

Thesis for the degree Master of Science

Submitted to the Scientific Council of the Weizmann Institute of Science Rehovot, Israel עבודת גמר (תזה) לתואר מוסמך למדעים

מוגשת למועצה המדעית של מכון ויצמן למדע רחובות, ישראל

By Ruthie Golomb <sub>מאת</sub> רותי גולומב

חקר דינמי של משחקים אבולוציוניים דרך ייצור של חלבוני טובין ציבוריים

Studying evolutionary game dynamics through public good protein production

Advisor: Prof. Yitzhak Pilpel מנחה: פרופ' יצחק פלפל

חודש ושנה עבריים אפריל, תש"פ

Month and Year April, 2020

### Abstract

The origin of cooperation is a key question in evolutionary biology. We investigate this question by studying the cooperative interaction of secretion of 'public good' proteins in microorganisms. 'Public good' proteins are secreted proteins that serve a beneficial function for the entire community. An established model to study this interaction in yeast is the invertase protein Suc2, a secreted enzyme that hydrolyzes sucrose into digestible monosaccharides outside of the cell. Thus, the producer of this enzyme benefits its neighboring cells and can be subject to exploitation from 'cheating' non-producing cells. While the dynamics of the producer and the 'cheater' have been previously studied, real world organisms may often exhibit non-dichotomous, continuously varying levels of cooperation. We built a large synthetic library that contains thousands of variants with a range of cooperative strategies, by potentially altering their secretion level of the invertase protein. We do so by altering protein secretion motifs naturally found in the SUC2 gene, such as the signal peptide and an RNA motif termed SECReTE that enhances protein secretion. Since we manipulate either an RNA motif or a signal peptide that will be cleaved, we can modulate invertase secretion levels without altering its active protein sequence. Although here we use these motifs as a potentially artificial tool to influence the public good secretion and thus cooperation, ultimately, this enables us to study the natural genetic basis for cooperation that may have allowed evolution to tune it.

We competed the synthetic library of the Suc2 SECReTE variants in a competition assay to study the population dynamics in this complex community in which different members feature different levels of cooperation with all others. We compared between the competition assay in two distinct conditions, the first being sucrose which enables the interactions and cooperation between variants, and the second being galactose which enables us to isolate the cost of each variant. In this comparison, we find a strong outlier population (~280 variants) that shows a higher fitness on the sucrose medium than expected based on their cost. Additionally, we find a smaller outlier population (~120) with the opposite trend, a lower fitness on sucrose than expected based on their cost. Based on features presumed to have an effect on the SECReTE motif and thus the Suc2 secretion level, it seems the 'successful' outlier population is cheater-like variants while the unsuccessful outliers are a population of exploited cooperators. While this is the picture we get from the experiment in the allotted time, this trend may change or fluctuate over longer evolutionary timescales due to the inherent instability of cheater-governed sub-populations.

Finally, we also investigated a potential public good protein that is not yet established as one. The catalase protein provides protection from oxidative stress by breaking down hydrogen peroxide into water and oxygen. We find that the catalase is released to the medium and provides protection to its neighboring cells however, it is unresolved whether this is a result of active secretion by the cell or a passive release due to lysis from the stress. This may present a new notion to further explore of public good cooperation through cell death and lysis.

## Table of Contents

| Abstract2   |
|---|
| Table of Contents   4   |
| List of abbreviations   |
| Introduction7   |
| Suc2 public good system   |
| Potential public good catalase  |
| Goals11   |
| Materials and methods12   |
| Strains, plasmids and media   |
| Growth experiments  |
| Creation of the Suc2 secretion variant sub-libraries                                |
| Design of the variant sub-libraries13   |
| Cloning of plasmid sub-libraries  |
| Bacterial transformation of the sub-libraries14                                     |
| Yeast transformation of the sub-libraries15   |
| Competition assay by serial dilutions16   |
| Competition assay by chemostat  |
| Sequencing of the competition assays  |
| Conditioned media transfer experiment   |
| Catalase activity assay   |
| Results   |
| Set-up of the Suc2 secretion library system   |
| Growth measurements of example Suc2 secretion variants                              |
| Simulation of co-culture competition assay between three Suc2 secretion variants 23 |
| Experimental co-culture competition assay between three Suc2 secretion variants 24  |
| Design of the large synthetic library of Suc2 secretion variants                    |
| Creation of the large synthetic library of Suc2 secretion variants                  |

| Appendix  | 63 |
|---|----|
| Acknowledgments   | 62 |
| References  | 59 |
| Future plans  |    |
| Catalase as a potential public good                       | 56 |
| Competition assay of the Suc2 SECReTE variant sub-library |    |
| Set-up of the Suc2 secretion library system               |    |
| Discussion  | 49 |
| Catalase as a potential public good                       |    |
| Variant-level analysis                                    |    |
| Population-level analysis                                 |    |
| Competition assay of the Suc2 SECReTE variant sub-library |    |

## List of abbreviations

**SRP:** Signal recognition particle

**ER:** Endoplasmic reticulum

**SECReTE:** Secretion-enhancing cis regulatory targeting element

UTR: Untranslated region

**SNP:** Single-nucleotide polymorphisms

tAI: tRNA Adaptation Index

## Introduction

Cooperation is a natural phenomenon that can be found at all levels of biological organization. However, the Darwinian theory of evolution is based on competition and survival of the fittest which should, therefore, reward selfish behavior. This raises a puzzling and fundamental question in evolutionary biology, how do systems of cooperation arise?

There are various mechanisms suggested to enable the preservation of cooperation in nature, such as kin selection<sup>1</sup>, reciprocity<sup>2</sup>, and group selection<sup>3</sup>. To help us further study this preservation, we can look at an important element of cooperation, which is the interaction by secretion of public good proteins. Public good proteins are proteins that serve a beneficial function for the entire community. While this interaction of secreting beneficial proteins is advantageous for all, it is quite a costly action for the individual secreting cell. This results in an inherent conflict for the producing cell between saving its own resources to increase its fitness and providing benefit to the entire community. Not only is the cell 'tempted' not to cooperate to save its own resources, but there is also a concern of other cells exploiting their costly secretion. These exploiters are called 'cheaters', nonproducing mutants that exploit the public goods.

#### Suc2 public good system

The dynamics between the cooperator and the cheater in a public good interaction has been studied in yeast with the protein Suc2.<sup>4–9</sup> Suc2 is an invertase that hydrolyses sucrose into glucose and fructose (Figure 1). There are two isoforms of the *SUC2* gene, one being the predominant form that is secreted to the periplasmic space and hydrolyzes the sucrose externally, and the second is an internal invertase.<sup>10</sup> In *S. cerevisiae*, sucrose hydrolysis occurs primarily extracellularly and in several of the common lab strains, it occurs exclusively extracellularly as the yeast cells are unable to import the sucrose (due to being maltose-negative).<sup>11</sup>

When the cell externally hydrolyzes sucrose, it retains ~1% of its monosaccharide products, while 99% of the products become public benefit.<sup>5</sup> This results in an environment that is susceptible to cheaters taking advantage. The cheaters can uptake the monosaccharides produced by the other cells without paying the cost of producing and secreting the Suc2 protein themselves. In previous studies, it has been shown that

the fitness of the cooperator versus the cheater is cell-density dependent.<sup>4</sup> In an environment of high cell-density, there is plenty of shared hydrolyzed products and as a result, the cheater has an advantage, as it is not paying the cost of cooperation. In contrast, in a low cell-density environment, with few shared goods, the cooperator has the advantage since it retains a percentage of its own products. Accordingly, depending on the conditions, while the cheater can invade a population of cooperators, the cooperators can invade a population of cheaters as well. Consequently, a steady-state of coexistence between the cooperator and the cheater can be found.<sup>5</sup>



**Figure 1. Extracellular hydrolysis of sucrose into glucose and fructose by the cell wall bound invertase protein.** The monosaccharide products diffuse into the shared medium while a small percentage is imported into the cell by a hexose transporter. Taken from <sup>12</sup>.

The research on the Suc2 system has focused on two strategies, cooperating and cheating. However, in nature, we may not expect to see a binary decision of yes or no to cooperation. In reality, an organism can have varying levels of production, secretion, and retention of the public good since regulation in a cell works on a continuous scale. The ability to represent many different strategies along this continuous scale of 'decision making' may enable us to study this interaction on a more complex level.

There are two motifs found in the *SUC2* gene that have evolved, and that can be mutated to alter the secretion level of Suc2 and thus, represent different levels of cooperation in this interaction. The first is the signal peptide, which is a commonly known targeting signal to the protein secretion pathway.<sup>13</sup> For proteins that have this amino acid sequence motif, translation begins in the cytoplasm and subsequently, due to a signal peptide that emerges from the ribosome, it is recognized by the signal recognition particle (SRP) protein which enables co-translational protein translocation to the endoplasmic reticulum (ER) (Figure 2A).<sup>14</sup> The signal peptide can even be artificially added to a protein that is not naturally secreted and cause its secretion.<sup>15</sup>

Indeed biotechnological efforts to ectopically express proteins are often based on adding a signal peptide to a protein of interest so it will be secreted and hence easily purified from cells.<sup>16,17</sup> However, it is not fully known which signal peptide sequence is optimal and which alterations to the sequence enhance or decrease the level of secretion. Further, while this is the classic model for protein secretion, other secretion pathways have been found as well.<sup>18–20</sup>

A second way to alter the secretion levels of the Suc2 protein is with a recent discovery by our lab, in collaboration with the Gerst lab, of an RNA motif that affects and regulates secretion of proteins from a cell. This motif was termed a secretion-enhancing cis regulatory targeting element (SECReTE). The motif consists of  $\geq 10$  nucleotide triplet repeats enriched with a pyrimidine (C/U) every third base. Since the motif is encoded in the RNA, it is a pre-translational signal which can be found in the coding region of the gene or in the 5' or 3' untranslated region (UTR). It was shown that when generating synonymous mutations that increase the number of SECReTE repeats, there is an increased level of protein secretion, and vice versa.<sup>21</sup> The SECReTE motif is presumed to serve as a signal for mRNA localization to the ER, meaning that besides the known co or post-translational mechanisms, mRNA translocation may be able to occur pre-translationally as well (Figure 2B). This motif is unique in that it enables us to alter the secretion level of the Suc2 protein without having to make any alterations on the protein level.



**Figure 2. Overview of the role of the signal peptide and the SECRETE motif in protein secretion. (A)** The translated signal sequence is recognized by the SRP protein that co-translationally targets the translating ribosome to the ER. Adapted from <sup>22</sup> (B) The mRNA transcript containing the SECRETE motif is recognized by a putative SECRETE binding protein that facilitates translocation of the mRNA to the ER. Presumably, the enhanced association of SECRETE containing transcripts to the ER promotes secretion. Adapted from <sup>21</sup>.

With these two secretion-affecting motifs, we can produce a large library of mutant variants in a systematic manner that tunes their secretion levels of our public good

protein Suc2. This will allow us to explore cooperation strategies at a continuum scale. This can be done by creating both synonymous and non-synonymous mutations in these motifs that will affect the secretion of the Suc2 protein. Next, we can compete the entire pooled synthetic library of variants in a competition assay. Throughout the competition, we can measure the relative frequency of each genotype by deep sequencing and thus assess each variant's fitness.

#### Potential public good catalase

While the invertase Suc2 is the most studied public good protein in yeast, there are other proteins that seem to be candidates for such cooperative interactions. One such example is the catalase protein, an important enzyme that protects the cell from oxidative stress by breaking down hydrogen peroxide into water and oxygen.<sup>23,24</sup> Although some fungal species are known to secrete catalase<sup>25,26</sup>, in *S. cerevisiae* it isn't commonly known to be a secreted protein and is considered to work intracellularly. However, there is evidence pointing to the existence of both cell-surface-bound and external catalase activity in *S. cerevisiae*, suggesting it is secreted.<sup>27</sup>

There is a key difference between Suc2 and the potential public good catalase in that the major function of Suc2 is extracellular, while the major function of catalase is intracellular.<sup>28</sup> Although there is also an intracellular isoform of the Suc2 protein, the sucrose hydrolysis occurs extracellularly since the cell typically does not import the sucrose.<sup>10</sup> In contrast, the main activity of catalase is found to be intracellular, majorly to protect cells from either internal byproducts or external factors that cause oxidative stress in the cell.<sup>29-31</sup> However, the secretion of catalase may be an advantageous protective mechanism to neutralize exogenous oxidative threats and when there is a collective secretion, we could hypothesize that it may provide a stronger protection. It is interesting to think about the possibility that, like in the case of Suc2 and other public goods, this set-up would give opportunity for cheaters to take advantage since the external catalase will help protect all cells in the vicinity. Yet when catalase remains intracellular, the benefits of the enzyme's activity are reserved only to the cell that produced it and does not allow exploitation. The potential ability to either share the benefits or retain them exclusively to the producing cell could result in an extra layer of complexity in this cooperative protein secretion interaction.

## Goals

The goal of this project is to advance our understanding of the cooperator-cheater interaction in public good secretion systems by examining it in a more complex yeast community, consisting of many different strategies ranging from exploiters to ultra-cooperators. By creating these various strategies through altering sequence signatures, such as the signal peptide and the SECReTE motif, we can investigate the genetic sequences that might govern this complex activity in public good interactions. In order to achieve this goal, I focused during my thesis on the following aims: <u>Aims</u>:

- Design and create a library of Suc2 secretion variants that could represent various strategies in the cooperative interaction.
- Compete the synthetic secretion variant library in a competition assay to study the dynamics and fitness of the variants throughout and investigate evolutionarily stable strategies, as well as delve into the genetic sequences that govern the observed dynamics.
- Explore the potential use of the enzyme catalase as an additional model system to study public good interactions.

## Materials and methods

Please note, all listed primers and PCR protocols can be found in the appendix in Table 2 and Table 3.

