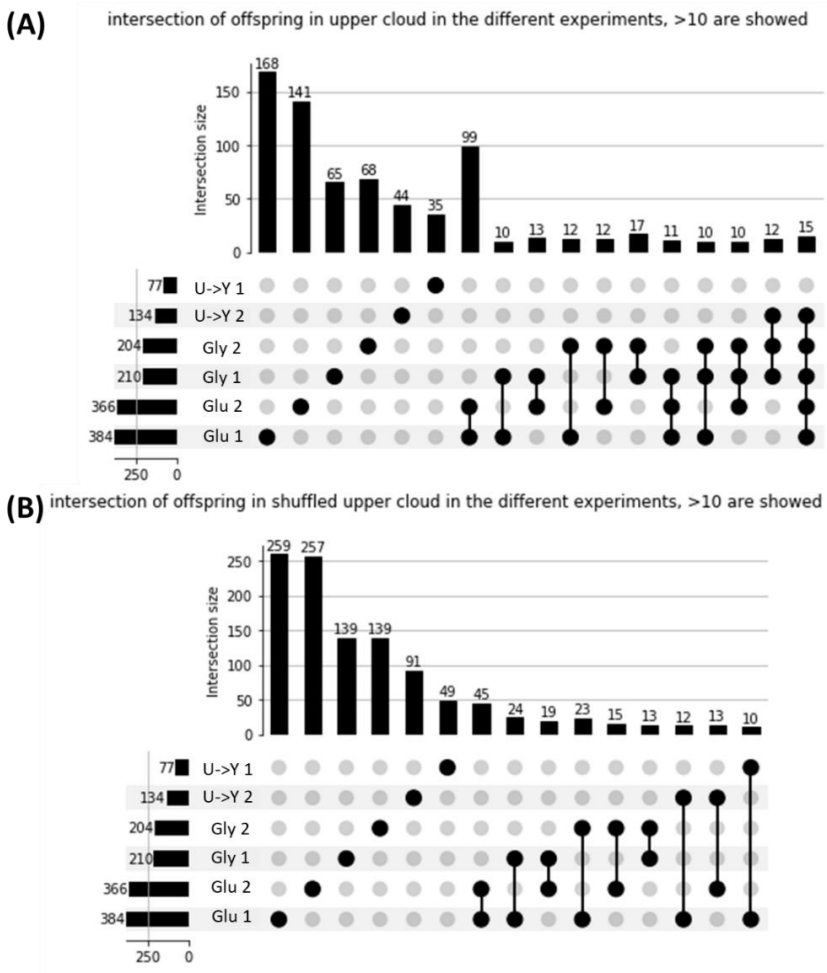


Figure 21 . Intersection between offspring in higher cloud is higher than resampled data
This figure shows an upset plot, which represents the intersection between groups written on the left side. horizontal bar plots, left to the names indicate the number of offspring in high cloud in each experiment. 6 first vertical bar plots, indicates offspring that are found only in one experiment (as indicated by the single black circle in the bottom of the figure). Other bar plots represents the intersection of offspring in the high cloud of different experiments (as indicated by the black circles). (A) True data, (B) the same analysis as in (A), with resampled data

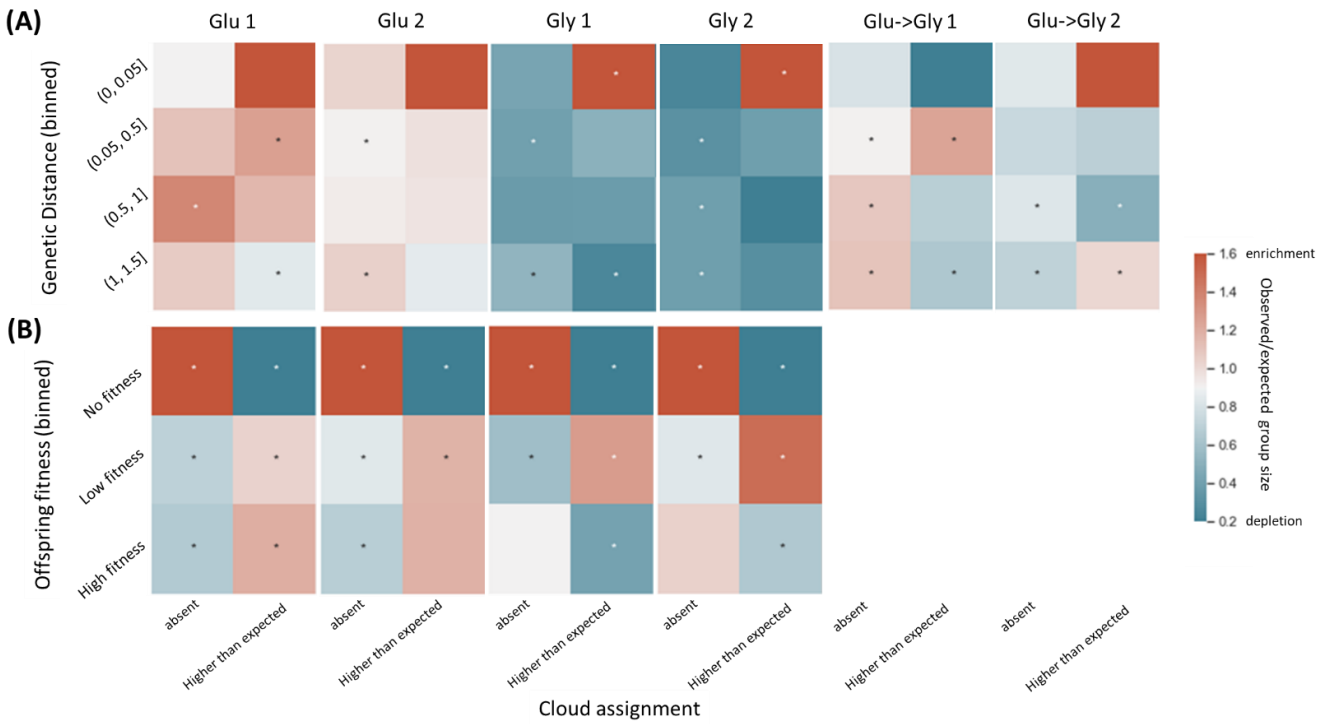


Figure 22 . different clouds are enriched for different characteristics
data was binned into the bins shown in Figure 20. In addition, data was binned based on the genetic distance between parents (A) or offspring fitness as measured from the competition experiments (B). observe/expected frequencies of data were calculated based on chi-square test, and additional residual analysis was done to verify significance (cells annotated with an asterisk are p-value<0.05). Color intensity represents the enrichment of genetic distance (A) or offspring fitness (B) in the different clouds (red - enrichment, blue - depletion)

9.2.10. Mating choices are clustered according to media

As explained above, the “All choose all” experiments were done in three different media; either media containing Glucose or Glycerol as a carbon source, and an additional condition in which cells were grown on Glycerol, but transferred into Glycerol for mating. Experiments were clustered according to offspring found in the upper cloud (using Jaccard metric), the two Glucose repeats are clustered together as well as the two Glycerol repeats. The Glu->Gly repeats are, again, showing a different pattern where one of the repeats (Glu->Gly rep2) is clustered with Gly, and the other is found as an outgroup (Glu->Gly1) (Figure 23). This result supports the previous results, that carbon source changes mate choice in yeast. Moreover, the fact that the Glu->Gly experiments are not clustered together, but rather clustered once with Glu and once with Gly suggests that mate choice in this context is not as clear as in the other two conditions.

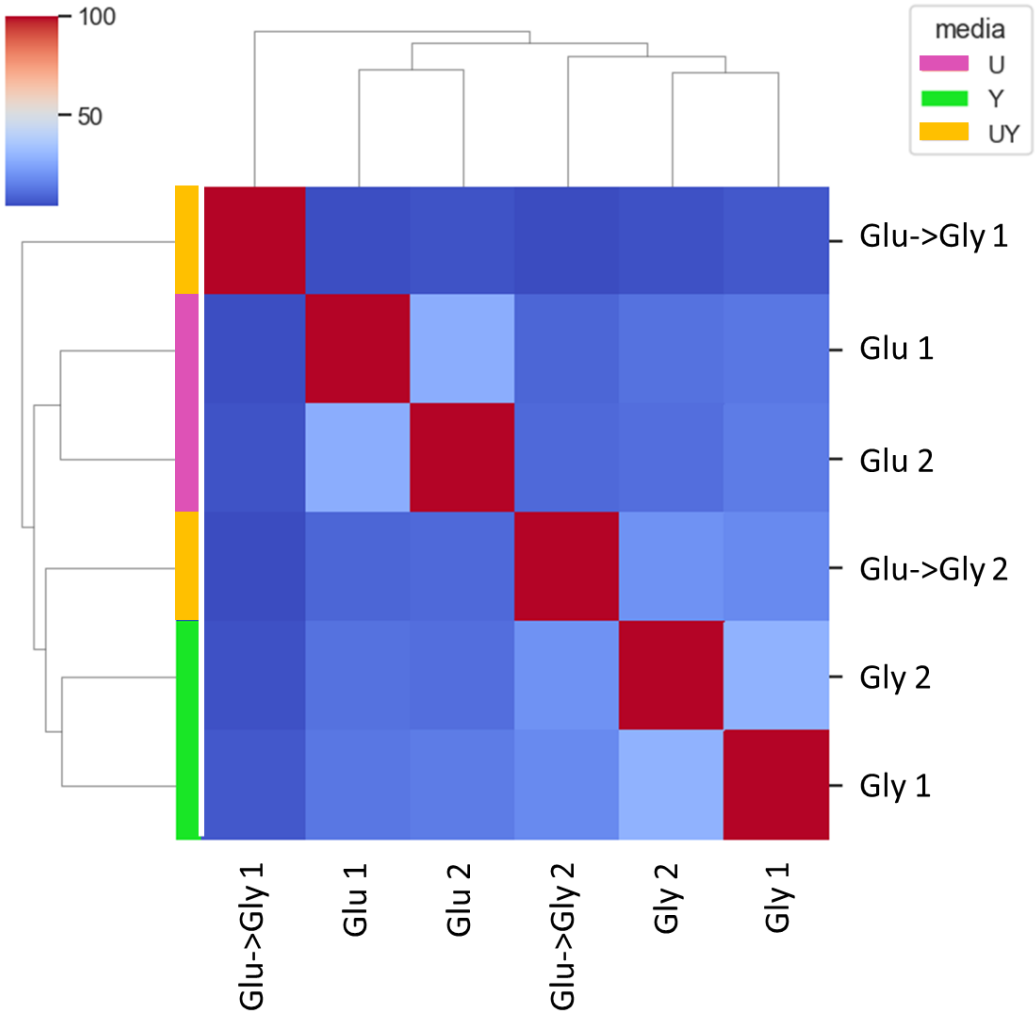


Figure 23 . clustering based on offspring in upper cloud reveals that Glucose and Glycerol are found in different clades, while Glu->Gly is spread out

The offspring in the upper cloud (Figure 20) were used to calculate the distance between experiments. Distance was calculated based on Jaccard metric (the intersect between the two groups divided by the union of them, in percentage). As seen in the Figure the two repeats of Glycerol (green on the left column) are clustered together, as well as the two repeats of Glucose (magenta on the left column), the Glu->Gly condition is not clustered together, and each repeat is clustered with an experiment done in different media.

