University of California, Irvine

Building tools to improve the scope of continuous directed evolution

Dissertation

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Mathematical, Computational, and Systems Biology

by

Ria Anand Deshpande

May 16, 2023

Dissertation Committee:
Dr. Chang Liu, Chair
Dr. Jun Allard
Dr. Fangyuan Ding
Dr. John Chaput

To Baba and Mummy

and to Arul and Rucha

Thank for you being an inexhaustible

source of encouragement and support!

Contents

List of Figures

List of Tables

"However great your dedication, you never win anything on your own"

- Rafael Nadal

1 Acknowledgements

First, I would like to thank my advisor, Dr. Chang Liu, for his constant support and encouragement throughout my PhD journey. I am eternally grateful to him for pushing me intellectually, teaching me how to think and for making me a better scientist.

I am also very thankful to my advancement and dissertation committees consisting of Dr. Jun Allard, Dr. Fangyuan Ding, Dr. John Chaput, Dr. Michelle Digman and Dr. Timothy Downing, who have always been very open to discussing both work and life at many points during my journey. Learning and being inspired after meetings with them have been an important part of my PhD journey.

I would like to thank members of the Liu Lab and my friends for being the brightest spots of my time at UC Irvine. They have made all the hours in lab very enjoyable through scientific discussions and all the antics otherwise. They are all fantastic scientists, and I am grateful to have spent time learning with and from, them. I started my time in the Liu Lab, working very closely with Dr. Theresa Loveless who has helped me since. She's always been available to talk about new projects and ideas and is always one of the first people I want to tell when things are working as well as when things aren't. She is an incredible scientist and I am very fortunate to have been mentored by her. I would also like to thank Dr. Helen Hobbs for her friendship and the joy she brings to all those around her.

My PhD journey would have been incomplete and a LOT less fun without Sumbul Haider, Vincent Hu and Jonathan Clements, all soon to be PhDs. I am grateful for their unending support through the many hours spent, eating, arguing and having fun.You have always inspired me and encouraged me when things were going well, but especially when we spent late hours in lab working on experiments. You never let me feel too bad when things weren't working and never encouraged e too much when I was contemplating drastic career moves. I am so lucky to have met you and that I get to be friends with you!

Big thanks to Samuel Morabito and Fairlie Reese, for all the fun times we have had so far. Thank you for feeding me and letting me play with your kitties. I am excited to continue exploring new places with you both! Also thank you to my cohort from the Mathematical, Computational and Systems Biology Program for their help with homeworks and for always being ready to banter. I am grateful to be part of such a supportive and diverse community that truly cares for each other.

I am so lucky to have people around the world who have been invested in me and have always been only a phone call way. Having friends in different timezones has ensured that I have always had someone to talk to no matter the time of day. I am immensely grateful for your support and company especially during the pandemic. I would to thank Pooja Hindocha for always being real with me, for Adwait Sadwilkar who I have complained to on numerous occasions and who has magically not gotten tired of me, for Bettina Lengger for keeping me motivated towards the end of this PhD journey. I am grateful to Srushti Paranjpe for being a patient ear and my soul sister who is always ready to have deep conversations and fight people for me.To Yash Deshpande, thanks for being a great friend, you can do this!

I am so so grateful to my local Irvine community for being there for me when I needed it. I want to thank Sharon Trang and Yuki Wong for constantly reminding that there is life outside of

lab and for making it so enjoyable. I have relied on you, sometimes too much for everything from boba to discussing life's issues. I am going to miss living this close to you. I am also glad that Naoki Kanagawa found his way onto our couch all those years ago and that he has stayed such a loving and gentle presence in my life since. I am indebted to Nalini Aunty and Sharad Uncle, and to Ramesh Uncle and Sudha Aunty for making sure I was fed and taken care of me especially when I wasn't well. You always made me feel welcome and part of your family. A thank you is not sufficient for what it has meant to me.

Most importantly, none of this would have been possible without my family. My family has been a constant source of support and inspiration. A huge thank you to all my cousins for checking in on me and for hanging out with me when I have visited and on zoom. I miss you all terribly and am very grateful for all the shenanigans. A huge thanks to Gayatri Buzruk and Anushka Deshande, for leading very colorful lives and for making me feel a part of them even from far away. To Chinmay Buzruk and Tejal Mehendale for all our conversations and all the stupid jokes all of my life.

To my brother Arul, I am so glad I get to go through life with you by my side. You've been the source of so much joy and many stupid decisions through all of it, and I wouldn't have it any other way. I am grateful for your sense of humor Thank you for bringing Rucha into all our lives and thank you to Rucha for bringing so much joy into his life and ours. You're the little sister I never had, and have always wanted. I am looking forward to many more adventures with the two of you and all the bad jokes that are inevitable.

To my mother, Sonali Deshpande, who tries very hard to understand what I do. Thank you for letting me be and for loving and supporting me through all the ups and downs. You've taught me to not be afraid and for reminding me that I can do hard things. Thank you for not making me feel too guilty for not being home as much as you would like me to be. I knew I wanted to do a PhD when I was very young because I was very inspired by my father, Dr. Anand Deshpande. You inspire me everyday to work hard and chase my dreams. Thank you seems very trivial for all that you have done for me. You've taught me to not get disheartened by failure and to be persistent in the face of adversity. To face every challenge with courage and to not compromise on my principles and to do all of this with grace and joy. For this and everything else, I'll always be grateful to you both.

2 Curriculam vitae

Ria Anand Deshpande

Education

University of California, Irvine — Irvine, CA **August 2018 - May 2023**
PhD in Mathematical, Computational and Systems Biology
Advisor: Prof. Chang Liu
Thesis: Building tools to expand the scope of continuous directed evolution

New York University — New York, NY **August 2016 - May 2018**
MS in Biology
Advisor: Prof. Lionel Christiaen
Thesis: Studying physical changes in the character of epidermal cells in Ciona embryos due to asynchronous cell division

Smith College — Northampton, MA **August 2012 - May 2016**
AB with High Honors in Biology and a minor in Math
Advisor: Prof. Nathan Derr
Honors Thesis: Biophysical and biochemical comparison between microtubule binding domains of human and yeast dynein

Teaching Experience

University of California, Irvine
BIO100 — Scientific Writing Winter 2020

University of California, Irvine
BIO94 — From Organisms to Ecosystems Winter 2021

University of California, Irvine
BIO98 — Biochemistry Winter 2022

3 Abstract of the Dissertation

Building tools to improve the scope of continuous directed evolution

by

Ria Anand Deshpande

PhD in Mathematical, Computational and Systems Biology

University of California, Irvine

Natural evolution occurs when mutations to DNA confer an advantage to an organism facing a new or changing environment. However, this process is slow, occurring over millions of years. There is a great need in the chemicals and pharmaceutical industries among others, to be able to engineer proteins to catalyze specific reactions, bind new substrates, or function at different temperatures. Directed evolution is a method that enables protein engineering on laboratory timescales. However, traditional methods of directed evolution are only able to sample a limited number of protein variants during the course of the experiment and require significant user involvement at each stage, making them difficult to apply at large scales. The methods of continuous directed evolution, as described here, in particular OrthoRep, provide an ideal way to evolve proteins faster while simultaneously allowing the user to test multiple conditions and sample large sequence space, thus overcoming the limitations of traditional directed evolution methods. OrthoRep can be used to easily connect sequence with structure and function, an important step in understanding what mutations can lead to altered function. Here, I describe the advantages of OrthoRep, best practices to design experiments, and the molecular biology underpinning OrthoRep function in cells. To enable easy and widespread adoption of OrthoRep, I developed a modular toolkit and curated detailed protocols pertaining to its use and best practices. Additionally, there are examples of the kinds of experiments that may be performed and the kinds that are currently at the edge of what is possible using this system.

4 Introduction

Natural evolution is the process of small changes accumulating over long periods of time that enable organisms to live under a given set of conditions most often for reasons of survival and reproduction. The world we live in today is a result of natural evolution. It is how a variety or organisms are able to coexist on the planet. Thermophiles have optimal growth at temperatures over 45^{o}C, acidophiles are able to maintain optimum growth at pH 3.0, and radioresistant organisms are able to grow even when exposed to high radiation, including ultraviolet and nuclear radiation (Taylor, Simon, Dickey, & Hogan, 2022). Even though the genomes of these organisms contain the Adenine (A), Thymine (T), Cytosine (C), and Guanine (G) bases, if you move a thermophile to an environment of pH 3.0 it will not survive. Thus, there must be small differences that enable growth of organisms in certain conditions.

There are five major contributors to natural evolution occurring in populations - mutations, genetic drift, gene flow, recombination and natural selection (Taylor et al., 2022). Mutations occur randomly in the individual ATGCs present in the DNA sequence and have an impact on the downstream expression of genes. This may lead to changes in amino acids that affect proteins by affecting their physical properties such as structure, stability, folding etc. Such mutations may appear to be small changes, but over time have a significant impact on how populations evolve (Taylor et al., 2022). Genetic drift occurs most frequently in small populations where inbreeding can occur that changes the frequencies of certain alleles leading to an imbalance or where a large chunk of the population gets wiped out leaving less diversity behind (Taylor et al., 2022). Gene flow occurs when new individuals and traits are introduced into the population leading to a change in the diversity in the gene pool with potential of introducing new traits (Taylor et al., 2022). Recombination in sexual reproduction leads to traits of both parents being inherited by their offspring, and in asexual reproduction, this may be caused by recombination within the genome during replication, or due to lateral gene transfer events (Taylor et al., 2022). These recombination also result in changes in allele frequencies per generation, so the effects are visible over time. Finally, natural selection is when environmental pressures such as temperature, salinity, predators etc. allow certain organisms with specific traits to survive over others leading to a directional change (Taylor et al., 2022). While all the other factors affect allele frequencies in populations, natural selection

affects phenotypic variation leading to evolution within populations.

We can think of evolution and the development of characteristics as a fitness landscape made up of all possible sequence space with peaks and valleys (Fig. 1)(Arnold, 1998). Organisms start at the plane or neutral level and as natural selection occurs, the organisms move around the landscape. As characteristics beneficial to the survival of organism develop, the organism moves up the hills towards the peaks while the mutations that don't give a survival advantage or adversely affect survival in any way move towards the valleys. The ultimate goal of an organism is to ensure survival and ability to reproduce with highest fitness in the environmental conditions it lives in. However, steps that lead to climbing the hill towards a peak must occur in small incremental steps and not large jumps. (Arnold, 1998)

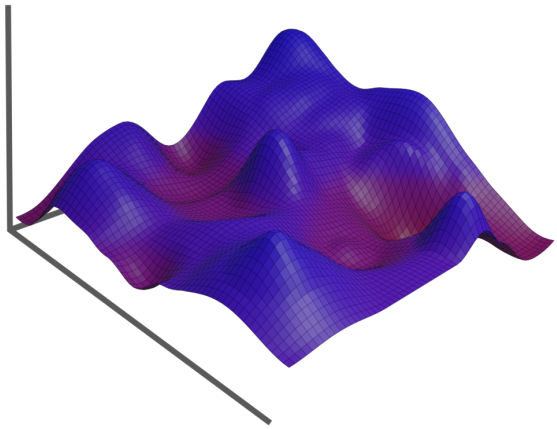


Figure 1: Example of a fitness landscape

This figure shows a large sequence space with peaks and valleys that denote different states of fitness. [Figure made by Vincent Hu]

Natural evolution has given us many proteins that are useful for a variety of reasons. However, there is great interest in studying more variation of proteins as well as improving their functions in the biotech industry for healthcare, agriculture, pharmaceuticals, chemicals among other industries. Often, to get enzymes to do specialized functions, that have not been evolved by nature, there are two approaches - ration design or engineering and evolution. These two approaches while seemingly different, have the same common goal and are often used in combination with each other.

4.1 The need for directed evolution

Evolution as a concept, is interesting to recreate in the lab to understand how the world around us came to be. In addition to curiosity, evolution, particularly that conducted in a lab offers solutions to produce proteins that may be beneficial to improving the lives we are able to lead today. There are examples of Cas9, and subsequent generations of gene editing technology, being evolved to improve healthcare (Hu et al., 2018). Directed evolution can be used to improve the resistance of plants to herbicides, and attempts to being made to improve crop yields (Engqvist & Rabe, 2019). Most often, directed evolution strategies are being used to improve enzyme properties including specificity towards particular substrates or products, activity, stability, folding etc.(Cobb, Chao, & Zhao, 2013). Other interests to design and test novel enzymes for possible functions that have not been evolved during natural evolution are also attractive. A new application due to the large number of tools available to create and analyze data, and the ability of directed evolution to run experiments at scale under varied conditions is an attractive proposition for continuing to invest in technologies to improve directed evolution.

Directed evolution tries to replicate natural evolution by utilizing a variety of methods available in the lab to provide opportunities for evolution. Of the five forces of evolution, natural selection is the easiest to apply in the lab, thus all methods of directed evolution involve the use of selection in some form or another. When considering the other four, mutations and recombinations can be done by synthetic processes as shown below while drift and gene flow may be more easily achieved in certain methods than others. The methods for creating such variation in the populations of enzymes to be evolved are described below.

4.1.1 Random Mutagenesis

Random mutagenesis is a simple way to introduce mutations in the gene to be evolved. There are several ways of performing mutagenesis including by use of chemicals, nucleotide analogs, and exploiting the errors created by polymerases. Cadwell and Joyce (1992) discovered amongst the polymerases they tested that Taq polymerase had an error rate ranging between 1 to 10 mistakes per million bases replicated which is ten-fold higher than the bacterial mutation rate that ranges between 1 mutation per 10 million to 1 mutation per 1 billion bases replicated (Chevallereau,

Meaden, van Houte, Westra, & Rollie, 2019). While this error-rate is not very high, it can be increased further by altering reaction conditions in which PCR is performed. Changing the ratios of the dNTPs added, increasing concentration of $MgCl_2$, decreasing concentration of $MnCl_2$ etc. causes an unbalanced reaction leading to more mistakes during replication (Cadwell & Joyce, 1992). Their data showed an approximate mutation rate of 2% per position per PCR which is a significantly higher error rate than the 0.001 - 0.02% by using Taq polymerase alone (Cadwell & Joyce, 1992).

4.1.2 Recombination

Homologous recombination occurs when genetic matter is exchanged between molecules of DNA that may be similar or different leading to a product that is different than either of the initial molecules of DNA and is another of the factors that influences evolution. A method created by Stemmer (1994) utilizes DNA shuffling to introduce mutations in the gene of interest. This technique was demonstrated by digesting TEM-1 by DNAseI to create small fragments that were allowed to recombine using a PCR-like process and then inserted into a vector sequence (Stemmer, 1994). These plasmids were then screened for increased levels of resistance against cefotaxime. These improved variants were then used a template and the process was repeated twice to get resistance to cefotaxime. The results showed a 0.7% rate of mutagenesis but with recombination which was similar to error-prone PCR. Improved variants with mutations were identified to have higher resistance as compared to the controls (Stemmer, 1994).

4.1.3 Rational design

Rational design or engineering is an another way to improve enzyme function. Rational design consists of understanding the structure and location of the active site of the enzyme in question to make mutations or variations that may result in modified or improved function of the enzyme (Chen, 2001). These changes may affect stability, folding and other physical aspects of the enzyme that may be beneficial. These approaches are usually used to identify specific sites on the crystal structure in areas where binding may occur to create variation before selection is applied (Chen, 2001). However, a major con is that variations only exist in the places that are identified by the user which may result in changes to the enzyme that prevent folding or may adversely affect function. The use of rational design methods for determining where mutations may be introduced require

prior knowledge of the structure and binding capabilities which is not known for novel proteins as well as for ones that are difficult to crystallize.

4.1.4 Site Saturation Mutagenesis

Site saturation mutagenesis is a technique often used when structural information about the protein is known. In this method, one or more sites each that contains one or more amino acids may be identified based on their location from the crystal structure and mutated to be one of the 19 available amino acids. We can study how mutations in those specific regions affect properties of these proteins (Siloto & Weselake, 2012). This method allows for the creation of proteins variants that would not occur via natural evolution. This approach is often paired with directed evolution to screen for variants with required activity and is usually carried out on a smaller scale with fewer variants when compared to the number of possible mutants that can be generated by error-prone PCR or recombination.

4.1.5 Limitations of Directed Evolution

Error-prone PCR and homologous recombination are straightforward ways of introducing mutations in the gene of interest that are easy to scale. However, for evolution to occur, these mutants must be tested under selective pressure for which they need to be cloned into a host, selection pressure applied and the best variants identified. For evolution to take place, this iterative process of error-prone PCR, transforming into a host, selection and screening must be repeated until a higher fitness peak has been achieved. To scalethis process, libraries must be created taking care to have sufficient diversity at every step, making this operation very labor intensive and not very convenient or cost-efficient. The process of recombination also leads to mutations at a rate that is similar to error-prone PCR and also has similar drawbacks. Both these process requires substantial user involvement and cannot be scaled up very easily leaving the need for better method to be used to introduce variation into populations.

4.2 Continuous directed evolution

Continuous directed evolution is the next step in improving methods of directed evolution to evolve proteins for specific functions. These systems aim to reduce user involvement by utilizing tools

available in nature making them easier to scale (Molina et al., 2022). Continuous evolution occurs in four steps, creation of mutants, followed by the translation to proteins, then selection or screening of the best performing mutants followed by their replication and these steps are repeated until better activity is evolved (Badran & Liu, 2015). Each of the continuous directed evolution systems utilize different methods to create mutations but try to use model organisms for the translation of proteins as well as for the replication of the selected mutants (Molina et al., 2022). Screening or selection may occasionally occur not automated but for classification purposes, we will still consider these systems as continuous since they require much less user involvement than the traditional methods of directed evolution.

There are some considerations when building a continuous directed evolution system. Outsourcing biological functions of replication and translation to organisms that have evolved to do so, makes these processes dependent on growth rates of model organisms and reduces user involvement. Often the concern with adding mutagens and introducing the ability to create mutations in organisms is that mutations in essential genes or in locations outside of the gene of interest may lead to death of the cell. Mutations that lead to inactivity or allow the strain to adapt to the environment without mutations may occur and these methods use a variety of ways to prevent that from happening (Molina et al., 2022). However, while mutations and natural selection are the stepping stones to directed evolution in the lab, other methods of incorporating variation like recombination, drift and gene flow provide an additional advantage to any system.

4.2.1 PACE

Phage-Assisted Continuous Evolution or PACE was invented by David Liu and colleagues (2011) and utilizes engineered bacteriophages that lack essential genes required to infect bacteria (Esvelt, Carlson, & Liu, 2011). These engineered phages then infect inducibly mutagenic E. coli cells that are able to produce the genes required for infection, only if the gene-of-interest has the appropriate mutations that allow for the selection circuit to be complete (Esvelt et al., 2011). The advantage of the system is that only the phages that acquire the ability to infect bacteria are then able to survive and continue infecting fresh bacteria until the required phenotypic results are observed (Esvelt et al., 2011). Thus, the selection mechanism is constantly refreshed allowing for evolution of functions that can only occur once per cell division, like cleaving of DNA. However, since this

system works in bacteria and phages, certain proteins that require more complex organisms to be expressed may not be evolved by PACE. Additionally, PACE requires use of a bioreactor and other specialized equipment that can be barriers to doing experiments at scale and for a lab that may not be equipped.

4.2.2 MutaT7

The MutaT7 continuous evolution system was created by Shoulders and colleagues (2018) in E. coli to perform mutations in a specific location downstream of a T7 promoter. The T7 polymerase is fused to a cytidine deaminase (rApo1) that can damage single-stranded DNA during replication. Since the system is pretty simple, it can be used in a variety of model organisms like yeast plants and mammalian cells in addition to E. coli where it was developed (Moore, Papa, & Shoulders, 2018; Molina et al., 2022). Selection and screening can be performed according to the gene that is being mutated. However, since the agent introducing mutations is a cytidine deaminase, the C $\rightarrow$ T mutations are observed most frequently though that can be balanced by adding a T7 promoter on the opposite strand (Moore et al., 2018). However, since the only factor that determine where mutations are caused is the T7 promoter, it can be integrated at any location in the genome but significant off-target mutations may be observed. This can prove fatal to the system by mutating essential genes or creating mutations that prevent evolution of the gene of interest (Moore et al., 2018).

4.2.3 EvolvR

The EvolvR system was created by Deuber and colleagues (2018) in E. coli to nick a strand of DNA using nCas9 and repairing it by use of error-prone polymerase fused to the nCas9. There is a strict requirement of an NGG PAM sequence since Cas9 is being used as a nicking agent (Halperin et al., 2018). A PolI3M polymerase was created that consisted of three mutations on the E. coli DNA polymerase that reduces fidelity ensuring that the nicks created by nCas9 are repaired in an error-prone manner (Halperin et al., 2018). There is a limited editing window of 56 bp that can be accessed by the PolI3M. While there are many locations that an NGG PAM motif can be found in the genome, off-target mutations are also possible though they were reported to be below the detection threshold (Halperin et al., 2018). The concern for EvolvR would be that the system is

only functional until the spacer sequences that binds sgRNA remains detectable and not mutated. Thus, this system may have limitations on the duration of the evolution campaigns that may be conducted and some restrictions on where mutations can occur also exist. In longer genes, where it may be desirable to get mutations in all regions, more than one sgRNA will need to be introduced.

4.2.4 VEGAS

Roth and colleagues described viral evolution of genetically actuated sequences or VEGAS, a continuous directed evolution platform that uses the a highly mutagenic Sindbis RNA virus that contains no proof-reading ability (English et al., 2019). Mammalian cells were then infected with an engineered Sindbis virus that contained a mammalian expression vector for the essential structural genes, and regulated by the gene being evolved (English et al., 2019). However, VEGAS has been subject to some criticism from scientists who have tried to use and it and discovered that cheater particles that contained the structural genes contaminated the cells preventing selection and subsequently evolution when carried on for multiple rounds (Denes et al., 2022).

4.3 OrthoRep

OrthoRep is a continuous evolution platform in yeast developed by Chang Liu and colleagues and described in detail in the next chapter (Ravikumar, Arrieta, & Liu, 2014). In brief, it uses an error-prone polymerase to replicate a cytoplasmic linear plasmid containing the gene of interest (Ravikumar, Arzumanyan, Obadi, Javanpour, & Liu, 2018). Selection and screening can be performed in a variety of ways using growth based assays or fluorescent reporters. OrthoRep has the ability to increase diversity within the populations by drifting populations of cells as well as encoding libraries made by error-prone PCR onto p1. Additionally, other methods of introducing variation as previously described can easily be teamed up with OrthoRep for the evolution of proteins.

5 OrthoRep Toolkit

5.1 Introduction

Natural evolution works over millions of years to evolve proteins that provide organisms with a survival or reproductive advantage in the environment they live in. Directed evolution has given scientists the ability to evolve proteins faster than nature to provide advantages beyond those for survival and reproduction. However, while directed evolution may be faster, it requires a lot of user involvement and interference at every step to improve function. These methods are limited in the diversity that can be sampled at every step thus making some functions unachievable even at laboratory scale evolutions. Continuous directed evolution utilizes model organisms to evolve proteins by creating mutations in vivo. This allows less user interference and the ability to perform evolution experiments at scale.

5.2 Development of Protocol

OrthoRep consists of two cytoplasmically located linear plasmids in yeast. The smaller of the two plasmids, p1, contains the gene of interest to be evolved and the markers to select for presence of p1 during the experiment (Ravikumar et al., 2014). The longer plasmid, p2, is necessary for the replication and maintenance of p1. The error-prone polymerase for p1, is either integrated in the genome or present on a CEN/ARS plasmid (Ravikumar et al., 2018). The native yeast genome located in the nucleus and the error-prone polymerase are replicated by the yeast native polymerases that have very low error rate. The error-prone polymerase only replicates p1, while p2 encodes its own DNA polymerase (Ravikumar et al., 2018). Thus, this system is completely orthogonal to the yeast genome.[Fig. 2]

In addition to the error-prone polymerase and the pair of linear plasmids, a selection system is necessary for evolving proteins using OrthoRep. An ideal selection system comprises a function that the protein being evolved contributes to directly or indirectly (Molina et al., 2022). Ideally a titratable selection, such that selection pressure can be increased over time should be used to ensure continuous selection of the highest performing variants (Molina et al., 2022). It is important for the protein being evolved to have some small starting activity as a point for selection, but this condition can be overcome by allowing accumulation of mutations into a large in vivo library prior

to the beginning of evolution (Molina et al., 2022).