#### Strains, plasmids, and media

All *Saccharomyces cerevisiae* strains used in this work are based on the BY4741 (MATa his $3\Delta 1$ , leu $2\Delta 0$ , met $15\Delta 0$ , ura $3\Delta 0$ ) genetic background and are listed below in Table 1. List of provided strainsTable 1.

| Species       | Strain                        | Genotype  | Reference |
|---------------|-------------------------------|---|-----------|
| S. cerevisiae | BY4741                        | MATa his $3\Delta 1$ leu $2\Delta 0$ met $15\Delta 0$ ura $3\Delta 1$   | 21        |
| S. cerevisiae | BY4741<br>Δsuc2::NAT          | MATa his3∆1 leu2∆0 met15∆0 ura3∆1<br>SUC2::NAT  | 21        |
| S. cerevisiae | BY4741<br>SUC2<br>SECReTE (+) | MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 met15 $\Delta$ 0 ura3 $\Delta$ 1<br>SUC2- altered to increase SECReTE<br>motif | 21        |
| S. cerevisiae | BY4741<br>SUC2<br>SECReTE (-) | MATa his3 $\Delta$ 1 leu2 $\Delta$ 0 met15 $\Delta$ 0 ura3 $\Delta$ 1<br>SUC2- altered to decrease SECReTE<br>motif | 21        |
| S. cerevisiae | BY4741<br>ΔCTT1               | MATa his $3\Delta 1$ leu $2\Delta 0$ met $15\Delta 0$ ura $3\Delta 1$<br>$\Delta CTT1$                              | 32        |

#### Table 1. List of provided strains

pGS2223 – Plasmid created for the construction of the Suc2 secretion variant library. It is based on the integrative plasmid pCfB2223 from the EasyClone 2.0 Yeast Toolkit, which includes a Kan resistance gene and homologous regions for an integrative site on chromosome X.<sup>33</sup> The WT *SUC2* gene was cloned into this plasmid to create pGS2223 using the NEBuilder HiFi DNA Assembly according to its reaction protocol.

**YPD** - Rich media composed of 10 g/L yeast extract, 20 g/L peptone and 20 g/L glucose.

**YP-Gly** - Media composed of 10 g/L yeast extract, 20 g/L peptone, 30 g/L glycerol.

**SD** - Media composed of 6.7 g/L nitrogen base, 1.5 g/L mix of all amino acids and 20 g/L glucose.

**SC-sucrose** - Media composed of 6.7 g/L nitrogen base, 1.5 g/L mix of all amino acids and 20 g/L sucrose.

**SC-low sucrose**- Media composed of 6.7 g/L nitrogen base, 1.5 g/L mix of all amino acids. The sucrose and glucose were added after the media was autoclaved for a final concentration of 0.05% and 0.005, respectively.

**SC-low galactose**- Media composed of 6.7 g/L nitrogen base, 1.5 g/L mix of all amino acids. The Galactose was added after the media was autoclaved for a final concentration of 0.1%.

LB – Media composed of tryptone 10 g/L, yeast extract 5 g/L, Nacl 10 g/L.

#### Growth experiments

Cells were grown for 48h on the indicated medium until reaching deep stationary phase and then were diluted into fresh medium (dilution factor 1:50). Growth experiments were done in 96-well plates with 150µl per well. In each plate, four strains were arranged in a checkerboard pattern to cancel out geographical effects. Plates were put in a shaking incubator set to 30°c and  $OD_{600}$  was measured every hour for 1-2 days by a plate reader (Spark). All measurements were done automatically using a robotic system (Thermo). The results of the growth experiment were analyzed using a designated MATLAB-GUI and when relevant, growth parameters were extracted, using the software Curveball.<sup>34</sup>

#### Creation of the Suc2 secretion variant sub-libraries

#### Design of the variant sub-libraries

The designed sequences were generated by custom codes written on MATLAB (MathWorks, Inc.).

#### Cloning of plasmid sub-libraries

~18,000 DNA fragments of 300bp each, that contain the designed sequences, were ordered from the company Twist. From these fragments, 200bp-long oligos were amplified into sub-libraries based on their location in the *SUC2* gene, with the middle 150bp being the variable region (See primers 1-12 and PCR 1). The 100bp that were not amplified are not part of the *SUC2* gene and were relevant for a previous method<sup>35</sup>

that we tried to create the secretion variant library, which proved not to work for our needs.

In parallel, the backbone plasmid pGS2223 (that the oligos will later be cloned into) was created by cloning the WT *SUC2* gene into the linearized integrative plasmid pCfB2223 from the EasyClone 2.0 Yeast Toolkit, using NEBuilder HiFi DNA Assembly cloning kit (See primers 13-14 and PCR 2). pGS2223 was then linearized, specific to each sub-library location, by performing a PCR reaction that amplified the vector using primers in opposite directions to each other, which then removed the relevant variable region (See primers 15-26 and PCR 3). The amplified oligo sub-libraries were cloned into the relevant linearized plasmids using NEBuilder HiFi DNA Assembly cloning kit with a 1:2 vector:insert ratio, replacing the removed region. The product was then cleaned with 1.5X SPRI beads (AMPure XP) according to the one-sided manufacture protocol.

#### Bacterial transformation of the sub-libraries

The clean product of the plasmid sub-libraries was transformed by electroporation into ElectroTen-Blue Electroporation Competent Cells (Agilent). For each sublibrary, between 4-8 electroporation reactions were performed (depending on the resulting efficiency) and 5.5ul of the SPRI-purified plasmids was added per reaction. The electroporation protocol was used from the Agilent guide called SureVector CRISPR Library Cloning Kit. In this protocol, the plasmid DNA was added to thawed ElectroTen-Blue Competent Cells and moved to a chilled cuvette which was then placed in the electroporator. After the cells were electroporated, they were immediately resuspended in a rich medium and moved to a fresh tube to recuperate in a shaking incubator set at 37°c for one hour. Next, the cells were plated on 20-30 LB+amp plates for selection. The plates were then scraped and all colonies from the same sub-library were pooled together reaching an estimated  $10^{12}$  cells suspended in ~40-60 ml of DDW (2ml was used to scrape each plate). The estimated cell count of  $10^{12}$  is based on the ~ $10^6$  single colonies grown per sub-library and estimating about  $10^6$  cells in each single colony. From the pooled culture, some glycerol stocks were made by suspending the culture to a final concentration of 30% glycerol, then being frozen and stored at -80°c. The plasmid sub-library was extracted directly from the rest of the pooled culture using Promega midiprep kit.

#### Yeast transformation of the sub-libraries

The purified plasmid sub-libraries were linearized using the restriction enzyme NotI. In each reaction, there was ~6ug of plasmid sub-library, 2ul NEB NotI-HF enzyme and 5ul CutSmart buffer in a total reaction volume of 50ul. The reaction was incubated at 37°c for 2 hours. Sixteen of these reactions were performed per sub-library to ensure sufficient material for the subsequent transformation. The linearized plasmid product was then cleaned with 1.5X SPRI beads and transformed into  $\Delta suc2$  yeast strain using heat shock.

For the heat shock transformation, cells were grown for 24h until reaching stationary phase and then diluted such that they reach mid-log ( $\sim 1 \times 10^7$  cells) the following morning. Then, cells were counted to verify they are in mid-log, and the appropriate amount of culture to contain  $\sim 1 \times 10^8$  cells was taken per sample. The culture was centrifuged at 3000g for 3 mins and the supernatant was removed. Next, the pellet of cells was washed twice with 1ml DDW per sample and was transferred to Eppendorf tubes. Then, 1µg of the plasmid library (in a total volume of 20 ul) was added and used to re-suspend the pellet. Next, the transformation mix was added to each sample and the cells were immediately re-suspended by vortex. The transformation mix contains 240ul of PEG 50% w/v, 36ul of Lithium Acetate 1M, 50 ul of boiled salmon sperm and 14ul water per sample. The cells were then incubated in a ThermoBlock at 42°C for 40mins. After the heat shock, the samples were centrifuged, and the transformation mix was removed. 1ml of YPD was added to each sample and the yeast cells were grown for 1hr (if the selection is Kan) or 3hrs (if the selection is Nat) at 30°c, to enable expression on the antibiotic resistant genes. After the growth, the cells were centrifuged at 3000g for 3 mins and approximately 800ul per sample was aspirated. The rest was re-suspended and was plated on plates with the relevant antibiotics.

In the case of the sub-libraries, 20 reactions were performed per sub-library and these transformed cells were plated on 40 plates of YPD + G418 + NAT to enable the growth of single colonies. G418 was added because a KAN resistance gene is transformed together with the variant SUC2 gene and NAT was added because the  $\Delta suc2$  strain has a NAT resistance gene in place of the original deleted SUC2 gene. The plates were then scraped and all colonies from the same sub-library were pooled together in ~80 ml of DDW (2ml was used to scrape each plate). From this pool, some

glycerol stocks were made by suspending the culture to a final concentration of 30% glycerol, then being frozen and stored at  $-80^{\circ}c$ 

Due to the transformation resulting in not only the expected colonies (with our insert in the correct location) but also in petite yeast cells, we grew the pooled sub-libraries on glycerol for one day to exclude the petite cells from the pool. For the glycerol growth, 200ul from the pooled yeast sub-library (which contains ~1.5x10<sup>9</sup> cells) was added to 100ml of YP-glycerol and grown in a shaking incubator at 30°c for ~ 24 hours. Then, glycerol stocks were frozen as previously described and stored at -80°c. To verify that in our yeast sub-libraries the vector inserts integrated into the correct location in the genome, a yeast colony PCR was performed. A sterile plastic inoculation loop was used to take cells from single colonies plated during the yeast transformation. These cells were resuspended in 20mM NaOH (50µl) and incubated for 20 minutes at 99°c to lyse the cells. 2µl of the lysate was used as a template for the subsequent PCR that amplifies the junctions between the genomic DNA and the vector insert (See primers 27-30 and PCR 4). The products were run on a 1% agarose gel.

#### Competition assay by serial dilutions

We performed two serial dilution competition assays. The first competition assay by dilutions was performed on the 3 strains, WT, SECReTE (+) and SECReTE (-), on the SC-low sucrose media with added Doxycycline (to prevent bacterial contaminations) at 30°c for 7 days. The competition was done with three different starting ratios of the strains, while each starting ratio was done in 5 independent repeats. Before starting the competition, each strain was grown separately in 5ml of SC-low sucrose media for 48 hours until reaching stationary phase. Then, the cells were counted and mixed together at the following ratios: **1.** equal ratios (1:1:1) **2.** low SECReTE (-) with equal ratios of WT and SECReTE (+) (1:1:0.85) **3.** low SECReTE (+) with equal ratios of WT and SECRETE (-) (1:0.85:1).

Cells were grown in SC-low sucrose media, 1.2ml cultures in a 24-well plate. Cells were diluted every 24 hours (dilution factor 1:120), once they had reached stationary phase, which allowed for ~7 generations per day. At each dilution, cells were frozen in a pellet for sequencing and at several time points, glycerol stocks were made as well. All samples were stored in -80°c.

The second competition assay by dilutions, performed on the SECReTE variant sublibrary, was done on SC-low sucrose and on SC-low galactose, both with added Doxycycline, at 30°c for 4 days. Before starting the competition, the pooled sublibrary was grown in 5ml of SC-low sucrose media and in 5ml of SC-low galactose media for 24 hours until reaching stationary phase. For each medium, the OD<sub>600</sub> of the culture was measured and then diluted to reach mid-log (in this case, OD<sub>600</sub> of ~0.6) after 12 hours. This mid-log culture was the starter for the competition experiment since we did not want to include the lag of exiting stationary phase in our experiment and was diluted to reach OD<sub>600</sub> of ~0.6 over the next 12 hours.

Throughout the experiment, for each medium, the cells were grown in 5ml cultures in glass 25ml starter tubes in a roller shaker. Every 12 hours, the  $OD_{600}$  of the culture was measured and based on the  $OD_{600}$ , the culture was diluted to reach the  $OD_{600}$  of 0.6 at the following time point. This dilution factor was typically approximately 1:16 which allowed for ~4 generations per time point. At each dilution, the remaining cells were frozen in pellets for sequencing and at the last time point, a glycerol stock was made as well. All samples were stored in -80°c.

#### Competition assay by chemostat

The lab evolution performed on the 3 strains, WT, SECReTE (+) and SECReTE (-), in a chemostat was done in the bacteriology unit run by Dr. Ghil Jona. The experiment was run on a DASGIP bioreactor system at 30°c. The volume of the bioreactor was 100 mL of SC-low sucrose with added Doxycycline. Before starting the competition, each strain was grown separately in 5ml of SC-low sucrose media for 48 hours until reaching stationary phase. Then, the cells were counted and mixed together at an equal ratio to create the starting co-culture. From this co-cultured starter, 3.5ml was added to the bioreactor and the first 24 hours was in batch mode, meaning there was no inflow of media or dilution. After 24 hours of incubated growth, the bioreactor was switched to continuous mode with a constant flow of media in and out. The flow rate was approximately 22 ml/hr and the shaking speed began at 150 RPM but throughout the experiment, due to yeast clumping, the RPM was gradually raised until it reached 450 RPM. Once a day, samples of ~3-4ml were extracted from the bioreactor and frozen in pellets for sequencing and in glycerol stocks to be stored in the -80°c.

#### Sequencing of the competition assays

A NGS library was prepared and sequenced at several time points for each of the competition assays performed, including the small scale competition assay with 3 strains done by dilutions and by chemostat, as well as the final large-scale competition assay of the SECReTE sub-library. Amplicon sequencing of the variable region of the *SUC2* gene, that contains the designed mutations, enables us to differentiate between variants. Primers with Ilumina adaptors were designed for the relevant *SUC2* variable region.

Based on various factors, each competition assay had a certain number of time points chosen to be sequenced, as well as the ancestor population of that assay. For the small-scale competition assay by daily dilutions with the 3 strains, WT, SECReTE (+) and SECReTE (-), three time points and the ancestor population of the assay were chosen to be sequenced (day 0, 2, 3 and 7 or generations 0, 14, 21 and 49), with 5 individual biological repeats per time point per starting concentration. For the small-scale competition assay by chemostat, the ancestor and every time point (=daily or every 7 generations) was sequenced, but with no repeats. For the final SECReTE library competition, the ancestor and six time points were chosen for sequencing (time points 0, 1, 2, 3, 4, 6, 8 or generations 0, 4, 8, 12, 16, 24, 32) while each time point had six individual biological repeats, in each of the two conditions.

The genomic DNA was purified from each sample with the MasterPure yeast DNA purification kit according to the manufacture protocol. After the purification of DNA from all the samples, each sample served as a template for PCR amplification of the *SUC2* variable region, in 3 (small-scale assay) or 8 (large-scale assay) different PCR technical repeats to control for PCR biases (See primers 31-38 and PCR 5). After the PCR amplification, all PCR technical repeats were pooled together, cleaned with 1.5X SPRI beads (AMPure XP), and their concentration was measured with Qubit. For indexing, the products from the first PCR, the amplified variable region with adaptors, served as a template for a second PCR amplification that added Ilumina indexes (See primers 39-41 and PCR 6). The products of the 2nd PCR were cleaned with 1.5X SPRI bead and their concentration was measured with Qubit. Then, each sample was diluted to 10ng/µl and all samples were pooled together in an equal ratio. This pool was then diluted further to 4ng/µl. For the small-scale competition, the pool was sequenced using a 150-bp paired end kit on a mini-seq (Ilumina) and for the large-

scale competition, the pool was sequenced using a 150-bp paired end kit on with NovaSeq.