9.3 Discussion

Sexual reproduction is wide spread and nature, and is the main reproduction method for most eukaryotes, and especially animals [82], [85], [86]. A key component in sexual reproduction is having a mate partner. While many works have shown that mate choice takes place in many organisms [103]–[105], [122]–[124] a study in much higher number is very much needed. I aimed to address several fundamental issues about sexual reproduction in evolution. First, I have examined fitness inheritance and observed a correlation of the offspring fitness to the average, maximal or minimal fitness of its parents

(Figure 15). The observed correlation is significant, however low, indicating that fitness is a complex trait, including dominant, recessive and codominant affects. As such, a simplified model such as the mean, minimum or maximum are not enough to explain all of the variation in the offspring.

In addition, the fact that fitness is affected by GD of the parents regardless of parents' fitness suggests that mating between haploid cells can "overwrite" recessive diminishing alleles. If one parent has a mutation that reduces fitness (maybe due to in-activation or reduced activity), high GD between parents increases the likelihood that the other parent doesn't contain this mutation; thus, after mating, the offspring will have one good copy of that gene, restoring its fitness. On the other hand, if GD between parents is low both parents probably share this mutation.

The second part of this project was aimed to undercover if yeast can choose their mating partners. In addition to the fact that yeast are convenient to work with, and thus serve as a great model organism for high-throughput experiments, the usage of yeast is more profound and can distinguish between two different philosophical theories. Since yeast are one of the lowest Eukaryotes that mate, if yeast do not choose mating partners, it indicates that sexual reproduction is always the better choice, regardless to whether the organism could choose a partner or not. On the other hand, if yeast choose mating partners and this trait was evolved pre- or along with sexual life cycle, it hints that mating without choice is not worthwhile. Notably, yeast such as *S. cerevisiae* contains a sophisticated mechanism of mating type switching. Whenever a haploid cell does not mate, its bud will have the opposite mating type, resulting in "self" mating, where one cell mates with a cell that is identical to it genetically.

An interesting combination of experiments in this report is the of "one-choose-all" experiments, vs "all-choose-all" experiments. One could think that "one-chooses-all" represents the real choice made by the "chooser" strain, as the "chooser" strain has no competition. The experiments set up were purposely executed such that each of the "candidate" strains have enough cells to allow mating to it only (see "Materials and Methods" in the section "One-chooses-all and All-choose-all"). On the other hand, in the "all-choose-all" experiment, there is no real "chooser" and "candidates", both mating types choose and being chosen. In that case, one can hypothesize that the strongest, or fastest strain to mate, could choose first, and others will have only the ability to mate with the remaining haploid cells. As such, comparing these two experiments is very intriguing. In the results shown here, for most strains the choices made in the "one-chooses-all" differ from the choices in the "all-choose-all" further analysis could perhaps shed light if specific strains are first to choose, while other mate with what remains.

Interestingly, Figure 16 shows that yeast fitness is optimized in large GD; however, Figure 22A shows that the high cloud in the Glycerol condition is enriched for low GD. Figure 22 also shows that for both Glucose

and Glycerol, the high cloud is enriched for low fitness offspring, while the absent cloud is enriched for high fitness offspring. Taken together, these results may indicate either that yeast do not select mating partners to optimize offspring fitness or that the offspring fitness calculated in this work as the result of the competition assay is not what yeast are trying to optimize.

Notably, the “all-choose-all” experiments were done in three conditions; in media containing either Glucose or Glycerol, and another condition were cells were grown on Glucose but mated on Glycerol. Although the repetitions in each of the conditions are not as correlated as highly as one would want, the fact that repetitions on the same medium show higher agreement than experiments done on different media support the notion that yeast choose differently in different conditions. I hypothesized that if yeast choose mating partners based on environmental conditions, then using the Glu->Gly condition will teach us if they aim to optimize the fitness of the offspring on the current media, or on the media they grew on before the emating encounter. The Glu->Gly repetitions were the least correlated among themselves (Figures 21, 23) yet one of them clustered with the two Glycerol condition repetitions. Interestingly, outcrossing in yeast occurs mainly after environment change, such as when spores germinates in gut of insects and then mate [125]. I suggest that the lack of correlation and clustering of the Glu->Gly condition teach us that when mating is performed after a radical environment change, the best option is to mate randomly to allow higher genetic variation in the offspring population, since past conditions (nor present conditions) can promise the future conditions. Since one the Glu->Gly repetitions (rep2) is clustered with the Glycerol condition, it is also plausible that yeast sense the current environment and choose accordingly, regardless of past conditions.

9.3.1. Future plans

This project opens enormous possibilities, which some of them are being explored currently or in the near future, and some of them will probably be studied years to come.

Each of the main question presented here can have follow up studies;

In fitness inheritance, we are currently performing Genome Wide Association Study (GWAS) analysis on the offspring strains aiming to identify SNPs that may explain fitness differences. Since the offspring genome is composed of the genomes of both of its parents, it's intriguing to see which loci, and from which parent, contributes to an offspring's fitness. it is also intriguing to make GWAS on other traits that I showed such as on the popularity of a strain (based on results from “one-chooses-all” experiments) or on mutual mate choice (from the “all-choose-all” experiments).

On mate choice questions, further analysis should be done. In the near future I aim to identify which are the strains that are being chosen more often, and why? Are they popular due to the parent fitness? Are

they performing well on many conditions? another option is to check for other similarities among the popular strains; their origin on earth, the ecological niche they arrived from etc. Also, I would like to study the exceptions, which strain chooses specific strains that no one else chose? Is their choice oriented for something else than the rest?

Another interesting question I would like to address in the future, is why do yeast choose mating partners? Is it indeed to optimize offspring fitness? to address it, I would like to propose the following experiment; let yeast mate *en masse*, and immediately measure fitness of the entire population of offspring (via growth experiment). In addition, conduct many pairwise mating of all the original strains participating in the *en masse* experiment and then pool their offspring to the same culture and measure their fitness (via growth experiment). Are the fitness of the two cultures (the first that was obtained after choosing partners) and the second (without choice) differ? Is the culture following choice perform better, and on which conditions?

These projects seem to compose a hierarchy; how much of mate choices done in "one-chooses-all" are solely govern by mating efficiency? And how much of the choices made in "all-choose-all" are based on "one-chooses-all"? This hierarchy will establish a novel understanding on the subject of sexual reproduction.

9.4 Materials and Methods

9.4.1. Strains and media

Strains in this project are natural isolates taken from Peter *et al*. [115]. Table S1 shows the list of strains used in this work.

SD Glu– 20g/L glucose, 6.7g/L nitrogen base, 1.5g/L amino acid mix and 2% Glucose (according to [72]).

SD Gly– 20g/L glucose, 6.7g/L nitrogen base, 1.5g/L amino acid mix and 2% Glycerol(according to [72]).

YPD - 10g/L yeast extract, 20g/L peptone, 20g/L glucose

YPA - 10g/L yeast extract, 20g/L peptone, 20g/L Potassium Acetate

SPO media - 2.5g/l yeast extract, 15 g/l potassium acetate

Antibiotic concentrations and initials:

Hygromycine B (Hyg) - 300mg/L,

Nourseothricin (NAT) – 100mg/L,

Kanamycin (G418) – 200mg/L

Zeocin (Zeo) – 200 mg/L

9.4.2. Plasmids and design construction

Strains were engineered by integrating into their genome a compound design that included the following parts; (Figure 12) (i) genome homology region: 210bp and 700bp homology to the HO locus on both ends of the construct (Figure 12A,E). The specific sequences of homology were chosen to have the least amount of SNPs between most strains, using multiple sequence alignment and BLAST search on strains' genomes published in Peter *et al*. [115]. Homology region was amplified from the genome of BY4741 strain (upstream homology region, chrIV:46062...46271, downstream homology region, chrIV: 47982...48682, (ii) Barcode Fusion Genetic (BFG) system (Figure 12B). This region includes (1) the barcodes region flanked with lox sites that was synthesized and cloned into the rest of the construct by Twist based on Fredrick Roth lab design, as well as (2) the Cre enzyme and rtTa inducer. This region was kindly given to us by the lab of Fredrick Roth [119], (iii) Constitutive markers and fluorescence proteins (Figure 12C) (kindly given to us by Naama Barkai's lab). Two combinations of constitutive fluorescence and antibiotic resistance was used, either a yeGFP gene under the control of the TDH3 promoter and Hygromycine resistance cassette with a TEF promoter, or a mCherry gene under the control of TEF2 promoter and Nourseothricin resistance cassette with a TEF1 promoter. (iv) Haploid selecting markers (Figure 12D). Zeocin or Geneticin (G418) resistance cassettes under the control of the of ste2 matA specific promoter or ste3 matα specific promoter, respectively. A synthetic terminator was added to both resistance cassettes [126]. Sections (iv) was synthetized by GenScript.