In OrthoRep, the error-prone polymerase replicates the p1 plasmid making one million times more mutations than the native yeast polymerase. Depending on the polymerase being used, the range of mutation rates varies from 10^{-4}–10^{-5} (Rix & et.al., 2023). Since the p1 and p2 plasmids exist in the cytoplasm in 50-100 copies per cell, the error-prone polymerase replicates all copies within the span of one cell division . As the cell divides, the copies of p1, some of which contain mutations, are randomly distributed to the daughter cells (Wesolowski, Algeri, Goffrini, & Fukuhara, 1982). As this process continues, the initial mutations get distributed to daughter cells and each daughter cells accumulates more mutations. When selection pressure is applied, only cells that contain mutations that give a growth advantage, will be selected for allowing certain mutations to fix. After several rounds of passaging, the surviving cells should have a substantial advantage when compared to the starting population. Next-generation sequencing (NGS) can help determine which mutations occur on the same strands, and if a variety of mutations exist. These mutations can then be cloned into a strain that is not evolving to identify which mutations provide highest growth advantage. Locating these mutations on the crystal structure of the protein will also determine which sorts of conformational changes, if any are occurring and how they might lead to a better protein.

OrthoRep was first validated by evolving the dihydrofolate reductase gene from Plasmodium falciparum (PfDHFR) in the presence of the PfDHFR inhibitor and antimalarial drug, pyrimethamine (Ravikumar et al., 2018). This evolution was conducted in 90 replicated with replicates being only 0.5 ml each. It is known that four mutations in the active site (N51I, C59R, S108N and I164L) can lead to resistance, and this was achieved by hand-passaging cultures at a 1:100 dilution 13 times for about 87 generations in total (Ravikumar et al., 2018). Of the 90 starting populations, 78 populations survived and were able to grow at the highest soluble concentration of pyrimethamine (3mM) (Ravikumar et al., 2018). Mutations that improved expression from cytoplasmic p1 were observed in addition to the mutations that led to resistance. These and subsequent improvements to OrthoRep have overcome limitations from that initial evolution.

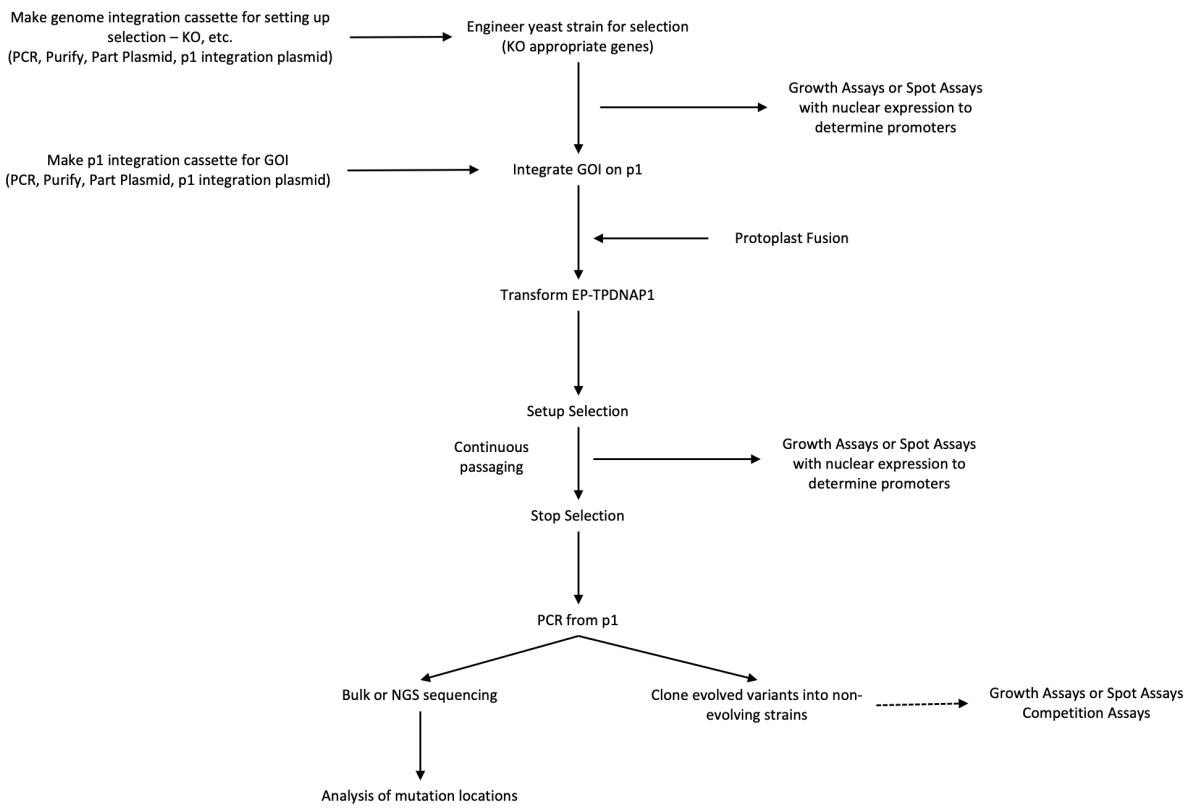


Figure 2: Schematic of an evolution comapign using OrthoRep

5.3 Applications

The application of OrthoRep was first demonstrated by evolving the Plasmodium falciparum dihydrofolate reductase (PfDHFR) in presence of the drug pyrimethamine (Ravikumar et al., 2018). Since then, OrthoRep has been used for a evolving enzymes and for FACS based selections (Molina et al., 2022). Screening libraries of antibodies for better binding and specificity is another application of this method (Wellner et al., 2021).

The evolution of TrpB from Thermatoga maritima (TmTrpB) to function at mesophilic temperatures and to accept indole to produce Tryptophan using OrthoRep is a great example of using the system to evolve functional enzymes (Rix et al., 2020). OrthoRep was also used to explore the possible evolution of THI4s, suicide enzymes that are degraded and synthesized after every reaction, to be energy efficient and perform multiple reactions. The MhTHI4 from Mucinivorans hirudinis, was evolved to function in more plant like environments that are similar to the environment of the cytoplasm in yeast cells (Garcia-Garcia et al., 2022).

BenM based biosensors that were better at binding, had high dynamic range and broad operational range for cis-cis-muconic acid (CCM) and adipic acid (AA) were evolved using FACS based selections (Javanpour & Liu, 2021). Antibodies can be evolved using OrthoRep by utilizing yeast surface-display methods to display antibody variants on the surface and using FACS based screening methods to identify high-affinity, high-quality antibodies (Wellner et al., 2021).

5.4 Comparison with other methods

There are few platforms that do continuous directed evolution and of those, EvolvR and MutaT7 exist primarily in E. coli, while PACE works with bacteriophages that infect E. coli. PACE utilizes M13 bacteriophage that are engineered to have the protein of interest encoded in place of the gIII gene that encodes pIII responsible for infecting E. coli (Esvelt et al., 2011; Miller et al., 2020). In absence of this gene, the bacteriophage is able to bind the E. coli which contains a mutagenesis plasmid and an accessory plasmid that has gIII encoded on it (Esvelt et al., 2011; Miller et al., 2020). When the protein of interest is expressed, it can trigger expression of gIII that allows the phage to detach and infect other host cells. During this process, the arabinose induced mutagenesis plasmid is randomly mutating everything including the host genome (Esvelt et al., 2011; Miller et al., 2020). This method allows proteins that can be expressed in phages to evolve and the host cells are constantly refreshed preventing the mutation of the bacterial genome from being an issue (Miller et al., 2020).

The EvolvR system utilizes a single plasmid that is made up of a nicking Cas9 (nCas9) fused to an E. coli DNA Polymerase I with three mutations (PolI3M) that reduced its fidelity and an sgRNA (Halperin et al., 2018). This system is able to target specific locations in the host genome due to the sgRNA and create a nick on the double-stranded DNA. This nick is repaired by the low fidelity PolI3M leading to mutations in the given 15-50 bp window (Halperin et al., 2018). To target multiple sections of a single gene, multiple sgRNAs would be needed. While off-target mutations are possible, the EvolvR system has not been able to detect them above the given threshold (Halperin et al., 2018).

The VEGAS system was adapted using a Sindbis virus that would carry the required plasmid and infect mammalian cells (English et al., 2019). Some structural genes were encoded onto a mammalian expression vector while other was encoded on a vector also encoding the gene to be

evolved (English et al., 2019). The gene of interest is responsible for regulating the expression of the structural genes needed for the Sindbis virus to keep infecting other cells. The mutation prone RNA Sindbis virus is responsible for creating mutations all across its genome including the gene of interest (English et al., 2019). However, there have been reports of this system not functioning as expected due to contamination resulting in the loss of the gene of interest leading to no evolution (Denes et al., 2022).

The MutaT7 systems fuses a cytidine deaminase (rApo1) to a T7 RNA polymerase that enables targeting to regions downstream of the T7 promoter (Moore et al., 2018). Deleting ung increases mutagenesis rates by preventing repair of deoxyuridine to deoxycytidine. The specificity of this system was further improved by enclosing the gene to be evolved on both sides by the T7 promoter (Moore et al., 2018). This system is fairly simple and can be used in multiple host organisms. Selection and screening of evolved clones can be determined by the gene being evolved. The advantage of this system is that it can be used for random mutagenesis of any gene since the only consideration is the presence of the the T7 promoter.

5.4.1 Limitations of OrthoRep

The main limitations of OrthoRep are that it is currently only works in yeast. As a result, certain genes of interest that may be toxic in yeast cannot be expressed and thus, cannot be evolved. Using yeast cells as host also means that genes that need host-specific post translational modification that yeast are unable to do cannot be evolved using OrthoRep (Molina et al., 2022). Yeast are really good at recombination, and even 20 bp located on the same chromosome can lead to recombination that may result in loss of part of your gene of interest or other genes essential for evolution. Yeast have an average doubling time of 90 minutes, but modifications to the strains required for evolving specific genes may result in slower doubling times leading to longer experiments for evolution. Currently, OrthoRep contains 50-100 copies of p1 in the cytoplasm, and thus, effects of mutations leading to better function may be diluted due to the presence of libraries within each cell. Additionally, proteins that work in complexes or as dimers may contain mutations on different copies that improve activity on a per cell basis but are ultimately difficult to detangle on a per sequence basis. At the current mutation rates, larger genes require larger bottlenecks per passage, and thus require larger volumes for evolution which may create logistical challenges for evolution in many replicates.

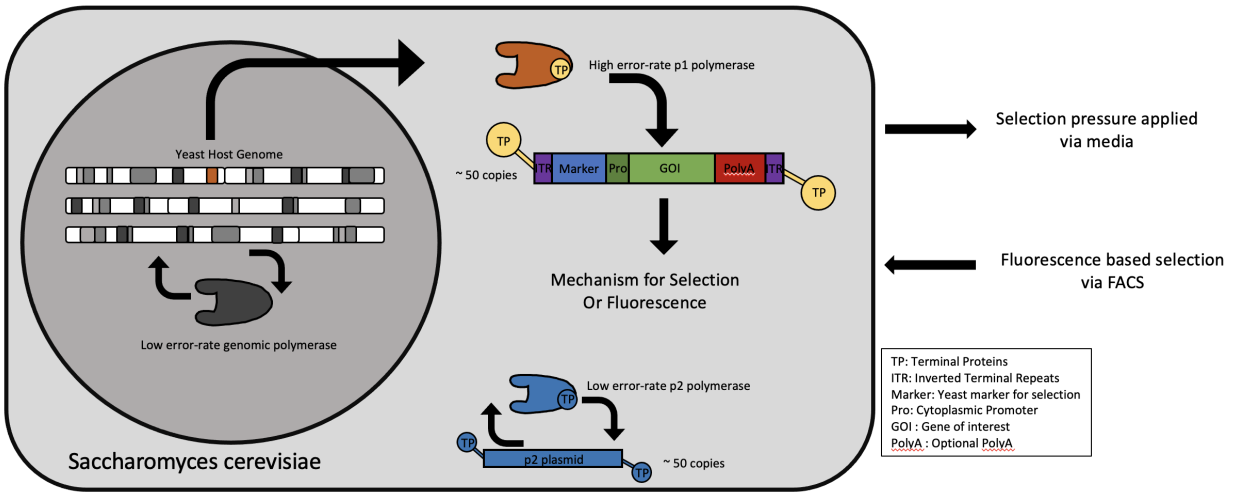


Figure 3: Schematic of OrthoRep mechanism

5.5 Experimental Design

There are many decisions that need to be taken when designing an OrthoRep experiment. A lot of these decisions may be determined by the experiment being performed and each comes with a unique list of pros and cons that may affect the results. However, this protocol will guide you through making the best decisions, and using basic molecular cloning techniques outlined below will enable you to change any of these variables when repeating the experiment.

5.5.1 Determining starting yeast strain

All OrthoRep experiments are done in S. cerevisiae. However, all strains of this baker's yeast are not created equal. The petite strains (F102-2) do not contain functional mitochondria, and thus grow slower than the commonly used lab strains (BY4741). Initial OrthoRep evolutions were done in F102-2 strains since the absence of functional mitochondrion allows for wild-type p1 and p2 to exist in the cytoplasm at high copy number (Ravikumar et al., 2018). The mitochondrial DNA is incompatible with the toxin-antitoxin system on p1, so alternatives need to be used. A landing pad, that may or may not contain the wild-type polymerase and a yeast marker such as MET15 may be integrated in place of wild-type p1 allowing for easy selection of the pair of linear plasmids while eliminating the toxin-antitoxin system that causes incompatibility (Kamper, Esser, Gunge, & Meinhardt, 1991). Thus, the first decision to be made is what starting strain to use.

The main advantage of using F102-2 strains is the presence of wild-type p1 of 8.9 kB length. However, this strain grows slower, and it might never fully get rid of wild-type p1 unless it is forced to. This strain contains the toxin-antitoxin system that may get activated at low pH. While yeast media for normal growth does not activate this system, certain evolution conditions may require media that may trigger the killer system. This will be an issue only if growing with or near strains that may be sensitive to this toxin. To integrate onto this wild-type p1, the part1 plasmid for short pol should be used if you wish to have wild-type polymerase still around and the full-del-pol flanks if there should be no WT polymerase on p1. If there is no other polymerase around, either wild-type or error-prone, then the wild-type p1 will stick around until another polymerase is available for p1 replication. The ratios of the recombinant p1 and the wild-type p1 will reach equilibrium within a few passages to levels such that the selection marker is sufficiently expressed for survival of cells and there are enough copies of WT-polymerase around to replicate both species of p1.

The alternative to using F102, is using the commonly used lab strain BY4741. This strain contains functional mitochondria, and thus is incompatible with the wild-type p1. To navigate this, we created several landing pads by replacing the toxin-antitoxin system with a yeast selectable marker while keeping the polymerase intact in some of them. These strains were created by transforming the wild-type p1 with a cassette that replaces ORF2-4 with Kanamycin resistance in the petite Kluyveromyces lactis strain and then protoplast fusing it to the BY4741 strain and selecting for the G418 resistance. The appropriate landing pad plasmid was then integrated, and the G418 resistance contained p1 was cleaved repeatedly using Cas9 until it disappeared. In this category, we created two kinds of strains, both of them contained wild-type polymerase in the genome but one set of them contained the wild-type polymerase on p1 and the other which did not. Instead, each strain contained p1 made up of MET15 with a split leu2 construct, or a MET15 by itself. The strains that contained the wild-type polymerase on p1 also had MET15 on p1 as the selectable marker with a split leu2 as well as one strain without it. One important decision to be made at this stage is whether to use a landing pad that contains the wild-type polymerase or one that does not. The presence of wild-type polymerase in addition to the error-prone one reduces the error rate as described in the next section.

BY4741 strains grow faster and are okay to use for most applications. OrthoRep works similarly in this strain as previously shown (Javanpour & Liu, 2019). A few alternations may be necessary

in the long miniprep to isolate p1/p2 as described below. The MET15 landing pad with the split leu2 marker is one of the most efficient landing pads we have for p1 integrations and we recommend their use in situations where the selection does not depend on MET15 and LEU2. In situations where LEU2 may be required for selection, the MET15 landing may be used. If however, MET15 and LEU2 are both needed for selection, use the landing pad containing the Kanamycin resistance gene. This landing pad doesn't have an auxotrophic marker that can be continuously selected for, thus the presence of p1 and p2 needs to be occasionally confirmed either in YPD + G418 media or agar plates to ensure that OrthoRep can still exist. This should be less of an issue once your gene of interest and an appropriate yeast marker have been integrated onto p1 for selection pressure is applied to cells.

5.5.2 Identifying best promoter for gene-of-interest

Once the strain in which evolution will occur has been determined, we need to build a cassette to integrate our gene-of-interest onto p1. For this purpose, we need to identify the cytoplasmic promoter that will be used for the gene-of-interest. Cytoplasmic promoters have different strengths and work differently than yeast nuclear promoters. Moreover, the cytoplasmic promoters exist in as many copies as copies of p1 exist in the cytoplasm. We worked on identifying promoters of varying strengths that allow proper expression of genes from p1.

We have included 7 promoters as part of this toolkit as shown in the table below. These promoters are a combination of three native p2 promoters, one engineered promoter, one evolved promoter and two rationally designed promoters. Of these, p2ORF1 has the lowest strength while p2ORF5 has the highest strength from the native cytoplasmic promoters and p2ORF10 is between these. The promoter p10B2 was a result of an engineered campaign while the promoter pGA evolved from p10B2 with two mutations (Zhong, Ravikumar, & Liu, 2018). Two additional promoters pGA-AA and pGA-AAAA since data suggested that the presence of a short polyA sequence results in higher expression as well as more capped mRNA sequences (Sykora et al., 2018). There is a slight increase with the longer string of As and since expression from p1 may make or break an evolution, even slight increases in expression may have a large benefit.

The first step to identifying the best promoter is to determine the lowest level at which the required gene needs to be expressed for functioning. Since cytoplasmic promoters are not very

strong, we recommend testing your system with promoters from the yeast toolkit ranging from low (pREV1) to medium (pRPL18b) (Lee, DeLoache, Cervantes, & Dueber, 2015). If the expression strength at pREV1 levels is sufficient for growth and selection, then any of the promoters may be used as they are as long as the gene-of-interest is not toxic at higher concentrations. The pGA promoter is the strongest promoter available without having to employ any tricks to increase promoter strength and we recommend using this if your gene of interest does not need to be very highly expressed. The addition of 2-4 As at the end of the promoter does slightly boost expression (Sykora et al., 2018). Additionally, PolyA tails can further increase expression strength (Zhong et al., 2018).

5.5.3 Choosing the appropriate error-prone polymerase

One of the key decisions that need to be made while designing the experiment is to determine what error-prone polymerase to use. The three main factors to consider while picking the best polymerase for your experiment are the mutation rate, the mutational spectra and the copy number. However, while designing the experiment, it is important to note that certain selections and genes-of-interest may be affected by the choice of polymerase, and may make the system unstable. To prevent this, we recommend that polymerases are encoded into the genome or added via a CEN/ARS plasmid as a last step so that a variety of them may be tested if necessary. Additionally, adding the error-prone polymerase as the last step ensures that no mutations are collected during the time from integrating the gene of interest on p1 to starting evolution.

The first set of error-prone polymerases for OrthoRep consist of pAR-Ec611 and pAR-Ec633 that have mutation rates in the $10^{-6} - 10^{-5}$. However, these polymerases are biased to make more transition mutations than transversions (Ravikumar et al., 2018). The newer polymerases as reported in Rix et.al., (in preparation) include BB-Tv, BadBoy2, Sgt. Kiss and BadBoy3 [Fig. 2]. All these polymerases have error-rate in the $10^{-4} - 10^{-5}$ and more balanced mutational spectra as shown in the figure. Amongst these, BB-Tv has lower error-rate when compared to the other new polymerases with a balanced mutational spectrum that allows both transition and transversion mutations to occur. The copy number of these is pretty high [waiting on exact data] and the presence of wild-type polymerase is not required to maintain stability of p1 in the cytoplasm. BadBay2 and BadBoy3 are comparable in terms of mutation rate and copy number and internal

evolution experiments have shown better results with BadBoy3 for yeast display. If unsure about wheather high mutation rates would negatively impact your experiment, start with BB-Tv and in parallel, with fewer replicates test out other error-prone polymerases that may be able to speed up evolution.

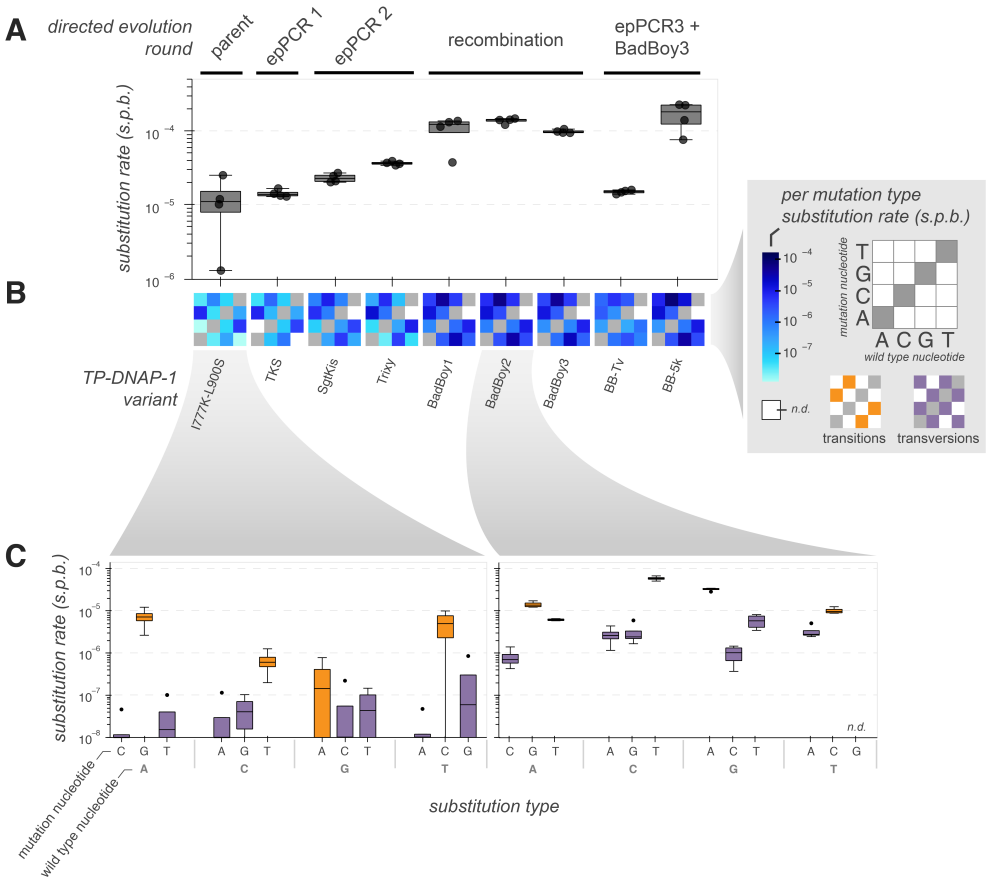


Figure 4: The new error-prone polymerases show higher mutation rate and balanced spectra. [Data and figure by Gordon Rix]

While the newer error-prone polymerases are superior, there are many nuances to using them that depend on individual experiments being performed. It seems likely that the size of p1 has an effect such that longer genes are more likely to suffer from genetic drift when using polymerases that have highest error rates, and these longer p1 plasmids make the system less stable over time. Strains that require substantial engineering pre-evolution affect general cell health and some polymerases with high error-rates and low copy number may affect the growth adversely. In such cases, BB-Tv is an ideal choice since. Growth based selections are easy to scale, and for such long-term experiments, lower error-rates may be preferred. When performing low throughput screens, higher

error-rate polymerases should be used to increase the probability of finding hits while for high throughput screens also, a higher error-rate polymerase may be beneficial since the bottleneck may be pretty severe with each passage. Varied copy number also has an impact on the way selections may work since each cell contains a library within itself and thus it takes a few replications for mutations to fix for an entire population. It is also possible that higher copy number may dilute the phenotypic effects observed.

5.5.4 Setting up selections and sustaining evolution

A selection may be designed in any way such that increasing selective pressure can be applied and no cheaters are able to thrive during the course of the evolution campaign. Ideally, in a continuous evolution system like OrthoRep selection should be linked to cell fitness or survival, like when the gene being evolved is essential for the cell (eg. TrpB to survive in -W media). These functions may be production of essential amino acids required for survival, producing antibiotic resistance in presence of increased concentrations of antibiotics, ability to grow at a different temperature etc. However, some applications may not be conducive to being tied to essential activities in the cell, in such a case, activation of a fluorescent marker may prove useful. While this does add a user directed step to screen for desirable traits, it offers an alternative that continues to be high-throughput.