The sequencing reads were de-multiplexed using bcl2fastq. Next, the paired-end reads were merged using Pear (http://sco.h-its.org/exelixis/web/software/pear/) and the primers were trimmed using cutadapt (http://cutadapt.readthedocs.io). In the small-scale competition, the reads were aligned with bwa to the three strains (WT, SECReTE (+) or SECReTE (-)). In the large-scale competition, reads were aligned with Bowtie2 to a file of reference sequences that contains all the variable region sequences found in the ancestor sample (that are represented with >100 reads). Then, the frequencies of all aligned variants were counted in each sample.

#### Conditioned media transfer experiment

WT and  $\Delta ctt1$  cells were each diluted from a deep stationary starter into two flasks of 30ml of fresh SD media (dilution factor 1:100), resulting in a total of 4 cultures. They were grown while shaking at 30°c for ~ 5.5 hours (~2-3 replications) until they reached mid-log. Next, the cells were centrifuged, the media was removed, and the cells were resuspended with fresh media. One culture of WT cells and one culture of  $\Delta ctt1$  cells were resuspended with SD + 2mM H<sub>2</sub>O<sub>2</sub>, while the remaining two cultures were resuspended with only SD. The SD medium was replaced with fresh SD, rather than just adding the H<sub>2</sub>O<sub>2</sub> to the relevant cultures, because this medium will later be passed on to fresh cells and there needs to be enough nutrients left in the medium for the fresh cells to grow on it. The resuspended cells were returned to shaking incubation at 30° c for 1.5 hours. After 1.5 hours, the cells were centrifuged, and the medium was separated from the cells. The separated medium was then transferred to a fresh tube. The media from the two cultures that underwent oxidative stress is called conditioned media and the media from the two cultures that did not go through oxidative stress is the non-conditioned media. To the conditioned media, an extra 0.5mM H<sub>2</sub>O<sub>2</sub> was added, to readjust for H<sub>2</sub>O<sub>2</sub> that was broken down in the previous stress so it should reach a final concentration of approximately 2mM H<sub>2</sub>O<sub>2</sub>. To the non-conditioned media, 2mM H<sub>2</sub>O<sub>2</sub> was added.

Next, fresh cells from a stationary phase starter, that had experienced no prior stress, were added to the conditioned or non-conditioned media at a dilution factor of 1:50. There were 4 experimental combinations for each type of medium. 1. WT cells added to WT conditioned/non-conditioned media. 2. WT cells added to  $\Delta ctt1$ 

conditioned/non-conditioned media. 3.  $\Delta ctt1$  cells added to WT conditioned/nonconditioned media. 4.  $\Delta ctt1$  cells added to  $\Delta ctt1$  conditioned/non-conditioned media. The growth of these cultures was measured in a 96-well plate as described in the growth experiments section.

#### Catalase activity assay

To measure the catalase activity in the conditioned or non-conditioned media, we used the QuantiChrom Peroxide assay kit to quantify the concentration of  $H_2O_2$  found in the medium. We began the experiment as we did in the conditioned media transfer experiment. The cells were grown until mid-log, their medium was replenished and  $H_2O_2$  was added to the relevant cells. After 1.5 hours, the medium was separated and the concentration of  $H_2O_2$  found in each medium was measured according to the protocol of the peroxide assay kit.

## Results

## Set-up of the Suc2 secretion library system

#### Growth measurements of example Suc2 secretion variants

As mentioned in the introduction, increasing or decreasing the SECReTE count in a gene can increase or decrease the secretion of that protein, respectively. Furthermore, this principle was demonstrated previously by our lab on the Suc2 protein by creating two new strains. The first strain has an increased *SUC2* SECReTE signal (SECReTE(+)), done by increasing the number of SECReTE repeats and increasing the pyrimidine content in the surrounding region, which results in higher Suc2 secretion. The second strain was altered in the opposite trend and has a decreased *SUC2* SECReTE signal (SECReTE(-)), resulting in lower Suc2 secretion.<sup>21</sup> With these strains, it can be tested whether the changes in the Suc2 secretion level can cause varied fitness advantages and potentially represent different cooperative strategies.

The first step was to measure the strains' growth on various media that are relevant to the biological function of Suc2, as well as search for a medium that prompts fitness differences between the strains. We compared the growth of 4 strains, the WT, SECReTE (+), SECReTE (-) and a deletion strain  $\Delta suc2$ . All of the upcoming experiments were done on liquid media, where the metabolites diffuse freely, as opposed to on solid media. While in nature, yeast are often found on a type of solid media, e.g. a grape skin, it has been shown in different works that the dynamics between the cooperator and cheater follow the same general trend on both liquid and solid media.<sup>4,5,36</sup>

Firstly, on YPD media, as we expected, no growth differences were detected between the strains since glucose represses the expression of the *SUC2* gene (Figure 3A). Additionally, when using the standard percentage of 2% sucrose in a synthetic complete (SC) media, no growth differences between the strains were found, except for the deletion strain (Figure 3B). Under these conditions, there is enough sucrose in the media that the SECReTE (-) is not affected by the disadvantage of its lower secretion level and it grows well. However, the  $\Delta suc2$  strain suffers significantly when the sole carbon source is sucrose and it has a much lower growth rate than the other strains. The deletion strain is still able to grow somewhat, even without the *SUC2* gene, which may be due to some spontaneous hydrolysis or possibly there is some leakage of sucrose into the cell which can then be hydrolyzed intracellularly by putative non-specific enzymes.

As we did not see fitness differences between the strains at a high sucrose concentration of 2%, we lowered this concentration to search for a more challenging medium that would differentiate between the strains. Accordingly, when lowering the sucrose percentage to 0.05% in a SC medium, fitness differences between the strains were observed. In addition, 0.005% glucose was added since it was previously demonstrated that a very low percentage of glucose can induce the expression of the Suc2 machinery.<sup>37</sup> On this SC medium, with 0.05% sucrose and 0.005% glucose, the SECReTE (+) and WT strain grew similarly with the highest growth rate, while the SECReTE (-) strain grew at a slightly lower rate. The  $\Delta suc2$  strain grew at the lowest rate with a much lower yield (Figure 3C). All following experiments were performed on this low sucrose medium that demonstrates the fitness differences.



Figure 3. Fitness differences between Suc2 secretion variants is found on a low sucrose media. Growth curve measurements of WT (red),  $\Delta$ suc2 (gray), SUC2 SECReTE (+) (blue), and SUC2 SECReTE (-) (green) are shown in OD values over time during continuous growth at 30°C. (A) YPD media. (B) SC media with 2% sucrose. (C) SC media with 0.05% sucrose and 0.005% glucose.

# *Simulation of co-culture competition assay between three Suc2 secretion variants*

Based on the growth curve data of each strain grown individually on the low sucrose media, a co-cultured competition was simulated between the 3 strains WT, SECReTE (+), and SECReTE (-). This was done with the Curveball competition simulator that uses growth parameters that are estimated by fitting growth models to the growth curve data.<sup>34</sup> In this simulation, there are no interactions assumed between the

competing strains. The fitness of the strain in the competition is purely dependent on its individual growth abilities and the carrying capacity of the medium and the simulation does not consider any cooperation or cheating between strains. Under these assumptions, there is a predicted sharp and immediate decline in the frequency of the SECReTE (-) strain due to its slight growth defect. Conversely, the WT and SECReTE (+) remain steady throughout the competition (Figure 4A).

# *Experimental co-culture competition assay between three Suc2 secretion variants*

Next, we wanted to compare the predictive simulation results to the results of actual biological experiments of a co-culture of the strains. There are two widely accepted methods to experimentally perform co-culture competition experiments, which are by daily dilutions and in a chemostat. They are both based on the main principle that strains with an advantage in the specific conditions will rise to a higher frequency throughout the competition. However, these two methods each have quite different environments from one another. When performing daily dilutions, typically every 24 hours a small subsample of the population (in our case  $\sim 1/100$ ) is transferred to a fresh medium, enabling regrowth. Each day, the population reaches stationary phase before being transferred and then needs to undergo lag phase before re-entering growth. In contrast, a chemostat enables continuous exponential-phase growth by constantly diluting the cells through addition of new media and simultaneous removal of old media and cells. As mentioned previously, in the Suc2 system, the fitness of the cooperator versus the cheater is very dependent on cell density. Therefore, we anticipate considerable differences between the two types of competition experiments and would ideally like to explore the dynamics in both methods.

Firstly, a biological experiment of the co-culture competition was performed by daily dilutions over 7 days. This was done with three different starting ratios of the three strains, with each ratio being done in 5 independent repeats. The planned starting ratios were an equal starting ratio between the strains, a low SECReTE (-) ratio, and a low SECReTE (+)) ratio. Each day, the cultures went through ~7 generations, resulting in a total of ~49 generations over the span of the experiment.

Since in this Suc2 system, cells externally hydrolyze sucrose and the monosaccharide products diffuse through the shared medium, we expect, in the biological experiment,

the strains to interact with each other by both sharing and taking these products. This ability to interact is not taken into account in the above-mentioned simulation. When comparing the results of the predictive simulation to the actual experiment, there is a clearly noticeable difference in the SECReTE (-) strain's behavior. In the biological experiment, the SECReTE (-) does not decline because of its growth defect (as it does in the simulation), but rather remains stable throughout the competition. This stable growth of SECReTE (-) was observed over several independent repeats, as well as in several different starting ratios. This suggests that the SECReTE (-) is able to interact with the other strains and use their products to compensate for its lower Suc2 secretion level (Figure 4B-D). Furthermore, while the strains remain relatively consistent and stable throughout the competitions, when the SECReTE (+) and WT begin at similar starting concentrations, the SECReTE (+) slightly increases its concentration, while the WT slightly decreases its concentration (Figure 4B-C). This is observed over multiple independent repeats.

In addition to performing the competition experiment by daily dilutions, it was also performed in a chemostat. In the chemostat, there are stronger changes in frequency of the strains and different trends than seen in the daily dilutions. The WT increases while the both the SECReTE (+) and SECReTE (-) decrease. It is important to note that there are no repetitions in this experiment due to lack of availability of chemostats (Figure 4E). It is also worth mentioning, we would expect the predictive simulation to resemble the competition by dilutions more than the by chemostat because it takes into account lag phase and yield.



**Figure 4.** Comparison of the co-culture competition assay prediction by Curveball simulation to the experimental results. Relative frequency of the strains WT (red), Suc2 SECRETE (+) (blue), and Suc2 SECRETE (-) (green) over a co-culture competition assay of 7-8 days. (A) Curveball simulation prediction. (B-D) Experimental results of competition assay performed by daily dilutions. The competitions vary in their starting ratios. There were aimed to be an equal starting concentration (B), a low ratio of SECRETE (-) while the WT and SECRETE (+) were equal (C), and a low ratio of SECRETE(+) while the WT and SECRETE (+) were equal (D). In graph (D), there was a mistake and the SECRETE (+) was added at high concentration instead. (E) Experimental results of competition assay performed by continuous growth in a chemostat.

In summary, the comparison between the predictive simulation to the experimental competition clearly indicates the existence of complex interactions between genetic variants in the Suc2 system. Additionally, the differences observed between the two

types of competition assays denotes that the biological settings of the competing culture are likely to affect the interactions and thus affect the frequency of the competing strains.

#### Design of the large synthetic library of Suc2 secretion variants

After establishing that Suc2 secretion variants can have varied fitness advantages and represent different strategies that interact with each other in a co-culture, we aim to study a more complex interaction by creating many potential strategies. To represent these various strategies in the Suc2 public good interaction, many different mutated variants were designed. Approximately 18,000 variants were designed computationally throughout 4 different locations in the *SUC2* gene. Each variant can contain up to ~30 single-nucleotide polymorphisms (SNPs) but these mutations will be found in only one of the four locations. These 4 locations are the naturally occurring SECReTE motif, the signal peptide, a control region and the 3'UTR (Figure 5A).

The first location of the variable sequence library within the *SUC2* gene is the naturally occurring SECReTE region that is found in the coding region of the gene. This location includes the SECReTE motif comprised of 13 nucleotide triplet repeats and the surrounding region of this motif, enabling us to lengthen, shorten or alter this naturally occurring motif.

The mutation types designed in the SECReTE site are as follows:

- SUC2 orthologs The SECReTE region was copied from various yeast SUC2 orthologs. (251 variants)
- <u>C2T/T2C synonymous mutations</u> A characteristic of the SECReTE motif is that it is made up of repeats of NNY and thus generally has a high pyrimidine content. In this mutation type, we created a bias in the pyrimidine content by either favoring a higher C content or a higher T content. This is to see if there is a difference between the two pyrimidines in their effect on the SECReTE motif. (180 variants)
- Y2R synonymous mutations Here we mutated pyrimidines (Y) to purines
   (R) without making any alterations to the protein sequence. As mentioned above, the pyrimidines are a crucial measure for the motif and by decreasing

the number of pyrimidines, we aim to decrease the Suc2 secretion level. (1933 variants)

- <u>R2Y synonymous mutations</u> Here we mutated purines (R) to pyrimidines (Y) without making any alterations to the protein sequence. By increasing the number of pyrimidines, we aim to increase the Suc2 secretion level. (1332 variants)
- Synonymous mutations (unbiased) This includes synonymous mutations that have no bias towards pyrimidines or purines, it retains the same Y to R ratio. We do not expect this to have an effect on the Suc2 secretion. (336 variants)
- Control synonymous mutations
  - Codon optimality This is used to control for the possibility that the mutations we are designing will affect secretion due to changes in codon optimality rather than changes in the SECReTE motif. It is thought that changes in the codon usage can affect the protein expression process.<sup>38</sup> We designed a spectrum of variants that range from very optimal to very suboptimal in their codon usage, without making changes that should affect the SECReTE motif. (340 variants)
  - <u>RNA structure</u> Another characteristic we wanted to control for is the secondary structure of the mRNA. Again, we designed a spectrum of variants ranging from strong to weak secondary structures, without making changes that should affect the SECReTE motif, to rule out the possibility that this is what is affecting Suc2 secretion. (342 variants)
- Non-synonymous mutations In this group, we included non-synonymous mutations, in addition to synonymous mutations. This was used to increase the SECReTE signal further by including R to Y mutations that affect the protein sequence as well. The tool SIFT<sup>39</sup> was used to predict whether the non-synonymous mutation would be tolerated by the protein. We only allowed for 1-2 non-synonymous mutations per variant, that were predicted not to hurt the protein's function. This group also includes a small subset that has non-synonymous mutations that are not biased towards pyrimidines or purines. However, we did not design variants that had non-synonymous mutations to increase Y to R mutations because we were able to diminish the Y content in the SECReTE region through synonymous mutations alone. (1901 variants)

The second location of designed variants, the signal peptide, is an established and well-known motif that can alter the secretion of a protein. The signal peptide is not a set sequence but rather varies between genes. However, there are certain characteristics that are commonly found in this motif, such as a hydrophobic core and a signal peptidase cleavage site in the C-terminal region. The designs included mainly non-synonymous mutations that generally retained these important characteristics.