All of the above parts, except for (ii1) were assembled together (in the order described in Figure 12) using restriction free methods and cloned into pET28a (Novagen #69864-3) plasmid by the cloning unit in Weizmann institute, generating one backbone plasmid to generate matA strains and another for matα. The barcode fusion genetics library section (ii1) was then cloned into the matA and matα plasmids by Twist Bioscience.

Figure 12, top panel shows the construct that was transformed to create matA cells, while the bottom panel shows the construct that was transformed to create matα cells (from now on referred to as matA construct and matα construct, respectively).

BFG construct and primers

The BFG system is composed of barcodes flanked by *lox* sites (based on [119]). While in the matA construct the order is *loxP-BC1-lox2272-BC2*, in matα construct the order is opposite *BC1-loxP-BC2-lox2272*. The other main component is the Tet-on system that is composed of an rtTA inducer under

constitutive promoter and an *Cre* enzyme under pTet promoter. The rtTA inducer is only active, and can mediate Cre transcription when tetracycline presents.

Following activation of Cre enzyme, a recombination event takes place, recombining the barcode regions of the two parents and results in two fused fragments, one on each chromosome; *BC1(matA)-loxP-BC1(matα)-lox2272* and *loxP-BC2(matA)-lox2272-BC2(matα)*

In addition to the barcodes and *lox* sites, these regions contain unique sequences of ~25nt that are either shared between the parents or are unique to each construct. Those regions can thus be used as primers for amplifying either matA barcodes construct only (Figure 13, primers B and F), matα barcodes construct only (Figure 13, primers H and K) , fused barcodes only (Figure 13, primers B and I or primers E and K), or all of the above (Figure 13, primers A and G).

Primers used for amplifying matA construct only, matα construct only, fused barcode only or all , will be termed matA primers, matα primers, fused primers or general primers, respectively.

9.4.3. Strain construction

transformation

Constructs (from the section *"plasmid and design construction"*) were amplified using the following primers; F: GGTGAAAACCTGTACTTCCAGGG, R: ATGCTAGTTATTGCTCAGCGGT. PCR was done using KAPA HiFi HotStart ReadyMix (Roche, KK2602) enzyme according to manufacturer instructions with the following details: primer annealing Tm of 64°C, elongation of 5 minutes, 30 cycles. PCR products were then transformed into the chosen strains as follows: For each transformation, 5 reactions of PCR were made (to increase complexity of the library). All reactions were ran on agarose gel (0.8% agarose) and size was verified (7kb).

To transform the cells, LiAc protocol [53] was adapted to allow high throughput transformation of many strains in a 96-well plate. In short, 48 strains were taken out of the -80 and inoculated in a 96-well plate in a checkerboard manner to avoid cross contamination between strains. Cells were grown overnight at 30°C with shaking (1200rpm), in a shaker incubator. Then, each strain was diluted 1:20000 in a 50ml tube, (0.5ul of culture into 10ml YPD) and grown for 16 hours at 30°C while shaking. Following the growth phase, four random strains were counted to estimate cell concentration. Cultures usually reached the late log stage (budded yeast, ~7E7cells/ml); each culture was diluted 1:50 into fresh YPD and allowed to grow for another couple of hours.

Cells were harvested by centrifugation (4000rpm for 5min) and washed twice; first with 5ml DDW then with 1ml LiAc 100uM. After the second wash, remaining liquid was vacuumed, and pellet was resuspended in 60ul DDW. Two 96-well plates for transformation, in a checkerboard manner (one for matA construct and one for matα construct) were made with 25ul of PCR product of the relevant construct. 30ul of each strain were suspended into each of the two transformation plates. Transformation mix (100ul PEG 50%, 15ul LiAc 1M, 4ul of 10mg/ml salmon sperm (Sigma Aldrich, D9156-1ML) were added to each well. Plates were incubated for 40 minutes at 42°C. Plates were centrifuges (3000rpm for 3 minutes), and liquid was vacuumed with a multi pipette vacuum adaptor. 150ul of YPD was added to each well and plates were incubated overnight at 30°C. The following morning, each well was plated on YPD agar containing the relevant antibiotic (Hyg for matA construct plate, and Nat for matα construct plate). Agar plates were incubated in 30°C for a couple of days until the appearance of colonies.

After colonies appeared, 4 colonies from each strain were picked, inoculated into liquid SD in 96-well plate (to verify fluorescence) and patched on YPD agar plate with the corresponding antibiotic (for continuing). 96-well plates were grown overnight at 30°C and diluted 1:50 into fresh SD. Plates were FACS analyzed to verify the correct fluorescence marker (cells with matA construct had yeGFP while matα construct corresponds to mCherry). One positive colony per strain was chosen to continue.

Sporulation and random spore analysis

Reagents for random spore analysis:

*TE buffer (Tris 10mM EDTA 1mM, pH 8.0), β-Mercaptoethanol (dilute 7.5ul into 1ml DDW), Triton-X100, 0.5%, **β**-Glucoronidase (Sigma- cat no. G7017-5ML).*

***β**-Glucoronidase is prepared as such; vortex container vigoursly, dilute 1:2 in DDW, centrifuge at maximal speed for 1 minute and remove sup into a syringe with filter (2mm mash).*

To generate engineered strains, positive diploid colonies (obtained as described in the section “transformation”) were inoculated into a 24-well plate with 1.2ml YPA and grown overnight at 30°C with shaking (1200rpm). Plates were centrifuged (4000rpm, 3 minutes) and washed with DDW, spin down and vacuumed. Pellets were resuspended in 1.2ml SPO media and incubated for 4-5 days in 25°C while shaking constantly. After 4 days, a couple of random cultures were observed under the microscope to verify sporulation, and score sporulation efficiency. In case of low sporulation efficiency plates were incubated an extra day.

50ul of each sporulated culture was transferred into a 96-well plate for random spore analysis as follows:

Pellets were resuspended with 50ul TE buffer (Tris 10mM EDTA 1mM, pH 8.0) + 2.5ul β-Mercaptoethanol and incubated on bench for 10 minutes. Plates were centrifuged and sup was discarded, pellets were then washed with 150ul DDW twice. Pellets were resuspended in 50ul **β**-Glucoronidase *(dilute 7.5ul into 1ml DDW)* and incubated for twof hours in 37°C, while shaking. Plates were centrifuged and pellets were washed with 200ul *Triton-X100, 0.5%*, this step was repeated twice. Plates were centrifuged, and pellets were resuspended with 120ul DDW and plated on YPD agar plates, with corresponding antibiotic for haploid of the correct mating type (matA construct were plated on Zeocin containing plates, while matα construct cells were plated on G418). Plates were incubated in 30°C for 48h until colonies appeared.

From each strain, 2 colonies were picked, inoculated into liquid SD in 96-well plate and patched on YPD agar plate with the corresponding antibiotic. 96-well plates were grown overnight at 30°C and diluted 1:50 into fresh SD. Plates were FACS analyzed to verify correct fluorescence marker (cells with matA construct had GFP while matα construct corresponds to mCherry).

Positive colonies (1-2 colonies) were continued to Sanger sequencing to recover barcode sequences.

One correct haploid colony per strain was freezed in a 96-well plate in -80°C.

9.4.4. *En masse* mating

All verified haploid strains were taken out from the -80°C into YPD media with corresponding antibiotic (Hyg or NAT for matA or matα respectively) in a 96-well plate using pinners. Strains were grown overnight (30°C, while shaking) and then diluted 1:1000 for another overnight incubation in 30°C. Strains were diluted 1:50 into fresh YPD and grown for another couple of hours in 30°C to reach mid-log phase. OD was measured to all plates by plate reads (infinite 200, Tecan). All matA strains and all matα strains were mixed (separately for matA and Matα), based on the measured OD, such that they will have equal representation in the mix to create to mixes, matA mix and matα mix. Mixes were centrifuged and resuspended in 0.2X volume to create a 5-fold increase in cell concentration (~1E8 cells/ml). *En masse* mating was conducted by mixing 50ul (~5E6 cells total) of each mating type mix in a 1ml medium (either SD-Glu or SD-Gly) with doxycycline (final concentration of 10ug/ml). Mating was executed for 20hours in 25°C without shaking.

Three repetitions of *en masse* mating were done per media.

One-chooses-all and All-choose-all experiments

En masse mating was performed roughly the same in all experiments with a couple of differences between “one-chooses-all” and “all-choose-all”. In “One-choose-all” experiments, in each mating reaction one mix of cells (either matA or matα) was inoculated with one strain of the opposite mating

type, while in “all-choose-all”, the two mixes (of matA and matα) were mixed together. In addition, in the all-choose-all experiments mating was performed in 24-well plate in 1ml volume, and in one-chooses-all mating was performed in 96-well plate in 150ul volume. In the “One-chooses-all”, the chooser strain, and each of the strains in the mix were mixed such they have the same number of cells in the mating reaction (i.e., chooser strain had 5E6 cells total in the mating, and each of the strain in the mix had also ~5E6 cells, so the mix had 5E6*#strains cells in the final mating reaction). Also, “One-chooses-all” experiments were done on SD-Gly only while “all-choose-all” were done on SD-Glu and SD-Gly.

9.4.5. Pooled competition

Each of the *en masse* mating was continued to a competition assay to measure the fitness of each offspring. Cuompetition was done in a daily dilution (1:240) manner in either SD-Glu or SD-Gly, with both Hyg and NAT to select for diploids. In the first dilution media also contained doxycycline (10ug/ml). Cultures were grown in 50ml tubes at 10ml volume. Cells were grown in 30°C while shaking. Dilution was carried out every day (in SD-Glu) or every two days (SD-Gly).

Cells were frozen in 30% Glycerol and kept in -80˚c every dilution. All competitions were carried for ~55 generations and freezed every 8 generations. All but the first dilutions were freezed to have 6 time points per competition.

Competition experiments for parents (either matA or matα) were conducted similarly.