An ideal selection must have high dynamic range and should be titrable to produce increasing selection pressure on the system. Such a selection would also be hard to cheat, and proper controls should be used to ensure that the probability of strain adaptation is low to confirm that evolution is occurring. This must also be kept in mind while determining the starting point for evolution. An ideal starting point is one where the strains have minimal activity but are able to grow. Over time, these strains should grow better and show mutations without strain adaptation.

A variety of assays depending on the kind of selection being used may be performed to track and determine a good stopping point. For growth based assays, or selections where resistance must be developed or tolerance is to be improved etc., spot assay and growth assays may be used. Additionally, competition assays to ensure that the evolved strains outcompete the pre-evolution strains may be used. Sequencing using Sanger or other NGS methods may also prove useful to identify mutations and to track them over time.

5.5.5 Characterizing evolved mutants

OrthoRep creates mutations as the error-prone polymerase replicates p1 depending on the error-rate of the particular polymerase being used. As these mutations accumulate on different copies of p1, cells will divide and segregate the p1's randomly in a non-Mendelian fashion. As cells continue to divide, the population of p1 will vary and some mutations that give an advantage and stick around will fix after a few passages. When bulk sequencing is performed, fixed mutations will be determined by nucleotide changes, that may translate to amino acid changes on the protein. Mixed peaks may be observed that show that some part of the population may be fixed. Mutations that are not majority of the population may not be detected by bulk Sanger sequencing. Next generation sequencing techniques can be beneficial to determine all the mutations that were created and barcoding can also help determine what mutations occurred on individual strands. The methods for barcoding will depend on the kind of Next-Generation sequencing being performed.

Once mutations have been identified, they need to be tested to determine how evolution has affected the gene of interest. Mutations should be cloned in a way that makes sense for the particular application, whether one at a time or in groups, and tested in a strain that is not evolving. This may be done by cloning them onto a CEN/ARS plasmid that relies on yeast native DNA polymerase or by retransforming onto p1 that contains no error-prone polymerase. While both these methods have their advantages, the decision may be made based on convenience. The same assays that were performed while sustaining evolution should be repeated and the best performing variants should be identified. Alternatively, barcoded libraries may be created and subject to competition assays to identify the best performing variants.

5.6 Parts of the OrthoRep Toolkit

5.6.1 Yeast strains

Name	Yeast Strain	p1 - p2
OTK-Y001	F102-2	WT p1, WT p2
OTK-Y002	BY4741	shortpol G418R on p1, WT p2
OTK-Y003	BY4741	shortpol MET15 on p1, WT p2
OTK-Y004	BY4741	shortpol MET15 - splitleu2 on p1, WT p2
OTK-Y005	BY4741	fulldelpol G418R on p1, WT p2
OTK-Y006	BY4741	fulldelpol MET15 on p1, WT p2
OTK-Y007	BY4741	fulldelpol MET15 - splitleu2 on p1, WT p2

Table 1: Yeast Strains included as part of the OrthoRep toolkit

5.6.2 E. coli plasmids

Part No.	Part Type	Description	Antibiotic Marker
pOTK001	1	Can1 upstream flank	Chloramphenicol
pOTK002	1	Ade1 upstream flank	Chloramphenicol
pOTK003	1	Lyp1 upstream flank	Chloramphenicol
pOTK004	1	ARS314 upstream flank	Chloramphenicol
pOTK005	1a	p1-ORF1-short pol flank	Chloramphenicol
pOTK006	1a	p1-ORF2-FDP flank	Chloramphenicol
pOTK007	1a	p1-ORF1 - split LEU2 flank	Chloramphenicol
pOTK008	1b	p2ORF1-MET15	Chloramphenicol
pOTK009	1b	p2ORF1-LEU2	Chloramphenicol
pOTK0010	1b	p2-ORF1-TRP1	Chloramphenicol
pOTK011	1b	p2-ORF1-URA3	Chloramphenicol
pOTK012	2	p2ORF1 promoter	Chloramphenicol
pOTK013	2	p2ORF5 promoter	Chloramphenicol
pOTK014	2	p2ORF10 promoter	Chloramphenicol
pOTK015	2	p10B2 promoter	Chloramphenicol
pOTK016	2	pGA promoter	Chloramphenicol
pOTK017	2	pGA-AA promoter	Chloramphenicol
pOTK018	2	pGA-AAAA promoter	Chloramphenicol
pOTK019	3	TmTrpB for dummy evolution	Chloramphenicol
pOTK020	3	WT-TPDNAP1	Chloramphenicol

pOTK021	3	BB-Tv	Chloramphenicol
pOTK022	3	BadBoy2	Chloramphenicol
pOTK023	3	BadBoy3	Chloramphenicol
pOTK024	3	Sgt. Kiss	Chloramphenicol
pOTK025	4	PolyA tail - No Ribozyme	Chloramphenicol
pOTK026	4	PolyA tail - HDV Ribozyme	Chloramphenicol
pOTK027	4	PolyA tail - HH Ribozyme	Chloramphenicol
pOTK028	45	p1 ORF4 integration flank	Chloramphenicol
pOTK029	5	p1 ORF4 integration flank	Chloramphenicol
pOTK030	5	Can1 downstream flank	Chloramphenicol
pOTK031	5	Ade1 downstream flank	Chloramphenicol
pOTK032	5	Lyp1 downstream flank	Chloramphenicol
pOTK033	5	ARS314 downstream flank	Chloramphenicol
pOTK034	678	ColE1-KanR GFP dropout backbone	Kanamycin
pOTK035	1	Trp5 upstream flank	Chloramphenicol
pOTK036	234	URA3 with nuclear promoter	Chloramphenicol
pOTK037	1	Trp5 downstream flank	Chloramphenicol
pOTK038	12345	Trp5 KO cassette	Kanamycin
pOTK039	12345	BadBoy2 Polymerase integration cassette	Kanamycin
pOTK040	12345	TmTrpB p1 integration cassette	Kanamycin

Table 2: Plasmids included as part of the OrthoRep toolkit

5.7 Materials

5.7.1 Reagents

- ADB (Zymo Research, D4001)
- Agarose (Fisher Scientific, cat. no. BP1356-500)
- Ampicillin Sodium Salt (Fisher Scientific, cat. no. BP1760)
- Bacteriological Agar (Genesee Scientific, cat. no. 20-273)
- 7x Blue Lysis Buffer (Zymo Research, cat. no. D4036-1-60)
- BsaI-v2 (New England Biolabs, cat. no. R3733)
- BsmBI-v2 (New England Biolabs, cat. no. R0739)
- Calcium Chloride Dihydrate (Sigma, cat. no. 223506)
- L-Canavanine salt (Millipore Sigma, cat. no. C9758)
- Citric Acid (Sigma, cat. no. 251275)
- Chloramphenicol (Sigma Aldrich, cat. no. C0387)
- rCutSmart Buffer (New England Biolabs, cat. no. B6004S)

- Dextrose (D-Glucose), Anhydrous (Fisher Scientific, cat. no. D16-10)
- DMSO (Fisher Scientific, cat. no. BP231-1)
- dNTPs (New England Biolabs, cat. no. N0446S)
- Dpn1 (New England Biolabs, cat. no. R0176S)
- EcoRI-HF (New England Biolabs, cat. no. R3101S)
- Ethidium Bromide (Milipore Sigma, E7673)
- Ethyl alcohol (Millipore Sigma, cat. no. 64-17-5)
- Ethylenediaminetetraacetic Acid [EDTA] (Fisher Scientific, cat. no. BP120-1)
- G418 Sulfate (Geneticin) (US Biological, cat. no. 108321-42-2)
- GeneRuler 1kb plus (Thermo Scientific, cat. no. SM1331)
- GeneRuler Mix (Thermo Scientific, cat. no. FERSM0332)
- Glacial acetic acid (Millipore Sigma, cat.no. A6283)
- Glycerol (Fisher Scientific, cat. no. BP229-1)
- Guanidine Hydrochloride (Fisher Scientific, cat. no. BP178)
- Hygromycin B (Gold Biotechnology, cat. no. H-270)
- Isopropanol (Fisher Scientific, cat. no. BP26184)
- Kanamycin Monosulfate (Gold Biotechnology, cat. no. K-120)
- LB Miller Agar (US Biological, cat. no. L1515)
- Lithium Acetate (Sigma Aldrich, cat. no. L6883)
- 2-Mercaptoethanol (Thermo Scientific, cat. no. 125472500)
- Mix & Go E. coli Transformation kit (Zymo Research, cat. no. T3002)
- Neuroseothricin Sulfate (Fisher Scientific, cat. no. 50-153-2817)
- Neutralization Buffer (Zymo Research, cat. no. 11-420D)
- Peptone (US Biological, cat. no. P3300)
- Phusion HS Flex Polymerase (New England Biolabs, cat. no. M0535)
- Polyethylene Glycol 3350 (Fisher Scientific, cat. no. P1463)
- Potassium acetate (Fisher Scientific, cat. no. P171-500)
- Potassium Chloride (Fisher Scientific, cat. no. P330-3)
- Proteinase K (Fisher Scientific, cat. no. BP1700-500)

- 6x Purple Gel Loading Dye (New England Biolabs, cat. no. B7024S)
- RNAseA (Fisher Scientific, cat. no. BP25391)
- Salmon Sperm DNA (Millipore Sigma, cat.no. D1626-5G)
- SC-HLUW Dropout mix w/o YNB (US Biological, cat. no. D9540)
- ScaI-HF (New England Biolabs, cat. no. R3122S)
- Sodium Chloride (Fisher Scientific, cat. no. BP358-212)
- Sodium L-Glutamate Monohydrate (Fisher Scientific, cat. no. G0188)
- Sodium Phosphate Dibasic Anhydrous (Fisher Scientific, cat. no. S375)
- Sodium Dodecyl Sulfate (SDS) (Fisher Scientific, cat.no. BP166)
- D-Sorbitol (Fisher Scientific, cat. no. AC132730010)
- Supercoiled DNA Ladder (New England Biolabs, cat. no. N0472S)
- T4 DNA ligase buffer (New England Biolabs, cat. no. B0202S)
- T4 ligase (New England Biolabs, cat. no. M0202)
- Top 10 Chemically Competent Cells (Invitrogen, cat. no. C404010)
- Tris base (Fisher Scientific, BP152)
- Trizma base (Millipore Sigma, cat. no. T1503)
- Tryptone (US Biological, cat. no. T8750)
- Yeast Extract (US Biological, cat. no. Y2010)
- Yeast Nitrogen base w/o AA, Carbohydrate & w AS (US Biological, cat. no. Y2025)
- Yeast Nitrogen base w/o AA, Carbohydrate & w/o AS (US Biological, cat. no. Y2030)
- Zymolyase 20T (US Biological, cat. no. 37340-57-1)

5.7.2 Equipment

- Attune NxT Flow Cytometer
- Axygen Microtubes (Fisher Scientific, cat. no. 14-222-168)
- Breathe-Easy Film [For plate reader](USA Scientific, cat. no. 9123-6100)
- ChemiDoc Imaging System (BioRad, cat. no. 12003153)
- 15 ml Conicals (Fisher Scientific, cat. no. 05-538-59B)
- 50 ml Conicals (Fisher Scientific, cat. no. 05-538-60)

- Cuvettes (FisherBrand, 14-955-127)
- 24-Well Deep Well Blocks (VWR, cat. no. 89080-534)
- 96-Well Deep Well Blocks (VWR, cat. no. 29445-130)
- Disposable Petri Dishes for E. coli (VWR 100mm x 10mm, cat. no. 25384-324)
- Disposable Petri Dishes for Yeast (VWR 100mm x 15mm, cat. no. 25384-164)
- EconoSpin RNA/DNA Micro Spin Columns[Gel extraction](Epoch Life Sciences, cat. no. 3010)
- EconoSpin All-In-One DNA/RNA Mini Spin Columns [Miniprep](Epoch Life Sciences, cat. no. NC0789236)
- 50 ml Filter (Thermo Scientific, cat. no. 09-741-88)
- 500 ml Filter (Thermo Scientific, cat. no. 09-741-202)
- 96-well Flat Clear Bottom Plates for Growth Assays (Corning, cat. no. 3904)
- Foil Seals (Genesee Scientific eXTReme FoilSeal, cat. no. 12-156)
- Gel Electrophoresis Power Supply (BioRad, PowerPac Universal Power Supply)
- Gel Electrophoresis System (BioRad, Mini-Sub Cell GT Cell)
- Glass Culture Tubes (VWR KIMAX-51, cat. no. 45060-650)
- 500 ml Glass Media Bottles (Fisher Scientific, cat. no. FB800500)
- 1 L Glass Media Bottles (Fisher Scientific, cat. no. FB8001000)
- Incubator Thermometers (Fisher Scientific, cat. no. 13-210-577)
- Incubator [one each needed at 30^{o}C, 37^{o}C, 60^{o}C](Thermo Scientific IGS60, cat. no. 51028063)
- Microcentrofuge (Eppendorf Centrifuge 5420, cat. no. 5420000245)
- Microporous Film (USA Scientific, cat. no. 2920-1000)
- Modular Heatblocks (VWR International, cat. no. 13259-000)
- NanoDrop One (Thermo Scientific, cat. no. ND-ONE-W)
- PCR Machine (BioRad, C1000 Touch Thermal Cycler)
- PCR Pull-Apart 8-Tube Strips (USA Scientific, cat. no. 1402-3900)
- Plate reader (Tecan infinite M1000 pro OR BioTek Epoch 2 Microplate Spectrophotometer)
- 4.5 mm Plating Beads (Zymo Research, cat. no. S1001)
- Promega Vac-man Vacuum Manifold (Fisher Scientific, cat. no. PR-A1331)
- Promega x-tracta Gel Extractors (Fisher Scientific, cat. no. PRA2122)

- PYREX Erlenmeyer Flasks (Fisher Scientific, cat. no. S63273)
- 10 ml Serological Pipets (Fisher Scientific, cat. no. 13-678-11E)
- Shaking Incubator (New Brunswick Innova 44, cat. no. M1282)
- Swinging Bucket Rotor Centrifuge (Backman Coulter or similar)
- Tube Revolver Rotator (Similar: Thermo Scientific, cat. no. 88881001)
- Vortex Mixer (VWR, cat. no. 10153-834)
- 42^{o}C Water bath (Similar: Thermo Scientific Precision General Purpose Baths, cat. no. TSGP02)

5.7.3 Reagent Setup

- **0.9% Agarose**
  To make 500ml of 0.9% agarose, add 4.5 g of Agarose and 10 ml of 50X TAE to 300 ml of ddH20. Microwave until boiling and agarose is fully dissolved. Needs to be stored at 60^{o}C – 65^{o}C oven until use.
- **1000x Ampicillin (100mg/ml)**
  Add 1g of Ampicillin to 10 ml of sterile water. Vortex mix to a homogenous solution and filter sterilize with a 0.22 um filter. Aliquot to smaller volumes and store at -20^{o}C. These aliquots are stable at -20^{o}C for 6 months to one year.
- **Buffer 1 [For Protoplast Fusion]**
  Combine 200 mL Citrate Phosphate buffer, 51.3 mL 3M KCl and 5.13 mL 0.5M EDTA (pH=8.0)
- **Buffer 2 [For Protoplast Fusion]**
  Combine 33.3 mL 50% PEG, 10 mL 3M KCl, 5 mL 0.5M CaCl2, and 1.7 mL ddH2O
- **Buffer 3 [For Protoplast Fusion]**
  Combine 20ml 3M KCl, 10 mL 0.5M CaCl2 and 70 mL ddH2O
- **Buffer PB**
  To make a 500 ml solution, dissolve 238.8g of Guanidine Hydrochloride (5M Gu-HCl) in 150 ml of isopropanol (30%) then add sterile ddH20 to 500 ml.
- **Buffer PE**
  To make 1 L solution, add 100 ml of 100mM Tris-HCl (pH 7.5) and 100 ml sterile ddh20 to 800 ml of 100% ethanol.
- **Canavanine Stock Solution**
  Dissolve 2.5 g of Canavanine salt in 50 ml of ddH20. Once dissolved, filter-sterilize. This solution should be stored at 4^{o}C. Use 600 mg/L concentration for yeast plates.
- **Citrate Phosphate Buffer [For Protoplast Fusion]**
  Add 34.8 mL 0.1 M Citric Acid to 65.2 mL 0.2 M Na2HPO4 and dilute to 200 mL with sterile water.

- **1000x Chloramphenicol (25mg/ml)**
  Dissolve 0.5 g of Chloramphenicol in 15 ml of 100% Ethanol. Bring to 25 ml total volume with 100% ethanol. Aliquot to smaller volumes and store at -20^{o}C. These aliquots are stable at -20^{o}C for 6 months to one year.

- **1000X G418**
  If using a 1g vial, calculate the amount of PBS (pH 7.4) to add to get a 200mg/ml 'active' concentration. Sterile filter the drug using a syrine and a 0.45 um filter. Aliquot into smaller volumes and store at -20^{o}C. 1000X G418 is stable at -20^{o}C for a few months.

- **333.3X Hygromycin B**
  Dissolve 1g of Hygromycin B powder in 10 ml of water, filter sterilize. Aliquot to smaller volumes and store at -20^{o}C. 333X Hygromycin B is stable at -20^{o}C for 6 months and upto one year.

- **1000x Kanamycin (50mg/ml)**
  Add 0.5 g of Kanamycin to 10 ml of sterile water. Vortex mix to a homogenous mixture and filter sterilize with a 0.22 um filter. Aliquot to smaller volumes and store at -20^{o}C. These aliquots are stable at -20^{o}C for 6 months to one year.

- **1M Lithium Acetate**
  Dissolve 10.g of Lithium Acetate in 100 ml of sterile ddH20, and filter sterilize. This solution can be stored at room temperature.

- **0.9% NaCl**
  Dissolve 9g of Sodium Chloride in 1000ml of sterile water. Autoclave or filter sterilize. Can be stored at room temperature for 6 months.

- **1000x Nat**
  Add 1g of Neuroseothin to 10 ml of sterile water. Vortex mix to a homogenous solution and filter sterilize with a 0.22 um filter. Aliquot to smaller volumes and store at -20^{o}C. These aliquots are stable at -20^{o}C for 6 months to one year.

- **Plating Buffer [For Protoplast Fusion]**
  3% agar 0.6M KCl and appropriate drop out media, 50 ml per fusion.

- **5M Potassium Acetate**
  To make a 50 ml solution, add 24.5g of potassium acetate to approximately 30 ml of sterile ddH20, dissolve completely at room temperature and then bring up to 50 ml volume. This solution can be stored at room temperature.

- **50x TAE**
  To make 1L of solution, add 242 gm of Tris base and 57.1 ml of glacial acetic acid and 100 ml of 0.5M EDTA (pH 8.0) to 500 ml of sterile water. Once Tris base has dissolved, bring up to volume.

- **TE Buffer**
  To make a 50 ml solution, add 2 ml of 0.5M EDTA and 2.5 ml of 1M Tris-HCl to 45.5 ml of sterile ddH20. This solution can be kept at room temperature.

- **Tris-HCl (100 mM, pH 7.5)**
  To make 1 L solution, dissolve 15.76g of Tris-HCl in 400 ml of sterile water. Once dissolved, bring to 1 L volume.

- **1X YPD**
  To make 500 ml solution, dissolve 5g of yeast extract, 10g of dextrose and 5g of peptone in 30 ml of sterile ddH20. Once dissolved, bring volume to 500 ml and either filter-sterilize or autoclave. If making 2X YPD, add 50 ml of 20% glucose to 450 ml of YP after autoclaving to prevent sugars from burning. This can be kept at room temperature.
- **2x YT**
  Add 8g of Tryptone, 5g of Yeast Extract and 2.5g of NaCl to 450ml of ddH20. Still well until fully dissolved, autoclave. Once cooled, this solution can be stored at room temperature.
- **Wash Buffer [For Protoplast Fusion]**
  Add 22.36 mg of KCl to 300 ml of ddH20, and stil until dissolved. Bring up to 500 ml volume. Filter-sterile or autoclave and store at room temperature.
- **Zymolyase Solution**
  To make a 50 ml solution, add 8.2g of Sorbitol (0.9M) , 10 ml EDTA (0.1 M, 8.0 pH) and 0.025g Zymolyase 20T, to 20 ml of sterile ddh20 and dissolve at room temperature and then bring up to 50 ml volume. Zymolyase may be stored at -20^{o}C in smaller aliquots or at 4^{o}C.

5.7.4 Primers

Primer No.	Sequence	Description
2744	CGGGTTCCTGGCCTTTTGCTGG	Seq. for Cam plasmid Fwd
2620	GGATTTGTTCAGAACGCTCGG	Seq. for Cam plasmid Rev
E704	GTGGATAACCGTGCGG	Seq. for Kan plasmid Fwd
B263	AATAAATTGCAGTTTCATTTGATGC	Seq. for Kan plasmid Rev
1881	CGAACTTGAAGAATAACCAAGG	Can1 Outside Fwd
1882	GTTTAGTGTATGTTGCGCTTAC	Can1 Outside Rev
8794	GCGCACTACCAGTATATCATCTCATTTCC	Ade2 Outside Fwd
8795	CGGACACTTATATGTCGAGCAAGAGC	Ade2 Outisde Rev
A461	CACCTCGTTCCCAGAAACAT	Lyp1 Outside Fwd
A462	AGTACTATCGCTGGCCGAAG	Lyp1 Outside Rev
D309	GAAGGTGTTTTAGAACCAAAGAGC	ARS314 Outside Fwd
D310	TCTATGGCCTTTTCGTCGG	ARS314 Outside Rev
2365	ACACATAACATAGGGGAGAGTACTAAAAG	p1 Outside Fwd
7313	ATCTATTTTTTCAGTACATAATTTAATTAAATC	p1 Inside Rev

Table 3: Primers to be used for sequencing or PCR

5.8 Procedures

5.8.1 Primer design for OrthoRep Toolkit plasmids

All the plasmids in the OrthoRep toolkit are made up of Golden Gate Assembly parts that are compatible with the yeast toolkit. The yeast toolkit will be necessary to build plasmids and utilize the full extent of this OrthoRep toolkit and system as described here.

Golden Gate Assembly utilizes Type IIs restriction enzyme where the cleavage site is different from the recognition site leading to an overhang (Engler, Kandzia, & Marillonnet, 2008). When using BsmBI or BsaI as our restriction enzymes, the overhang created is 4 bp in length (Engler et al., 2008). We can modularize our ability to make plasmids by having each category of parts required to make a plasmid have the same set of overhangs. For example, a standard set of parts required for a plasmid will consist of a promoter, a coding sequence, a terminator as well as a bacterial origin of replication and a resistance marker for selection. If this plasmid needed to be integrated into the yeast genome, or in this case, p1, we would also need a set of parts that would act as integration flanks and a part that would be a yeast marker for selection. Similarly, if we wanted to add a CEN/ARS plasmid to our yeast strain, we would need a yeast origin of replication as well. Many these parts have been created for this purpose and included in the yeast toolkit (Lee et al., 2015). The plasmids specifically required for evolution using OrthoRep are included here as an OrthoRep toolkit that is to be used in addition to the yeast toolkit.

One of each part from 1 – 8 mentioned below must be included to build a complete plasmid using the Golden gate Assembly method. Most of the parts, other than the coding sequence (your gene of interest) are included in this toolkit. For genomic integrations or p1 integrations, a yeast origin of replication is not necessary, so use the 678 part that contains only the bacterial origin and resistance marker for selection. These integration plasmids must be digested to linearize before transformation. Appropriate restriction sites, ScaI for p1 integration and EcoRI for genomic integrations are already included in the flanks mentioned. When yeast plasmids need to be made, a yeast origin of replication in the form of low copy CEN/ARS or a high-copy 2 micron must be included.