The mutation types designed in the signal peptide site are as follows:

- <u>Natural signal peptides</u> This includes all the natural signal peptides found *S. cerevisiae* which amounts to 329.
- Synonymous mutations A handful of variants were made with only synonymous mutations, thus the original amino acid sequence of the SUC2 signal peptide is retained. Since the signal peptide works on the protein level, we would not expect synonymous mutations to have an effect here. (100 variants)
- ➢ <u>Non-synonymous mutations-</u>
  - Retain chemical characteristics Throughout the signal peptide, amino acids were replaced by other amino acids that possess the same chemical characteristics. For example, a basic amino acid can be switched to a different basic amino acid, but not to an acidic amino acid. The number of mutated amino acids differs between variants. This is to differentiate between the importance of the chemical characteristic and the specific amino acid. (810 variants)
  - Enrich hydrophobicity The hydrophobic core is a central feature of the signal peptide and is important for the secretion process. Here we increased the hydrophobic core either by performing insertions to increase the length, or mutations that mutate the surrounding amino acids to hydrophobic ones. We believe this possibly could increase Suc2's secretion. (1622 variants)
  - Loose hydrophobicity Here we decreased the hydrophobic core either by performing deletions to decrease the length, or by mutations that mutate the amino acids found on the edges of the hydrophobic core to hydrophilic ones. We believe this possibly could decrease Suc2's secretion. (1621 variants)

The third location for designed variants is a control site which is a neutral location in the coding region of *SUC2*, outside of the *SUC2* identified SECReTE motif or signal peptide. Many of the synonymous mutation manipulations performed on the natural SECReTE region were performed also on the control region. This was designed to show that while these manipulations affect secretion levels in the natural SECReTE region, they should not affect secretion when altering a neutral, un-related region.

The mutation types designed in the control site are as follows:

All the types of synonymous mutations designed for the SECReTE region were used to design the mutations for the control site as well. For example, variants were designed with synonymous mutations that mutate pyrimidines to purines (Y2R). Decreasing the number of pyrimidines in this region, that does not have a SECReTE motif, should not influence the secretion of Suc2. (1800 variants)

The fourth and final location designed is the 3'UTR. The SECReTE motif can be found anywhere throughout an mRNA, in the coding region or in the untranslated region. It has been shown that adding a SECReTE motif to the 3'UTR of a gene can increase the protein's secretion level.<sup>21</sup> An additional advantage is that, since the 3'UTR is not part of the coding region, we do not have the same coding constraints we have in other regions. This allows us much more freedom in designing different variants. Therefore, in our design, we inserted a synthetic SECReTE region in the 3'UTR, in addition to the naturally occurring SECReTE in the coding region.

The mutation types designed in the 3'UTR site are as follows:

- Naturally occurring SECReTE in yeast A subset of the naturally occurring SECReTE motifs found throughout various genes in yeast were added to our 3'UTR site. These SECReTE motifs could be from the 3'UTR, 5'UTR or coding region of the gene it originally is found in. (911 variants)
- Variants from the natural SECReTE site We randomly selected a subset of the designed sequences that were made by our manipulations on the naturally occurring SECReTE site in the coding region and added them into the 3'UTR. (1430 variants)
- Various Y-content High Y-content is a common characteristic of the SECReTE motif and since we do not have coding constraints in the 3'UTR, we have the liberty to design sequences with very high amounts of

pyrimidines. We refer to these sequence stretches of high Y-content as polyYs. In this group, there are multiple factors that we can adjust to create polyYs, such as the percentage of pyrimidines (ranges from 50-100%), the ratio between the two pyrimidines C and T, the length of the polyY stretch, and its distance from the beginning of the 3'UTR. (2153 variants)

Natural SECReTE reverse compliment – When designing the library, we discussed the idea that if we insert into the 3'UTR a reverse compliment of our natural SECReTE site, possibly it could work as a sort of silencer. Perhaps the reverse compliment could base pair with the SECReTE motif in a Suc2 mRNA and silence its signal. We added 6 such variants into the pool.



Figure 5. Design of large synthetic library containing  $\sim 18,000$  Suc2 secretion variants. (A) A schematic of the SUC2 gene showing where the 4 sites that were modified are found along the gene. (B) A more in-depth schematic where each site along the SUC2 gene that was altered is marked with a different colored circle- signal peptide in red, SECReTE in orange, 3'UTR in purple and the neutral site in blue, with N representing the number of variants designed in that location. The rectangular blocks branching out from the circle represent a function that creates a specific type of mutation and N represents the number of variants designed with that type of mutation.

#### Creation of the large synthetic library of Suc2 secretion variants

For the creation of the designed synthetic library, ~18,000 DNA fragments of 200bp that contain the planned mutations were ordered from the company Twist. These fragments were amplified into sub-libraries based on their location and each sub-library was then cloned into an integrative plasmid called pGS2223 that contain the *SUC2* gene (Figure 6A). The plasmids were transformed by electroporation into electro-competent bacteria, with 4-8 reactions per sub-library, depending on the batch efficiency. The resulting efficiencies ranged from  $9x10^5-2x10^6$  bacterial cells per sub-library. Since each sub-library contains a few thousand variants, this gives a high coverage of roughly  $10^3$  per variant. All the bacterial colonies from the same sub-library were pooled together and the plasmid libraries were extracted. All the sub-libraries have reached this step.

Next, the plasmid libraries were linearized by the restriction enzyme NotI, that exposes the homologous region that will undergo homologous recombination in the yeast. The linearized plasmid sub-libraries were transformed into  $\Delta suc2$  yeast strain using heat shock. Twenty reactions per sub-library were performed to help increase the efficiency of this step. The yeast transformation was completed for the SECReTE and the signal peptide sub-libraries. The SECReTE sub-library transformation resulted in ~90,000 colonies, giving a coverage of ~18x per variant. The signal peptide sub-library resulted in ~55,000 colonies, giving a coverage of ~12x per variant. Currently, the control and 3'UTR sub-libraries are at the plasmid library stage and need to be transformed into yeast.

A phenomenon we observed after the yeast transformation was that we had a mixture of colony types on the plates. There were big round colonies, as we would expect to see, and there were also lots of very small colonies. We found these small colonies to be petite cells. Petite yeast cells are cells that have a block in their aerobic respiratory chain pathway due to mutations in mitochondrial related genes.<sup>40</sup> As a result, they cannot grow on nonfermentable carbon sources (such as glycerol) and form small colonies when grown on fermentable carbon sources (such as glucose). In these observed petite colonies, we found that the insert did not enter the correct and intended location. Our insert may not have been specific enough to our neutral target location, and some inserts went through homologous recombination in a different locus that may have affected the mitochondria. Therefore, we grew the pooled sub-

libraries on glycerol for one day to exclude the petite cells from the pool. It is important to note, the count of colonies listed above to measure coverage per variant in the transformation included only the correct large colonies and not the petite colonies. For the following experiments, we used only the pooled sub-library after it was grown on glycerol. Sixteen colonies from the SECReTE sub-library were then verified for the correct genomic location by amplifying the junctions between the yeast genome and the inserted plasmid (Figure 6B). In addition, the *SUC2* region of these colonies was sent to Sanger sequencing. Later, the entire SECReTE sub-library was deep sequenced. Unexpectedly, it was found upon sequencing that in addition to our designed variants, there were many undersigned variants included in our library. Since they are already found in the oligo stage, before cloning and transformation, we believe it occurred during the synthesis of the DNA fragments at the Twist laboratories. This resulted in our SECReTE sub-library containing a total of 10,274 variants, with approximately half being the designed planned variants.



**Figure 6. Verification of the SECRETE sub-library variants. (A)** The integrative plasmid pGS2223 containing the SUC2 gene that the variable region is cloned into. The plasmid sub-libraries were digested by Notl before yeast transformation. **(B)** After the yeast transformation, colonies were tested for containing the SUC2 gene in the correct location. This is one example colony out of 16. Lanes 1 and 2 show the junctions between the yeast genome and the inserted plasmid which are supposed to be 1059bp and 667bp respectively. Lane 3 is the entire SUC2 gene (including flanking regions) which is supposed to be 3,261bp.

#### Competition assay of the Suc2 SECReTE variant sub-library

Once the SECReTE sub-library was completed and verified, we performed a competition assay to study the growth rate and potential interactions between variants and examine their fitness differences. The type of competition assay that was decided upon for the sub-library was a combination of the two methods discussed above. It was done by dilutions, which enables working in a high-throughput manner, but the dilutions were done at mid-log, twice a day, instead of once per day at the stationary phase. Once the cells reach stationary phase, the media is saturated and most of the nutrients have already been broken down. We believe mid-log to be a better representation of expected interactions since the cells are still paying the price of breaking down sucrose, as well as reaping the benefits, rather than slowing down growth due to saturation.

The assay was performed on two different mediums, a low sucrose medium that was discussed earlier and a galactose medium. Galactose is a sugar source that does not repress the expression of the Suc2 machinery but Suc2 is unable to break down galactose. Consequently, galactose enables us to isolate the cost of each variant since in this environment, each variant is paying the cost of production of Suc2 but does not reap any benefits. If a variant behaves similarly on both mediums, this indicates that the fitness differences we see on sucrose is merely due to cost differences. Therefore, by comparing our competition assay results on sucrose to the results on galactose, we aim to substantiate differences in variant fitness between mediums and reinforce that there are biological interactions on sucrose.

We plan to substantiate galactose as a proper measure for cost by performing deep RNA sequencing of the variable region in the *SUC2* mRNA to check the variants' expression level in the sub-libraries. Higher expression levels of variant *SUC2* mRNAs are expected to be found for variants with lower fitness on galactose, as this would suggest a higher cost. Although we cannot measure the variants' protein level for our sub-libraries, as the variants are pooled together and many of the mutations are synonymous, we plan to isolate a subset of variants and measure their invertase protein expression to further verify the cost measurement.

The competition assay was performed by dilutions every 12 hours, to keep the cells in mid-log, and in six independent repeats for each of the two mediums. The assay lasted for a total of 4 days (8 dilutions) with ~4 generations between each dilution, resulting

in a total of ~32 generations throughout. We then deep-sequenced the variable SECReTE region of the pooled sub-library throughout various time points and calculated the fitness of each variant. To calculate the fitness, a generalized linear model with Poisson distribution was used to create a regression line that best fits the variant's frequency over the time points of the experiment (MATLAB's glmfit). The slope of this regression line represents the variant's fitness. It is important to note, this fitness calculation does not consider the interactions that we anticipate to occur on the sucrose media. Therefore, while on galactose we expect the fitness calculations to be accurate (as there are no interactions), the fitness calculations on sucrose may not be as precise. We will keep in mind that, in these results, we are using a preliminary calculation for fitness on sucrose. We aim to improve this and are currently working on a model to calculate fitness with all the relevant considerations.

#### Population-level analysis

On the low sucrose medium, each variant's fitness is well correlated between repeats, with a correlation ranging from 0.38 to 0.5 (Figure 7A). On the galactose media, while the fitness of each variant is correlated between repeats, the correlation is a bit lower, ranging from 0.27-0.34 (Figure 7B). This result suggests that our sub-library has more measurable differences between variants on sucrose than on galactose. The fitness variation due to changes in just cost, as on galactose, is smaller and less considerable than the fitness differences we see on sucrose, which should be influenced also by Suc2's effects on the sucrose metabolism and interactions with other variants. Furthermore, when comparing each variant's fitness on different media, the correlation between repeats is even lower, ranging from 0.16 to 0.22 (Figure 7C). This strengthens our speculations that the fitness differences observed on sucrose is more than just cost differences, as it shows the dissimilarity of the variant's fitness between the two mediums.





Figure 7. Comparison of the variants' fitness between the independent repeats of the competition assay shows positive correlation on each media, with a stronger correlation on the low sucrose media. (A) Comparison of fitness estimation of all library variants between the different repeats of the experiment on sucrose media. Each dot represents a variant in the SECRETE sub-library and the Pearson correlation coefficient is marked on each comparison. (B) Comparison of fitness estimation of all library variants between the different repeats of the experiment on galactose media (C) Heatmap comparing the Pearson correlation coefficients between repeats on the same media, as well as between repeats on different mediums.

After showing that the independent repeats correlate, we next compared, between the two mediums, the mean fitness of the entire population over the generations of the experiment. To calculate the population fitness over generations, the frequency of each variant at a time point was multiplied by that variant's fitness, and the sum of that is the mean fitness of the population at that time point. This was done for each time point and each repeat.

While on both media, the fitness increases over generations, the slope is steeper on the sucrose medium, resulting in a higher final population fitness in sucrose than in galactose (Figure 8). Since the experiment is not long enough to expect mutational changes to occur and fixate, increase in population fitness is likely due to advantageous strains taking over. This result further suggests that the fitness differences between variants are stronger on sucrose than galactose.



Figure 8. The mean fitness of the population over the generations of the experiment has a steeper increase on sucrose media (red) than on galactose media (blue). Shown are all 6 repeats for each medium.

#### Variant-level analysis

Next, important for our analysis is to identify the variants that did not behave the same on both media, so the mean fitness of each variant across all repeats was calculated for each medium. The mean fitness of the variants on sucrose was plotted against the mean fitness of the variants on galactose (Figure 9A). Since we would not expect, in this comparison, the absolute values of the fitness to be the same, as we are comparing between different media, we used a linear regression line to estimate the relationship between the two conditions. The plot shows that there is a sizable population of variants that deviate from the regression line and show an especially high fitness on sucrose while retaining a high fitness on galactose as well. This population could contain potential cheater-like strategies, as they supposedly pay a low cost since their fitness is high on galactose, presumably due to lower expression or secretion of Suc2, but they still gain a lot of benefit by utilizing the product of sucrose breakdown done by other variants, as they have a high fitness on sucrose. In contrast, we have much less variants on the other side of the spectrum that strongly deviate below the regression line. These variants potentially represent a cooperator or ultra-cooperator strategy and thus have an especially low fitness on sucrose while showing a low fitness on galactose as well. Meaning, they seem to pay a high cost but do not manage to gain enough benefit to cover these costs.