9.4.6. Library preparation, sequencing and read analysis

For all experiments, library preparation was done the same. Primer used in the different experiments vary, and will be specified here. Also, DNA extraction methodology was different and will be specified here.

Libraries for sequencing the barcode region were constructed by designing Plate-Row-Column PCR methodology; in which a first PCR is done using primers targeting the barcode region, plate barcode and tails that match Illumina adapters (F: ACGACGCTCTTCCGATCT**NNNNNN***BFGprimer*, R: AGACGTGTGCTCTTCCGATCT**NNNNNN***BFGprimer*

Capital letters corresponds to Illumina adaptors, N correspond to plate index and BFGprimer is the primer shown in Table 2. First PCR was done in 25ul final volume with 2ul of template DNA (either genomic DNA or after cell blow up in 20mM NaOH and bioloing for 15 minutes). PCR program: Tm of 60°C, elongation of 10 seconds, ~20 cycles. Usually, 4 PCR reactions were done per experiments and they were pooled together after PCR (to avoid PCR biases). 2ul of the first PCR was used as a template for the second PCR.

A second PCR (F: AATGATACGGCGACCACCGAGATCTACACTCTTTCC CTACACGACGCTCTTCCGATCT, R: CAAGCAGAAGACGGCATACGAGAT***NNNNNNNN***GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT. ***N*** corresponds to Illumina index for library multiplexing) was carried out to attach the adapters for the Illumina run. PCR was done in 25ul volume. PCR program: Tm of 62°C, elongation of 10 seconds, ~20 cycles. After second PCR libraries were cleaned using SPRI beads to eliminate unspecific bands and primer dimers.

Amplicons were sequenced using paired-end methodology, on the NovaSeq platform (Illumina) (see kits used in Table 2).

Table 2. DNA extraction method and primers used in each of the described experiments. (*) MasterPure Yeast DNA Purification Kit by epicenter, (**) Blow up in 20mM NaOH + bioling for 15 minutes

Experiment	DNA extraction method	Amplified region	primers used for PCR I of lib prep	F	R
pooled competition	DNA extraction kit (*)	fused barcodes only	fused primers (B+I)	taacccttagaaccgagagtgtg	gttatcagaggtatgcgagttag
all-choose-all	DNA extraction kit (*)	all combinations	general primers (A+G)	cctcataagcagcaatcaattctatctatactttaaa	ggccgttacttacttagagctt
one-chooses-all	Blow up (**)	all combinations	general primers (A+G)	cctcataagcagcaatcaattctatctatactttaaa	ggccgttacttacttagagctt
BFG efficiency	Blow up (**)	all combinations	general primers (A+G)	cctcataagcagcaatcaattctatctatactttaaa	ggccgttacttacttagagctt

After initial de-multiplexing by the Illumina platform, libraries were further separated based on the plate index using cutadapt [127]. All reads were further analyzed by cutadapt to leave only the barcode region in the file. For alignment, I created a synthetic genome from all strains' barcodes using bowtie2 (bowtie2-build command). Alignment was performed using bowtie2 as well. To recover read counts per fused barcode, I used in house script.

9.4.7. Fitness estimation based on pooled competition

Fitness was derived by employing a Maximum-Likelihood (ML) algorithm on all read count measurements along the competition experiment per variant (fitness was calculated only for strains with more than 10 reads in the beginning of the competition). Briefly, first, each variant fitness is estimated by using a simple loglinear regression over the first three time points. Based on these estimations, the initial relative frequencies of each variant, and a noise model that accounts for experimental errors [47], expected trajectory of each variant is estimated and compared to the measured trajectory. Next, small changes are made to our fitness estimates, comparison is repeated, fitness is updated if they better fit the data (higher likelihood). This procedure is performed iteratively until fitness estimates are stable (maximized likelihood).

9.4.8. Mating efficiency

Relevant strains (see Table S2) were taken from -80°C and were grown over weekend at 30°C in YPD while shaking. Cells were diluted 1:1000 into SD-Glu and incubated overnight in 30°C while shaking. Cells were diluted 1:50 and allowed growing for a couple of hours to reach mid-log. OD was measured using a plate reader (infinity 200, Tecan). Two strains (one matA and one matα cells) were inoculated into the same well such that their cell number is equal. Mating was executed for 20 hours in 25°C without shaking. All mating reactions were done in 3 repetitions.

Mating efficiency for each pair of strains was determined using FACS (Attune) using 96-well plate module. Cultures were diluted 1:200 before FACS into SD with 5mM EDTA. Each culture was divided into 3 populations based on GFP/mCherry ratio; high GFP and low mCherry were considered matA cells, high mCherry and low GFP were the matα cells, and high mCherry and GFP are the offspring. mating efficiency was calculated based on the following formula: $ME = \frac{\#offspring}{\min(matA, mat\alpha)}$ (ME: mating efficiency). Mating efficiency was calculated as the number of offspring divided by the minimum of matA and matα number of cells.

9.4.9. Sorting for offspring

Mating was performed as described in *en masse* mating part. Mating was done in three conditions; cells growth and mating on SD-Glu, cells growth and mating on SD-Gly or cells growth on SD-Glu followed by mating in SD-Gly (termed Glu->Glu experiment).

Sorting experiment was done in three consecutive days. matA and matα mixes were kept in 4°C and used for mating in all three days. 12 mating reactions were performed for each media type, in a 24-well plate. Mating was carried out for 20 hours in 25°C without shaking in all 3 media type. After 20 hours of mating, all mating reactions of the same media type were combined, and EDTA was added for a final concentration of 5mM.

Flow cytometry analysis and sorting were performed on a FACSAria Fusion instrument (BD Biosciences) equipped with a 405, 488, 561 and 640 nm lasers, using a 100 mm nozzle, controlled by BD FACS Diva software v8.0.1 (BD Biosciences), at The Weizmann Institute of Science Flow Cytometry Core Facility. Further analysis was performed using FlowJo software v10.2 (Tree Star). Cells were gated according to FSC and SSC to avoid debris and big aggregates. Another gate of high GFP and high mCherry was determined for sorting of offspring only.

Each day, the order of experiments to be sorted was changed to avoid bias. Cultures were kept in 4°C before and while sorting of other experiments took place.

Approximately 5E7 cells were sorted per experiment, resulting in ~1E6 offspring cells.

9.4.10. BFG efficiency

Chosen strains (Table S3) were used for measurement and calculation of BFG efficiency. Strains were removed from -80°C and grown at 30°C overnight while shaking. Cells were diluted 1:1000 and grown at 30°C overnight while shaking. OD was measured using plate reader (infinite 200, Tecan), and all strains were diluted to equal OD values. Cells were mixed in 96-well plate on SD-Glu with 10ug/ml Doxycycline and allowed to mate for 20 hours at 25°C without shaking. Cells were diluted 1:120 into fresh SD-Glu with 10ug/ml doxycycline and Hyg and NAT, and continued growing. Cells were grown in a daily dilution manner in SD-Glu + Hyg + NAT (without Doxy) to select for offspring only. After 5 days (~30 generations), plates were FACS analyzed to verify offspring ratio from culture by using GFP/mCherry gating, in most wells offspring were <90% of the culture. One row and one column had no offspring, indicating a problem with that strain (data not shown). At this point, DNA was extracted by boiling the cells in 20mM NaOH for 15 minutes and library were constructed with the most upstream and downstream primers (primers A + G Figure 13, and Table 2). BFG efficiency was calculated by dividing the number of reads with fused barcodes with number of reads with original barcodes for each pair.

10. Thesis Discussion

In my three PhD projects I aimed to study evolution and evolvability. Evolthon was a first of its kind community effort to study lab evolution. It represents how scientists from all over the world think of lab evolution, and on the methods thought to be a major force in evolution. Each of the strategies employed in Evolthon could be further studied; in my PhD I focused on only two, evolution mediated by retrotransposons, and the evolution of mating. These two projects shed light on two interesting, and different, aspects of evolution. In the first, I studied a mechanism that is considered to be Lamarckian. I studied if retro elements, which considered to be selfish elements and that are found in the genome of most organisms, contribute to the evolution of their host. Based on the results, it seemed that yeast retro elements have the potential of affecting evolution, and perhaps indeed they do.

In the second project, I focused on mating. On the one hand, I studied it effect on evolution by investigation how fitness is inherit, but in addition I studied the evolution of mating.

11. Bibliography

[1] T. Dobzhansky, “Nothing in Biology Makes Sense Except in the Light of Evolution.” American Biology Teacher, pp. 125–129, 1973.

[2] M. Pigliucci, “Is evolvability evolvable?,” *Nat. Rev. Genet.*, vol. 9, no. 1, pp. 75–82, Jan. 2008, doi: 10.1038/nrg2278.

[3] B. Szamecz *et al.*, “The Genomic Landscape of Compensatory Evolution,” *PLoS Biol.*, vol. 12, no. 8, p. e1001935, Aug. 2014, doi: 10.1371/journal.pbio.1001935.

[4] J. L. Payne and A. Wagner, “The causes of evolvability and their evolution,” *Nat. Rev. Genet. 2018 201*, vol. 20, no. 1, pp. 24–38, Nov. 2018, doi: 10.1038/s41576-018-0069-z.

[5] O. Tenaillon, B. Toupance, H. Le Nagard, F. Taddei, and B. Godelle, “Mutators, Population Size, Adaptive Landscape and the Adaptation of Asexual Populations of Bacteria,” *Genetics*, vol. 152, no. 2, 1999.

[6] D. M. Weinreich, R. A. Watson, and L. Chao, “PERSPECTIVE:SIGN EPISTASIS AND GENETIC CONSTRAINT ON EVOLUTIONARY TRAJECTORIES,” *Evolution (N. Y).*, vol. 59, no. 6, p. 1165, 2005, doi: 10.1554/04-272.

[7] R. J. Woods, J. E. Barrick, T. F. Cooper, U. Shrestha, M. R. Kauth, and R. E. Lenski, “Second-Order Selection for Evolvability in a Large Escherichia coli Population,” *Science (80-.).*, vol. 331, no. 6023, 2011.