In the first step of making parts compatible with the toolkit, and efficiently using them, we start by defining the number of the part we are trying to create. Use the table above as reference. For example, to make a part3 plasmid for the coding sequence, we will PCR the gene of interest with overhangs on either side. These overhangs will contain BsmBI and BsaI sites with the BsmBI sites being on the outside and the BsaI sites closer to the gene of interest. The 4 bp overhang that define the part plasmid will be at the cleavage site for the BsaI enzyme. The amplicon should be designed as follows:

PartNo.	Category	TTTT	tttt	bbbb	BBBB
1	Genomic Integration flank	CCCT	AACG	GGGA	TTGC
1a	p1 integration upstream flank	CCCT	GTTT	GGGA	CAAA
1b	p1 promoter and marker	GTTT	AACG	CAAA	TTGC
2	Nuclear and p1 promoters	AACG	TATG	TTGC	ATAC
3	Nuclear and p1 coding sequences	TATG	ATCC	ATAC	TAGG
4	Nuclear terminator	ATCC	GCTG	TAGG	CGAC
4	PolyA for p1	ATCC	GCTG	TAGG	CGAC
45	p1 integration downstream flank	ATCC	TACA	TAGG	ATGT
5	p1 integration downstream flank	GCTG	TACA	CGAC	ATGT
5	Genomic integration downstream flank	GCTG	TACA	CGAC	ATGT
6	Yeast marker	TACA	GAGT	ATGT	CTCA
678	Bacterial origin of replication + resistance marker	TACA	CCCT	ATGT	GGGA
7	Yeast origin of replication	GAGT	CCGA	CTCA	GGCT
8	Bacterial origin of replication + resistance	CCGA	CCCT	GGCT	GGGA

Table 4: Categories of parts for OrthoRep Toolkit

5'-GCATCGTCTCATCGGTCTCATTTT –part– ttttTGAGACCTGAGACGGCAT-3'

3'-CGTAGCAGAGTAGCCAGAGTbbbb -part- BBBBACTCTGGACTCTGCCGTA-5'

The first part of the primer must consist of the two restriction sites. Replace the TTTT for the forward primer with the appropriate TTTT code as given in the table above. Similarly, for the reverse primer, both restriction sites must be added and then the part specific overhang as denoted in the table in the bbbb column. The second part of the primer should contain 15-20 bp of overlap with the end of the part to ensure amplification. The PCR for amplification should be done in two steps, the first step to amplify the part and the second step to include the enzyme binding sites as well.

5.8.2 PCR gene of interest

This is a general protocol for PCR using Phusion Enzyme but can be substituted with another PCR enzyme and the appropriate protocol. For ordering primers, please read the section above.

PCR Reagent	1x Reaction
5x Phusion GC Buffer	10. 0 ul
dNTPs	1.0 ul
Fwd Primer (100ul)	0.25 ul
Rev Primer (100ul)	0.25 ul
Template (Plasmid: 1ng/ul or gDNA)	1.0 ul
DMSO	1.5 ul
Phusion HS Flex Polymerase	0.5 ul
ddH2O	35.5 ul (Upto 50.0 ul)

Note: If amplifying from a plasmid, it is important to dilute the plasmid to a concentration between 1pg – 10 ng/ul concentration and then using 1ul of that dilution as the template. If using a plasmid as template, digest with Dpn1 for 1 hour before proceeding with the next steps. This prevents the plasmid from being visible on the gel and thus accidentally extracting it with the newly amplified amplicon. It also ensures that most of the DNA is the new amplicon and not the template.

Thermocycler protocol:

Temperature	Time	No. of cycles
98^{o}C	30 sec	
98^{o}C	10 sec	
45^{o}C - 72^{o}C	30 sec	5 cycles
72^{o}C	30 sec/kb	
98^{o}C	10 sec	
45^{o}C - 72^{o}C	30 sec	30 cycles
72^{o}C	30 sec/kb	
72^{o}C	5 min	
12^{o}C	∞	

Note: This protocol is a two-step PCR protocol, which is used when using primers with overhangs that need to be added to the final amplicon including when making golden gate parts. The first annealing temperature should be calculated where the primer will be annealed to the starting amplicon. This starting temperature should ideally be above 45^{o}C and the primer should overlap at least 15bp. The second temperature should be the annealing temperature for the entire primer and may be 72^{o}C. If using a plasmid as template and the primers entirely bind to it, you can follow the first step followed by the second loop and a final extension as the PCR protocol.

5.8.3 Purification of PCR

This is a general protocol for purification of PCR. All steps were performed at room temperature unless otherwise mentioned.

- Pour 30 ml of 0.9% Agarose into the casting tray with the correct size of combs. Ensure that the agarose makes it about 80% of the way to the top of the combs.
- Add 5ul of ethidium bromide to the tray and stir it around to ensure it is properly mixed.

  Note: Ethidium bromide is a known carcinogen and must be handled carefully.
- Let the gel dry for 20-30 minutes at room temperature.
- When dry, place the gel in the gel box and add 1X TAE until the buffer covers flows over the gel such that no part of the gel is exposed.
- Mix your DNA sample (PCR, restriction digest etc.) with loading dye by pipetting up and down a few times, then gently pipette each sample in to an individual well. Add the appropriate ladder

  Note: When only one band is expected on the gel, run 5 ul of the PCR on the gel and purify the rest over a column. Add 200 ul of ADB to each PCR sample and follow protocol after with no changes.
- Run gel at 120V for 30-35 minutes or until the loading dye has made it to 80% of the length of the gel.
- Image gel with UV light to visualize bands.
- Once imaging has confirmed that the bands are the required size, use either a plastic gel extractor tool or a fresh razor blade to excise the correct band, and add it to a sterile labelled centrifuge tube.
- Add 3 volumes of ADB for the gel slice and place it on a 65^{o}C heatblock until the gel slice is fully melted. This should take 15-20 minutes.

  Note: Measure the amount of gel that has been extracted and add 300ul of ADB for each 100 mg of gel slice cut. Additionally, 50 ul of PB may be added for better binding.
- Place a white column on the vacuum manifold. Gently pipette the liquid onto the membrane of the column making sure to not disturb the membrane.

Note: A spin at 4000g for 3 minutes may be used in place of a vacuum manifold.

- Once the liquid has flowed through, add 500ul of PB buffer to the column to wash. Repeat the PB Buffer wash.
- Add 500ul of PE Buffer and let it flow through. Repeat the PE buffer wash.
- Move the column to a collection tube and centrifuge at 12,000g or max speed for 3 minutes.
- Place the column in a fresh centrifuge tube and gently pipette 15 ul of sterile water onto the membrane. Incubate at room temperature for 3 minutes.
- Centrifuge at 12,000g or max speed for 3 minutes to elute. Measure concentration using NanoDrop

5.8.4 Assembling the part plasmid

This is a general protocol for assembling a part plasmid using Golden Gate Assembly Protocols using Type II Restriction Enzymes. This protocol was performed at room temperature unless otherwise described.

- Use a freshly miniprepped aliquot of pYTK001 the entry vector for a part plasmid that has a GFP dropout.
- For each reaction, 50 fmol of the backbone and 100 fmol of the insert must be used. The most straightforward way is to make a stock solution such for each such that the backbone is at a 50 fmol/ul solution and each of the inserts are at 100 fmol/ul solution.
- Setup a 20 ul reaction on ice as follows:
  - 1 ul backbone (pYTK001)
  - 1 ul insert
  - 1 ul BsmBI
  - 1 ul T4 Ligase
  - 2 ul T4 ligase buffer (10X)

- 14 ul Sterile Water

- Incubate at 37^{o}C for 30 – 45 minutes.

- Transform 5 ul into a 25ul aliquot of Top 10 chemically competent cells. Recover in 200 ul of 2YT for 1 hour and plate on a LB + Chloramphenicol plate.

- Incubate overnight (14-18 hrs) at 37^{o}C until colonies are formed.

- Pick colonies that are not green in liquid 2YT + Chloramphenicol.

  Note: To increase efficiency of this reaction, a cycling protocol can be used. Alternately, a longer isothermal protocol can be used for applications where there is low tolerance for background green colonies.

5.8.5 Chem Comp Cell Preparation

This is a general protocol for most E. coli lab strains. Top 10 cells were used in this protocol. This protocol can be scaled to higher volumes as needed. All steps except the overnight culture must be done between 0^{o}C to 4^{o}C. A cold room may be used. The Mix and Go E. coli transformation Kit and Buffer Set was used for this protocol.

- Streak Top 10 cells onto a plate made of either LB with no antibiotic, or 2YT with no antibiotic.

  Note:This plate can be parafilmed and stored in the fridge for upto one month.

- Start a 2 ml culture of Top10 cells in 2YT media with no antibiotics in a glass culture tube. Grow overnight at 37^{o}C in a shaking (200 rpm) incubator.

- Inoculate 0.5 ml of the saturated culture in 50 ml of SOB media in a 500ml glass culture flask. Shake at 200rpm and 37^{o}C until OD600 is 0.4-0.6.

- Pour equal volumes of the culture into two 50ml Falcon tubes and incubate on ice for 10 minutes. Centrifuge at 2500g for 10 minutes at 4^{o}C. Discard supernatant.

- Resuspend the cells in one of the falcon tubes gently in 5 ml ice cold 1X Wash Buffer. Transfer the entire volume into the other falcon tube and resuspend cells in the same liquid.

- Repeat centrifugation. Discard supernatant. Resuspend the cells gently in 5ml ice cold Competent Buffer.

- Aliquot on ice or in the cold room, 25ul of cells into sterile Eppendorf tubes and store at -80^{o}C until ready to use.

  Note: These cells are ready to use as necessary and can be stored at -80^{o}C for up to 6 months. They will remain efficient as long as they do not undergo freeze-thaw cycles.

5.8.6 Chem Comp Cell Transformation

This is a general protocol for most E. coli lab strains. Top10 chemically competent cells were used for this protocol. This protocol is performed at room temperature unless otherwise noted.

- Thaw an aliquot of Top10 chemically competent cells for each transformation on ice for 5-15 minutes.

- Add 5-15 ul of golden gate reaction to per tube and incubate on ice.

- Heat shock for 30 seconds in 42^{o}C water bath. Put back on ice.

- Add 200ul of 2YT to each tube and keep in shaking incubator at 200rpm and 37^{o}C to recover for 1 hour.

- Plate on Petri dishes containing LB Agar + appropriate antibiotic and keep overnight at 37^{o}C.

  Note: Cells transformed with plasmids containing ampicillin or carbenicillin resistance do not require recovery and can be plated immediately following heat shock.

- Colonies should appear in 14-18 hours after plating.

5.8.7 Picking colonies to start cultures

This is a general protocol for most E. coli and S. cerevisiae lab strains. This protocol is performed at room temperature unless otherwise noted.

- With the Bunsen burner turned on, and near the flame to maintain sterility, prepare a block with appropriate number of wells or the appropriate number of glass culture tubes with 3 mls of 2YT + correct antibiotic. Ensure that there is one extra well or culture tube containing media as the control.

- With the burner still on, touch the tip of a pipette tip or a sterile reusable wood stick or reusable toothpick from a sterile container, touch the colony you want to pick while being careful to not pick any other nearby colonies and drop into an individual well or culture tube.

- Once all colonies have been picked into their own well or tubes, gently tap the pipette tip or wooden stick or reusable toothpick to the wall of the well or tube and discard into a separate container.

- Once ready, is using a block, stick the breathable film onto the block ensuring all the well that have been occupied have been properly covered. Label the block or tubes and put in a 37^{o}C shaking incubator at 200 rpm for 14-16 hours.

5.8.8 Bacterial cell miniprep

This is a general protocol for most E. coli lab strains. Top10 chemically competent cells were used for this protocol. This protocol is performed at room temperature unless otherwise noted.

- From the 3 ml culture growing in 2YT, stock 500ul of culture with 500ul of 50% glycerol in a cryovial. Make sure the culture is well-mixed before storing at -80^{o}C. Strains can be stored at -80^{o}C for long periods of time in absence of frequent freeze-thaw cycles.

- Aliquot 1.5 ml of each bacterial culture into a fresh sterile Eppendorf tube.

- Centrifuge at 5000g for 10 minutes.

- Toss supernatant.

- Resuspend in 600ul of 0.9% NaCl.

- Add 100ul of 7X Blue Lysis Buffer and invert the tube several times until the liquids are well mixed.

- Add 400ul of chilled Neutralization Buffer and shake to mix properly. The culture will start to become clumpy.

  Note: Neutralization Buffer must be kept chilled and in the 4^{o}C fridge to ensure efficient bacterial DNA extraction. Please remember to put away Neutralization Buffer in the fridge as soon as you're done using it.

- Centrifuge tubes at 12,000g for 15 minutes. While the centrifuge is running, place the required number of columns on vacuum manifold.

  Note: Vacuum manifolds were used in this protocol but are not necessary. A centrifuge can be used instead.

- Pipette the supernatant into the columns, making sure that the pellet is avoided and allow it to flow through the column.

- Add 500ul of PB buffer – twice.

- Wash with 500ul of PE buffer – twice.

- Add the column to a collection tube and centrifuge between 10,000g and max speed for 3 minutes.

- Move the column to a fresh Eppendorf tube and add 50ul of sterile water. Incubate on the benchtop for 3 minutes at room temperature.

  Note: When adding sterile water to the membrane in the column, make sure to not touch the membrane and to pipette the water in the middle of the membrane and not the wall of the column.

- Centrifuge between 10,000g and max speed for 3 minutes – the elute contains plasmid DNA. The concentration can be measured on NanoDrop or another similar instrument and used for downstream steps.

- For a plasmid, before sequencing, run a small volume on a 0.9% agarose gel to check for size using a supercoiled ladder or plasmids can be digested with appropriate restriction sites and then run with appropriate ladder to check for size.

5.8.9 Building transcriptional units and CEN/ARS plasmids

This is a general protocol for building transcriptional units and CEN/ARS plasmids using Golden Gate Assembly. This may be scaled up as necessary.

Note: To setup golden gates, it is important to start with clean fresh minipreps. The quality of the miniprep makes a huge difference to the efficiency of the golden gate reaction.

- Make 10 nm aliquots for each part that will be in the golden-gate reaction. Ensure that all parts from 1-8 to form a complete plasmid are used. Pre-assemble the parts using NEB Golden gate Assembly tool to confirm.

  Note: NEBiocalculator may be used for this purpose. The amount of DNA is given in mg/L, and should be scaled down to 10-50 ul depending on number of reactions to be performed.

- Setup a 20ul reaction on ice as follows:

  - 1 ul per part to be used (10 nm concentration)
  - 1 ul BsaI-HF-v2 [This will be BsmBI-HF-v2 if making a CEN/ARS plasmid]
  - 1 ul T4 ligase
  - 2 ul T4 ligase buffer (10X)
  - Sterile water upto 20 ul

  Note: Setup reactions on ice for best results.

  Thermocycler protocol:

Temperature	Time	No. of cycles
36^{o}C	3 min	
16^{o}C	1:30 min	Repeat 30 cycles
60^{o}C	5 min	

  Note: If the reaction is being performed overnight, repeat the 60^{o}C for 5 minutes step before transforming

- transform 5 ul into a 25 ul aliquot of Top 10 chemically competent cells. Recover in 200 ul of 2YT for 1 hour and plate on appropriate agar plate.

- Incubate the plate overnight (14-18) hours at 37^{o}C until colonies are formed.
- Pick colonies into liquid 2YT + Antibiotic media and grow overnight.
- These cultures can be miniprepped and used as needed after.

5.8.10 Yeast transformations

This is a general protocol for most S. cerevisiae lab strains. The same protocol can be used for integrating into the genome, integrating onto p1 as well as transforming circular plasmids like CEN/ARS and 2 micron. The genetic material to be added and the quantities to be added per transformation will vary. The protocol is done at room temperature unless otherwise noted.

- Start a 2 ml culture of the starting parent strain in a glass culture tube or one well of a plastic 24-well block until saturated in the appropriate media (YPD or SC). Culture should be grown in a shaking incubator at 200rpm and 30^{o}C until culture is saturated Depending on the strain being used, this should be between 2-4 days.
- Once saturated, back dilute the culture 1:30 in YPD the morning of the transformation. Consider 10 ml YPD culture per transformation. Let cells grow until each culture is at a saturation of 2 x 10-7 cells/ml. This takes approximately 3-5 hours based on the types of cells used.
- Once the culture is at an appropriate saturation, harvest it into centrifuge tubes and centrifuge at 3000g for 5 minutes.
- Pour off the supernatant being careful not to dislodge the pellet. Resuspend the pellet in 1 ml of sterile water per transformation and aliquot into individual Eppendorf tubes for each transformation.
- Spin the tubes as before, at 3000g for 5 minutes. Pour off supernatant. The cells are now really for transformation mix and the required DNA to be added.
- To each labelled tube, add the following in the order specified:
  - 240 ul of PEG (50% w/v)

- 36 ul of 1.0 M LiAc
- 10 ul of 5.0 mg/ml salmon sperm DNA
- 65 ul of sterile water + required DNA

  Note: For genomic integrations, digest 2ug of the circular plasmid per transformation with EcoRI, for p1 integrations, digest 2ug of the circular plasmid per transformation with ScaI, and for circular plasmids like CEN/ARS and 2 micron, use 200 ng – 2 ug of plasmid per transformation. For control, use water instead of plasmid DNA.

- Vortex each tube vigorously until the cell pellet is dislodged and properly mixed such that no chunks remain. This usually takes one minute.

- Incubate in a shaking incubator at 200 rpm and 30^{o}C for 30 minutes. Note: This 30-minute timepoint can be allowed to go longer for up to 2-3 hours if necessary.

- At the end of 30 minutes, place tubes in a 42^{o}C water bath for 20 minutes. Note: This timepoint is important to stick to as closely as possible.

- Once done, centrifuge at 1500g for 5 minutes. Pipette the supernatant out being careful to not dislodge the pellet. If cells have antibiotic resistance, cells should be resuspended in 600ul of YPD and rotated at 30^{o}C for at least 5 hours but can be left overnight if necessary. Alternatively, transformations can be plated onto YPD plates and replica-plated the next day onto appropriate antibiotic containing plates.

- Gently resuspend the cells in 200ul of 0.9% NaCl and plate using beads or an alternate method on plates with appropriate dropout to select for newly added plasmid or integration cassette. Store in a plastic box in a 30^{o}C incubator until colonies form. CEN/ARS plasmid colonies can be expected in 2-3 days, genomic integrations take between 3-5 days to appear and p1 integrations xtake between 4-7 days to appear.

- Once colonies are visible, for genomic and p1 integrations, single colonies should be streaked to singles on the same places previously to ensure that single colonies have been properly isolated. This second round of colonies typically appears in 3-4 days. The colonies can then be picked into appropriate liquid media and grown to saturation. They must then be

stocked at -80^{o}C in equal volumes of culture and 50% glycerol until they are needed again. If downstream activities need to be completed, they can be performed using the saturated culture.

Note: When making plates using yeast media, antibiotics can be used as normal in YPD media. If using Synthetic Complete or dropout media with antibiotics, please use Yeast Nitrogen Base WITHOUT Ammonium Sulfate, add 1 g/L monosodium glutamate as the nitrogen source instead.

5.8.11 Protoplast fusion to transfer p1 into desired strain

This is a general protocol for most S. cerevisiae strains. This protocol can be used to transfer p1 into strains that do not contain it. It is important to note that it is essential to have a selectable marker on the p1 that is to be transferred, and p2 will be transferred automatically. Wild-type p1 containing the toxin-antitoxin subunits are not stable in strains that contain functional mitochondria. All steps are performed at room temperature unless otherwise mentioned.

- 3 days before fusion: Start cultures from single colonies or glycerol stocks in 3ml of appropriate selective media. Grow at 30^{o}C in shaking incubator until fully saturated.
- 1 day before fusion: Add 1 ml of saturated culture to 50 ml of YPD in glass flask for each fusion partner separately. If one donor strain is being fused with two acceptor strains, start 2x 50 ml cultures for the donor and 1x 50 ml culture for each of the acceptor strains.
- Day of fusion: Prepare 300 ml of plating buffer and autoclave for 15 minutes.
- Pour saturated cultures into 50 ml conicals and centrifuge at 3000g for 5 minutes using a swinging bucket rotor. Discard supernatant.
- Resuspend each pellet in 10 ml of sterile water and transfer to 15 ml conical.
- Centrifuge using swinging bucket rotor at 3000g for 5 minutes. Discard supernatant completely. Tap onto a KimWipe to ensure that all liquid has been drained, and that the cells are firmly set at the bottom of the tube.

- Measure 0.3-0.4 g of cells into a fresh 15 ml conical for each fusion partner per reaction and discard excess. Use a portable flame to maintain sterility.

- Resuspend cells in each conical in 1.8 ml of beta-mercaptoethanol and EDTA solution (0.2% beta-mercaptoethanol, 0.06M EDTA in water).

- Place in a 30^{o}C incubator for 30 minutes with no shaking. After 15 minutes, gently rock the tubes a few times and return to incubator.

- For each sample, prepare 5 ml of Zymolyase solution with 1.5 mg of 20T Zymolyase dissolved in Buffer 1.

- Wash with 3 ml of Wash Buffer to get rid of excess BME and centrifuge at 3000g for 5 minutes. Repeat twice.

  Note: Use fresh 10 ml serological pipettes for each sample.

- Resuspend each sample with 5 ml of Zymolyase solution.

- Incubate samples at 30^{o}C for 1 hour with no shaking. Gently rock the samples halfway through the incubation and return to the incubator.

  Note: This is the time to set a centrifuge at 0^{o}C for a future centrifugation step. The plating buffer should be autoclaved and put in the 60^{o}C oven until later.

- Centrifuge at 2000 rpm at 0^{o}C for 10 minutes.

  Note: The supernatant may be turbid.

- Wash the samples with Wash Buffer (at RT). Centrifuge at 2000 rpm at 0^{o}C for 10 minutes. Repeat wash and spin.

- Resuspend each sample in 3 ml of Buffer 1.

- Create fusion mixes by adding 1.5 ml of each fusion partner to a fresh 15 ml conical. Use the rest as the control.

- Centrifuge to mixes at 2000 rpm at 0^{o}C for 10 minutes.

Note: Make sure the centrifuge is balanced.

- Resuspend each sample in 5 ml of Buffer 2 and use 2.5 ml for each control.
- Incubate at 30^{o}C for 30 minutes with occasional shaking.
- Centrifuge at 2000 rpm at 0^{o}C for 10 minutes.

  Note: Transfer the plating buffer to a 42^{o}C water bath ensuring that the entire volume of the plating buffer is submerged within the water. Media containing agar will solidify at temperatures below 60^{o}C. However, when yeast is combined with media at 60^{o}C, it will kill the yeast.

- Resuspend each fusion sample in 5 ml of Buffer 3 and each control in 2.5 ml of Buffer 3.
- Label one plate for each fusion and control. Transfer 500 ul from each sample (fusion and control) into a fresh 15 ml conical.
- Quickly pour 10 ml of plating buffer into one of the tubes, gently mix and pour before agar solidifies in the conical. Spread agar as much as possible before it solidifies. Repeat for each fusion and control.

  Note: As the agar cools quickly, ensure that this step is done quickly. The agar may solidify before it is uniformly spread over the plate. This will not affect the protoplast fusion.

5.8.12 Short miniprep for genomic DNA isolation

This is a general protocol for most S. cerevisiae strains. This protocol may be used to PCR DNA after either genomic integrations or p1 integration. This protocol cannot be used instead of the long miniprep to visualize p1 and p2.

- Spin 1.5 ml of saturated culture at 5000g for 1 minute. Discard supernatant.
- Wash the pellet with 1 ml of 0.9% NaCl. Repeat spin at 5000g for 1 minute. Discard supernatant.
- Resuspend in 125 ul of Zymolyase solution and incubate at 37^{o}C for 1 hour with rotation.

- Spin at 3000g for 5 minutes, Discard supernatant.
- Resuspend in 140.25 ul of Proteinase K solution.

  Note: Proteinase K solution is made with 125 ul TE + 12.5 ul 10%SDS + 2.75 ul Proteinase K (10mg/ml)
- Place in 60-65^{o}C oven for 30 minutes with rotation.
- Add 75ul of 5M Potassium Acetate to each tube and incubate on ice for 30 minutes.
- Spin at 12,000 g for 10 minutes.
- Transfer supernatant to a fresh sterile Eppendorf tube and add 400 ul of DNA binding buffer (eg. ADB or PB Buffer).
- Place the miniprep columns on a vacuum manifold.
- Mix properly and pipette the liquid onto the membrane of the miniprep column. Allow liquid to flow through.
- Wash with 500 ul of Buffer PB twice
- Wash with 500 ul of Buffer PE twice.
- Place the miniprep columns in a fresh collection tube and centrifuge at 12,000g for 3 minutes.
- Place the column in a fresh labelled Eppendorf tube and add 30 ul of sterile water to the membrane. Let sit at room temperature for 3 minutes.
- Centrifuge at 12,000g for 3 minutes to elute DNA from the column.
- This eluted DNA can now be used for PCR.

  Note: This DNA cannot be run on a gel to visualize p1/2. To be able to visualize p1/2 on gel, the extracted DNA needs to be treated with proteinase K to digest the terminal proteins at the ends of p1/2.

5.8.13 Long miniprep for p1/2 visualization

This is a general protocol for most S. cerevisiae strains. This protocol is necessary to isolate and to visualize p1/2 linear plasmids on a gel. This protocol is performed at room temperature unless otherwise noted. For best results, follow the protocol and the incubation times as closely as possible. Appropriate locations for pausing are mentioned below.