We further investigated these deviant variants and characterized them as positive or negative fitness residuals (Figure 9B, C). We define fitness residual as the difference

between a variant's observed fitness on sucrose media from their expected fitness based on just their cost, meaning fitness on galactose. While Figure 9 shows the mean fitness of each variant, we can also compare each variant's fitness between the individual repeats on either medium. Since there are six repeats on each of the mediums, we can make 36 of these comparisons for each variant. Finally, we defined a high or low fitness residual variant as being a variant that has a fitness residual above a certain threshold in at least 32 out of these 36 comparisons. In Figure 9B, we see these fitness residual variants that pass the threshold of the absolute value of 0.002 while in Figure 9C, the threshold is the absolute value of 0.004. The threshold of 0.002 was chosen because it is able to remove variants that are very close to the regression line but is also low enough to include relatively many negative fitness residual variants are retained (except for those still close to the regression line) however, we lose most of the negative fitness residual variants.



**Figure 9. Comparing the mean fitness on sucrose versus the mean fitness on galactose for each variant shows some deviation from the regression line.** (A) The mean fitness of the variants on sucrose media plotted against their mean fitness on galactose media. Each dot represents a variant in the SECReTE sub-library and the Pearson correlation of the regression line is r=0.4192. (B) The same plot showing the positive and negative fitness residuals variants. In this figure, the fitness residual of these variants must consistently pass a threshold of the absolute value of **0.002**. (C) The same plot showing the positive and negative fitness residuals variants. In this figure, the fitness residual of these variants must consistently pass a threshold of the absolute value of **0.004**.

Defining the populations of fitness residual variants enables us to search for features that are unique to one of the populations. The first feature we investigated in this context is the Y content. As previously mentioned, in the SECReTE motif, a high Y content is thought to be related to a higher secretion level, while a lower Y content is suspected to be related to a lower secretion level. Therefore, we compared between the Y content in positive fitness residual variants, negative fitness residual variants and the entire library.

In this comparison, we see a lower Y content in the positive fitness residual variants than in the entire library. However, this difference was found to be statistically significant (p-value < 0.05) only at the higher threshold of 0.004 (Figure 10A). This

result supports the notion that the positive fitness residuals contain cheater-like strategies. As mentioned, lower Y content is associated with a lower secretion so these cheater-like variants may secrete less of the Suc2 protein, which results in a high fitness on sucrose, as they can consume other's products without paying the cost (i.e. maintaining high fitness on Galactose). Although this fits nicely with our theory, we wanted to further investigate the significance of this result by increasing the threshold and checking its p-value. Ideally, as we increase the threshold, we will continue to see a significant result. However, in Figure 10B, we see the p-value does not always remain below the standard threshold of 0.05. Though the p-value does remain consistently low and around the value of 0.05, only at threshold 0.004 and threshold 0.008 it is considered to be a significant result.

Another consideration is that in the positive fitness residual variants, we have a mixture of variants with non-synonymous mutations and variants with exclusively synonymous mutations, with an approximate ratio of 1:2 respectively. The non-synonymous mutations may cause unknown additional effects that alter the protein and thus the phenotype of the variant may not correlate with its SECReTE characteristics. Therefore, we decided to look exclusively at the synonymous mutation variants in our population. In this case, when comparing between the positive fitness residual variants and the entire library, we see the Y content is significantly lower in the positive variants at the threshold 0.002 (Figure 10C) but not as we further increase the threshold (Figure 10D). It is important to note, that as demonstrated in Figure 10B and Figure 10D, the number of all the positive fitness residual variants is larger than the number of variants, the statistical power is lowered as well, and we get higher p-values. Possibly this is the reason that we do not retain the statistically significant result.



**Figure 10.** Comparison of Y content found in positive fitness residual variants, negative fitness residual variants and the entire library shows a lower Y content in positive variants. (A) The distribution of the Y content in the variants of each group, positive fitness residual variants (blue) and entire library (gray) at the **threshold 0.004**. (We do not show the negative fitness residual variants as there are very few variants at this threshold.) (B) On the left Y-axis, the log<sub>10</sub>(p-value) of the t-test performed on the Y content found in the entire library verses in the positive fitness residual variants is plotted against the applied threshold. The gray dotted line represents log<sub>10</sub>(p-value=0.05) and a red asterisk represents a significant p-value. On the right Y-axis, the number of positive fitness residual variants with exclusively synonymous mutation, grouped by positive fitness residual variants, negative fitness residual variants with exclusively synonymous mutation.

We performed the same analysis for the NNY score. The NNY score is defined as the proportion of NNY nucleotide triplet repeats throughout the relevant sequence. This feature is similar to the Y content in that the higher the score, the more pyrimidines there are, but this has to follow the NNY pattern that characterizes the SECReTE

motif. When looking only at variants with exclusively synonymous mutations, we find a significantly lower NNY score in the positive fitness residual variants than the entire library at both thresholds 0.002 and 0.004 (Figure 11). This significant result is not retained as we increase the threshold further, but it is worth remembering, the number of variants and thus statistical power is decreasing as well.

We also looked into the features such as delta G, which represents the strength of the secondary structure of the mRNA, and the tRNA adaptation index (TAI) score, which is a translation efficiency measure. Neither feature showed significant differences between these defined variant groups.



**Figure 11.** Comparison of NNY score found in positive fitness residual variants and the entire library shows a possible lower NNY score in positive variants. This includes only variants with exclusively synonymous mutation. (A) The distribution of the NNY score in the variants with exclusively synonymous mutation, grouped by positive fitness residual variants (blue) and entire library (gray) at the **threshold 0.002.** (B) The distribution of the NNY score in the variants with exclusively synonymous mutation, grouped by positive fitness residual variants, negative fitness residual variants and entire library at the **threshold 0.004.** (C) On the left Y-axis, the log<sub>10</sub>(p-value) of the t-test performed on the NNY score found in the entire library verses in the positive fitness residual variants is plotted against the applied threshold. The gray dotted line represents log<sub>10</sub>(p-value=0.05) and a red asterisk

# represents a significant p-value. On the right Y-axis, the number of positive fitness residual variants found at each threshold is plotted.

In addition to looking at general features found over the whole SECReTE variable region and their enrichment in positive/negative fitness residual groups, we can also look at specific locations along the variable region. We explored whether the positive or negative fitness residual groups (threshold 0.002) were enriched for certain mutations in specific locations. To do so, we computed the proportion of R2Y and Y2R mutations along the SECReTE variable region in each of these groups, the positive variant group, negative variant group, and the entire library. Next, we checked if this proportion found in the positive or negative fitness residuals is significantly different than the proportion found in the entire library. In Y2R mutations, we did not find significant differences. However, in R2Y mutations we did. In Figure 12 we can see that, immediately upstream to the natural SECReTE location, the negative fitness residual variants are enriched with R2Y mutations. It seems that these variants have an enhancement to their existing SECReTE signal by introducing pyrimidines upstream to the motif. Interestingly, we see the opposite trend in the positive fitness residual variants, as they have a depletion of the R2Y mutations in this same upstream location. When looking even further upstream, we observe a curious switch in the trend, and the negative fitness residuals are now depleted for these mutations.

These findings (Figure 10Figure 11, and Figure 12) support our running theory that negative fitness residual variants are represented by cooperators and positive fitness residual variants are represented by cheater-like variants. The negative fitness residual variants have an enhancement to their SECReTE motif by increasing the neighboring Y content and thus, potentially increase their secretion and cooperative behavior, while the positive fitness residuals decrease their general Y content and NNY score, potentially decreasing their secretion.



Figure 12. The measurement of R2Y mutation enrichment throughout each nucleotide position along the SECReTE variable region shows an enrichment of R2Y mutations in negative fitness residual variants immediately upstream to the natural SECReTE motif. The entire variable region is 231bp long and each position along this region is represented on the X-axis. Each line extending from the zero line signifies enrichment of this mutation at that location in either the positive fitness residual group (blue) or the negative fitness residual group (red). The value of the mutation enrichment on the y-axis is the proportion of variants with this R2Y mutation in the fitness residual group subtracted by the proportion of variants with this mutation in the entire library. The black line located at position 138-177 represents the natural SECRETE location.

#### Catalase as a potential public good

The Suc2 system is the only well-established public good system found in yeast. It would be ideal to test the dynamics of this complex cooperative interaction not only on Suc2, but on additional genes as well. As a result, we investigated whether catalase could be a candidate protein to study this. As previously mentioned, catalase is an enzyme that protects the cell from oxidative damage by catalyzing the decomposition of hydrogen peroxide to water and oxygen. The yeast *S. cerevisiae* has two types of catalase enzymes, a peroxisomal catalase, encoded by the *CTA1* gene, and a cytosolic catalase, encoded by the *CTT1* gene. While in *S. cerevisiae* catalase has been suggested to be a secreted protein and could thus potentially provide protection to surrounding cells,<sup>41</sup> this is not yet proven. The strong advantage to pursing catalase as a public good is that it provides a very interesting possibility. Since it is considered to mainly work intracellularly but can show activity in the shared media as well, perhaps the cell could use the protein both intracellularly and extracellularly. In this case, the cell could either share this protein's protection, or keep it to itself, or do both. This is a complexity that we do not find in the Suc2 system.

We ran multiple experiments to try and answer this question of whether catalase is a secreted public good. The first experiment was a conditioned media transfer experiment. This entails applying a stress on yeast cells, in this case an oxidative stress of 2mM H<sub>2</sub>O<sub>2</sub>, and then transferring only the media to fresh cells that haven't received any prior stress. These fresh cells, with the conditioned media, are then exposed to the same oxidative stress. If the cells that received conditioned media survive the stress better than cells that received non-conditioned media, this suggests there is something protective in the media. In this case, we refer to media transferred from cells after an oxidative stress as conditioned media, while we refer to media transferred from cells that were not exposed to any stress as non-conditioned media.

The basis for performing this media transfer experiment is from our lab's previous work which showed that yeast have an 'auto-protection' in oxidative stress.<sup>42</sup> This means that if cells were exposed to a mild oxidative stress, they then survived better in a subsequent encounter with an otherwise severe oxidative stress. This demonstrates that cells can be 'primed' to be better prepared for an upcoming stress, which could be due to internal molecular mechanisms, but it could also potentially be due to external signals released from the cells. While in the experiments of this previous work there was no notion of media transfer and cooperation between cells, it raised the idea for our current work since we are looking for the release of signals to the shared medium and their benefits under oxidative stress.

To identify the contribution of catalase to the protection in the media transfer, we compared the protection of conditioned media from the WT strain and from a catalase deletion strain. If the catalase contributes to the protection, we expect to see that conditioned media from a deletion strain will not have the same level of protection as the conditioned media from the WT strain. In these experiments, we chose to focus on the cytosolic catalase (Ctt1) since it was previously shown that the Ctt1 protein level strongly increases when exposed to H<sub>2</sub>O<sub>2</sub>, while the Cta1 protein level does not.<sup>31</sup> Therefore, we used a  $\Delta ctt1$  strain that was kindly provided by the Schuldiner Lab.

Our results show that the WT strain grows faster during oxidative stress when it has received conditioned media from the WT, which produces Ctt1, compared to its growth upon receiving media from the  $\Delta$ ctt1 strain that does not produce Ctt1 (Figure 13a). This suggests that the catalase does contribute to the protective effect. We observe even bigger response when the receiving cells are the  $\Delta$ ctt1 strain.

Importantly, the  $\Delta ctt1$  strain does poorly during the oxidative stress in comparison to the WT but when it receives the WT conditioned media, this defect is significantly rescued (Figure 13A). On the other hand, when performing the same experiment with non-conditioned media, meaning the first cells did not receive an oxidative stress, we do not see as strong of a protective effect. The  $\Delta ctt1$  strain grows slower than the WT regardless of whether it received media from the WT or the deletion strain (Figure 13B). This indicates that when there is no first stress that induces the production of catalase in the original cells, the transferred conditioned media provides less protection.



Figure 13. Transfer of conditioned media from catalase-producing cells provides a protective effect to fresh cells during oxidative stress. Growth curve measurements of fresh cells in oxidative stress that received either (A) conditioned media or (B) non-conditioned media are shown in OD values over time during continuous growth at 30°C. WT cells receiving media from WT strain is in light blue, WT cells receiving media from  $\Delta$ CTT1 strain is in dark blue,  $\Delta$ CTT1 cells receiving media from WT strain is light green, and  $\Delta$ CTT1 cells receiving media from  $\Delta$ CTT1 strain is in dark green.

To investigate whether we can detect catalase activity in the conditioned media, which would further suggest catalase is secreted to the media, we performed a catalase activity assay on the conditioned media. At mid-log, WT and  $\Delta ctt1$  cells were exposed to varying concentrations of H<sub>2</sub>O<sub>2</sub>. After the oxidative stress, the media was separated from the cells and the concentrations of hydrogen peroxide were measured by a peroxide assay kit to derive catalase activity. As we may expect, in the conditioned media with the H<sub>2</sub>O<sub>2</sub> stress concentrations of 2mM, 1.5mM and 1mM, there are lower concentrations of H<sub>2</sub>O<sub>2</sub> in the media separated from WT cells, that produce Ctt1, than from  $\Delta ctt1$  cells. However, at the low stress concentration of 0.2mM H<sub>2</sub>O<sub>2</sub>, we no longer see that trend, as they both have the same concentration of hydrogen peroxide (Figure 14). Additionally, to isolate the natural breakdown of  $H_2O_2$ , we measured the  $H_2O_2$  concentration in a media containing no cells. We see that both strains break down more  $H_2O_2$  than is naturally broken down, which is in accordance with the fact that a cell has additional methods besides catalase to deal with oxidative stress.

These results support the hypothesis that catalase could work as a public good. However, there is a concern that the protection given by the catalase is not due to its active secretion, but rather due to cell lysis that occurs during the oxidative stress. If cells indeed are lysed their content could be spilled over, realizing a catalase that might serve for unintentional protection, rather than "intentional" cooperation. To explore this possibility, we performed a plate assay to check cell survival after an oxidative stress of 2mM H<sub>2</sub>O<sub>2</sub>, which was the concentration used during the media transfer experiment. By plating cells before and after the stress, a 20% survival rate of cells was observed. The high death rate strengthens the possibility of the catalase being a protective enzyme in the transferred media due to cell death and lysis. An open question that remains is whether, in our system, the dying cells are going through lysis and passively releasing the catalase or if it is an active secretion.