[8] E. V Koonin and Y. I. Wolf, “Is evolution Darwinian or/and Lamarckian?,” *Biol. Direct*, vol. 4, no. 1, p. 42, Nov. 2009, doi: 10.1186/1745-6150-4-42.

[9] P. Hammerstein, “Darwinian adaptation, population genetics and the streetcar theory of evolution,” *J. Math. Biol. 1996 345*, vol. 34, no. 5, pp. 511–532, 1996, doi: 10.1007/BF02409748.

[10] D. Graur, M. Gouy, and D. Wool, “In Retrospect: Lamarck’s treatise at 200,” *Nat. 2009 4607256*, vol. 460, no. 7256, pp. 688–689, Aug. 2009, doi: 10.1038/460688a.

[11] D. L. Hull and R. W. Burkhardt, “The Spirit of System: Lamarck and Evolutionary Biology.,” *Syst. Zool.*, vol. 27, no. 2, p. 248, Jun. 1978, doi: 10.2307/2412986.

[12] E. V. A. Jablonka and G. A. L. Raz, “Transgenerational epigenetic inheritance: prevalence, mechanisms, and implications for the study of heredity and evolution,” *Q. Rev. Biol.*, vol. 84, no. 2, pp. 131–176, Jun. 2009, doi: 10.1086/598822/ASSET/IMAGES/LARGE/FG3.JPEG.

[13] A. Bošković and O. J. Rando, “Transgenerational Epigenetic Inheritance,” *https://doi.org/10.1146/annurev-genet-120417-031404*, vol. 52, pp. 21–41, Nov. 2018, doi: 10.1146/ANNUREV-GENET-120417-031404.

[14] J. BEISSON and T. M. SONNEBORN, “CYTOPLASMIC INHERITANCE OF THE ORGANIZATION OF THE CELL CORTEX IN PARAMECIUM AURELIA,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 53, no. 2, p. 275, 1965, doi: 10.1073/PNAS.53.2.275.

[15] C. Terry and J. D. F. Wadsworth, “Recent advances in understanding mammalian prion structure: A mini review,” *Front. Mol. Neurosci.*, vol. 12, p. 169, Jul. 2019, doi: 10.3389/FNMOL.2019.00169/BIBTEX.

[16] Y. O. Chernoff, “Stress and prions: Lessons from the yeast model,” *FEBS Lett.*, vol. 581, no. 19, pp. 3695–3701, Jul. 2007, doi: 10.1016/J.FEBSLET.2007.04.075.

[17] O. Rechavi and I. Lev, “Principles of Transgenerational Small RNA Inheritance in Caenorhabditis elegans,” *Curr. Biol.*, vol. 27, no. 14, pp. R720–R730, Jul. 2017, doi: 10.1016/J.CUB.2017.05.043.

[18] R. S. Moore *et al.*, “The role of the Cer1 transposon in horizontal transfer of transgenerational memory,” *Cell*, vol. 184, no. 18, pp. 4697-4712.e18, Sep. 2021, doi:

10.1016/J.CELL.2021.07.022.

[19] M. F. Perez and B. Lehner, “Intergenerational and transgenerational epigenetic inheritance in animals,” *Nat. Cell Biol. 2019 212*, vol. 21, no. 2, pp. 143–151, Jan. 2019, doi: 10.1038/s41556-018-0242-9.

[20] J. D. Boeke, “Retrotransposons,” *RNA Genet.*, pp. 59–103, Jan. 2018, doi: 10.1201/9781351076432-4.

[21] R. A. Elbarbary, B. A. Lucas, and L. E. Maquat, “Retrotransposons as regulators of gene expression,” *Science (80-.).*, vol. 351, no. 6274, Feb. 2016, doi: 10.1126/SCIENCE.AAC7247/ASSET/2B881D09-FE4B-4740-BF52-4566654237B4/ASSETS/GRAPHIC/351_AAC7247_F5.JPEG.

[22] M. Tatalovic, “Evolution of raised guarding behavior in meerkats, Suricata suricatta,” 2010.

[23] J. R. Litman, B. N. Danforth, C. D. Eardley, and C. J. Praz, “Why do leafcutter bees cut leaves? New insights into the early evolution of bees,” *Proc. R. Soc. B Biol. Sci.*, vol. 278, no. 1724, pp. 3593–3600, 2011, doi: 10.1098/RSPB.2011.0365.

[24] G. Zhang *et al.*, “Comparative genomics reveals insights into avian genome evolution and adaptation,” *Science (80-.).*, vol. 346, no. 6215, pp. 1311–1320, Dec. 2014, doi: 10.1126/SCIENCE.1251385/SUPPL_FILE/ZHANG-TABLE_S9.TXT.

[25] E. V. Koonin, L. Aravind, and A. S. Kondrashov, “The Impact of Comparative Genomics on Our Understanding of Evolution,” *Cell*, vol. 101, no. 6, pp. 573–576, Jun. 2000, doi: 10.1016/S0092-8674(00)80867-3.

[26] H. Ellegren, “Comparative genomics and the study of evolution by natural selection,” *Mol. Ecol.*, vol. 17, no. 21, pp. 4586–4596, Nov. 2008, doi: 10.1111/J.1365-294X.2008.03954.X.

[27] A. H. Yona *et al.*, “tRNA genes rapidly change in evolution to meet novel translational demands,” vol. 2, p. 1339, 2013, doi: 10.7554/eLife.01339.001.

[28] R. E. Lenski, M. R. Rose, S. C. Simpson, and S. C. Tadler, “Long-Term Experimental Evolution in Escherichia coli. I. Adaptation and Divergence During 2,000 Generations,” *Am. Nat.*, vol. 138, no. 6, pp. 1315–1341, Dec. 1991, doi: 10.1086/285289.

[29] D. Papadopoulos, D. Schneider, J. Meier-Eiss, W. Arber, R. E. Lenski, and M. Blot, “Genomic evolution during a 10,000-generation experiment with bacteria.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 96, no. 7, pp. 3807–12, Mar. 1999, doi: 10.1073/PNAS.96.7.3807.

[30] A. H. Yona *et al.*, “Chromosomal duplication is a transient evolutionary solution to stress.,”

Proc. Natl. Acad. Sci. U. S. A., vol. 109, no. 51, pp. 21010–5, Dec. 2012, doi: 10.1073/pnas.1211150109.

[31] M.-M. Perrineau, E. Zelzion, J. Gross, D. C. Price, J. Boyd, and D. Bhattacharya, "Evolution of salt tolerance in a laboratory reared population of *C hlamydomonas reinhardtii*," *Environ. Microbiol.*, vol. 16, no. 6, pp. 1755–1766, Jun. 2014, doi: 10.1111/1462-2920.12372.

[32] E. Toprak, A. Veres, J.-B. Michel, R. Chait, D. L. Hartl, and R. Kishony, "Evolutionary paths to antibiotic resistance under dynamically sustained drug selection," *Nat. Genet.*, vol. 44, no. 1, pp. 101–105, Jan. 2012, doi: 10.1038/ng.1034.

[33] S. Kaminski Strauss *et al.*, "Evolthon: A community endeavor to evolve lab evolution," *PLOS Biol.*, vol. 17, no. 3, p. e3000182, Mar. 2019, doi: 10.1371/journal.pbio.3000182.

[34] A. H. Yona *et al.*, "Chromosomal duplication is a transient evolutionary solution to stress," doi: 10.1073/pnas.1211150109.

[35] G. Liu *et al.*, "Gene Essentiality Is a Quantitative Property Linked to Cellular Evolvability," *Cell*, vol. 163, no. 6, pp. 1388–1399, Dec. 2015, doi: 10.1016/j.cell.2015.10.069.

[36] J. E. Barrick *et al.*, "Genome evolution and adaptation in a long-term experiment with Escherichia coli," *Nature*, vol. 461, no. 7268, pp. 1243–1247, Oct. 2009, doi: 10.1038/nature08480.

[37] V. S. Cooper, A. F. Bennett, and R. E. Lenski, "EVOLUTION OF THERMAL DEPENDENCE OF GROWTH RATE OF ESCHERICHIA COLI POPULATIONS DURING 20,000 GENERATIONS IN A CONSTANT ENVIRONMENT," *Evolution (N. Y).*, vol. 55, no. 5, pp. 889–896, May 2007, doi: 10.1111/j.0014-3820.2001.tb00606.x.

[38] F. Taddei, M. Radman, J. Maynard-Smith, B. Toupance, P. H. Gouyon, and B. Godelle, "Role of mutator alleles in adaptive evolution," *Nature*, vol. 387, no. 6634, pp. 700–702, Jun. 1997, doi: 10.1038/42696.

[39] R. Miralles, P. J. Gerrish, A. Moya, and S. F. Elena, "Clonal interference and the evolution of RNA viruses.," *Science*, vol. 285, no. 5434, pp. 1745–7, Sep. 1999, doi: 10.1126/SCIENCE.285.5434.1745.

[40] D. Gresham and J. Hong, "The functional basis of adaptive evolution in chemostats," *FEMS Microbiol. Rev.*, vol. 39, no. 1, p. n/a-n/a, Sep. 2014, doi: 10.1111/1574-6976.12082.

[41] R. Axelrod, "A report on the computer tournament for the iterated Prisoner's Dilemma: tit for tat wins the first round," 1978.

[42] E. Bender, “Challenges: Crowdsourced solutions,” *Nature*, vol. 533, no. 7602, pp. S62–S64, May 2016, doi: 10.1038/533S62a.

[43] C. D. Smolke, “Building outside of the box: iGEM and the BioBricks Foundation,” *Nat. Biotechnol.*, vol. 27, no. 12, pp. 1099–1102, Dec. 2009, doi: 10.1038/nbt1209-1099.

[44] Y. Ram, E. Dellus-Gur, U. Obolski, M. Bibi, J. Berman, and L. Hadany, “Predicting competition results from growth curves,” *bioRxiv*, p. 022640, Jul. 2015, doi: 10.1101/022640.