- Spin down 1.5 ml of saturated cultures in individual sterile Eppendorf tubes at 5000g for 1 minute. Discard supernatant.

  Note: For respirating strains like BY4741/ BY4742 or other, use only 750ul of a fully saturated culture since these strains saturate at higher ODs. The presence of mitochondrial DNA increases the amount of DNA per cell and downstream processes may be affected. Alternatively increasing the volume and time for Zymolyase, Proteinase K and RNase may help.

- Resuspend in 1 ml of 0.9% NaCl. Repeat spin at 5000g for 1 minute. Discard supernatant.

- Resuspend in 250ul of Zymolyase solution and incubate at 37^{o}C for 1 hour with rotation.

  Note: Zymolyase solution is prepared sing 0.9M Sorbitol + 0.1M EDTA (pH 8.0) + 10 units/ml Zymolyase. This solution can be made in a 50ml conical and stored at 4^{o}C, or it can be aliquot into smaller volumes for more longer-term storage at -20^{o}C.

- Spin down at 3000g for 5 minutes and discard supernatant.

- Resuspend the pellet in 280.5 ul of Proteinase K solution

  Note: Proteinase K solution consists of 250ul TE + 25ul of 10% SDS + 5.5 ul 10mg/ml Proteinase K. The 10mg/ml of Proteinase K solution must be prepared in water and can be stored in the 4^{o}C fridge for 6 months. This solution needs to be refrigerated immediately after use.

- Incubate at 60-65^{o}C for 30 minutes. During the incubation time, fill an ice bucket for the next step.

- Add 75ul of 5M Potassium acetate and mix the tubes well. Incubate on ice for 30 minutes.

Note: The tubes can stay on ice for longer than 30 minutes if necessary. This is a good pause point and tubes can stay on ice for up to 3 hours.

- Spin down 1.5 ml tubes at 12,000g for 10 minutes. Transfer supernatant to fresh labelled sterile tubes.
- Add 2 volumes (700ul) of 100% Ethanol and invert tubes a few times.

  Note: This is a good pausing / stopping point. Samples can be stored overnight in the refrigerator before continuing with the rest of the protocol.
- Spin down at 3500g for 15 minutes. Discard supernatant.
- Dry pellet at room temperature for 10 minutes and then in the 60-65^{o}C oven for 20 minutes.

  Note: Flip tubes containing pellets over to allow the liquid to drain out onto a paper towel. Tape the lids of the tubes to the paper towel, which can then be transferred to the oven for drying. Alternately, pellets can be dried for 20 minutes at 60-65^{o}C only while skipping the 10 minutes at room temperature. Over drying of the pellet can lead to difficulty in dissolving the DNA.
- Resuspend pellet in 150ul TE by vortexing.

  Note: Be vigorous during vortexing and ensure that most if not the entire pellet has been properly resuspended. Not properly vortexing leads to most of the DNA being in the pellet and not being released into the supernatant for the next step.
- Spin down at 12,000g for 15 minutes. Transfer supernatant to fresh labelled tubes.
- Add 8ul of 1 mg/ml RNaseA and incubate at 37^{o}C for 30 minutes.

  Note: Aliquot RNaseA into small workable volumes and store at -20^{o}C. Do not refreeze the aliquots after use. Use a fresh aliquot of RNase A each time this protocol is performed.
- Add 1 volume (150ul) 100% isopropanol. Invert tubes a few times.
- Spin down at 12,000g for 15 minutes. Dump supernatant.

- Dry pellet at room temperature for 10 minutes and then in the 60-65^{o}C oven for 20 minutes as previously described.
- Resuspend in 30ul of sterile water.
- Run at least 15ul of on a 0.9% agarose gel at 80V for 120 minutes.

  Note: For best results, let the gel dry after casting at room temperature for 1 hour before loading. The gel may be run the day after completing the long miniprep.

5.8.14 Setting up evolutions

This is a general method to calculate optimal evolution volumes for setting up evolution as needed. These calculations are not exact and are only meant to serve as a guide while making key decisions while setting up evolutions.

For these calculations, the following numbers are important:

Error-rate of polymerase (pe)

Length of gene-of-interest (l)

Copy number of p1 plasmid (n)

Number of possible options for single nucleotide mutations (m)

Number of possible mutations = l x m = l x 3 = 3l

Here, m = 3

Number of bases replicated per cell division = n x l

Number of mutations per cell division = n x l x pe

Number of cells required to cover all single mutations = 3l x n x pe

Number of cells required to cover all double mutations = $(3lxnxpe)^2$

If OD600 1.0 = 1.5 x 107 cells,

Number of cells per passage = $\frac{1.5*10^7}{3l*n*pe}$

Note: We recommend ensuring that the bottleneck between passaging is at least large enough to cover all single mutations each time to ensure sufficient diversity.

5.8.15 Growth Assays

This is a general protocol to determine growth rate and may be used as a method for comparison of growth rates among samples and over time using non evolved and evolved variants. This method is not ideal when dealing with strains that have a doubling time over 5 hours.

- Normalize the freshly saturated strains by measuring OD600 of the cultures required and back dilute in 200 ul – 2 ml of media ensuring that the number of cells is approximately equal per sample. Ideally, the OD600 after back dilution is between 0.2 and 0.3.

  Note: While back dilution is not strictly necessary, it allows cells to be in log phase and reduces the lag time before collecting useful data once on the plate reader.

- Allow these back diluted cells to grow for 4 hours at 30^{o}C in a shaking incubator. This should allow 2-3 doublings.

- To plan a growth curve, use a 96 well plate. Ensure that all conditions and samples to measure growth rate are at least in duplicate, but ideally in triplicate. This will allow between 36 and 48 samples per 96 well plate.

- In the 4 hours, prepare appropriate media for all the conditions to be tested ensuring appropriate controls.

- Add 99 ul of appropriate media per well. Add 1 ul of the back diluted culture to each well containing appropriate media. Add a transparent plate seal to prevent evaporation ensuring that there are no wrinkles in the seal.

  Note: The ideal starting OD600 is between 0.05 and 0.1. If the back dilution is too dilute to be able to start at that OD600, spin down the cells after back dilution and resuspend in smaller volume to ensure appropriate starting OD.

- When using a plate reader, setup the program such that an OD600 measurement is taken every 30 minutes for a period of 24-48 hours. Ensure that the temperature is maintained at 30^{o}C and that the plate is shaking in a double-orbital motion (figure of eight motion) throughout the duration of the experiment.

- Once experiment is complete, analyze data to determine growth rate of each sample before averaging over the replicates.

5.8.16 Spot Assay

This is a general protocol to compare differences in growth between different samples. This is not a quantitative assay but is a useful tool to detect changes in growth rate at a higher throughput than a growth assay.

- Make agar plates of all the conditions needed for the spot assay. One 15mm petri dish with 15ml of agar can comfortably fit 24-30 spots while one large plate can comfortably fit 96-100 spots. Have at least one control where all samples will grow (eg. YPD) to ensure that cells have been transferred to plates and dilutions look as expected. This will also give a sense of how long colonies will take to appear on the plates.

- Normalize the OD600 of the freshly saturated cultures such that there are approximately equal number of cells to start out with. Alternately, the cultures can be spun down and resuspended in 0.9% NaCl such that the OD600 of the samples is approximately the same.

- Make four serial dilutions of each sample, 1:10, 1:100, 1:1000 and 1:10,000, ensuring 10ul for each plate they need to be plated on, with 10% excess.

- Use a multichannel P20 pipette to pipette 5-10ul spots equidistant from each other on all plates.

  Note: The spots will be bigger in size on plates made from YPD media and smaller on SC media. To prevent the spots from combining, smaller volumes may be pipetted.

- Allow spots to be fully dried before transferring to 30^{o}C until there is growth. Depending on the strains and the media conditions, this may take from 2 – 7 days.

- Once colonies have appeared, image using a ChemiDoc with the agar side facing the top. The protein gel with the silver stain gives highest contrast and allows for proper visualization of colonies from individual spots.

5.8.17 Competition Assay using fluorescence

This is a general protocol to identify the fastest growing strains from a mixed population of pre-evolution and post-evolution strains. This assay can determine whether the post-evolution strains are faster at growing or surviving in selection conditions better than the pre-evolution strains.

- Start with a strain that will not evolve the gene of interest. This may be using a wild-type TP-DNAP1 in the genome to amplify p1 or by encoding the mutants on a CEN/ARS plasmid and expressing them in the nucleus as detailed above.
- Create a transcriptional unit using different fluorophores as Part3 plasmids for strains in different stages of evolution. The simplest version of this experiment will use two transcriptional units each one containing a different fluorophore eg. mKate and GFP.
- If CEN/ARS plasmids are not being used in the experiment for any purpose, then use parts from the Yeast toolkit to assemble a CEN/ARS plasmid for each fluorophore. If CEN/ARS plasmid is already being used, then integrate the fluorophore at a location that is available in the strain.
- Transform the yeast strain pre-evolution with a fluorophore and the post-evolution strain with a different fluorophore. Ensure that both strains will grow in the same dropout media.
- Once colonies appear, pick several colonies and allow them to saturate.
- Mix pre-evolution and post-evolution strains in a known ration (50:50, 80:20 etc). and subject to similar conditions as the evolution.
- Select the top mutants, which will depend on your selection. Use a flow cytometer or FACS to determine the percentage or ratios of the top performers using the fluorophores in the strains. The post-evolution strains should outcompete the pre-evolution strains.
- Identify the best mutants and the corresponding mutations.

5.9 Anticipated Results

A range of results may be anticipated that vary based on the specific conditions being used, the length of the gene, the starting activity, the volume of passaging, the strictness of the bottleneck,

the particular error-prone polymerase being used etc. However, when phenotypic differences are observed, whether through growth rate assays, spot assays or when new population are detected or populations have shifted when using FACS selections, mutations should be observed. If mutations have fixed by this point, they may be detectable using bulk sequencing. The number of mutations will vary based on the factors mentioned above. The absence of mixed peaks on bulk sequencing traces or any mutations when NGS is performed is usually a clue that there may be issues with how the selection was designed.

In continuous directed systems, like OrthoRep, while there is less user involvement, it is important to note that occurrence of recombination can go undetected for longer durations. Yeast are very prone to recombination and even two sets of 20 bp located on the same chromosome can recombine. If this recombination occurs in a region required in the evolution scheme, the evolution will not work as intended. Additionally, sustained exposure to chemicals or small molecules used to increase selection pressure during the experiment that is used to slow down to prevent growth of cells may prove to be ineffective. Yeast are extremely good at finding ways to survive, and may find ways to be immune to chemicals or small molecules added to the media by preventing uptake. This may lead to no selection pressure being applied thus no evolution over the course of the experiment.

Selections over the course of evolution that depend on survival in the given environment are highly prone to contamination, since controls or other cheaters will survive better in the given media environment. It is important to keep controls separate and that extra care is taken to prevent contamination to the extent possible. An additional selection marker is a useful way to periodically check for contamination as well as to prevent or get rid of contamination should it occur. One potential pitfall is when selection is designed such that it is easy to bypass, by using alternate mechanisms that allow for growth and cell division without evolution. It is unlikely that during the course of evolution, a few mutations will lead to significant phenotypic changes so, assays used must be sensitive enough to be able to detect small changes in the parameters used to monitor evolutions.

There are some reasons why evolutions may fail that are not related to OrthoRep, but are a result of continuous directed evolution. One reason amongst them is that smaller genes have slower rates of evolution since more bases need to be copied to get mutations, whereas larger genes require larger populations to generate and maintain enough diversity. Larger genes will also require

larger bottlenecks and thus larger passaging volumes making them more prone to contamination. Mutation rates that are too high, with many mutations occurring over a short period of time are not ideal but either are mutation rates too slow, with mutations occurring rarely. The mutation rate should be considered while designing an experiment to ensure that too much time does not occur between the point at which mutations start accruing and the first selection. If many mutations are created during this time without selection, it may lead to a loss of diversity as many inactivating mutations may collect on the gene of interest which are hard to reverse to lead to productive mutations.

Finally, when evolving genes that are not natively found in yeast, adaptation to the yeast cellular environment before evolving for function is an important step in the evolution process. However, the process of adaptation may alter the fitness landscape irreversibly in a way where the starting point for functional evolution has now shifted such that the best possible solution may be inaccessible during evolution. There are many ways adaptation can occur that may change the fitness landscape allowing for better and more accessible starting points. However, this cannot be predicted and can only be discovered during or after an evolution campaign.

5.10 Troubleshooting

Step	Problem	Reason	Possible Solution
PCR	No bands	All reagents not added	Redo PCR, check that all reagents are added in correct concentrations and enough template is used.
		Primers are incorrect	Make sure that primers bind to the correct region on plasmid map
		Temperatures and times on thermocycler incorrect	Depending on the polymerase being used, make sure that the elongation time is sufficient. For longer PCRs, longer extension times are preferred. Also confirm that the melting temperatures are correct for primers being used. Use a 2-step PCR if necessary.
		No ethidium bromide added	If ladder and PCR band are not visible, enough ethidium bromide has not been added, soak gel in EtBr or rerun ensuring appropriate amount of EtBr has been added.
	Bands are incorrect size	Primers bind to wrong spot	Recheck primer design, if they are correct, make them longer to ensure higher specificity
		Template is incorrect	Ensure that template is being added in the correct concentration, and the amplicon being used for PCR exists in the strain.
	Too many bands	Primers bind at more than one location	Increase the length of the primers to increase specificity and reduce off-target binding
		(Yeast) Strain is diploid	When checking for genomic integration, and a wild-type band and recombinant band exist from the same sample, strain is diploid, use colonies that are not diploid.
	Bands very faint	PCR is not efficient	Increase number of cycles, or run the reaction in replicates and pool before imaging and purification
PCR Purification	Concentration is low	PCR is not efficient	Run more cycles, or run PCR in replicates or with higher volume
		Gel extraction not efficient	If only one band expected, run 5 ul of sample on gel for diagnostic purposes and purify the rest over a column. Alternately, add 10ul of PBS to ADB and allow gel to melt fully before putting on column.

		Columns are faulty	Test columns with PB Buffer and allowing to flow through, use a centrifuge and rerun the flow through after adding ADB + sample
		Elute is too dilute	Elute in lower volume (10-15ul)
	Ratios are off	Product is not clean	Wash properly with PB and PE (500ul each, wash twice).
Assembling the Part Plasmid (after transformation)	All colonies are green	Too much starter plasmid	Use correct dilution of pYTK001 such that it 1ul of a 10 nM stock is added.
		Insert ends are incorrect	Assemble the plasmid with appropriate ends using a software like NEB Golden Gate tool or other to ensure that the ends are correct, and the plasmid can be assembled.
		Not enough enzyme	Ensure that the correct enzymes have been added in the correct concentrations.
		T4 ligase buffer is inactive	Replace T4 ligase buffer if it has been through several freeze-thaw cycles or has cooled to room temperature.
	No colonies on the plate	Wrong antibiotic used	Check that the plasmid antibiotic resistance marker, and the plate that the transformation is plated on matches. Sequence resistance part plasmid to confirm.
	No colonies on the plate	Ligase or ligase buffer may be inactive	Use a fresh tube of the T4 ligase and the T4 ligase buffer and repeat assembly.
Bacterial Cell Miniprep	No cells grown	Wrong antibiotic used	If cultures are not saturated after 16-18 hours at 37^{o}C with shaking, check that the correct antibiotic was used in the 2YT media, and the concentration is correct.
	Cells grown in control well	Media is contaminated	If media is contaminated, restart cultures from colonies and toss out the contaminated media and the cultures growing with it
	Low concentration of elute	Cells did not grow properly	If 16-18 hours of growth did not happen, allow cells to grow longer before miniprepping. If cells grew for longer and did not saturate, investigate for contamination.

		Lysis buffer and neutralization buffer not properly mixed	If cells grew properly, ensure proper mixing after adding 7x lysis buffer. Ensure that the neutralization was cold, replace and redo miniprep. Mix properly after adding each buffer.
	Ratio of A260/280 very high	Prep is unclean	Make sure to wash with at least 500ul of PB and PE twice ensuring that the liquid has fully drained before adding more to ensure a clean prep.
	Plasmid bands are wrong size	Wrong ladder used	If a supercoiled ladder is used with digested plasmid, or vice versa, the appropriate band size cannot be detected.
		Plasmid not properly assembled	The backbone may have re-ligated. Check more colonies and check ends to ensure no repeated overhangs and no regions for recombination.
Building transcriptional units (after transformation)	No colonies	Wrong antibiotic used	Recheck antibiotic being used for plating. Start a culture of the backbone that is being used in the media with antibiotic to confirm it grows.
		Ligase and/or Ligase Buffer is inactive	Use a fresh tube of ligase and ligase buffer and redo assembly
	All colonies are green	Backbone not fully digested	Ensure correct concentration of backbone is added. Add more enzyme if problem persists. If making libraries, digest and purify backbone before adding to golden gate reaction. Digest backbone with 1ul of an enzyme that has a restriction site in the original plasmid, but not the final one (eg. In the GFP dropout).
		Parts not properly assembled	Ensure that the correct parts are being used. Assemble the whole plasmid virtually using the NEB online tool to ensure that parts fit together.
		Wrong restriction enzyme used	Confirm that the correct restriction enzyme is being used and that the site is available
	Plasmids are the wrong size	Wrong ladder used	Using a supercoiled ladder for digested plasmid and regular ladder for undigested plasmid can lead to incorrect determination of the size of the plasmid. Rerun gel with correct ladder.

	Plasmids are the wrong size	Plasmid not properly assembled	Confirm using a virtual tool that the parts fit together and that the correct parts are being used. Confirm that there is no likelihood of recombination. If recombination is possible, use a different method of cloning.
Yeast transformations	No colonies	Plated on wrong plates	Confirm the genotype of the strain and the dropout media it can grow in. Ensure that all the required nutrients are added, and the transformation has been given enough time to produce antibiotic resistance if being plated on antibiotic.
		Cells need longer to grow	Genomic integrations and p1 integrations typically take longer for colonies to appear. However, the more transformed and sicker the strain, the longer for colonies to appear. It is not unusual to get colonies in strains where growth rate is severely affected after 10 days. Ensure that cells are given enough time to recover if plating on antibiotic plates.
		Concentration of reagents incorrect	While the yeast transformation is robust, not adding one of the ingredients or adding the incorrect concentration can lead to no colonies. Check reagents before repeating.
		Plasmid for integration not digested	For genomic or p1 integrations, a linear piece of DNA is necessary, not linearizing DNA can lead to no colonies on plate after transformation.
	No colonies or few colonies	Cells not in log phase	Redo transformation, allowing the back dilution to be at least two doublings before transformation. This can be checked using OD600 measurements after dilution and before starting transformation. Wait at least 4 hours between those two times.
	Small colonies in the background	Plasmids not properly digested	Plasmids when not fully digested can lead to small background colonies, especially seen when using a auxotrophic marker. Streak to singles on plates with same media. The colonies that grow can be checked further.

		Incorrect concentration of ingredients in selection media	This can occur when concentration of antibiotics or canavanine is low enough for incorrect transformants to survive. Check concentrations and restreak before next steps.
	Lawn instead of single colonies	Plated on wrong media	Confirm that media is selective enough to only allow correct transformants to grow.
		Marker previously used	Check that the marker for selection has not been used by streaking the parent strain on a plate with the same media. If colonies appear, design experiments to use a different marker
		Contamination of parent strain	Streak the parent onto a plate with the same media. If few colonies grow, parent strain is likely contaminated. Streak the parent to a plate that should allow only the parent to grow. Alternatively, streak parent strain on plate that allows both parent and contaminant to grow. Pick colonies into 10ul of 0.9% NaCl, and spot half that volume on a plate where only parent will grow and one plate where only contaminant will grow. Pick colonies of the parent that do not grow on the plate allowing growth of contaminants.
		Contamination of plate	Old plates when kept at room temperature may get contaminated. Always have a control with water to confirm that plates and plating media are correct.
		Too many cells plated	Streak to singles on a plate with the same media and then pick single colonies. For a CEN/ARS transformation or another known high efficiency transformation, plate a smaller volume or plate entire transformation across several plates.

	Colonies growing on control plates	Contamination of strain	Streak the parent onto a plate with the same media. If few colonies grow, parent strain is likely contaminated. Streak the parent to a plate that should allow only the parent to grow. Alternatively, streak parent strain on plate that allows both parent and contaminant to grow. Pick colonies into 10ul of 0.9% NaCl, and spot half that volume on a plate where only parent will grow and one plate where only contaminant will grow. Pick colonies of the parent that do not grow on the plate allowing growth of contaminants.
		Plates and media are incorrect	Confirm that plates are selecting for the strains after transformation, there should be colonies on the control plates.
		Antibiotic concentrations incorrect	Confirm antibiotic concentrations as well as concentration of canavanine or other similar chemicals.
Protoplast fusion	No colonies on any plates	Plating media too hot or incorrect	Confirm that plating madia was prepared correctly to select for the correct markers in the final strain. Plating media that is too hot can kill cells, make sure it is at a reasonable temperature before plating
	Colonies on controls	Selection is not stringent enough	Restreak from all plates on more stringent plates if available, if not, pick several colonies and confirm by long miniprep. May need to screen a large number of colonies.
		Strains don't have proper knockout	Confirm that the controls cannot grow in the plating media. Kncokout of different kinds may allow growth in other less specific media.
Yeast Long Miniprep (Gel imaging)	No bands	Ethidium Bromide not added	Confirm that Ethidium bromide was added to the gel or TAE before runnung, if not, add stain to TAE and run for 15-20 minutes.
	Only genomic DNA visible	p1 and p2 not present in strain	Confirm that the strain should contain p1 and p2 plasmids
		Bands too faint to be visible	Passage the clones a few times in selective media so that copy number has the opportunity to balance out before redoing miniprep.

	Gel is smeary	Ethanol not fully evaporated	Ensure that the pellets are fully dry before resuspending, dry at RT for 10 minutes, and 65^{o}C for 20 minutes.
		Check pH of reagents	Confirm that the pH of the reagents being used is correct, before repeating
		Proteinase K or RNAse reaction incomplete	When redoing the miniprep test by allowing cells to be in Proteinase K and RNAse for 20-50% longer to get rid of the smearing.
	p2 band not visible	Too much genomic DNA	Especially for By4741 strains, where a lot of genetic and mitochondrial DNA exists, p2 may not be visible, but if p1 is detected, p2 must exist. Starting with fewer cells can help see the p2 band more clearly.
	Wrong p1 size observed	Recombination may occur	Confirm that there are no identical sites that can allow recombination leading to smaller plasmid. Do an outside-outside PCR using ITR primers to amplify and sequence.
	Bands very faint	Not enough cells	Scale up volume or start with more cells and redo miniprep.
Growth assay	No growth observed	Check media	Confirm that appropriate media is grown
		Cells not healthy	Passage cells once to ensure they are healthy before starting growth curves.
	OD v Time graphs look choppy	Strains grow slowly	This may not be the ideal assay if the strains have very long doubling times.
Spot assay	No growth observed	Check media	Confirm that multiple plates are being used with a variety of selections including one plate that is YPD where all spots should grow.

6 p2 Project

6.1 Introduction

While OrthoRep has the potential to be an ideal system for evolving proteins, the current limitations prevent evolution of all proteins since the system is limited by polymerases that have a transition bias, cytoplasmic promoters that have low expression and inability to tune copy number in individual cells to range from low to high. The ability to tune each of these components individually has advantages, but together they will make OrthoRep an even more powerful system with fewer limitations when evolving proteins of choice.

We know that the ORFs encoded on p2 are overlapping and have low expression strengths based on previous data. We also know that over expression of certain components (RNA polymerase) leads to instability in the system leading to loss of p1 [unpublished data]. We hypothesize, that the components expressed off p2 must be in specific ratios to allow the replication and maintenance of this system. If the ratios are out of balance, it can lead to instability leading to loss of p1 or both p1 and p2 plasmids. We think that by refactoring the individual components into the genome, we are will be able to increase expression of the individual components, at the same ratio to maintain stability but also increase expression. We believe that the tuning of expression will also let us tune copy number and will reduce the two plasmid system to a one plasmid system that will be easier to transfer to other host organisms. Freeing up the p2 plasmids of essential ORFs will also let us have another plasmid that can encode genes that can be evolved with an error-prone polymerase specific to p2 enabling evolution of multiple genes and potentially pathways at once.