Figure 14. A catalase activity assay shows there is activity of catalase in the media of cells that undergo an oxidative stress of 1mM  $H_2O_2$  and higher. The measurement of catalase activity is derived from the hydrogen peroxide concentration found in media separated from cells that have been exposed to different severities of oxidative stress. The  $H_2O_2$  concentration was measured in media separated from WT cells and  $\Delta$ ctt1 cells, as well as in media that contained no cells, to represent natural breakdown of hydrogen peroxide.

## Discussion

In this work, we delve into understanding the complexity of public good interactions and we present a novel approach to further investigate these interactions.

We explored two different yeast proteins as public-good proteins. The first is the wellestablished public good protein Suc2, that has been studied in multiple previous works,<sup>4–9</sup> and has been shown to generate cooperative behavior by releasing nutritional products into the shared media. The second is a putative public good, catalase, that seems to be released into the shared media either by active secretion or passive lysis and provides protection to fellow cells from oxidative stress. Since the catalase protein is not yet recognized as a public good protein, we focused on demonstrating that it is indeed one. However, as the Suc2 protein is already wellestablished as a public good, it enabled us to further study the complex dynamics of this type of interaction.

Our approach to explore this system in a novel way was by designing and creating thousands of variants with various Suc2 secretion levels to represent a diverse range of strategies in this interaction. We then competed them in different conditions to follow the fitness of the various strategies. We find outlier populations that either have a higher or lower fitness on the sucrose condition than expected based on their fitness in the galactose condition, which represents the variant's cost. By exploring the features of these outlier variants, such as Y content, NNY score and localized enrichment of mutation types, it seems to show that the high fitness outliers are cheater-like variants while the lower fitness variants are cooperators. From the results of the SECReTE variant competition, we see that making alterations to this motif can have effects on the 'strategy' of the variant and push it towards more cooperator or cheater-like activity. By studying the genetic 'grammar' that results in these different strategies by being artificially altered, we can learn in return about the natural 'grammar' of cooperation that could perhaps have been tuned by evolution.

#### Set-up of the Suc2 secretion library system

We will first focus on the Suc2 protein system. To establish our approach, we began with a small-scale investigation. We worked with 2 strains that were created by altering the SECReTE motif and were made to represent extreme strategies. First, the SECReTE (+) is a strain with 56 synonymous mutations surrounding the natural SECReTE signal, altering purines to pyrimidines, to strengthen the potential secretion signal as much as possible. On the other side of the spectrum is the SECReTE (-) strain which has had the natural SECReTE signal removed by making 129 synonymous mutations surrounding that region, that alter pyrimidines to purines, to eliminate the secretion signal. Our findings show that when competing these two strains together with the WT, in co-culture competition assay, the SECReTE (-) overcomes the fitness defect it has when grown alone in a monoculture (Figure 4B-D). This indicates that the SECReTE (-) is able to compensate for its lower secretion level and thus lower production rate of the monosaccharide products by taking advantage of the shared products from the other strains. This result gave us motivation to further pursue this system as we were able to observe the interactions and 'exploitation' between strains.

We performed this small-scale co-culture competition assay with both methods, daily dilution transfer and chemostat. While by dilutions we have many repeats and different starting ratios that all show the same trend of the SECReTE (-) strain remaining stable throughout the assay, the chemostat does not show this trend (Figure 4E). There are many differences between the two systems that could explain the conflicting results, such as population size and shaking speed which can affect cell clumping and biofilm tendency, which can then affect physical proximity between cells. Also, the chemostat retains the cells at mid-log while the daily dilutions system does not. Furthermore, the assay performed in the chemostat was done with no repeats, due to technical reasons, thus the results are not as reliable. This experiment would need to be done in multiple repeats and from different starting cultures, as was done in the assay by dilutions. Although we do not believe we can rely on the results from this chemostat experiment, it gave us valuable information on how to further adjust our competition assays. In the chemostat, already in the first few days, there are strong differences between the strains however, in the daily dilutions we do not see these strong trends. As previously mentioned, in the chemostat, the cells are maintained at mid-log, whereas by dilutions, the cells reach stationary phase each day before being serially transferred. This led us to perform the succeeding competition assay by dilutions (to allow for high-throughput work) but to dilute at mid-log instead of at stationary phase. At mid-log, the cells are still paying the cost of breaking down

sucrose and reaping the benefits, as opposed to stationary phase where the growth is slowing down due to saturation.

After exploring the system on a small-scale, we designed and created our large synthetic library of Suc2 secretion variants (Figure 5). We first attempted to create the library using a relatively recent technology termed CRISPEY.<sup>35</sup> It is a CRISPR-based high-throughput genome editing method that can introduce thousands of precise mutations in parallel in different clones in a population. This method combines a guide RNA, which encodes the 'address', together with the donor DNA, which encodes the desired mutation and each yeast cell gets one of these constructs. This method would have enabled us to create the entire library relatively quickly and all in parallel. Meaning, contained in one pool would be all the variants mutated in the signal peptide, as well the variants mutated in the SECReTE region and so on. However, along the way we discovered that the efficiency of this method was significantly reduced when attempting to introduce more than a point mutation. Although the donor DNA can be a 200bp long fragment, it is optimized to introduce only point mutations and cannot accommodate multiple mutations throughout the donor template. Optimizing this technology to be able to introduce multiple nearby mutations in the same variant would be very helpful for creating various types of yeast libraries in the future. In retrospect, we believe it was to our advantage that we did not create one large pool containing all of the Suc2 secretion variants but rather created a sub-library for each of the four separate locations in the SUC2 gene. This is because the phenotype of signal peptide mutants might be stronger than the phenotype of SECReTE signal mutants and as a result, the more delicate variation between the SECReTE variants may be masked by the more extreme variants.

#### Competition assay of the Suc2 SECReTE variant sub-library

In the results of the SECReTE sub-library competition assay, we see a positive correlation for the variants' fitness between independent repeats, on each media separately (Figure 7A, B). This being said, the correlation isn't as high as we may have expected it to be, e.g. based on previous related experiments.<sup>43</sup> This limited correlation between biological repeats of the evolution suggests that in our library there is a lot of genetic drift occurring relative to selection effects. This means that for many of the variants, the mutations we made on the *SUC2* gene are relatively neutral and do not have a strong effect in these conditions. Therefore, in comparison to

previous experiments the extent of genetic drift, rather than selection, is larger for many of the variants. Since drift is random, the drift on a specific variant is not reproducible and thus, the variant's fitness may not be consistent over repeats. However, the correlation value is similar along all the pairs of repeats being compared on the same medium and this means that the extent of drift, versus selection, is consistent. Still, we see some variants in the population that are found to be consistently under selection, the majority being under a positive selection (Shown by the relatively larger group of variants found to be above the regression line in Figure 7A,B). It is encouraging that the highest correlation is found on the sucrose media, compared to galactose, signifying a relatively lower drift and a stronger selection on the former sugar. Since on sucrose we anticipate seeing the strongest effects of the mutations we introduced, we would expect to see this stronger selection here. Fittingly, the correlation slightly lowers on galactose, suggesting a slightly higher drift and lower selection than on sucrose. On galactose, we expect our mutations to have a relatively lower effect as the fitness differences should be caused solely by cost differences. Finally, the correlation is very low when comparing between the repeats on the two different mediums (Figure 7C). In this case, we expect to see this nonreproducibility and high drift as they are different conditions with different selections. When looking at the fitness of the entire population throughout the experiment, we reassuringly see that the population fitness on the sucrose media has a steeper slope and reaches a higher final fitness than the population on galactose (Figure 8). This result further supports that there is a stronger selection on the sucrose media. Advantageous variants in the population can increase their frequency in the population and thus increase the fitness of the entire population. This result we need and plan to validate experimentally as well, by comparing the growth of the population before and after the competition assay by conventional, e.g. OD-based, growth assays. If we see a faster growth in the population after the competition assay on sucrose compared to the galactose media, this will confirm that the population has increased its fitness on sucrose.

Since this experiment was run for only  $\sim$ 32 generations, it raises the question- how would the population behave on a longer timescale evolution? We have since continued the experiment for another  $\sim$ 32 generations and are currently processing these results but perhaps an even longer timescale is required. If cheater-like variants temporarily take over the population due to their significant fitness advantage,

perhaps this could eventually lead to the collapse of the population. However, I would hypothesize that over a long timescale we would see fluctuations in the population fitness. In this case, the cheater-like variants would increase in frequency and increase the overall population fitness, as I believe we see in this experiment. However, at some point, once the frequency of the cheater-like variants increases too much, there will be less products shared in the medium, and this will result in the fitness advantage returning to the cooperator variants. The cooperators would then increase in the population, bringing down the overall fitness, but this would eventually again lead to the rise of the cheaters and the overall fitness. This cycle would continue reaching a steady state of fluctuations in the population fitness as has been similarly shown when competing between just two strains, WT and  $\Delta suc2$ .<sup>5</sup>

From looking into the fitness on a variant level and comparing each variant's mean fitness between each of the mediums, we were able to identify smaller populations of variants that act differently between the two mediums (Figure 9A). As mentioned in the results, we suspect the cheater-like variants to be found significantly above the regression line (the positive fitness residual variants). This is because cheaters have a higher fitness on sucrose than expected based on their cost, which is derived from their fitness on galactose. This describes a variant that does not pay a high cost since it does not produce and secrete a lot of the Suc2 protein but it is able to use the products that other cells provide so it has a very high fitness on sucrose. Although we expect the cheater-like variants to display such fitness residual values, there are other putative types of variants that could also make up this group. For example, variants that are generally adapted, meaning they have a beneficial mutation unrelated to the Suc2 machinery or to the type of sugar in the medium. This could be investigated by performing the same competition assay of the sub-library on YPD. Since the expression of Suc2 is under glucose repression, we expect to see no fitness differences between the variants unless they are generally adapted. Another possible type of variant that could be found in this population is a putative variant that has a nonsynonymous mutation in the Suc2 protein that has an effect on the protein's activity. If, by chance, it increases the efficiency of the protein, the variant will have a higher benefit without necessarily paying a higher production cost. To further investigate these discussed populations, we defined them as either positive (blue) or negative (red) fitness residual variants. By adding a threshold, we aim to remove variants that are simply under drift and keep only the variants under selection.

Although the following analysis focuses on this grouping of positive and negative fitness residual, another possible way to group the variants is by diving them into quadrants along the regression line. As seen in Figure 9B and C, the plot is divided into the following quadrants- 1. top-left, 2. bottom-left, 3. top-right, and 4. bottom-right.

- 1. Top-left: These variants have a beneficial strategy shown by their relatively high fitness on sucrose but also pay a high cost shown by their low fitness on galactose. This suggests they are cooperators that are able to make their efforts worthwhile. There are not many variants that are found to have a fitness residual above the threshold in this quadrant, which indicates that it may be rare for cooperators to not be at a disadvantage due to their high cost.
- 2. Bottom-left: These variants still have the high cost, but they also have a low fitness on sucrose, showing that their expensive strategy does not pay off. Here we find many variants with a fitness residual above the threshold. This suggests that the cooperator or ultra-cooperator strategies more often do not pay off (represented by quadrant 2) rather than be beneficial (represented by quadrant 1).
- 3. Top-right: These variants have a beneficial strategy shown by their high fitness on sucrose, but do not pay a high cost for it, as we see from their high fitness on galactose. This describes the cheater-like strategy. In this quadrant, there are the most variants found to have a fitness residual above the threshold. This shows that we have a relatively large population of potentially cheater-like variants and that they are the most 'successful' strategies on both mediums.
- 4. Bottom-right: These variants have a poor strategy, as they have a low fitness on sucrose but also do not pay a high cost, since they have a high fitness on galactose. This is the only quadrant that should not logically occur according to the Suc2 system. Since we have shown that the variants are capable of interacting and using shared products from other cells, if the cell is not paying the cost to produce the Suc2 machinery, it should still be able to live of its neighbors. Accordingly, we see almost no fitness residual variants in this quadrant.

It is worth noting that the cost of a variant can be influenced by multiple unknown factors besides just the variation in its secretion level and could thus potentially add

further variability into the population. A cell's cost could be impacted by the processing, e.g. glycosylation, or stability of the invertase protein, as well as the expression and regulation of hexose transporters and hexose kinases.

Two important features of the SECReTE motif that we would anticipate having an effect on the variants are their Y content and NNY score. These two features are intertwined, and thus we would expect to see similar trends between them. Both features show a lower content/score in the positive fitness residual variants (Figure 10 and Figure 11), which are characteristics that should decrease the strength of the SECReTE signal. Presumably, by lowering the secretion signal, these variants lower their Suc2 secretion level and consequently, their cost. This result points towards the positive fitness residual variants indeed being cheater-like variants. However, as we also see in the results, the signal of this finding is not very strong and does not remain significant throughout all thresholds (Figure 10B, D and Figure 11C). In Figure 10B, the p-value of the comparison between the Y content in the positive fitness residual variants to the entire library is significant at some thresholds but not at all of them. We do see that the p-value remains low and keeps around the cut-off value of 0.05 throughout the thresholds. In contrast, once we move to Figure 10D, where we see the p-value of the same comparison but only in variants with exclusively synonymous mutations, the p-value jumps to high and non-significant values as we increase the threshold. It is worth noting that the number of variants included in this group after filtering only the synonymous mutation variants is much lower than the previous comparison. This may lower the statistical power and be a cause of the jump in the pvalue. Similarly, when looking at the NNY score of the exclusively synonymous mutation variants, the p-value of this comparison is significant at the 2 lower thresholds but jumps up as we increase the threshold and lower the number of variants included in the comparison (Figure 11C).

Perhaps looking for a significant signal consistently throughout the entire variable region will dilute true signals. Therefore, we next moved to looking at the individual positions throughout the variable region. When doing so, we can see strong trends that are found in only certain locations. The enrichment of R2Y mutations in negative fitness residual variants immediately upstream to the natural SECReTE motif, as seen in Figure 12, supports our notion that these variants are cooperators. As previously shown, increasing the number of repeats in the SECReTE motif can enhance the secretion level of the protein.<sup>21</sup> These variants have an enhanced and lengthened

SECReTE motif, possibly enabling them to be stronger cooperators by secreting more Suc2. However, this signal switches at about ~80bp upstream to the natural SECReTE motif and these variants have a depletion for this type of mutation. In this location, it may no longer be important for increasing the secretion signal. The switching of the signal in different locations along the variable region could explain why we would not see this result when looking at the general Y content or NNY score of the variant, as in Figure 10 and Figure 11. Also, interestingly there is a depletion in R2Y mutations in the positive fitness residuals, suggesting their avoidance of increasing the SECReTE signal. This further fits with idea that this population is made up of mainly cheater-like variants.