[45] E. Oxman, U. Alon, and E. Dekel, “DEFINED ORDER OF EVOLUTIONARY ADAPTATIONS: EXPERIMENTAL EVIDENCE,” *Evolution (N. Y).*, vol. 62, no. 7, pp. 1547–1554, Jul. 2008, doi: 10.1111/j.1558-5646.2008.00397.x.

[46] I. Frumkin *et al.*, “Gene Architectures that Minimize Cost of Gene Expression,” *Mol. Cell*, 2017, doi: 10.1016/j.molcel.2016.11.007.

[47] S. F. Levy, J. R. Blundell, S. Venkataram, D. A. Petrov, D. S. Fisher, and G. Sherlock, “Quantitative evolutionary dynamics using high-resolution lineage tracking,” *Nature*, vol. 519, no. 7542, pp. 181–186, Mar. 2015, doi: 10.1038/nature14279.

[48] M. J. McDonald, D. P. Rice, and M. M. Desai, “Sex speeds adaptation by altering the dynamics of molecular evolution,” *Nature*, vol. 531, no. 7593, pp. 233–236, Mar. 2016, doi: 10.1038/nature17143.

[49] C. R. Otey, G. Bandara, J. Lalonde, K. Takahashi, and F. H. Arnold, “Preparation of human metabolites of propranolol using laboratory-evolved bacterial cytochromes P450,” *Biotechnol. Bioeng.*, vol. 93, no. 3, pp. 494–499, Feb. 2006, doi: 10.1002/bit.20744.

[50] J. Sylvestre, H. Chautard, F. Cedrone, and M. Delcourt, “Directed evolution of biocatalysts,” *Org. Process Res. Dev.*, vol. 10, no. 3, pp. 562–571, 2006, doi: 10.1021/op050243h.

[51] T. Bosma, J. Damborský, G. Stucki, and D. B. Janssen, “Biodegradation of 1,2,3-trichloropropane through directed evolution and heterologous expression of a haloalkane dehalogenase gene.,” *Appl. Environ. Microbiol.*, vol. 68, no. 7, pp. 3582–7, Jul. 2002, doi: 10.1128/AEM.68.7.3582-3587.2002.

[52] A. L. Goldstein and J. H. McCusker, “Three new dominant drug resistance cassettes for gene disruption inSaccharomyces cerevisiae,” *Yeast*, vol. 15, no. 14, pp. 1541–1553, Oct. 1999, doi: 10.1002/(SICI)1097-0061(199910)15:14<1541::AID-YEA476>3.0.CO;2-K.

[53] R. D. Gietz and R. A. Woods, “Transformation of yeast by lithium acetate/single-stranded carrier DNA/polyethylene glycol method.,” *Methods Enzymol.*, vol. 350, pp. 87–96, 2002.

[54] T. Baba *et al.*, “Construction of Escherichia coli K-12 in-frame, single-gene knockout mutants: the Keio collection.,” *Mol. Syst. Biol.*, vol. 2, p. 2006.0008, Feb. 2006, doi: 10.1038/msb4100050.

[55] J. Kim, A. M. Webb, J. P. Kershner, S. Blaskowski, and S. D. Copley, “A versatile and highly efficient method for scarless genome editing in Escherichia coli and Salmonella enterica.,” *BMC Biotechnol.*, vol. 14, p. 84, Sep. 2014, doi: 10.1186/1472-6750-14-84.

[56] M. Lynch, “Evolution of the mutation rate,” *Trends Genet.*, vol. 26, no. 8, pp. 345–352, Aug. 2010, doi: 10.1016/J.TIG.2010.05.003.

[57] J. F. Roth, “The yeast Ty virus-like particles.,” *Yeast*, vol. 16, no. 9, pp. 785–95, Jun. 2000, doi: 10.1002/1097-0061(20000630)16:9<785::AID-YEA550>3.0.CO;2-L.

[58] D. J. Garfinkel, “Genome evolution mediated by Ty elements in Saccharomyces,” 2005. [Online]. Available: http://www.karger.com.ezproxy.weizmann.ac.il/Article/Pdf/84939. [Accessed: 02-Jul-2014].

[59] D. J. Eichinger and J. D. Boeke, “The DNA intermediate in yeast Ty1 element transposition copurifies with virus-like particles: Cell-free Ty1 transposition,” *Cell*, vol. 54, no. 7, pp. 955–966, Sep. 1988, doi: 10.1016/0092-8674(88)90110-9.

[60] K. B. Chapman, a S. Byström, and J. D. Boeke, “Initiator methionine tRNA is essential for Ty1 transposition in yeast.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 89, no. 8, pp. 3236–40, Apr. 1992.

[61] M. J. Curcio and D. J. Garfinkel, “Single-step selection for Ty1 element retrotransposition,” *PNAS*, vol. 88, no. 3, pp. 936–940, 1991.

[62] P. H. Maxwell, C. Coombes, A. E. Kenny, J. F. Lawler, J. D. Boeke, and M. J. Curcio, “Ty1 mobilizes subtelomeric Y’ elements in telomerase-negative Saccharomyces cerevisiae survivors.,” *Mol. Cell. Biol.*, vol. 24, no. 22, pp. 9887–98, Nov. 2004, doi: 10.1128/MCB.24.22.9887-9898.2004.

[63] G. Zheng *et al.*, “Efficient and quantitative high-throughput tRNA sequencing,” *Nat. Methods*, vol. 12, no. 9, pp. 835–837, Sep. 2015, doi: 10.1038/nmeth.3478.

[64] Y. Arava, Y. Wang, J. D. Storey, C. Long Liu, P. O. Brown, and D. Herschlag, “Genome-wide analysis of mRNA translation profiles in Saccharomyces cerevisiae,” 2003.

[65] D. J. Hogan, D. P. Riordan, A. P. Gerber, D. Herschlag, and P. O. Brown, “Diverse RNA-Binding Proteins Interact with Functionally Related Sets of RNAs, Suggesting an Extensive Regulatory System,” *PLoS Biol.*, vol. 6, no. 10, p. e255, Oct. 2008, doi: 10.1371/journal.pbio.0060255.

[66] Y. Xia, E. A. Franzosa, and M. B. Gerstein, “Integrated Assessment of Genomic Correlates of

Protein Evolutionary Rate," *PLoS Comput. Biol.*, vol. 5, no. 6, p. e1000413, Jun. 2009, doi: 10.1371/journal.pcbi.1000413.

[67] L. Loison, "Epigenetic inheritance and evolution: a historian's perspective," *Philos. Trans. R. Soc. B*, vol. 376, no. 1826, Jun. 2021, doi: 10.1098/RSTB.2020.0120.

[68] J. Gumna, A. Andrzejewska-romanowska, D. J. Garfinkel, and K. Pachulska-wieczorek, "RNA Binding Properties of the Ty1 LTR-Retrotransposon Gag Protein," *Int. J. Mol. Sci. 2021, Vol. 22, Page 9103*, vol. 22, no. 16, p. 9103, Aug. 2021, doi: 10.3390/IJMS22169103.

[69] J. Ashley, B. Cordy, D. Lucia, L. G. Fradkin, V. Budnik, and T. Thomson, "Retrovirus-like Gag Protein Arc1 Binds RNA and Traffics across Synaptic Boutons," *Cell*, vol. 172, no. 1–2, pp. 262-274.e11, Jan. 2018, doi: 10.1016/J.CELL.2017.12.022.

[70] A. Rein, S. A. K. Datta, C. P. Jones, and K. Musier-Forsyth, "Diverse interactions of retroviral Gag proteins with RNAs," *Trends Biochem. Sci.*, vol. 36, no. 7, pp. 373–380, Jul. 2011, doi: 10.1016/J.TIBS.2011.04.001.

[71] J. Clever, C. Sassetti, and T. G. Parslow, "RNA secondary structure and binding sites for gag gene products in the 5' packaging signal of human immunodeficiency virus type 1," *J. Virol.*, vol. 69, no. 4, pp. 2101–2109, Apr. 1995, doi: 10.1128/JVI.69.4.2101-2109.1995.

[72] G. R. Christine Guthrie;Fink, *Guide to Yeast Genetics and Molecular and Cell Biology, Part C | 978-0-12-182254-5 | Elsevier*. 2002.

[73] R. Trivers, " The Evolution of Sex The Masterpiece of Nature: The Evolution and Genetics of Sexuality. Graham Bell ," *Q. Rev. Biol.*, vol. 58, no. 1, pp. 62–67, Mar. 1983, doi: 10.1086/413059.

[74] C. M. Thomas and K. M. Nielsen, "Mechanisms of, and Barriers to, Horizontal Gene Transfer between Bacteria," *Nat. Rev. Microbiol. 2005 39*, vol. 3, no. 9, pp. 711–721, Sep. 2005, doi: 10.1038/nrmicro1234.

[75] C. L. Schneider, "Bacteriophage-Mediated Horizontal Gene Transfer: Transduction," 2021, doi: 10.1007/978-3-319-41986-2_4.

[76] "Bacterial Conjugation - Google ספרים." [Online]. Available: https://books.google.co.il/books?hl=iw&lr=&id=hmjmBwAAQBAJ&oi=fnd&pg=PA1&dq=conjugation+bacteria&ots=ihHInFSYp5&sig=PQur8arvU_yuXiNxBP1fZtkWQ9s&redir_esc=y#v=onepage&q=conjugation bacteria&f=false. [Accessed: 13-Dec-2021].

[77] T. Lawley, B. M. Wilkins, and L. S. Frost, "Bacterial Conjugation in Gram-Negative Bacteria,"

Plasmid Biol., pp. 203–226, Apr. 2014, doi: 10.1128/9781555817732.CH9.

[78] D. Dubnau, "Genetic competence in Bacillus subtilis," *Microbiol. Rev.*, vol. 55, no. 3, pp. 395–424, Sep. 1991, doi: 10.1128/MR.55.3.395-424.1991.

[79] M. S. Fox, "SOME FEATURES OF GENETIC RECOMBINATION IN PROCARYOTES," 1978.

[80] P. Stenberg and A. Saura, "Cytology of Asexual Animals," *Lost Sex Evol. Biol. Parthenogenes.*, vol. 9789048127702, pp. 63–74, 2009, doi: 10.1007/978-90-481-2770-2_4.