6.1.1 pGKL1 and pGKL2

OrthoRep uses the system of two linear plasmids, pGKL1 and pGKL2, from yeast for continuous directed evolution of enzymes. The presence of two novel linear plasmids,pGKL1 and pGKL2, in the petite yeast Kluyveromyces lactis was first reported by Gunge et.al., in 1981 though their localization within the cell was still unknown (Gunge, Tamaru, Ozawa, & Sakaguchi, 1981). These plasmids were subsequently discovered to be in multiple copies and were found to be localized to the cytoplasm by using DAPI, a DNA-staining dye in petite yeast that did not contain functional mitochondria (Fujimura, Yamada, Hishinuma, & Gunge, 1987) (Kamper et al., 1991) (Gunge,

Fukuda, Takahashi, & Meinhardt, 1995). A killer phenotype and subsequent immunity was observed in cells that contained both these linear plasmids (Gunge et al., 1981). Gunge and Sakaguchi (1981) showed that pGKL1 and pGKL2 could be transferred from K. lactis to S. cerevisiae by protoplast fuse without losing functionality of the killer plasmid suggesting that these plasmids were not dependent on specific genetic elements found in K. lactis for their propagation (Gunge & Sakaguchi, 1981). Further data also shows that these elements are inherited in a non-Mendelian manner (Wesolowski et al., 1982).[Fig. 4]

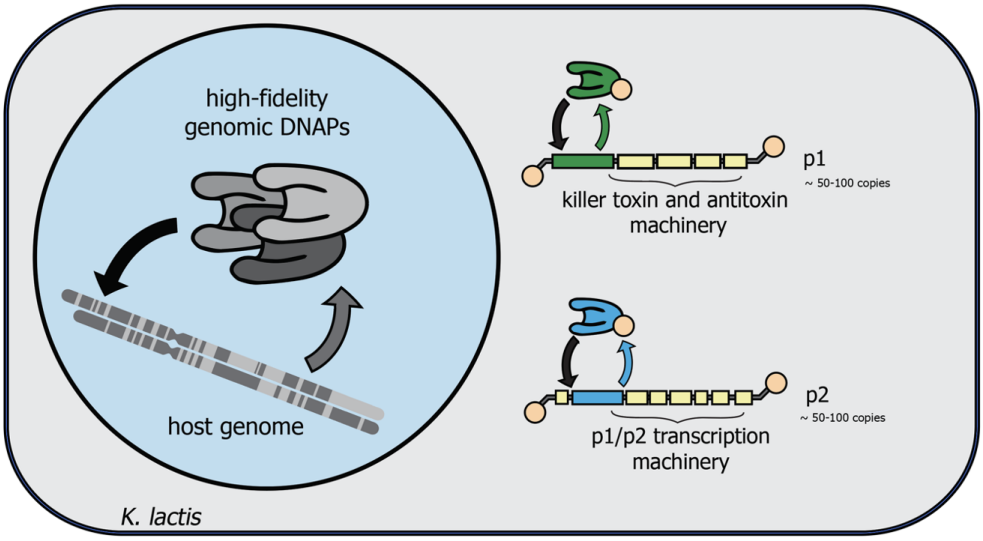


Figure 5: K. lactis showing presence of two cytoplasmic linear plasmids

The cytoplasmic linear plasmids pGKL1 and pGKL2 were found to be 50 - 100 copies per cell and were found to have different size (Gunge, Murata, & Sakaguchi, 1982; Gunge, 1986). The smaller plasmid is the pGKL1 plasmid and is about 8.9 kb in length and electron microscopy analysis showed the linear DNA to be 2.7 um in length (Gunge, 1986). Similarly, the longer plasmid pGKL2 plasmid was about 13.4 kb in length and electron microscopy analysis showed the linear DNA to be 4.2 um in length (Gunge, 1986; Fujimura, Hishinuma, & Gunge, 1987). Since the plasmids exist in higher copy, and are different in size, agarose gel electrophoresis can be used to detect them as well as understand their size. The buoyant densities were found to be different and lower than those of nuclear (1.699g/cm3) and mitochondrial DNA (1.692g/cm3) using a neutral Caesium Chloride solution at 1.687g/cm3 (Gunge et al., 1981). These plasmids also contain high A+T content at over 74% as compared to the nuclear genome which is approximately 50% A+T

content (Gunge, 1986).

Analysis of the sequences of the two plasmids do not show presence of a TATA-motif that is typical of a promoter in the nucleus. The ORFs from either of the plasmids could not expressed when a nuclear promoter was used in a conventional yeast vector and a nuclear gene could not be expressed when it was cloned onto the linear plasmid with the native promoter (Gunge et al., 1995). This suggests a replication mechanism that is different that in the nucleus. There are structural similarities between the linear plasmids and those of adenoviral DNA suggesting that replication follow a similar process called protein primed replication using terminal proteins at either end of the linear plasmids (Fujimura, Hishinuma, & Gunge, 1987; Romanos & Boyd, 1988). Protein primed replication uses an autonomously replicating sequence (ARS) as the replication initiation site allowing the polymerase to attach and add dNTPs using the proteins as a primer for elongation (McNeel & Tamanoi, 1991).

6.1.2 Genomics of pGKL1 and pGKL2

The smaller 8.9 kb linear plasmid pGKL1, here p1, encodes 4 ORFs of known function and occupy 94% of the genome (Gunge, 1986; Jeske, Tiggeman, & Meinhardt, 2006). These ORFs contain the DNA polymerase that is responsible for replicating p1 and the remaining three ORFs encode the toxin and antitoxin subunits that characterize the killer phenotype (Kikuchi, Hirai, & Hishinuma, 1984). A TAA codon is the termination signal for each ORF (Gunge, 1986). It has been identified that killer character and immunity is present on p1 (Wesolowski et al., 1982). ORF 2 encodes the alpha and beta subunits and ORF4 encodes the gamma subunit of the toxin while ORF3 encodes the protein required for immunity against the killer phenotype (Tokunaga, Wada, & Hishinuma, 1987). The killer and the immunity ORFs are non essential to the survival of the plasmids . The DNA polymerase that is encoded in ORF1 is essential, since that is the only element required for the replication of p1 located on p1 (Ravikumar et al., 2014). This system was used to invent OrthoRep as previously described.

The longer 13.4 kb linear plasmid, pGKL2, here p2, is known to encode 10 ORFs but the functions of some of them are still unknown (Fujimura, Hishinuma, & Gunge, 1987). The ORFs comprise 97% of the plasmid and each of the ORFs are at leads 95 amino acids in length (Stark, Boyd, Mileham, & Romanos, 1990). Experiments that checked to see if p1 and p2 could survive

independently of each other showed that p1 could not survive by itself, but p2 could suggesting that the ORFs on p2 encode for proteins essential for the maintenance of both p1 and p2 (Tommasino, Ricci, & Galeotti, 1988). The ORFs on p2 encode for a plasmid specific DNA polymerase, a capping enzyme, a helicase, a single-stranded DNA binding protein, two subunits forming the RNA polymerase and a terminal recognition factor (Jeske et al., 2006; Larsen, Gunge, & Meinhardt, 1998; McNeel & Tamanoi, 1991). Each of the p1 and p2 plasmids contain inverted terminal repeats (ITRs) that differ from each other both in terms of length and sequence (Kikuchi et al., 1984).

Ten ORFs have been discovered on p2 so far. Of these, ORF1 has unknown function and was shown to be not essential for replication and maintenance of p1 and p2 (Jeske et al., 2006). ORF2 encodes the TP-DNAP2, which is the DNA polymerase required for replicating p2 and navigating the terminal proteins (Jeske et al., 2006). This polymerase is specific to p2 and cannot be used to replicate p1. The protein encoded in ORF3 is cytoplasmic capping enzyme that is responsible for capping mRNA transcripts produced in the cytoplasm (Jeske et al., 2006; Larsen et al., 1998). The sequence for ORF3 shows homology with capping enzymes that are known to have a viral source suggesting that p1 and p2 are viral in origin (Larsen et al., 1998). The protein encoded in ORF4 is the helicase and is mainly required for unwinding of DNA during replication of p1 and p2 (Jeske et al., 2006). It seems likely that ORF5 encodes for the single-stranded binding protein that is also required for binding to intermediates during replication (R & Meacock, 1995). ORF6 and ORF7 encode two subunits of the RNA polymerase that synthesize mRNA from the ORFs encoded on both p1 and p2 (Jeske et al., 2006; Schaffrath, Meinhardt, & Meacock, 1997). Experimental data shows that the protein encoded on ORF9 is essential, but its function is yet unknown (Jeske et al., 2006). ORF10 encodes a Terminal Recognition Factor (TRF) that is important for initiation of replication and is also responsible for recognizing terminal sequences of both p1 and p2 (McNeel & Tamanoi, 1991; Tommasino et al., 1988). While 10 ORFs have been identified on p2, and some information is known about cytoplasmic promoters, there is still unknown information about the p1 and p2 system of cytoplasmic linear plasmids.

The first version of OrthoRep included error-prone polymerases 611 and 633 that has a transition bias towards creating mutations (Ravikumar et al., 2018). This made it so that very few transversion mutations would occur making a lot of selections harder if not impossible. To improve OrthoRep, there is a need for polymerases that have higher error rates and are able to make equal amounts of

transition and transversion mutations. Currently the polymerases that we have are able to replicate few copies of p1 during each cell cycle, but the most error-prone polymerase we have so far is unable to sustain the p1-p2 system in the absence of wild-type polymerase. In addition, there are very few cytoplasmic promoters that can express genes highly enough off cytoplasmic plasmids if they are expressed at medium or high levels in the nucleus. We have employed certain tricks like using PolyA tails to increase expression, but that is only possible to a certain limit (Zhong et al., 2018). We believe that an attempt at refactoring genes off p2 into the genome will have several advantages. Reducing one plasmid and decoupling the plasmids from each other will ensure that the system is simpler and can be transferred to other host systems. By fine tuning the individual components from p2, the expression of the genes required for replication and translation, we hope that we can adjust the copy number of p1 as desired.

6.1.3 Project Blueprint

We considered two broad strategies for this project, one was to knock out one ORF at a time, and the other was to create a library of promoters for each of the ORFs to make a giant CEN/ARS plasmid with all possible combination for each ORF and then knock out the entire p2 plasmid all at once. The first approach while slower would be more methodical and would allow us to gather the most information in this journey as compared to the second approach where it could work, but given the unknown it was unlikely to work in the first shot. We were concerned about the unknowns starting with the number of ORFs that exist on p2 and their corresponding promoter strengths. If the second approach did not work, we would have very little information to make the next decisions. So we decided to go with the more conservative approach that would allow us to potentially add the second approach as more information became available to us over the course of the experiments that would be performed.

6.2 Identifying and Annotating all ORFs

6.2.1 Introduction

The first step to refactoring ORFs from p2 to the genome was to identify the number of ORFs on p2 and their start codons and the promoters that lead to their expression. Some information

is available about the cytoplasmic promoters from literature. The motif TA^nTNTGA, where the T is followed by a number of As followed by TNTGA, where N is any nucleotide, is found to be the start site of the cytoplasmic promoter (Gunge et al., 1995). The distance from this upstream conserved sequence (UCS) to the start codon should be the promoter of the following ORF. We used this information to proceed with identifying all possible ORFs that may be located on the p2 plasmid. Additionally, the promoter element also consists of a transcription initiation region (INR) that enable binding of the two subunits of the RNA polymerase and the capping enzyme that is responsible for capping the mRNA transcripts produced .

6.2.2 Annotating all ORFs

We looked through the entire 13.4 kB of the p2 plasmid and looked for the upstream conserved sequence, in particular the TNTGA subsection where each of the A, T, G, and C nucleotides were substituted in place of the N. This search gave us 53 results for the TATGA sequence, 21 results for the TCTGA sequence, 44 results for the TTTGA sequence and 11 results for the TGTGA sequence. We next looked for where these upstream conserved sequences were unlikely to give us a possible ORF. These locations included the UCSs on the bottom strand facing towards the end of the plasmid and on the top strand facing the end of the plasmid as well. Next, we looked for possible start codons and stop codon to determine the coding sequences that lie on p2 and exceeded 100 bp which still gave us over 100 possible ORFs. We narrowed down the number of possibly ORFs by looking for the TA^n part of the upstream conserved sequence and narrowed down further the number of possible ORFs. Ultimately, we were left with 12 ORFs with a few having multiple start sites. We then compared the ORFs we had annotated to the ones that had been reported in the literature. Stark, Boyd, Mileham and Romanos (1990) and Jeske, Tiggeman and Meinhardt (2006) have extensively characterized the promoter for the ORFs that had been identified them. When comparing our annotations to theirs, our findings overlap for the most part. The key differences we noticed were the identification of 12 ORFs as compared to the 10 identified by Stark et.al., and the 11 ORFs identified by Jeske et.al. It is important to note that while we identified 12 ORFs from the sequence of p2, we are unsure still of whether all these ORFs exist. [Fig. 5]

Another key component to identifying the ORFs was to also determine their function. The literature said that some of the functions had been identified previously but that some functions

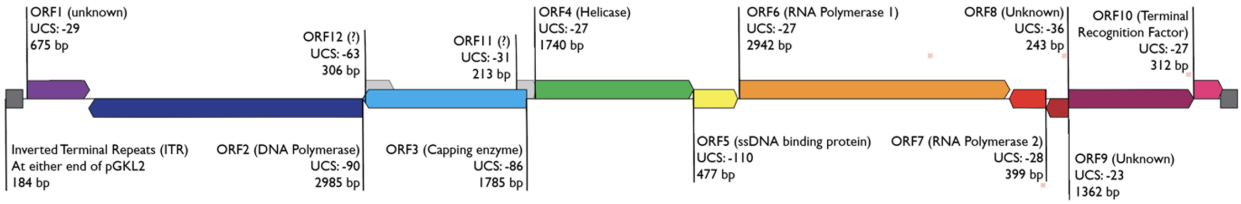


Figure 6: Linear map of p2 showing all possible ORFs on p2

ORF No.	Function	Size (bp)	UCS Location	UCS
ORF1	Unknown	675	-29	TAATATGA
ORF2	DNA Polymerase	2985	-90	gcATATGA
ORF3	Capping Enzyme	1785	-86	TAATCTGA
ORF4	Helicase	1740	-27	TATCTGA
ORF5	ssDNA Binding Protein	477	-110	TAATGTGA
ORF6	RNA Polymerase I	2942	-27	TATATGA
ORF7	RNA Polymerase II	399	-28	TATGTGA
ORF8	Unknown	243	-36	TATTTGA
ORF9	Unknown	1362	-23	TATTTGA
ORF10	Terminal Recognition Factor	312	-27	TATATGA
ORF11	Unknown	213	-31	AgATTTGA
ORF12	Unknown	306	-63	TATTTGA

Table 6: Annotations of ORFs on p2, their UCS and the location

were still unknown. We used BLAST to check if any structures had regions of similarity with previously known structures. However, this did not yield any new information. Similarly, we used alpha-fold to determine structure of these individual ORFs, but while alpha-fold was able to give us the best guess for structures, it was unable to give us any structures that had homology with previously described structures. We also used alpha-fold to check the differences between predicted structure of the two ORFs that had multiple start sites. The only difference was the presence of an extra alpha-helix between the two methionine coding amino acids. The details of the identified ORFs, their Upstream Conserved sequences and the lengths of the predicted ORFs are given in the table below.

6.3 Building KO and Complementation Plasmids

6.3.1 Introduction

To get started with knocking ORFs off of p2, we needed a base strain that would contain a way to track if p1 was still around while p2 was being modified without selecting for it. We decided to use a fluorophore, mKate to detect presence of p1. We integrated mKate with the pGA promoter and the LEU2 marker on p1 and preserved the p1 polymerase since we will not be integrating that in the genome. The LEU2 marker on p1 allows us to select for it when necessary and the wild-type polymerase at its original location means that there are no p1 elements integrated elsewhere. [Fig. 6] However, integrating over the toxin-antitoxin elements of p1 means that, in the cytoplasm there exist two kinds of p1 plasmids. Selecting for the newly recombinant p1 would ideally allow the wild-type p1 to be lost, but the timescale of that can't be determined definitively.

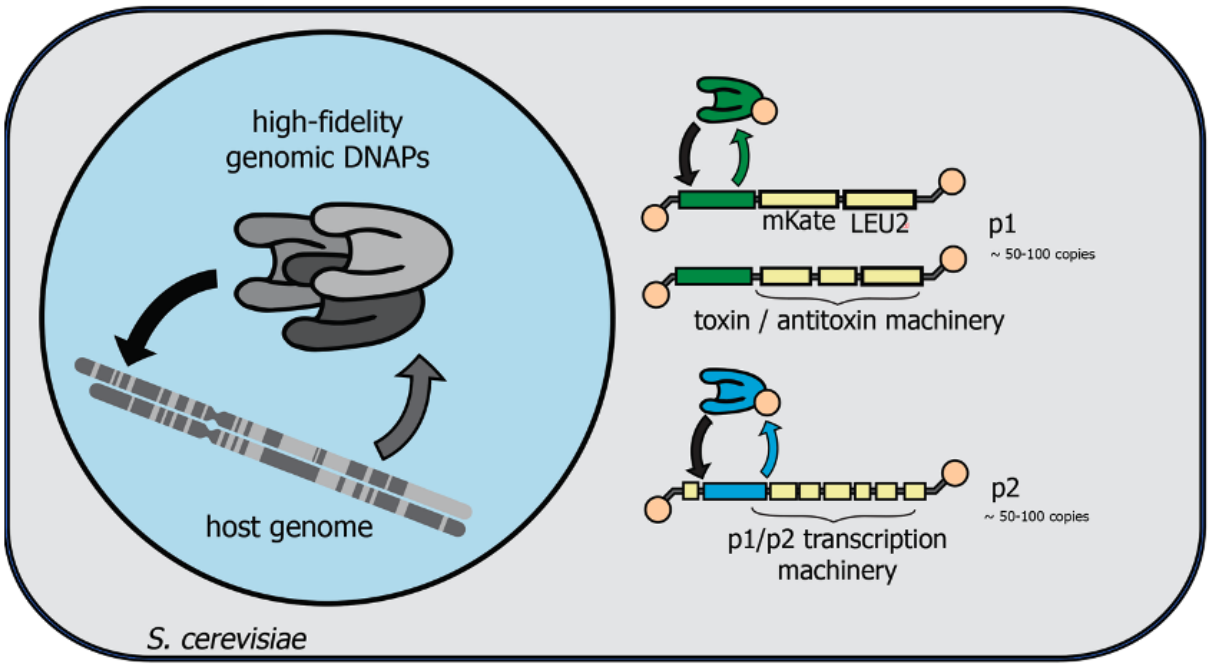


Figure 7: Base strain created containing mKate and LEU2 on p1

6.3.2 Building KO plasmids

Once all the p2 plasmids have been annotated and identified it was time to get rid of the ORFs one at a time. To do this, we created p2 integration plasmids that had the coding sequence of one

ORF replaced by a cytoplasmic promoter that drives expression of a URA3 gene off p2. Upstream and downstream flanks that were 500 bp in length were used while ensuring that promoters and coding sequences of other ORFs located upstream or downstream were not disturbed. These three individual parts were created using golden gate overhangs such that they could be assembled into an E. coli plasmid using the ColE1 origin of replication and a Kanamycin resistance marker. We also added ScaI sites at the ends of the the flanks to linearize the plasmid before integration onto p2.[Fig. 7A]

The parts were put in a golden gate reaction and then transformed into Top10 cells. Colonies that were not green were picked into 2YT + Kan media and grown overnight until saturated. The plasmids were then miniprepped, digested with ScaI, and transformed into yeast strains that contained mKate on p1 and a wild-type p2. The tranformation were then plated on SC - LU plates to select for the recombinant p1 containing mKate and LEU2 and the newly recombinant p2 that contained URA3 in place of an ORF. This transformation led to the creation of two sets of p2 that coexist in the cytoplasm, one of which is the wild type p2 and the other being the newly recombinant one. The wild type p2 will not disappear in this case since the recombinant p2 is missing an ORF, however, passaging will adjust copy number to stably maintain both of these plasmids. The presence of these recombinant p2 plasmids can be confirmed by long miniprep and then running them on an agarose gel for 2 hours. This gel will allow separation of the linear plasmids making visualization of all the linear plasmids possible. However, since some of the ORFs being replaced by URA3 are about the same size, it is not possible to differentiate between those two species of p2 by gel alone. To confirm integration, we used PCR where one primer binds in the integration and the other primer binds wild-type p2 and sequenced the results. This can also be done by using primers that bind outside the integration spot and then sequencing subsequent bands. We integrated and knocked out each of the ORFs in individual strains to create several strains of yeast. We knocked out all the ORFs from 2-10 since ORF1 is known to be unessential and ORF11 is inaccessible to be deleted due to its location between ORFs 3 and 4 and their cytoplasmic promoter. Similarly, ORF12 is fully contained within ORF3 and cannot be knocked out. ORFs 11 and 12 will be dealt with at the end.

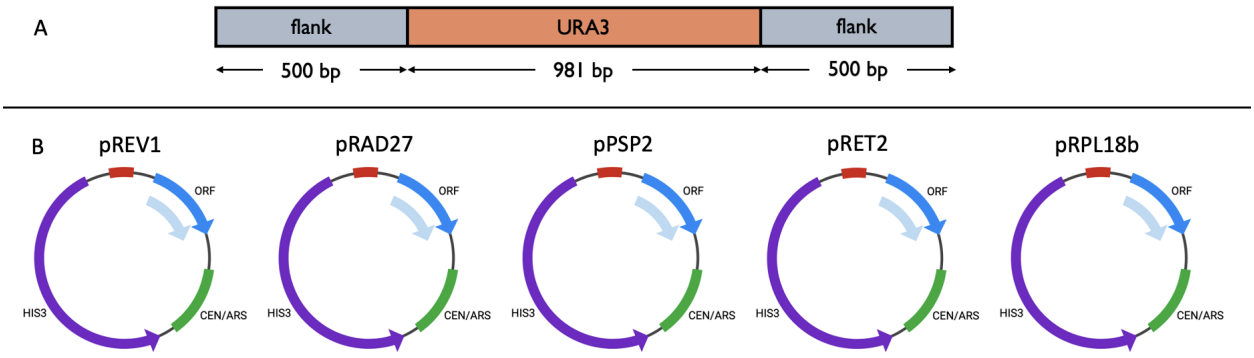


Figure 8: ORF KO and complementation plasmids

[A] The KO plasmid consists of 500 bp flanks and a URA3 cassette with a cytoplasmic promoter.

6.3.3 Building complementation plasmids

Knocking out individual ORFs from p2 leads to formation of two types of p2 plasmids in each cell. One set of these plasmids are the wild-type p2 plasmids that have no change at all and the second set of plasmids is where the knockout cassette has been integrated and one of the ORFs has been replaced by URA3. To determine if the knocked out ORF can be complemented in the nucleus with nuclear promoters, we need to make CEN/ARS plasmids with a variety of nuclear promoters for each ORF. The choosing of these nuclear promoters does not need to be blind because some information is available from the promoter strength of each ORF. This experiment is discussed later on. Using the yeast toolkit, we identified that the range of expression between pREV1 and pRPL18b is the relevant range for expression of ORFs located on p1. Initially we made CEN/ARS plasmids with containing the pREV1 and pRPL18b promoters and then added in three more promoters, pRET2, pRAD27 and pPSP2 with varying strength between low and medium (Lee et al., 2015). A marker that was not present for selecting the recombinant p1 (LEU2) or the recombinant p2(URA3) was needed and so HIS3 was used for this purpose.[Fig. 7B]

The CEN/ARS plasmid were made in a similar way as described above for the knockout plasmids. The transcriptional units were created in backbone with KanR and then a two part Golden Gate Assembly reaction was performed with the HIS3 backbone containing an AmpR marker as well as a CEN/ARS part and a ColE1 was used as the bacterial origin of replication. This plasmid was then transformed into Top10 cells and non green colonies were picked. After miniprep, the plasmid was run on a gel and the promoter was sequenced to ensure no contamination and the

correct promoter was being used. The CEN/ARS plasmids containing an ORF were transformed into the respective appropriate yeast strains containing the respective ORFs knocked out. Once the strains have been prepared with the ORF knocked out and then complemented on the CEN/ARS plasmid using a variety of promoters, we can get rid of the wild-type p2 to check if the the complementation was successful. This is discussed in more detail in a later section. Thus, each individual ORF knockout strain after being transformed with a complementation plasmid made somewhere between 2 and 10 strains depending on whether there were one or two possible start codon options.