#### Catalase as a potential public good

In the catalase experiments, we were able to show that catalase in the media is able to provide protection to surrounding cells. Furthermore, catalase is released to the media when cells are exposed to oxidative stress and can then pass on this protection to fresh cells that had no prior stress (Figure 13). However, we have not yet determined if the release of catalase to the media is due to active secretion or passive release during cell lysis. Originally, we were hoping to show that catalase is actively secreted to the media so that this protein can fit the classic set-up of a public good that we are already familiar with. However, due to the high cell death we measured after the oxidative stress, there is a good chance the catalase is released into the media through cell death. To resolve this currently un-answered question, we plan to perform a non-destructive cytolysis assay. In this assay, it is designed to measure the release of the enzyme adenylate kinase, a robust protein present in all eukaryotic cells, which is released into the culture media when cells die.<sup>44,45</sup> If the cells are unselectively releasing proteins due to cell death, this would strongly support that the catalase is found in the media due to permeable membranes and lysis.

If we do not detect cell lysis, then we can determine our catalase protein as a public good in the classic and established understanding and further investigate this interaction as we did in the Suc2 system. However, if we find there is lysis and that catalase is passively released into the shared medium, it does not necessarily mean we cannot consider it as a public good, though potentially in a new sense of the term. Possibly, we have stumbled upon an unfamiliar form of public good interaction where the cooperative action is performed through cell death and lysis. Although we want to avoid anthropomorphism, perhaps the cells 'sacrifice' themselves to provide protection to others. This may be un-intended and just a byproduct of the unavoidable cell death due to stress. Alternatively, it could be some sort of group selection, where a certain percentage of cell death in the population is advantageous for the group and enables their survival from the stress and thus, cell death is in part selected for. It would be very interesting to pursue these theories and further investigate this protein in the context of cooperation.

#### **Future plans**

There are many avenues left to pursue and future experiments we plan to perform to advance our understanding in this work.

First, we have recently continued the SECReTE sub-library competition assay for another ~32 generations to observe how the frequencies and fitness of the variants change over a longer span of time. Based on these findings, we may decide to perform the assay on an even longer timescale. For example, as previously discussed, we'd like to follow the population fitness over a longer timescale to see if the cheater-like population continues to increase in frequency eventually leading to collapse of the population or if they would fluctuate and decrease, giving opportunity to the cooperators to rise, which would enable a steady state.

Additionally, we plan to perform the same competition assay on the rest of the sublibraries (signal peptide, control region and 3'UTR). We will soon begin the competition assay of the signal peptide library. This is an interesting sub-library since, although the signal peptide is very well researched, a large-scale study of the effects of various mutations throughout the signal peptide will provide new information on what effects the secretion level of this signal. While there are some main determinants known to be preserved,<sup>13</sup> maybe this library could enlighten us further about the pattern that determines the efficiency of this signal and sequence-function characteristics. Also, as the signal peptide has a strong influence on the secretion of the protein, maybe these variants will have stronger phenotypes, ensuing stronger fitness effects in our assays.

Moreover, it would be interesting to perform these competition assays on solid agar plates, as well as on the liquid media. On liquid media, there is a homogenous distribution of the nutrients caused by shaking of the media, which gives an advantage to the cheaters as the products are easily accessible. However, on a solid media the

57

products do not diffuse the same and proximity is much more important. If the cheater is not physically near a cooperator, their 'success' in the population may be different. Performing this competition assay at various cell densities could shed light on this. Another very important assay to substantiate our results is measuring the ER localization of the variant mRNAs. To remind, the SECReTE motif is presumably a signal to localize mRNA to the ER. We will perform ER fractionation on the various sub-libraries to isolate the fraction of the ER from the cytosol. We will then follow by deep RNA sequencing of the variable region of the *SUC2* mRNA. By comparing the localization of the variant mRNA to the ER versus the cytosol, we can examine the direct effects of the mutations we have generated. Since increased localization to the ER of variant mRNAs is associated with their increased secretion, we can compare this to the fitness outcome of the corresponding variants to learn more about the dynamics of the interaction.

Finally, we discussed above the advantage of pursing catalase as a public good protein due to the potential that it can provide protection both intracellularly and extracellularly. We explored the notion of a cell being able to either retain its public good intracellularly and keep the benefits private or secrete the protein and share the public good with others. Although it seems this may not be the case with catalase, since catalase is possibly released during cell death, perhaps we could explore this in the Suc2 system. There is a maltose proton symporter, encoded by the AGT1 gene, that can also transport sucrose into the cell.<sup>46</sup> In s288c derived strains, there is a mutation in the maltose genes' regulator thus these strains cannot express this transporter. However, if we constitutively express this symporter among all strains in our library, we could create variants that could both use their extracellular invertase to cooperate with others, as well as import the sucrose and keep the products for themselves. On a similar note, in a very interesting and recent work, a 'private metabolizer' strain was created by introducing a sucrose transporter to a yeast strain containing only the internal invertase (no signal peptide) and showed how the strategy of private metabolizing can cause population decline.<sup>47</sup> I believe it would be fascinating to explore the dynamics of this Suc2 interaction if the cells had the ability to go back and forth between privatizing their products and publicly sharing them. The spectrum of strategies would now have three extremes, cheating, fully cooperating and fully privatizing.

## References

- 1. Hamilton WD. The genetical evolution of social behaviour. II. *J Theor Biol*. 1964;7(1):17-52. doi:10.1016/0022-5193(64)90039-6
- 2. Smukalla S, Caldara M, Pochet N, et al. FLO1 is a variable green beard gene that drives biofilm-like cooperation in budding yeast. *Cell*. 2009;135(4):726-737. doi:10.1016/j.cell.2008.09.037.FLO1
- 3. Wilson DS. A theory of group selection. *Proc Natl Acad Sci U S A*. 1975;72(1):143-146. doi:10.1073/pnas.72.1.143
- 4. Greig D, Travisano M. The Prisoner's Dilemma and polymorphism in yeast SUC genes. *Proc R Soc B Biol Sci.* 2004;271(SUPPL. 3). doi:10.1098/rsbl.2003.0083
- 5. Gore J, Youk H, Van Oudenaarden A. Snowdrift game dynamics and facultative cheating in yeast. *Nature*. 2009;459(7244):253-256. doi:10.1038/nature07921
- 6. Koschwanez JH, Foster KR, Murray AW. Sucrose utilization in budding yeast as a model for the origin of undifferentiated multicellularity. *PLoS Biol.* 2011. doi:10.1371/journal.pbio.1001122
- MacLean RC, Fuentes-Hernandez A, Greig D, Hurst LD, Gudelj I. A Mixture of "Cheats" and "Co-Operators" Can Enable Maximal Group Benefit. Keller L, ed. *PLoS Biol.* 2010;8(9):e1000486. doi:10.1371/journal.pbio.1000486
- 8. CRAIG MACLEAN R, BRANDON C. Stable public goods cooperation and dynamic social interactions in yeast. *J Evol Biol*. 2008;21(6):1836-1843. doi:10.1111/j.1420-9101.2008.01579.x
- Sanchez A, Gore J. Feedback between Population and Evolutionary Dynamics Determines the Fate of Social Microbial Populations. Ellner SP, ed. *PLoS Biol*. 2013;11(4):e1001547. doi:10.1371/journal.pbio.1001547
- Carlson M, Botstein D. Two differentially regulated mRNAs with different 5' ends encode secreted and intracellular forms of yeast invertase. *Cell*. 1982;28(1):145-154. doi:10.1016/0092-8674(82)90384-1
- Vidgren V, Ruohonen L, Londesborough J. Characterization and functional analysis of the MAL and MPH loci for maltose utilization in some ale and lager yeast strains. *Appl Environ Microbiol*. 2005;71(12):7846-7857. doi:10.1128/AEM.71.12.7846-7857.2005
- Koschwanez JH, Foster KR, Murray AW. Improved use of a public good selects for the evolution of undifferentiated multicellularity. *Elife*. 2013;2013(2). doi:10.7554/eLife.00367
- 13. von Heijne G. The signal peptide. *J Membr Biol*. 1990;115(3):195-201. doi:10.1007/BF01868635
- Akopian D, Shen K, Zhang X, Shan S. Signal Recognition Particle: An Essential Protein-Targeting Machine. *Annu Rev Biochem*. 2013;82(1):693-721. doi:10.1146/annurev-biochem-072711-164732
- 15. Liu S, Bugos RC, Dharmasiri N, Su WW. Green fluorescent protein as a secretory reporter and a tool for process optimization in transgenic plant cell cultures. *J Biotechnol*. 2001;87(1):1-16. doi:10.1016/S0168-1656(00)00421-1
- 16. Young R, Rance J. Mammalian expression vector with a highly effient secretory signal sequence. *Pat No WO 2008/148519 A2*. 2008.
- Freudl R. Signal peptides for recombinant protein secretion in bacterial expression systems. *Microb Cell Fact.* 2018;17(1). doi:10.1186/s12934-018-0901-3

- 18. Schuldiner M, Metz J, Schmid V, et al. The GET Complex Mediates Insertion of Tail-Anchored Proteins into the ER Membrane. *Cell*. 2008;134(4):634-645. doi:10.1016/j.cell.2008.06.025
- 19. Aviram N, Ast T, Costa EA, et al. The SND proteins constitute an alternative targeting route to the endoplasmic reticulum. *Nature*. 2016;540(7631):134-138. doi:10.1038/nature20169
- 20. Rabouille C. Pathways of Unconventional Protein Secretion. *Trends Cell Biol.* 2017;27(3):230-240. doi:10.1016/j.tcb.2016.11.007
- 21. Cohen-Zontag O, Baez C, Lim LQJ, et al. A Secretion-Enhancing Cis Regulatory Targeting Element (SECReTE) Involved in MRNA Localization and Protein Synthesis. Vol 15.; 2019. doi:10.1371/journal.pgen.1008248
- 22. Akopian D, Shen K, Zhang X, Shan S-O. Signal Recognition Particle: An essential protein targeting machine. doi:10.1146/annurev-biochem-072711-164732
- 23. Izawa S, Inoue Y, Kimura A. Importance of Catalase in the Adaptive Response to Hydrogen Peroxide : Analysis of Acatalasaemic Saccharomyces Cerevisiae. Vol 320.; 1996.
- 24. Switala J, Loewen PC. Diversity of properties among catalases. *Arch Biochem Biophys*. 2002;401(2):145-154. doi:10.1016/S0003-9861(02)00049-8
- Chaga GS, Medin AS, Chaga SG, Prath JO. Isolation and characterization of catalase from Penicillium chrysogenum. *J Chromatogr A*. 1992;604(1):177-183. doi:10.1016/0021-9673(92)85542-2
- 26. Garre V, Tenberge KB, Eising R. Secretion of a fungal extracellular catalase by Claviceps purpurea during infection of rye: Putative role in pathogenicity and suppression of host defense. *Phytopathology*. 1998;88(8):744-753. doi:10.1094/PHYTO.1998.88.8.744
- Gille G, Sigler K, Hofer M. Response of catalase activity and membrane fluidity of aerobically grown Schizosaccharomyces pombe and Saccharomyces cerevisiae to aeration and, the presence of substrates. *J Gen Microbiol*. 1993;139(7):1627-1634. doi:10.1099/00221287-139-7-1627
- 28. KAPLAN JG. The alteration of intracellular enzymes. II. The relation between the surface and the biological activities of altering agents. *J Gen Physiol*. 1954;38(2):197-211. doi:10.1085/jgp.38.2.197
- Hiltunen JK, Mursula AM, Rottensteiner H, Wierenga RK, Kastaniotis AJ, Gurvitz A. The biochemistry of peroxisomal β-oxidation in the yeast Saccharomyces cerevisiae. *FEMS Microbiol Rev.* 2003;27(1):35-64. doi:10.1016/S0168-6445(03)00017-2
- Lushchak O V., Lushchak VI. Catalase modifies yeast Saccharomyces cerevisiae response towards S-nitrosoglutathione-induced stress. *Redox Rep.* 2008;13(6):283-291. doi:10.1179/135100008X309037
- 31. Martins D, English AM. Catalase activity is stimulated by H2O2 in rich culture medium and is required for H2O2 resistance and adaptation in yeast. *Redox Biol.* 2014;2(1):308-313. doi:10.1016/j.redox.2013.12.019
- 32. Winzeler EA, Shoemaker DD, Astromoff A, et al. Functional characterization of the S. cerevisiae genome by gene deletion and parallel analysis. *Science (80-)*. 1999;285(5429):901-906. doi:10.1126/science.285.5429.901
- 33. Stovicek V, Borja GM, Forster J, Borodina I. EasyClone 2.0: expanded toolkit of integrative vectors for stable gene expression in industrial Saccharomyces cerevisiae strains. *J Ind Microbiol Biotechnol*. 2015;42(11):1519-1531. doi:10.1007/s10295-015-1684-8

- 34. Ram Y, Dellus-Gur E, Bibi M, et al. Predicting microbial growth in a mixed culture from growth curve data. *Proc Natl Acad Sci U S A*. 2019;116(29):14698-14707. doi:10.1073/pnas.1902217116
- 35. Sharon E, Chen S-AA, Khosla NM, Smith JD, Pritchard JK, Fraser HB. Functional Genetic Variants Revealed by Massively Parallel Precise Genome Editing. *Cell*. 2018;175:544-557.e16. doi:10.1016/j.cell.2018.08.057
- Celiker H, Gore J. Competition between species can stabilize public-goods cooperation within a species. *Mol Syst Biol.* 2012;8:621. doi:10.1038/msb.2012.54
- Özcan S, Vallier LG, Flick JS, Carlson M, Johnston M. Expression of the SUC2 gene of Saccharomyces cerevisiae is induced by low levels of glucose. *Yeast*. 1997;13(2):127-137. doi:10.1002/(sici)1097-0061(199702)13:2<127::aid-yea68>3.0.co;2-%23
- 38. Plotkin JB, Kudla G. Synonymous but not the same: The causes and consequences of codon bias. *Nat Rev Genet*. 2011;12(1):32-42. doi:10.1038/nrg2899
- 39. Vaser R, Adusumalli S, Ngak Leng S, Sikic M, Ng PC. SIFT missense predictions for genomes. 2015. doi:10.1038/nprot.2015.123
- 40. Day M. Yeast petites and small colony variants: For everything there is a season. In: *Advances in Applied Microbiology*. Vol 85. Academic Press Inc.; 2013:1-41. doi:10.1016/B978-0-12-407672-3.00001-0
- 41. Gille G, Sigler K, Hofer M. Response of catalase activity and membrane fluidity of aerobically grown Schizosaccharomyces pombe and Saccharomyces cerevisiae to aeration and the presence of substrates. *J Gen Microbiol*. 1993;139(7):1627-1634. doi:10.1099/00221287-139-7-1627
- 42. Mitchell A, Romano GH, Groisman B, et al. Adaptive prediction of environmental changes by microorganisms. *Nature*. 2009;460(7252):220-224. doi:10.1038/nature08112
- 43. Frumkin I, Schirman D, Rotman A, et al. Gene Architectures that Minimize Cost of Gene Expression. *Mol Cell*. 2017;65(1):142-153. doi:10.1016/j.molcel.2016.11.007
- 44. Jacobs AC, DiDone L, Jobson J, Sofia MK, Krysan D, Dunman PM. Adenylate kinase release as a high-throughput-screening-compatible reporter of bacterial lysis for identification of antibacterial agents. *Antimicrob Agents Chemother*. 2013;57(1):26-36. doi:10.1128/AAC.01640-12
- 45. Crouch SPM, Kozlowski R, Slater KJ, Fletcher J. The use of ATP bioluminescence as a measure of cell proliferation and cytotoxicity. *J Immunol Methods*. 1993;160(1):81-88. doi:10.1016/0022-1759(93)90011-U
- 46. Stambuk —y<sup>\*</sup>y BU, Da Silva MA, Panek AD, De Araujo PS. *Active K-Glucoside Transport in Saccharomyces Cerevisiae*. doi:10.1111/j.1574-6968.1999.tb13361.x
- 47. Lindsay RJ, Pawlowska BJ, Gudelj I. Privatization of public goods can cause population decline. *Nat Ecol Evol*. 2019;3(8):1206-1216. doi:10.1038/s41559-019-0944-9