[81] T. De Meeûs, F. Prugnolle, and P. Agnew, "Asexual reproduction: Genetics and evolutionary aspects," *Cell. Mol. Life Sci. 2007 6411*, vol. 64, no. 11, pp. 1355–1372, Mar. 2007, doi: 10.1007/S00018-007-6515-2.

[82] W. M. Lewis, "The cost of sex," *Experientia. Suppl.*, vol. 55, pp. 33–57, 1987, doi: 10.1007/978-3-0348-6273-8_2.

[83] L. E. Hawker, "The physiology of reproduction in fungi.," *Physiol. Reprod. fungi.*, 1956.

[84] B. O. Bengtsson, B. O. Bengtsson, and A. Ceplitis, "The balance between sexual and asexual reproduction in plants living in variable environments," *Artic. J. Evol. Biol.*, 2000, doi: 10.1046/j.1420-9101.2000.00187.x.

[85] A. Ruvinsky, "Sex, Meiosis and Multicellularity," *Acta Biotheor. 1997 452*, vol. 45, no. 2, pp. 127–141, Jun. 1997, doi: 10.1023/A:1000334022255.

[86] S. P. Otto, "Selective Interference and the Evolution of Sex," *J. Hered.*, vol. 112, no. 1, pp. 9–18, Mar. 2021, doi: 10.1093/JHERED/ESAA026.

[87] N. Colegrave, "Sex releases the speed limit on evolution," *Nat. 2002 4206916*, vol. 420, no. 6916, pp. 664–666, Dec. 2002, doi: 10.1038/nature01191.

[88] Y. Kim and H. A. Orr, "Adaptation in Sexuals vs. Asexuals: Clonal Interference and the Fisher-Muller Model," *Genetics*, vol. 171, no. 3, pp. 1377–1386, Nov. 2005, doi: 10.1534/GENETICS.105.045252.

[89] M. J. McDonald, D. P. Rice, and M. M. Desai, "Sex speeds adaptation by altering the dynamics of molecular evolution," *Nature*, vol. 531, no. 7593, pp. 233–236, Mar. 2016, doi: 10.1038/nature17143.

[90] T. F. C. MacKay, E. A. Stone, and J. F. Ayroles, "The genetics of quantitative traits: Challenges and prospects," *Nature Reviews Genetics*, vol. 10, no. 8. Nature Publishing Group, pp. 565–577, Aug-2009, doi: 10.1038/nrg2612.

[91] M. Lynch and B. Walsh, "Genetics and Analysis of Quantitative Traits."

[92] M. Lynch, “THE GENETIC INTERPRETATION OF INBREEDING DEPRESSION AND OUTBREEDING DEPRESSION,” *Evolution (N. Y).*, vol. 45, no. 3, pp. 622–629, May 1991, doi: 10.1111/j.1558-5646.1991.tb04333.x.

[93] J. A. Zeh and D. W. Zeh, “The evolution of polyandry I: Intragenomic conflict and genetic incompatibility,” *Proc. R. Soc. B Biol. Sci.*, vol. 263, no. 1377, pp. 1711–1717, 1996, doi: 10.1098/rspb.1996.0250.

[94] M. Turelli and H. A. Orr, “Dominance, epistasis and the genetics of postzygotic isolation,” *Genetics*, vol. 154, no. 4, pp. 1663–1679, Apr. 2000.

[95] A. J. Dagilis, M. Kirkpatrick, and D. I. Bolnick, “The evolution of hybrid fitness during speciation,” *PLoS Genet.*, vol. 15, no. 5, p. e1008125, 2019, doi: 10.1371/journal.pgen.1008125.

[96] R. Shapira, T. Levy, S. Shaked, E. Fridman, and L. David, “Extensive heterosis in growth of yeast hybrids is explained by a combination of genetic models,” *Heredity (Edinb).*, vol. 113, 2014, doi: 10.1038/hdy.2014.33.

[97] R. Shapira and L. David, “Genes with a Combination of Over-Dominant and Epistatic Effects Underlie Heterosis in Growth of Saccharomyces cerevisiae at High Temperature,” *Front. Genet.*, vol. 7, May 2016, doi: 10.3389/fgene.2016.00072.

[98] D. Hedgecock and J. P. Davis, “Heterosis for yield and crossbreeding of the Pacific oyster Crassostrea gigas,” *Aquaculture*, vol. 272, no. SUPPL. 1, pp. S17–S29, Jan. 2007, doi: 10.1016/j.aquaculture.2007.07.226.

[99] R. H. Herbst *et al.*, “Heterosis as a consequence of regulatory incompatibility,” *BMC Biol.*, vol. 15, no. 1, p. 38, Dec. 2017, doi: 10.1186/s12915-017-0373-7.

[100] D. Bar-Zvi, O. Lupo, A. A. Levy, and N. Barkai, “Hybrid vigor: The best of both parents, or a genomic clash?,” *Current Opinion in Systems Biology*, vol. 6. Elsevier Ltd, pp. 22–27, 01-Dec-2017, doi: 10.1016/j.coisb.2017.08.004.

[101] J. P. Bernardes, R. B. Stelkens, and D. Greig, “Heterosis in hybrids within and between yeast species,” *J. Evol. Biol.*, vol. 30, no. 3, pp. 538–548, Mar. 2017, doi: 10.1111/jeb.13023.

[102] X. Wei and J. Zhang, “The optimal mating distance resulting from heterosis and genetic incompatibility,” *Sci. Adv.*, vol. 4, no. 11, p. eaau5518, Nov. 2018, doi: 10.1126/sciadv.aau5518.

[103] B. N. Willson, Mary F., “Mate Choice in Plants: Tactics, Mechanisms, and Consequences,” *Princeton University Press, Princeton, NJ*, 1983. [Online]. Available:

https://books.google.co.il/books?hl=iw&lr=&id=BOU9DwAAQBAJ&oi=fnd&pg=PA3&ots=hshYf__yE1&sig=w4foOSgKQMyID6NX4qHce8Y56ac&redir_esc=y#v=onepage&q&f=false. [Accessed: 31-Mar-2020].

[104] R. D. HOWARD, R. S. MARTENS, S. A. INNIS, J. M. DRNEVICH, and J. HALE, "Mate choice and mate competition influence male body size in Japanese medaka," *Anim. Behav.*, vol. 55, no. 5, pp. 1151–1163, May 1998, doi: 10.1006/ANBE.1997.0682.

[105] L. Aquiloni and F. Gherardi, "Mutual mate choice in crayfish: large body size is selected by both sexes, virginity by males only," *J. Zool.*, vol. 274, no. 2, pp. 171–179, Feb. 2008, doi: 10.1111/j.1469-7998.2007.00370.x.

[106] M. D. Jennions and M. Petrie, "VARIATION IN MATE CHOICE AND MATING PREFERENCES : A REVIEW OF CAUSES AND CONSEQUENCES," *Biol. Rev.*, vol. 72, pp. 283–327, 1997, doi: 10.1017/S0006323196005014.

[107] M. L. THOMAS and L. W. SIMMONS, "Crickets detect the genetic similarity of mating partners via cuticular hydrocarbons," *J. Evol. Biol.*, vol. 24, no. 8, pp. 1793–1800, Aug. 2011, doi: 10.1111/j.1420-9101.2011.02319.x.

[108] M. J. E. Charpentier, J. C. Crawford, M. Boulet, and C. M. Drea, "Message 'scent': lemurs detect the genetic relatedness and quality of conspecifics via olfactory cues," *Anim. Behav.*, vol. 80, no. 1, pp. 101–108, Jul. 2010, doi: 10.1016/J.ANBEHAV.2010.04.005.

[109] B. D. NEFF and T. E. PITCHER, "Genetic quality and sexual selection: an integrated framework for good genes and compatible genes," *Mol. Ecol.*, vol. 14, no. 1, pp. 19–38, Nov. 2004, doi: 10.1111/j.1365-294X.2004.02395.x.

[110] M. Israeli, D. Kristt, Y. Nardi, and T. Klein, "Genetic Considerations in Human Sex-Mate Selection: Partners Share Human Leukocyte Antigen but not Short-Tandem-Repeat Identity Markers," *Am. J. Reprod. Immunol.*, vol. 71, no. 5, pp. 467–471, May 2014, doi: 10.1111/aji.12213.

[111] S. Takayama and A. Isogai, "SELF-INCOMPATIBILITY IN PLANTS," *Annu. Rev. Plant Biol.*, vol. 56, no. 1, pp. 467–489, Jun. 2005, doi: 10.1146/annurev.arplant.56.032604.144249.

[112] A. R. Isles, M. J. Baum, D. Ma, E. B. Keverne, and N. D. Allen, "Urinary odour preferences in mice," *Nature*, vol. 409, no. 6822. Nature Publishing Group, pp. 783–784, 15-Feb-2001, doi: 10.1038/35057323.

[113] S. Jacob, M. K. McClintock, B. Zelano, and C. Ober, "Paternally inherited HLA alleles are

associated with women's choice of male odor," *Nat. Genet.*, vol. 30, no. 2, pp. 175–179, Feb. 2002, doi: 10.1038/ng830.

[114] I. Herskowitz, "Life cycle of the budding yeast Saccharomyces cerevisiae.," *Microbiol. Rev.*, vol. 52, no. 4, pp. 536–53, Dec. 1988.

[115] J. Peter *et al.*, "Genome evolution across 1,011 Saccharomyces cerevisiae isolates," *Nature*, vol. 556, no. 7701, pp. 339–344, Apr. 2018, doi: 10.1038/s41586-018-0030-5.

[116] F. A. Cubillos, E. J. Louis, and G. Liti, "Generation of a large set of genetically tractable haploid and diploid Saccharomyces strains," 2009, doi: 10.1111/j.1567-1364.2009.00583.x.

[117] J. E. Haber, "Mating-Type Genes and MAT Switching in Saccharomyces cerevisiae," *Genetics*, vol. 191, no. 1, pp. 33–64, May 2012, doi: 10.1534/genetics.111.134577.