6.4 Cleaving wild-type p2 using Cas9

6.4.1 Introduction

Our strain at this stage consists of two species of p1, the wild-type and the recombinant containing mKate and LEU2 as well as two species of p2, the wild-type and recombinant containing the ORF knockout and replaced with URA3. Passaging while selecting for p2 using URA3 in absence of the complementation will not get rid of the wild-type p2 plasmid as a component necessary for replication and maintenance is encoded only on the the wild-type p2 plasmid. Even with perfect complementation, we now that passaging will not be able to cure out wild-type p2. An external way to get rid of it must be introduced into the cytoplasm and a Cas9 plasmid is an ideal solution. A Cas9 plasmid with no NLS sequence can be used to selectively target the wild-type p2 in the cytoplasm and all its copies (Arzumanyan, Gabriel, Ravikumar, Javanpour, & Liu, 2018). We can make these plasmids specific for each individual ORF by finding a target within the ORF that has been knocked out from the wild-type so that the only copy that exists in the cytoplasm must be on wild-type p2 (Arzumanyan et al., 2018). Thus, we need an efficient sgRNA for each ORF from 3 to 10.

6.4.2 Building the Assay

We needed to build an assay that would tell us how efficient each of the single guide RNAs (sgRNAs) was at cleaving the high copy-number wild-type p2 in the cytoplasm. The first assay we tried used 2 sgRNAs that were specific to each ORF. We transformed our yeast cells (p2Y1) that contained pGA > mKate : LEU2 on p1 and had only wild-type p2 with each of these two sgRNAs to create

two individual strains. We plated the transformations on SC - L plates. If colonies formed on SC - L plates, it was likely that those colonies had not completely lost the wild-type p2 present in the cytoplasm. Since p2 is required for replication and maintenance of p1, loss of p2 would result in loss of p1 which encoded LEU2. Our initial assay would assume that the higher the number of colonies, the less efficient the guide while number of cells and amount of plasmid containing Cas9 and sgRNA were equal. We decided to replica plate colonies that grew on the SC-L plate onto another SC-L plate to passage the growing cells and we expected to see fewer and fewer colonies per passage until we have no more colonies growing signaling the complete cleavage and subsequent loss of wild-type p2 and thus p1.

Several colonies grew on SC - L plates so we replica-plated them onto more SC - L plates. However, instead of seeing fewer colonies, we saw more colonies. We believe that cells spread out during the replica-plating in the process of transferring them onto the velvets and then from the velvets back to the plates. This led to the number of colonies increasing with each subsequent passage. This assay does not work, but it is unclear if that's because the guides are inefficient or the assay is unable to determine efficiency of each sgRNA.

The results of the above assay showed that it was unlikely we would get to zero colonies so we decided to use a different strategy. Our new strategy utilized five distinct sgRNAs for each ORF. We used a negative control that contained Cas9 but did not contain a functional sgRNA and would not result in any cleave of wild-type p2. We also used a positive control that had been previously tested to be very efficient which resulted in a near perfect cleavage of wild-type p2 resulting in complete loss of the linear plasmid system. In this assay, we transformed the base strain p2Y1 with each of the five sgRNAs and each of the two controls. We plated one half of each transformation on YPD + G418 plates and the other half of each transformation on SC - L + G418 plates. Once grown up (6 days), we expected that there would be a difference in the number of colonies that would grow on the YPD plates when compared to the SC-L plates. The SC-L plates select for propagation and maintenance of p1, and thus also p2, while YPD is the non-selective media. We also expect that the most number of colonies exist when using the negative control, the plasmid with a functional Cas9 but no guide RNA, and that the ratio of colonies on the YPD plates to the SC-L plates is about equal. On the contrary, for the positive control, with the most efficient guide, we expect there to be ideally, several fold more colonies on the YPD plates when compared the

SC-L plates. We also expected that the guide efficiency would be higher if the ratios more closely matched the positive control and the efficient would be worse if they matched the negative control.

6.4.3 Results and Discussion

Colonies took about 6 days to grow up on both plates. When we counted colonies for the samples on all of the plates we observed that there were no major differences in the number of colonies on any plates. most importantly, the controls did not show a large enough difference in colony count to determine efficiency. We rationalized that if the sgRNAs being tested had high efficiency, then the colonies growing on YPD plates would have lost p2 and subsequently p1 leading to a loss of fluorescence when measured on the flow cytometer. It is possible that colonies that grow on SC-L plates lose some p2 but only colonies that do not fully get rid of p2 and subsequently p1 to a level where enough LEU2 is produced are able to survive. We utilized the ability to detect fluorescence and loss of fluorescence using the flow cytometer and picked 12 colonies into 2 mls of YPD off both YPD + G418 plates and SC-L+G418 plates. After they grew for a day, we added 20 ul of cells to 200 ul of 0.9% NaCl and ran the samples on the flow cytometer. This data was collected and analyzed in Flow Jo and final figures to determine Cas9 efficiency were made in R.

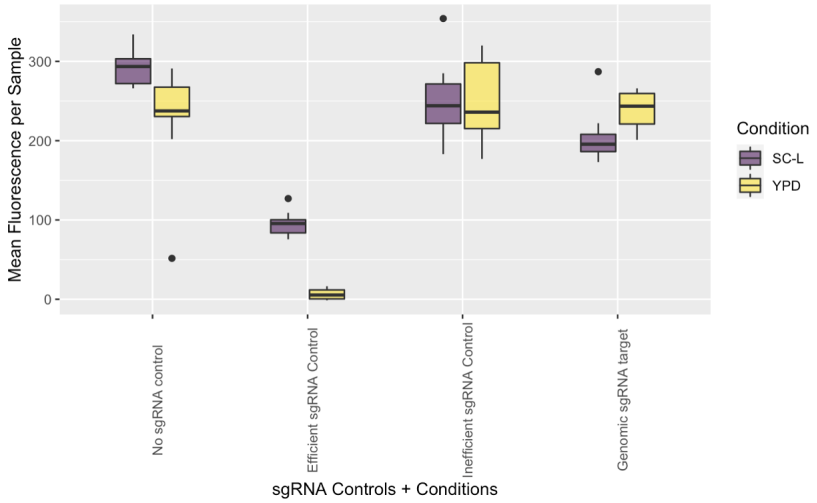


Figure 9: Testing sgRNA efficiency for controls

This assay was adapted from Arzumanyan et.al., (2018) as a way to determine the efficiency of several sgRNA's in parallel in a way that was scalable and would be simple and robust enough to replicate. In that paper, sgRNAs to efficiently cleave at ORF1 and ORF2 on p2 were discovered. We

identified all the location with an NGG PAM within each ORF and used ChopChop to determine the most efficient sgRNA (Labun et al., 2019). We tested the inefficient and the efficient sgRNA for ORF2 as mentioned to confirm that this assay gave the same results [Fig. 8]. The data show that there is a difference in levels of mean fluorescence across media and samples. The strains that do not contain the sgRNA spacer, show high fluorescence in both SC-L media and YPD media. They show noticeably lower fluorescence in YPD media when compared to SC-L media and this can be attributed to a loss of the LEU2 containing p1 plasmid in YPD over several generations of growing in YPD. The efficient sgRNA show lower levels of fluorescence in SC-L media when compared to the no sgRNA control, but higher then the fluorescence in YPD. This difference can be attributed to the repeated cleavage of the wild-type p1 plasmid. However, selection of LEU2 on p1 ensures that only cells that have not completely cured the p2 plasmids can grow. This is not true for the YPD plates, allowing for complete loss of p2 resulting in complete loss of p1 ensuring no fluorescence in those samples. Conversely, if fluorescence is observed in colonies picked off the YPD plates, we can say that the corresponding sgRNA has lower efficiency as seen in the inefficient sgRNA data.

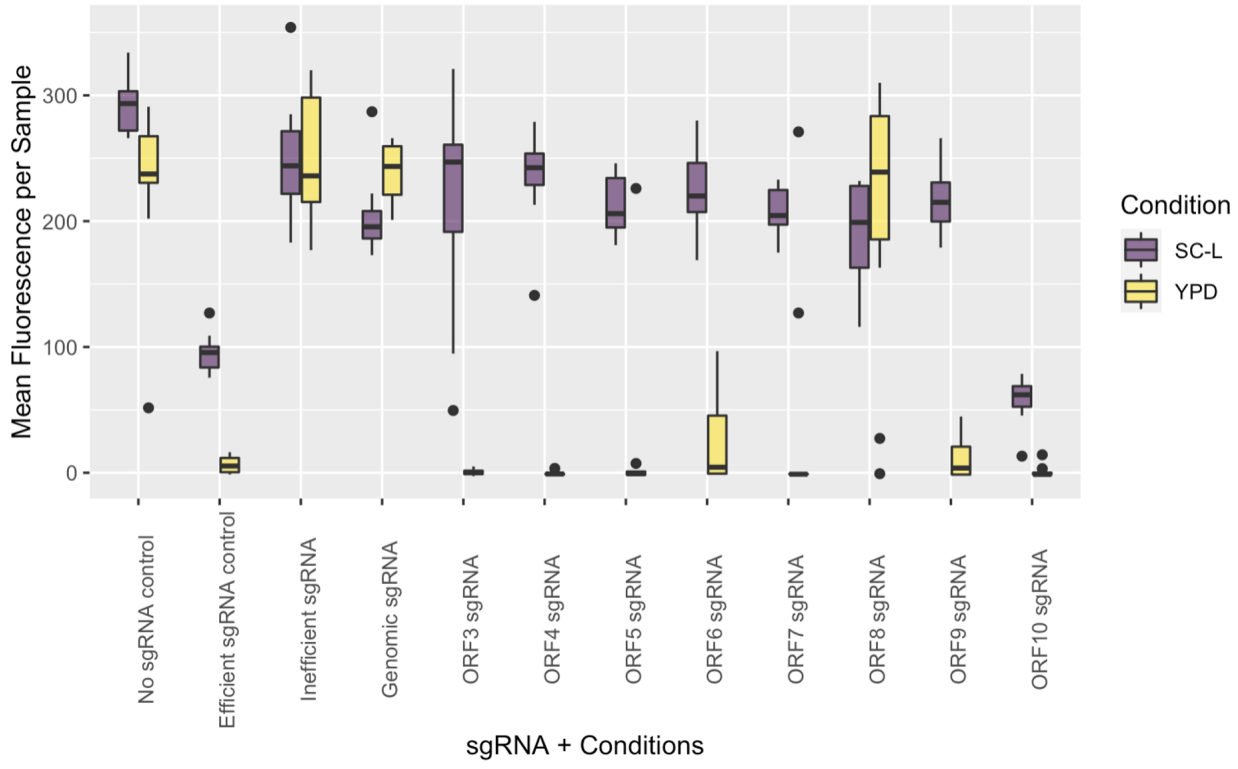


Figure 10: Most efficient sgRNAs for each ORF

We can use the same logic as above to identify the most efficient sgRNAs for each of the ORFs. We used 5 distinct sgRNA for each ORF. However, some transformations did not yield colonies on plates for all the sgRNAs. If an efficient sgRNA was identified from the available data, the experiment was not repeated to include the missing sgRNA since one efficient sgRNA was sufficient for our purposes. The data for ORF3 shows that the sgRNA2 and sgRNA4 are the most efficient as signified by the loss of fluorescence in YPD media when compared to the controls. The data for ORF4 shows sgRNA1 and sgRNA4 are the most efficient, while sgRNA2 and sgRNA5 are able to almost get rid of all fluorescence. The data for ORF5 shows that only sgRNA3 is efficient from the ones tested while sgRNA1 and sgRNA2 are pretty efficient. The data for ORF6 shows that sgRNA1 and sgRNA2 are most efficient closely followed by sgRNA5. The data for ORF7 shows that in the first five sgRNAs tested, there are no sgRNAs that are efficient at cleaving wild-type p2. The results from ChopChop showed that there was a possible twelve spacers located within the wild-type ORF7. We cloned the remaining 7 spacer sequences to check if there are any efficient sgRNAs that may exist in that set.However, this second set of sgRNAs proved to be even more inefficient that the first set. Similarly for ORF8, the first set of five sgRNA did not show as high efficiency as the control. ChopChop gave us a possible 11 spacers located within ORF8 so we cloned more sgRNAs and tested them, but found that none of those were efficient either. The data for ORF9 show that sgRNA2 , sgRNA4 and sgRNA5 are very efficient at cleaving wild-type p2. Lastly, for ORF10, sgRNA4 is the most efficient while sgRNA5 is pretty close.

We picked the most efficient sgRNA for each of the ORFs, along with the controls and performed this assay to ensure reproducibility and to reconfirm the efficiency of the sgRNAs. The data showed lower levels of fluorescence for the colonies picked off SC-L plates across the board and similar levels of fluorescence for colonies picked off the YPD plate. While not ideal, the difference in fluorescence expression was found to be significant enough to continue with the experiment. [Fig. 9]

6.5 Estimating Cytoplasmic Promoter Strength

6.5.1 Introduction

Cytoplasmic promoters on p2 are not isolated are are often overlapping with other ORFs. They have low expression strengths since the plasmids exist in high copy number. All the genes expressed on p2

have different expression levels that vary depending on the function of the ORF they are promoters for. Once we had identified the 12 ORFs on p2, we also identified the promoters preceeding them. However, the length of the promoter and the upstream conserved sequence cannot help determine the strength.

Data from literature shows that when identified properly, the promoters are able to express other genes that may be located on p1 or p2. While they are cytoplasmic promoters, it has been confirmed that p2 promoters can function on p1 and vice versa. Experiments using the promoter of p2ORF5 to drive expression of LEU2 gene showed that LEU2 was expressed from p2 using the p2ORF5 promoter, but in the absence of the promoter, LEU2 was not expressed from the promoter of other genes. This tells us that each ORF must have it's own promoter and that each promoter determines how much of the downstream gene is expressed.

6.5.2 Setting up Experiments

To understand the promoter strength, we needed to look at them individually, in isolation to be able to compare them to each other and to the nuclear expressed promoters as well as to those on p1. To isolate these promoters, we cloned each of the cytoplasmic promoters consisting of the Upstream Coding Sequences (UCS) until the downstream located start codon onto p1. This distance ranged from 27 bp upto 110 bp. Downstream from the promoter, we encoded mScarletI and LEU2 was encoded downstream to select for p1. The nuclear promoters were encoded onto a CEN/ARS plasmid upstream of mScarletI and used a PGK1 terminator. LEU2 with its native promoter was used as a selection marker for the CEN/ARS plasmid. [Fig. 10]

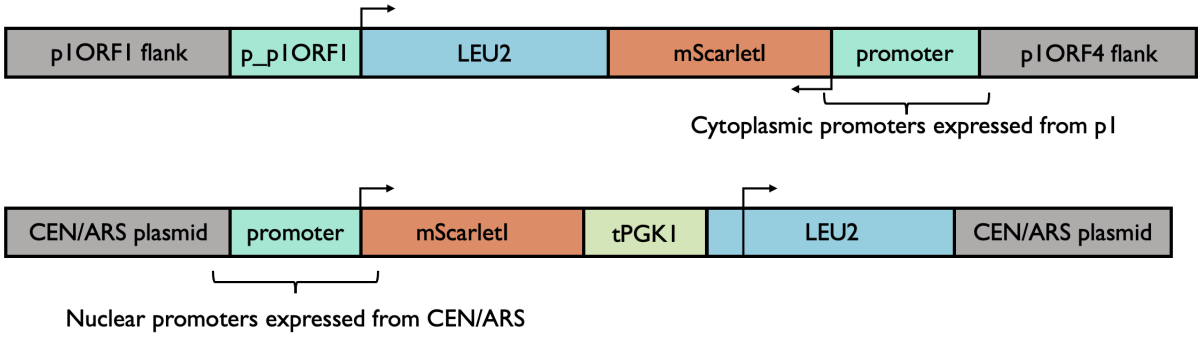


Figure 11: Experimental setup for testing promoter strength

We integrated the cassette containing the cytoplasmic promoters with mScarletI and LEU2 onto wild-type p1 in a strain that also had wild-type p2. and transformed the parent strain containing wild-type p1 and p2 with the CEN/ARS plasmids using the nuclear promoters in the genome. Colonies for the CEN/ARS plasmid appeared in 3 days whereas the p1 integrations appeared in about four days. We picked 12 colonies from each plate into a 96 well block with 0.5 ml of SC-L media. The blocks grew overnight and were imaged the next day using the Attune in an appropriate dilution of 0.9% NaCl. The data was then exported to FlowJo and the geometric means of each population were analyzed in R.

6.5.3 Results and Discussion

The cytoplasmic promoters, of all different lengths had different strengths. The data show that the strenght of the promoter is not determined by the length or by the upstream conserved sequence. However, we observed that the promoter with highest strength was the p2ORF5 promoter, where the ORF5 encodes for the single-stranded binding protein that is essential for DNA replication, repair and recombination. The mScarletI gene expressed using the p2ORF10 promoter had second highest fluorescence, and the ORF10 encodes the terminal recognition factor (TRF). The TRF recognizes the a specific sequence located within the ITR of both p1 and p2 plasmids and helps in unwinding for replication. The two lowest strength promoters - p2ORF1 and p2ORF8 have lowest strength, and the functions of both these ORFs are currently unknown. ORFs 2, 3, 4, 6, and 7 lie in the middle and all of these ORFs have functions related to DNA replication and translation with ORF2 being the DNA polymerase, ORF3 being the capping enzyme, ORF4 being the helicase, and ORFs 6 and 7 together making up the RNA polymerase.[Fig. 11]

This data that gives us an estimate of promoter strength both in comparison to other promoters on p2 as well as when comparing to certain promoters of known expression levels, we can better estimate the ratios of components required to express the gens on p2. Additionally we can use the comparison to the nuclear promoters to estimate what promoters are ideal for complementation when we knock out ORFs and also get rid of wild-type p2. However, while estimating these promoters, it may lead to some instability as we are complementing ORFs one at a time. This instability will likely not be able to be resolved when only one ORF is being complemented at a time, however this should ideally be resolved when knocking out larger chunks off p2 or when

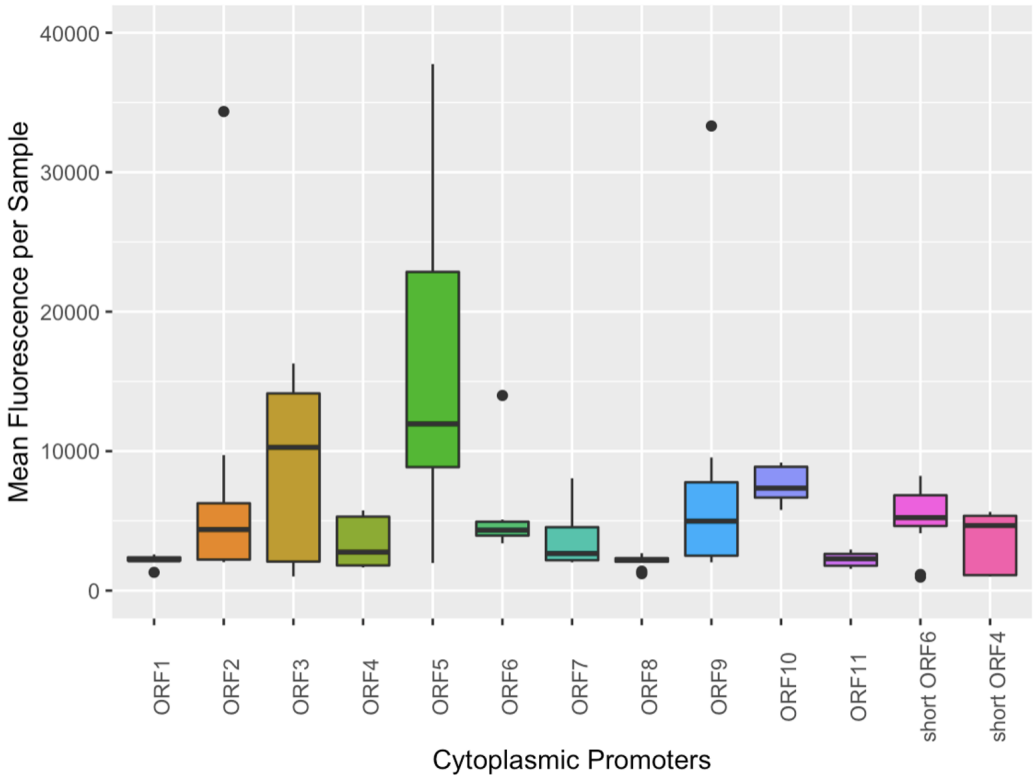


Figure 12: Promoter strength data for cytoplasmic promoters

complementing the entire plasmid.

6.6 Knocking out p2 ORFs individually

6.6.1 Introduction

The final step is to put together all the parts. We use the starting strain with two species of p1, the wild-type and the recombinant consisting of mKate and LEU2 to enable selection by auxotrophy and also by fluorescence. We then integrate the knockout plasmid that replaces an individual ORF with URA3. Once knocked out, the ORF needs to be complemented in the nucleus on a CEN/ARS plasmid. Anywhere between 2 and 10 CEN/ARS plasmid were used to complement each of the ORFs depending on the number of alternate start codon possible. Each of these straisn was then transformed with a 2 micron plasmid that contained a Cas9 with no NLS and the most efficient guide identified to cleave wild-type p2. The real test of survivability was determined after all these transformations were complete.

6.6.2 Expected Results

Once all the transformations are complete, we can expect one of several possible results. We expect that when there is no complementation, and an efficient sgRNA is added, the only time there is mKate expression is when the recombinant p1 is selected for on SC-L plates. When selecting on SC-UH for the empty CEN/ARS plasmid we expect that the efficient sgRNA will cleave all copies of wild-type p2, and since one of the ORFs will be missing, p1 will not be able to survive. When growing on YPD plates, since there is no selection for any of the plasmids, any number of plasmids may be lost, including recombinant p1, wild-type p1, recombinant p2 and wild-type p2. Additionally, it is possible to lose CEN/ARS plasmids within a few generation when not being selected for using Histidine. We expect that in presence of an inefficient sgRNA or no sgRNA and no complementation, we expect that the copy numbers of any of the four species of cytoplasmic plasmids will not be affected. Thus, on colonies picked from any plates, there will be approximately equally high levels of fluorescence.[Fig. 12]

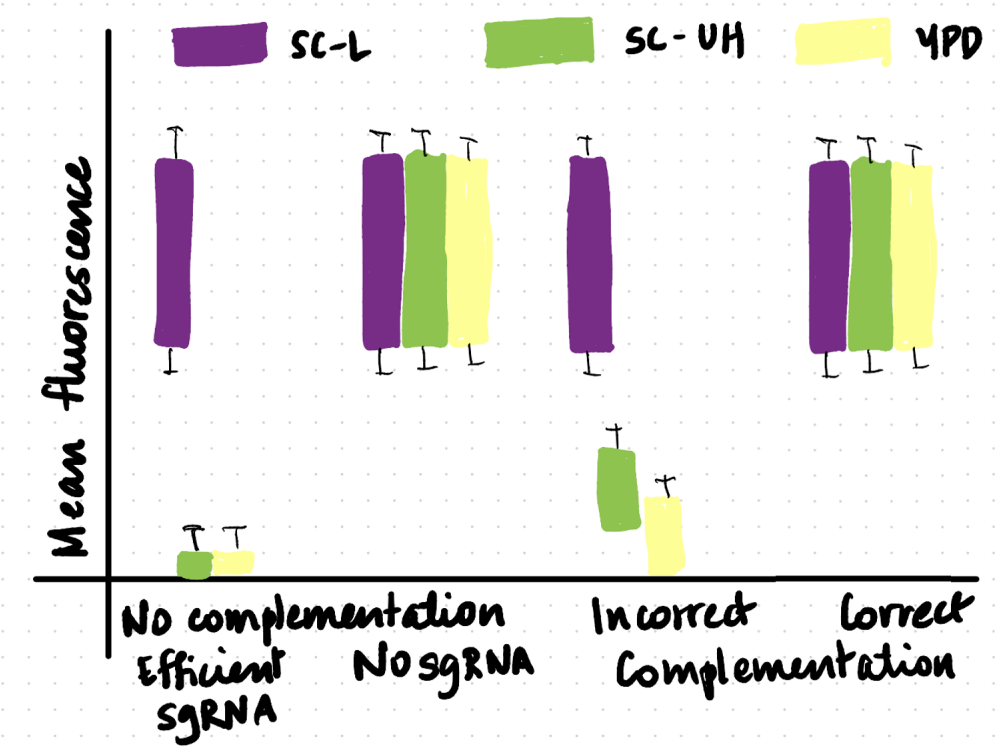


Figure 13: Expected results for p2 ORF complementation pn CEN/ARS

When there is complementation however, we expect things to look a little different. In presence

of the complementation plasmid, and an inefficient sgRNA or no sgRNA we expect the fluorescence to remain uniformly high through all colonies picked from all conditions of plates. When the complementation plasmids exist, and an efficient sgRNA does too, the complementation may or may not be successful. If the complementation is unsuccessful we can expect that there is low to no fluorescence when p1 is not selected for, so on colonies picked from SC-UH or YPD plates. If the complementation is successful, then we will observe similar levels of fluorescence that are relatively high and match the no sgRNA controls to the extent possible. This should be the first sign that the complementation may be successful. With these data in mind, we performed this assay for all of the ORFs and with their relevant complementation plasmids. The data are described below.