## Acknowledgments

First and foremost, I would like to thank my mentor, Prof. Yitzhak Pilpel, who has provided endless guidance, support, and so many stimulating discussions! I would like to thank Dr. Orna Dahan for her valuable input and boundless patience throughout this project. A special thanks to Dvir Schirman, who has been helping me navigate this project since day one and has taught me so much. To all the members of the lab, thank you for all your helpful insights, as well as making the lab such a welcoming environment to work in each day. Finally, on a personal note, I'd like to thank my continuously supportive family and friends.

# Appendix

## Table 2. List of primers.

| ID # | Primer name      | Used for                         | Sequence                    |  |
|------|------------------|----------------------------------|-----------------------------|--|
| 1    | LibAmp_SP_F      | Oligo amplification              | GCTTTTCTTTTCACTAACG         |  |
| 2    | LibAmp_SP_R      | Oligo amplification              | GAAAGTACAGATGCCATTTG        |  |
| 3    | LibAmp_ins1_F    | Oligo amplification              | GAGTGGGTTTTTCAATGATAC       |  |
| 4    | LibAmp_ins1_R    | Oligo amplification              | GAGAAGGTTCATACCAGAACAC      |  |
| 5    | LibAmp_ins2_F    | Oligo amplification              | AGATCAACCCATTGCTATCG        |  |
| 6    | LibAmp_ins2_R    | Oligo amplification              | TAAAAGTGTAACCACCATCAAG      |  |
| 7    | LibAmp_ins3_F    | Oligo amplification              | TTATAACACTCCTGAAAGTG        |  |
| 8    | LibAmp_ins3_R    | Oligo amplification              | AGGAGTAAATTTCAATTTTGTAG     |  |
| 9    | LibAmp_control_F | Oligo amplification              | CGAACCAATATTGAACATTAG       |  |
| 10   | LibAmp_control_R | Oligo amplification              | CTAAACCCTTGAACCAAAG         |  |
| 11   | LibAmp_3UTR_F    | Oligo amplification              | CAAGTTCCAAGTAAGGGAAG        |  |
| 12   | LibAmp_3UTR_R    | Oligo amplification              | CATAAAGAACAAATTCCAGGTAAC    |  |
| 13   | IntVec_gibson_F  | pCfB2223 linearization           | AAGCTATCTGTAACAGGAGCATCGCG  |  |
|      |                  |                                  | TGCATTCATCCG                |  |
| 14   | IntVec_gibson_R  | pCfB2223 linearization           | GATTGTGACGTGTGGATGCTATCGC   |  |
|      |                  |                                  | ACGCATTCCGTTG               |  |
| 15   | pGS2223_SPlib_F  | pGS2223 linearization            | CAAATGGCATCTGTACTTTC        |  |
| 16   | pGS102_SPlib_R   | pGS2223 linearization            | CATATACGTTAGTGAAAAGAAAAGC   |  |
| 17   | pGS102_ins1lib_F | pGS2223 linearization            | AAGGTGTTCTGGTATGAACC        |  |
| 18   | pGS102_ins1lib_R | pGS2223 linearization            | ATCAATAGTATCATTGAAAAAACCC   |  |
| 19   | pGS102_ins2lib_F | pGS2223 linearization            | TCTCTTGATGGTGGTTACAC        |  |
| 20   | pGS102_ins2lib_R | pGS2223 linearization            | CTTGGGAGCGATAGCAATG         |  |
| 21   | pGS102_ins3lib_F | pGS2223 linearization            | GACTACAAAATTGAAATTTACTCC    |  |
| 22   | pGS102_ins3lib_R | pGS2223 linearization            | TTGCTCTTCACTTTCAGGAG        |  |
| 23   | pGS102_cntllib_F | pGS2223 linearization            | TTATCACTTTGGTTCAAGGG        |  |
| 24   | pGS102_cntllib_R | pGS2223 linearization            | AGCATTACTAATGTTCAATATTGG    |  |
| 25   | pGS102_3UTRlib_F | pGS2223 linearization            | AGTTACCTGGAATTTGTTC         |  |
| 26   | pGS102_3UTRlib_R | pGS2223 linearization            | CCAAGTAAGGGAAGTAAAATAG      |  |
| 27   | IntVec_X3_2220   | Verification of yeast            | CCTGCAGGACTAGTGCTGAG        |  |
|      |                  | (down junction)                  |                             |  |
| 28   | IntVec_X3_904    | Verification of yeast            | CCGTGCAATACCAAAATCG         |  |
|      |                  | sub-libraries<br>(down junction) |                             |  |
| 29   | IntVec_X3_2221   | Verification of yeast            | GTTGACACTTCTAAATAAGCGAATTTC |  |
|      |                  | sub-libraries                    |                             |  |
| 30   | IntVec_X3_903    | Verification of yeast            | TGACGAATCGTTAGGCACAG        |  |
|      |                  | sub-libraries                    |                             |  |

|    |                      | (up junction)           |                               |
|----|----------------------|-------------------------|-------------------------------|
| 31 | Suc2_F_illumina      | PCR1 in amplicon        | AGACGTGTGCTCTTCCGATCTTTTGGT   |
|    |                      | sequencing of small-    | CCACTTCACACCC                 |
|    |                      | scale competition       |                               |
| 32 | Suc2_R_illumina      | PCR1 in amplicon        | ACGACGCTCTTCCGATCTATGTATT     |
|    |                      | sequencing of small-    | GCTCTTCACTTTC                 |
|    |                      | scale competition       |                               |
| 33 | SECReTE_illuminaF1   | PCR1 in amplicon        | ACGACGCTCTTCCGATCTAGATCAACC   |
|    |                      | sequencing of large-    | CATTGCTATCGC                  |
|    |                      | scale competition       |                               |
| 34 | SECReTE_illuminaF2   | PCR1 in amplicon        | ACGACGCICITCCGATCITAGATCAAC   |
|    |                      | sequencing of large-    | CCATTGCTATCGC                 |
| 25 | SECD TE '11 E2       | scale competition       |                               |
| 35 | SECRETE_IIIuminaF3   | PCR1 in amplicon        | ACGACOCICITCCGATCIACAGATCA    |
|    |                      | sequencing of large-    | ACCCATTGCTATCGC               |
| 36 | SECRATE illuminaE4   | PCP1 in amplicon        | ACGACGCTCTTCCGATCTCTCAGATCA   |
| 50 | SECKETE_IIIuIIIIIa14 | sequencing of large-    |                               |
|    |                      | scale competition       | ACCCATTGCTATCGC               |
| 37 | SECReTE illuminaE5   | PCR1 in amplicon        | ACGACGCTCTTCCGATCTTTAGAGATC   |
|    |                      | sequencing of large-    |                               |
|    |                      | scale competition       | AACCCATTOCTATCOC              |
| 38 | SECReTE_illuminaR    | PCR1 in amplicon        | AGACGTGTGCTCTTCCGATCTGAGAA    |
|    |                      | sequencing of large-    | GGTTCATACCAGAACAC             |
|    |                      | scale competition       |                               |
| 39 | Illumina_i5_F        | PCR2 in amplicon        | AATGATACGGCGACCACCGAGATCTA    |
|    |                      | sequencing of small-    | CACTCTTTCCCTACACGACGCTCTTCC   |
|    |                      | scale competition       | GATCT                         |
| 40 |                      |                         |                               |
| 40 | Illumina_15_F2       | PCR2 in amplicon        | AATGATACGGCGACCACCGAGATCTA    |
|    |                      | sequencing of large-    | CACATGCTCCGACACTCTTTCCCTACA   |
|    |                      | scale competition       | CGACGCTCTTCCGATCT             |
| 41 | Illumina_i7_R        | PCR2 in amplicon        | CAAGCAGAAGACGGCATACGAGAT      |
|    |                      | sequencing with indexes | [index]GTGACTGGAGTTCAGACGTGTG |
|    |                      |                         | CTCTTCCGATCT                  |
|    |                      |                         |                               |

## Table 3. List of PCR reactions.

| ID | PCR reaction   | Primers | PCR program   | Used for                                   |
|----|--|---------|---|--|
| #  |  | used    |   |  |
| 1  | Phusion High-Fidelity PCR<br>Master Mix: 25 ulF primer: 7.5 ulR primer: 7.5 ulDNA: varied*<br>ddW: complete volume to 50ul*The amount of oligo DNA<br>added was dependent on the<br>percentage of the sub-library<br>out of the entire library.SP – 2ul (dil 1:10) and<br>increased primers to 11.5ul<br>each due to complications with<br>reaction.Ins1,2,3 – 7.5ul (dil 1:10)<br>Control – 1ul (no dil)<br>3UTR - 4ul (dil 1:10) | 1-12    | <ol> <li>98°C 30 sec</li> <li>98°C 10 sec</li> <li>3. X°C 20 sec **</li> <li>72°C 30 sec</li> <li>5. Repeat steps 2-4 for<br/>20 cycles</li> <li>72°C 10 min</li> <li>** SP, control, ins3 - 47°C</li> <li>3UTR - 50°C</li> <li>Ins1,2 - 51°C</li> </ol>  | Oligo<br>amplification                     |
| 2  | <u>KAPA HiFi HotStart</u><br><u>ReadyMix:</u> 25 ul<br><u>F primer:</u> 1.5 ul<br><u>R primer:</u> 1.5 ul<br><u>Plasmid:</u> ~25ng<br><u>ddW:</u> complete volume to 50ul  | 13-14   | <ol> <li>95°C 3 min</li> <li>98°C 20 sec</li> <li>56°C 15 sec</li> <li>72°C 5 min</li> <li>Repeat steps 2-4 for 25 cycles</li> <li>72°C 9 min</li> </ol>  | pCIB2223<br>linearization                  |
| 3  | KAPA HiFi HotStart<br>ReadyMix: 25 ul<br><u>F primer:</u> 1.5 ul<br><u>R primer:</u> 1.5 ul<br><u>Plasmid</u> : ~30ng<br><u>ddW:</u> complete volume to 50ul   | 15-26   | <ol> <li>95°C 3 min</li> <li>98°C 20 sec</li> <li>50°C 15 sec</li> <li>72°C 5 min</li> <li>Repeat steps 2-4 for<br/>25 cycles<br/>72°C 9 min</li> <li>** for primer 15+16 of SP<br/>sub-library, steps 3 +4 were</li> <li>changed to:         <ol> <li>49°C 15 sec</li> <li>72°C 1 min</li> </ol> </li> </ol> | pGS2223<br>linearization                   |
| 4  | KAPA HiFi HotStart<br>ReadyMix: 25 ul<br><u>F primer:</u> 1.5 ul<br><u>R primer:</u> 1.5 ul<br><u>DNA</u> : 2ul of lysate<br><u>ddW:</u> complete volume to 50ul   | 27-30   | <ol> <li>95°C 3 min</li> <li>98°C 20 sec</li> <li>53°C 15 sec</li> <li>72°C 20 sec</li> <li>Repeat steps 2-4 for<br/>30 cycles<br/>72°C 3 min</li> </ol>  | Verification of<br>yeast sub-<br>libraries |

| 5 | KAPA HiFi HotStart   | 31-38 | 1. 95°C 3 min   | PCR1 in    |
|---|--|-------|---|------------|
|   | ReadyMix: 12.5 ul  |       | 2. 98°C 20 sec  | amplicon   |
|   | F primer: 0.75 ul  |       | 3. 55-58°C 15 sec **  | ·          |
|   | <u>R primer:</u> 0.75 ul   |       | 4. 72°C 15 sec  | sequencing |
|   | <u>DNA</u> : ~30ng   |       | 5. Repeat steps 2-4 for   |            |
|   | ddW: complete volume to 25ul                                       |       | 20 cycles   |            |
|   | -  |       | 72°C 1 min  |            |
|   |  | 20.11 | ** 58 °C in small-scale<br>competition and 55 °C in large-<br>scale competition | DCD2 :     |
| 6 | Phusion High-Fidelity PCR  | 39-41 | 1. $98^{\circ}C$ 30 sec   | PCR2 in    |
|   | Master Mix: 25 ul  |       | 2. $98^{\circ}C 10 \sec$  | amplicon   |
|   | <u>F primer:</u> 2.5 ul  |       | 3. $58 - 64^{\circ}C$ 20 sec **   | sequencing |
|   | <u>R primer:</u> 2.5ul   |       | 4. $72^{\circ}$ C 15 sec  | sequencing |
|   | DNA: Iul (ranged from 1-10   |       | 5. Repeat steps 2-4 for   |            |
|   | ng)  |       | 15 cycles   |            |
|   | $\frac{\mathrm{d}\mathbf{d}\mathbf{W}}{\mathrm{O}_{\mathrm{rel}}}$ |       | /2°C 5 min  |            |
|   | 901  |       | ** 58 °C in small-scale   |            |
|   |  |       | competition and 64 °C in large-   |            |
|   |  |       | scale competition   |            |
|   |  |       | _   |            |
|   |  |       |   |            |