[118] A. H. Y. Tong and C. Boone, "High-Throughput Strain Construction and Systematic Synthetic Lethal Screening in Saccharomyces cerevisiae," 2007, doi: 10.1016/S0580-9517(06)36016-3.

[119] N. Yachie *et al.*, "Pooled-matrix protein interaction screens using Barcode Fusion Genetics.," *Mol. Syst. Biol.*, vol. 12, no. 4, p. 863, Apr. 2016, doi: 10.15252/MSB.20156660.

[120] F. Galton, "Regression Towards Mediocrity in Hereditary Stature.," *J. Anthropol. Inst. Gt. Britain Irel.*, vol. 15, pp. 246–263, 1886.

[121] Y. S. Aulchenko *et al.*, "Predicting human height by Victorian and genomic methods," *Eur. J. Hum. Genet. 2009 178*, vol. 17, no. 8, pp. 1070–1075, Feb. 2009, doi: 10.1038/ejhg.2009.5.

[122] R. Chaix, C. Cao, and P. Donnelly, "Is mate choice in humans MHC-dependent?," *PLoS Genet.*, vol. 4, no. 9, p. e1000184, Sep. 2008, doi: 10.1371/journal.pgen.1000184.

[123] H. A. Murphy, H. A. Kuehne, C. A. Francis, and P. D. Sniegowski, "Mate choice assays and mating propensity differences in natural yeast populations," *Biol. Lett.*, vol. 2, no. 4, pp. 553–556, Dec. 2006, doi: 10.1098/rsbl.2006.0534.

[124] M. L. Head, G. M. Kozak, and J. W. Boughman, "Female mate preferences for male body size and shape promote sexual isolation in threespine sticklebacks.," *Ecol. Evol.*, vol. 3, no. 7, pp. 2183–96, Jul. 2013, doi: 10.1002/ece3.631.

[125] I. Stefanini, L. Dapporto, L. Berná, M. Polsinelli, S. Turillazzi, and D. Cavalieri, "Social wasps are a *Saccharomyces* mating nest," *Proc. Natl. Acad. Sci.*, vol. 113, no. 8, pp. 2247–2251, Feb. 2016, doi: 10.1073/pnas.1516453113.

[126] K. A. Curran, N. J. Morse, K. A. Markham, A. M. Wagman, A. Gupta, and H. S. Alper, "Short Synthetic Terminators for Improved Heterologous Gene Expression in Yeast," *ACS Synth. Biol.*,

vol. 4, no. 7, pp. 824–832, Jul. 2015, doi: 10.1021/SB5003357.

[127] M. Martin, “Cutadapt removes adapter sequences from high-throughput sequencing reads,” *EMBnet.journal*, vol. 17, no. 1, pp. 10–12, May 2011, doi: 10.14806/EJ.17.1.200.

12. Appendices

Table S1. List of strains made and used in the project. All letters are the Standarized names of strains as describe in Peter et al. [115]. All strains except for PAR1 are taken from [115]. For some strains, two colonies with two different barcodes were made; the 1 or 2 at the end of the standardized name represent the colony taken and used in the research.

matA	matα
BKF1 , BKT1 , BKH1 , BLA1 , BHN1 , BHB1 , BKM1 , AKB1 , ALG1 , AIK1 , ALH1 , AIM1 , AKS1 , AKG1 , AKV1 , ANE1 , AVK1 , BBI1 , BCB1 , BAS1 , BBA1 , BEE1 , BHV1 , BIM1 , BKE1 , BIA1 , BLV1 , BQC1 , BMA1 , BNT1 , BQR1 , BMB1 , BPM1 , BLE1 , BLE2 , BHC1 , BHC2 , BKA1 , BKA2 , BLH1 , BHF1 , BHT1 , BMC1 , BLL1 , BQG1 , BNK1 , BPQ1 , BQH1 , BRA1 , BQI1 , BLP1 , BMH1 , BNE1 , BPT1 , BRD1 , BNP1 , BPV1 , BQM1 , BLS1 , BMK1 , AST1 , BES1 , BSD1 , BSD2 , BTN1 , BTN2 , BVV1 , BSL1 , BSL2 , BVK1 , CAD1 , BSB1 , CAE1 , BTH1 , CAF1 , CCG1 , CPK1 , CQD1 , CQF1 , CQG1 , CQP1 , SACE-YBG1 , SACE-YBH1 , SACE-YBW1 , CCQ1 , CCV1 , CDF1 , CDQ1 , CEQ1 , CFE1 , CGL1 , CGV1 , CHC1 , CHD1 , CHD2 , CFR1 , SACE-YAB1 , SACE-YAM1 , SACE-YCH1 , SACE-YCH2 , SACE-YCM1 , SACE-YCQ1 , PAR1	AKB1 , AKP1 , ALG1 , ANG1 , AIM1 , AKV1 , BBI1 , BAS1 , BAK1 , BEH1 , BIN1 , BLA1 , BHM1 , BHN1 , BHB1 , BLE1 , BHF1 , BKR1 , BLL1 , BNI1 , BQG1 , BNK1 , BPQ1 , BQH1 , BRA1 , BLN1 , BQI1 , BRB1 , BLP1 , BRC1 , BNE1 , BPT1 , BRD1 , BMK1 , BQP1 , BLV1 , BQQ1 , BMA1 , BQR1 , BVK1 , CAD1 , BSB1 , BTH1 , AHP1 , CQD1 , APV1 , AST1 , CCV1 , CEK1 , SACE-YAR1 , SACE-YBU1 , CEQ1

Table S2. List of pairs in the mating efficiency experiment.

pair#	matA	matα	pair#	matA	matα	pair#	matA	matα	pair#	matA	matα	pair#	matA	matα
1	AKP	AKV	41	AKV	AIK	81	BLE	CQG	121	BQG	CQG	161	YBU	BLE
2	AKV	AKP	42	BLP	CFR	82	BLV	CQG	122	BLH	BBA	162	YBU	CFR
3	BLV	BQG	43	BHF	YAM	83	BRA	BLN	123	BRC	BNE	163	BHB	YAM
4	BSB	CFR	44	YAR	APV	84	APV	BLV	124	APV	BRD	164	BHB	BBI
5	BPQ	BRD	45	BQI	APV	85	AKP	BLN	125	BLH	BRD	165	BLN	BRD
6	BRC	BTH	46	BMA	BPM	86	CQD	CCQ	126	BRC	CPK	166	BAK	CQG
7	BRC	BHC	47	AKP	BMC	87	AKP	AKI	127	BLV	CQF	167	YAR	CAE
8	BPQ	BTH	48	AKV	BMC	88	AKV	AKI	128	BMA	AKP	168	YAR	BBA
9	BTH	CBG	49	BLV	ALH	89	BLP	AKS	129	BMA	AKV	169	YAR	BNE
10	BRC	BBI	50	BSB	AKG	90	BHF	AKS	130	BMA	BQG	170	CQD	BRD
11	BTH	BNE	51	BPQ	BLP	91	YAR	AKG	131	BMA	BMC	171	BHN	BBI
12	BQI	BAS	52	BRC	BLP	92	BQI	BKH	132	AHP	YAM	172	BQG	BRD
13	AKP	BHC	53	BRC	BRD	93	BQI	BLL	133	AHP	CAE	173	BAK	CPK
14	BHC	BQM	54	BPQ	BNP	94	BMA	YBW	134	AHP	CCQ	174	BAK	CQF
15	CQD	CAE	55	BTH	BMA	95	CAD	BMC	135	AHP	ANE			
16	BQR	BQM	56	BRC	BMA	96	BTH	YAM	136	BAK	CFR			
17	APV	BBA	57	BTH	AKG	97	BTH	CAE	137	AKB	BBI			
18	BQI	CAE	58	BQI	CDF	98	BVK	CCQ	138	AKB	BAS			
19	APV	BHF	59	AKP	CAE	99	BRB	ANE	139	AKB	BBA			
20	BLH	CAE	60	BHC	BNP	100	BTH	BBI	140	BLP	BBI			
21	YAR	BQG	61	CQD	CDF	101	BTH	BPT	141	BHC	BPT			
22	BHC	BRF	62	BQR	BHB	102	BVK	BTH	142	BLP	BTH			
23	BTH	BLS	63	APV	CDF	103	BHC	BBI	143	BVK	BBI			
24	BQH	BRF	64	BQI	ALH	104	BLP	YAM	144	BAK	CAE			
25	BQR	APV	65	APV	CCQ	105	BQI	CFR	145	BAK	APV			
26	BLH	YBG	66	BLH	CCG	106	BHC	BTH	146	BAK	BBI			
27	BSB	BBA	67	YAR	CDF	107	BVK	CFR	147	CCV	BHF			
28	BPQ	BBI	68	BHC	BHB	108	BQR	BRD	148	BRB	YAM			
29	BAK	BKE	69	BTH	BQG	109	BLE	CAE	149	BQG	BTH			
30	BMA	BKE	70	BQH	BHN	110	BRB	BPT	150	BHB	CFR			
31	BRB	AIM	71	BQR	CCG	111	BHN	BQM	151	CCV	BBA			
32	BQP	BLH	72	BLH	CCQ	112	BQI	BPT	152	CCV	BRD			
33	BRC	BLA	73	BSB	ANE	113	BQI	CQD	153	CCV	CAE			
34	BLE	BAS	74	BPQ	CDQ	114	BHF	BBA	154	AKB	YAM			
35	BLV	BIM	75	BAK	CCQ	115	APV	BQM	155	BHB	BPT			
36	BRA	BLS	76	BMA	AKG	116	BLH	BLE	156	BRB	BQM			
37	APV	BHN	77	BRB	YBW	117	BLN	CFR	157	BHN	BPT			
38	AKP	BAS	78	BQP	AIM	118	BQG	YAM	158	BAK	CQD			
39	CQD	BLV	79	BQR	ANE	119	BLN	BBI	159	YBU	BBA			
40	AKP	BNE	80	BRC	CDQ	120	BHN	BRD	160	YBU	BQM			