6.6.3 Results and Discussion

We initially started with ORFs 2 and 5 and used two complementation plasmids each with a low strength promoter, pREV1 and a medium strength promoter, pRPL18b. For ORF2, with no sgRNA, SC-L and YPD showed high fluorescence as expected, but the plate for SC-UH was fully dried out and no colonies were ale to picked from it. When an efficient sgRNA was added, with no complementation, the data also looked as expected with some fluorescence observed for SC-L but no fluorescence on the SC-UH and YPD plates. For the pREV1 promoter, and comparing between no sgRNA and efficient sgRNA conditions, we observed that for the efficient sgRNA there was no fluorescence observed on the SC-UH and YPD plates showing that pREV1 was unlikely to complement ORFs sufficiently. Similarly, when comparing for the pRPL18b promoter, in presence of the efficient sgRNA, some fluorescence is returned but not at the SC-L levels. This shows that pRPL18b is unlikely to complement ORF2. On looking at the promoter expression data closer, we can see that ORF2 promoter is a medium strength promoter when compared to the other cytoplasmic promoters on p2. Thus, the likely complementation is at a level between pREV1 and pRPL18b. Thus, we need to add more promoters in the middle range to improve our chances of better estimating complementation.[Fig. 13]

The data for ORF5 showed that when ORF was not complemented, the fluorescence measurements looked similar across the colonies picked from the SC-L plates. For the efficient sgRNA, the fluorescence dropped for colonies from SC-UH and YPD plates where in absence of the sgRNA, the fluorescence levels were much higher. When ORF5 was complemented using pREV1, in the

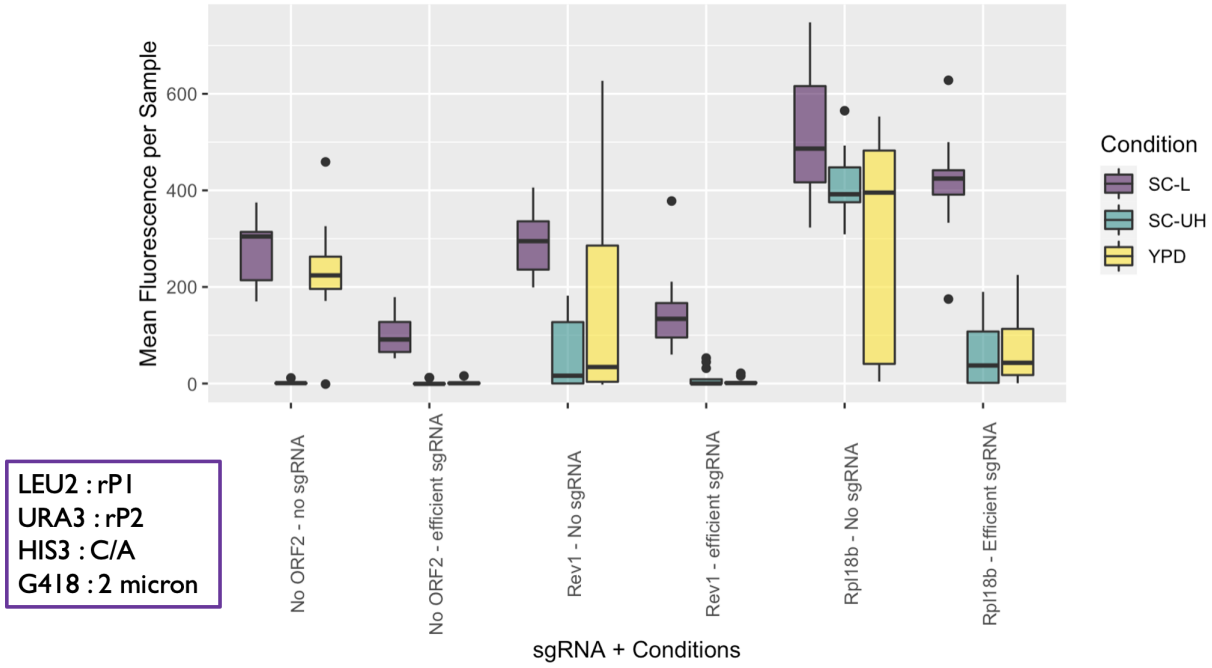


Figure 14: Testing sgRNA efficiency for controls

presence of an efficient guide, there was no fluorescence observed for colonies that were growing on SC-UH and YPD plates when fluorescence was observed in the absence of an sgRNA. When ORF5 was complemented by pRPL18b, and no sgRNA was used, there was fluorescence observed for colonies picked from all the plates. However, when an efficient sgRNA was used, the colonies from the SC-L plate howed fluorescence but the colonies from the SC-UH plate also showed fluorescence. While these levels were lower than those observed from colonies picked off the SC-L plate, it was more than the corresponding control with no complementation. This gives us hope that pRPL18b may be able to complement ORF5 in the nucleus. [Fig. 14]

We decided to perform a long miniprep to investigate these results further. For the samples - No complementation ORF5: no sgRNA, No complementation ORF5: efficient sgRNA, pREV1 > ORF5: no sgRNA, pREV1 > ORF5: efficient sgRNA, pRPL18b > ORF5: no sgRNA and pRPL18b > ORF5: efficient sgRNA - we would expect to see the genomic DNA band on all of them. We also expect to be able to see one p2 band, and at least one p1 band from either the recombinant p1 or the wild-type p1 for samples with no complementation when no sgRNA was used. We expect that all these bands should disappear when using the efficient sgRNA is used in the non-complementation conditions. When complemented with pREV1, we expect no bands when the sgRNA is efficient

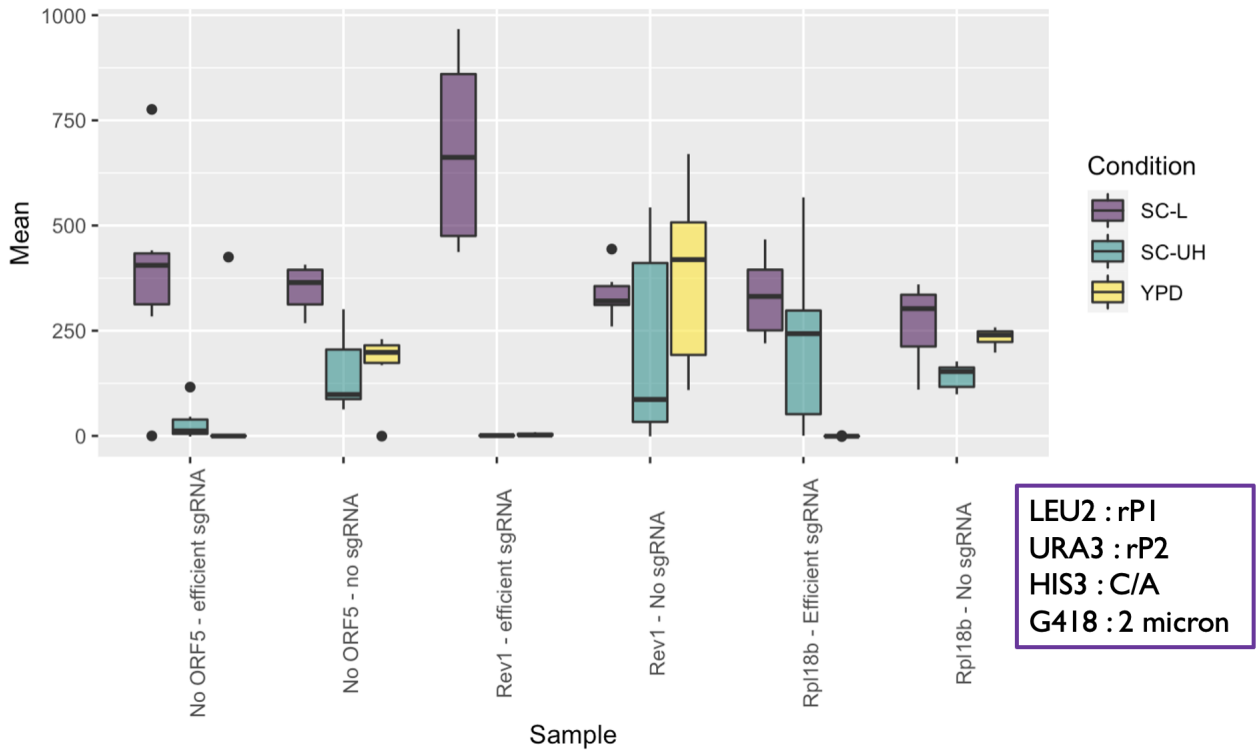


Figure 15: Testing sgRNA efficiency for controls

and at most three bands when there is no sgRNA. However, for the pRPL18b complementation, from the fluorescence data, we expect that we will be able to observe at most three bands for both the no sgRNA and the efficient sgRNA conditions. When we performed the long miniprep, we observed a strong band for genomic DNA on all the lanes. In the absence of complementation, we observed the recombinant p2 band as well as the wild-type band. The recombinant p1 band was observed in the no sgRNA condition, but the band had disappeared in the efficient sgRNA condition but the wild-type p2 band was observed which does not align with the expected bands. For the pRPL18b complementation, we observed the bands as expected. These results enhance the possibility that ORF5 has been complemented but the presence of the wild-type p1 band in the no complementation but efficient sgRNA lane shows the need for a more sensitive assay. [Fig. 15]

We performed the fluorescence assay for ORFs 6, 7, 8, 9, and 10. The data for ORF6 showed that the controls looked as expected, however even using 5 different promoters for complementation, none of them looked as expected making it harder to determine whether the ORF was complemented. The sgRNA plasmids for ORFs 8 and 9 were not very efficient making the data for that difficult to interpret since the controls did not look very clean. For ORF10, the data is the original assay

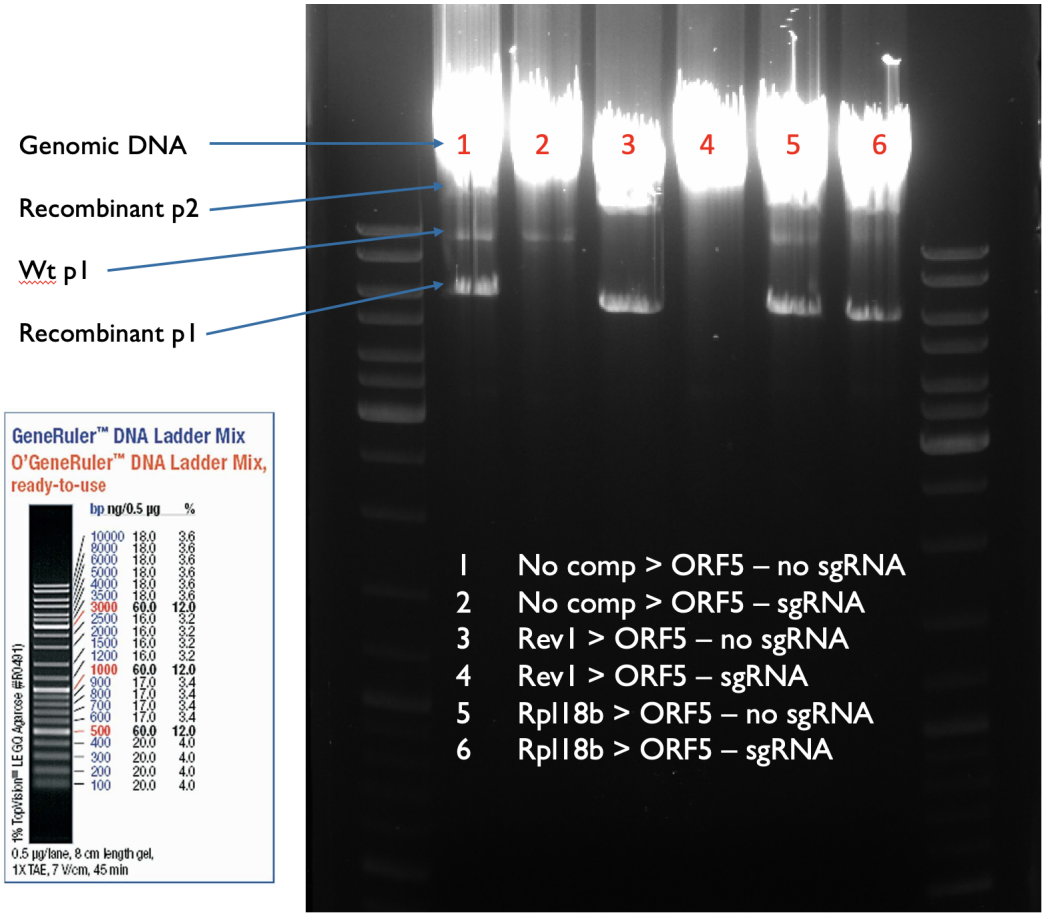


Figure 16: Testing sgRNA efficiency for controls

showed low fluorescence even in colonies from the SC-L plates and in the actual experiment, there was no loss of fluorescence in the no complementation and efficient sgRNA case. This suggested that we needed to revisit the assay to make it more sensitive to the differences and align the original Cas9 assay in a way to make it more interpretable to the actual experiment we will be conducting for the confirming complementation.

6.6.4 Next steps

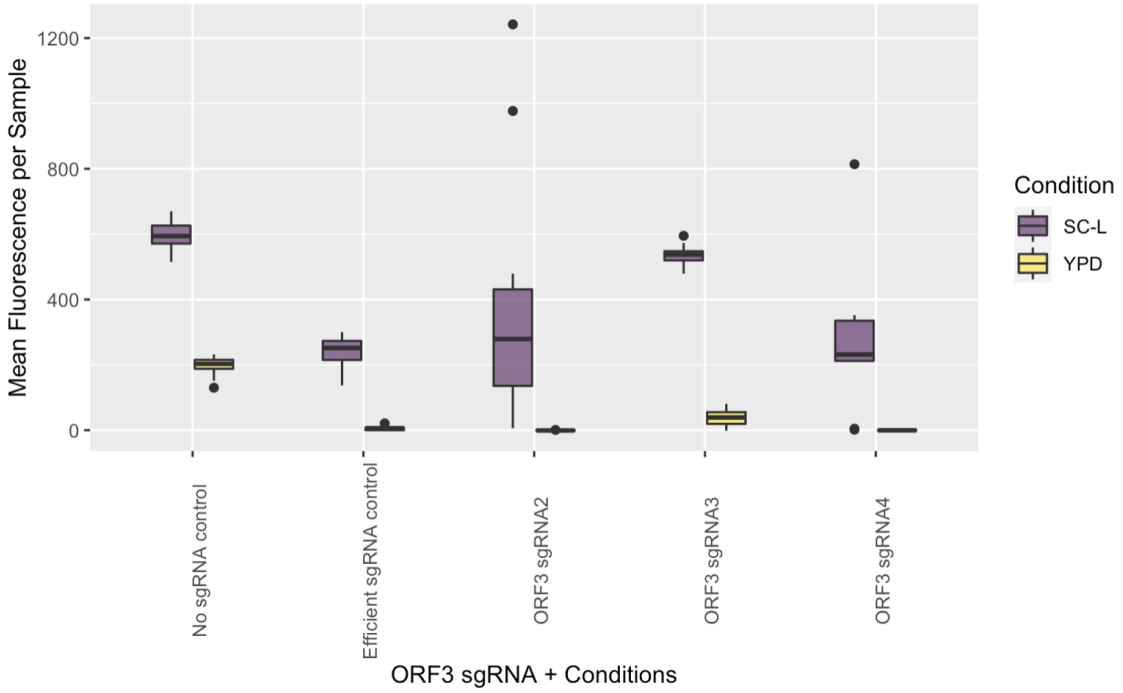
The next steps for this project are to redesign the assay for Cas9. The simplest way to try first would be to put an empty CEN/ARS plasmid with a HIS3 marker in all the strains and to observe the fluorescence data that would be generated as a result on plates that select for LH. Once the assay is confirmed to work, it will make it easier to interpret data from the actual experiment. Additionally, it may be useful to focus on the ORFs that have efficient guides and have them

fully complemented. The ones that do not can then be estimated and more than one ORF can be knocked out at once to confirm complementation using the efficient sgRNA.

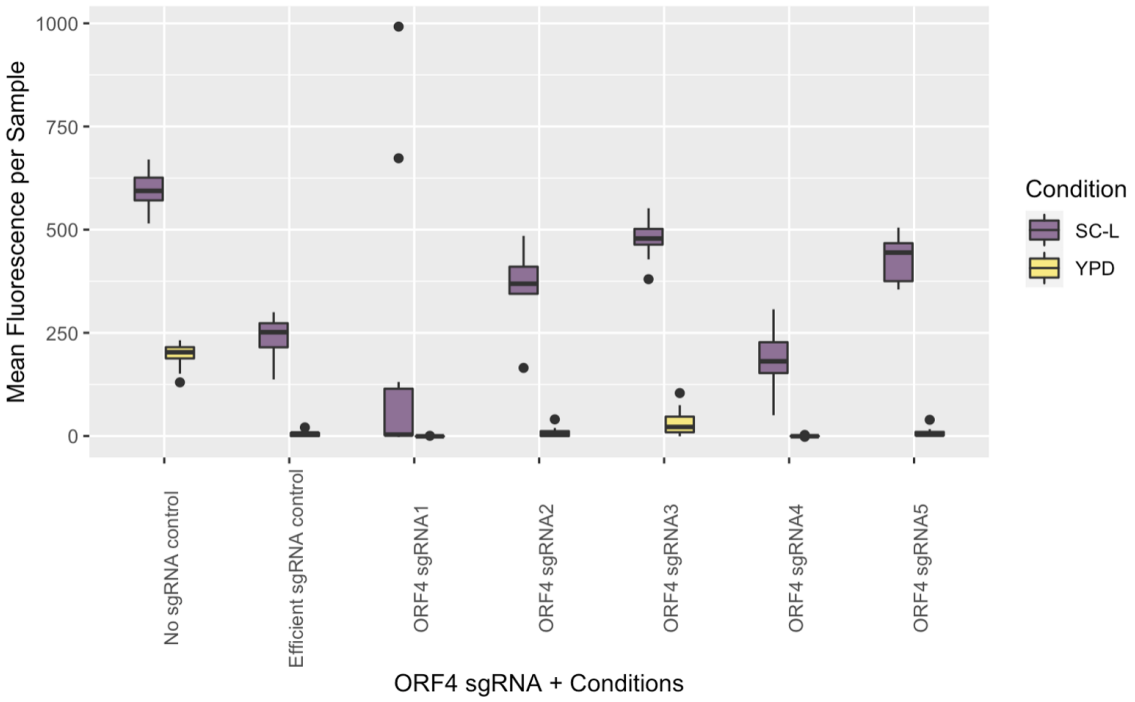
Once complete, the refactoring of p1 will prove useful for upgrading OrthoRep. The need for only one plasmid will enable the system to be broadly transferred to other organisms. It will make the system more tunable since the components required to replicate and maintain p2 will be controlled by nuclear promoters. Refactoring p2 will also open the door to potentially using a different error-prone polymerase on p2 to evolve genes on there as part of a more complex pathway that is currently hard to access for most systems.

6.7 p2 Project Appendix

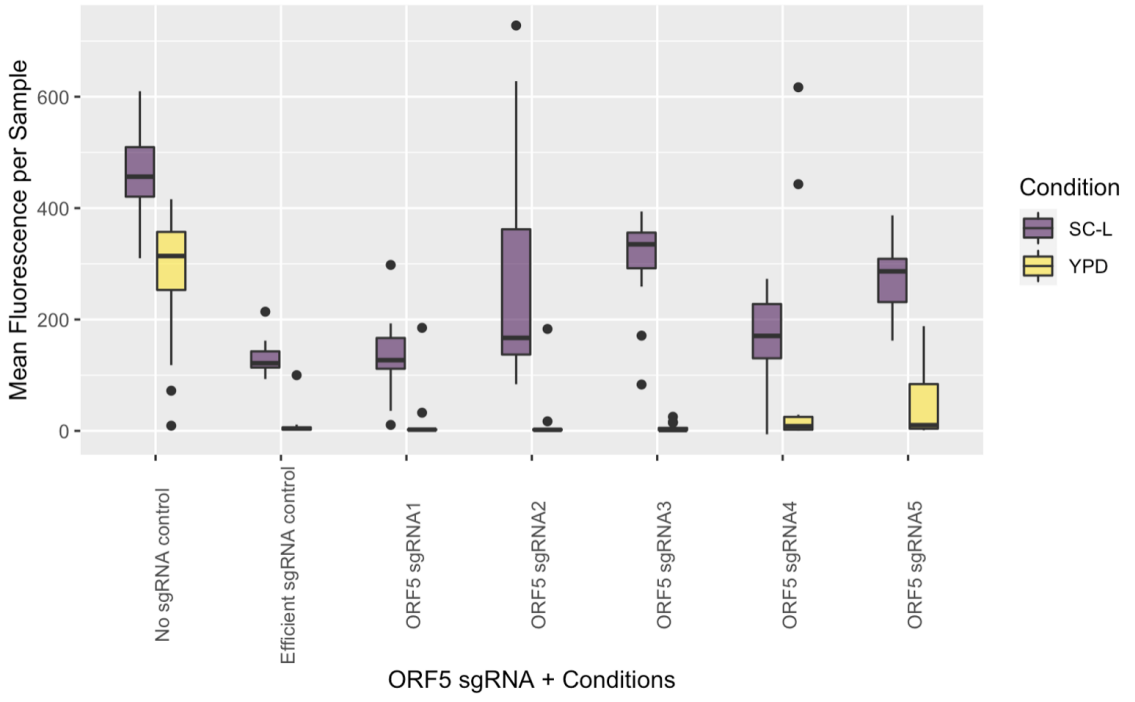
6.7.1 sgRNA efficiency data for all ORFs



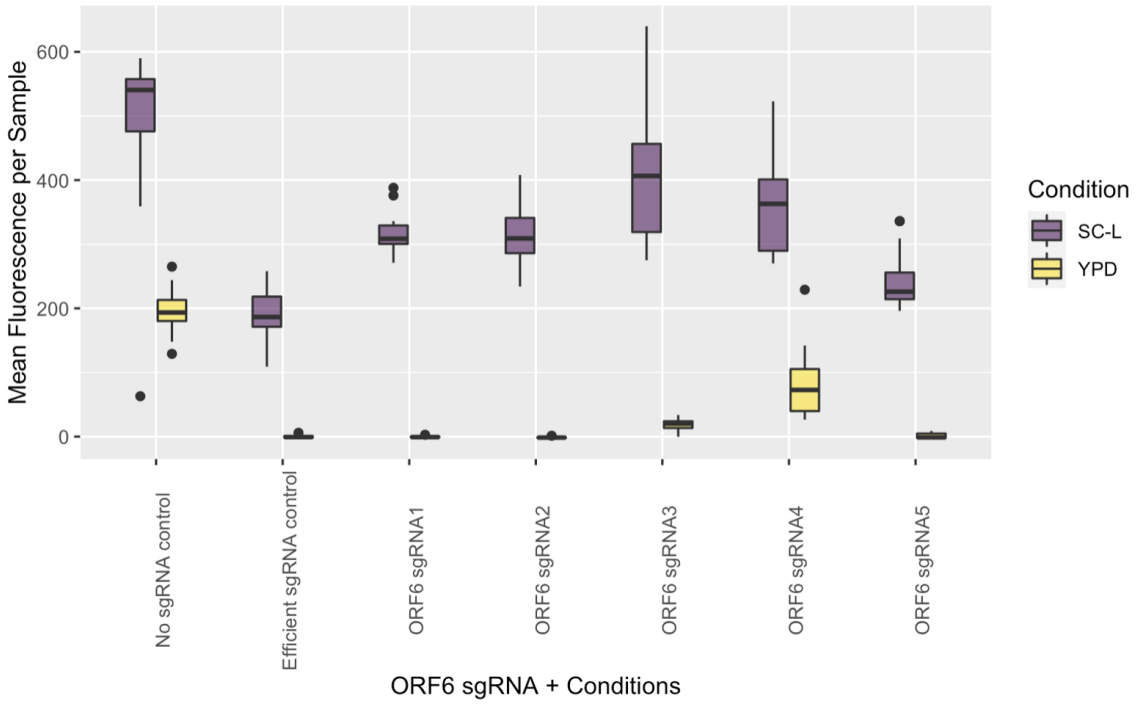
Determining most efficient sgRNA for ORF3



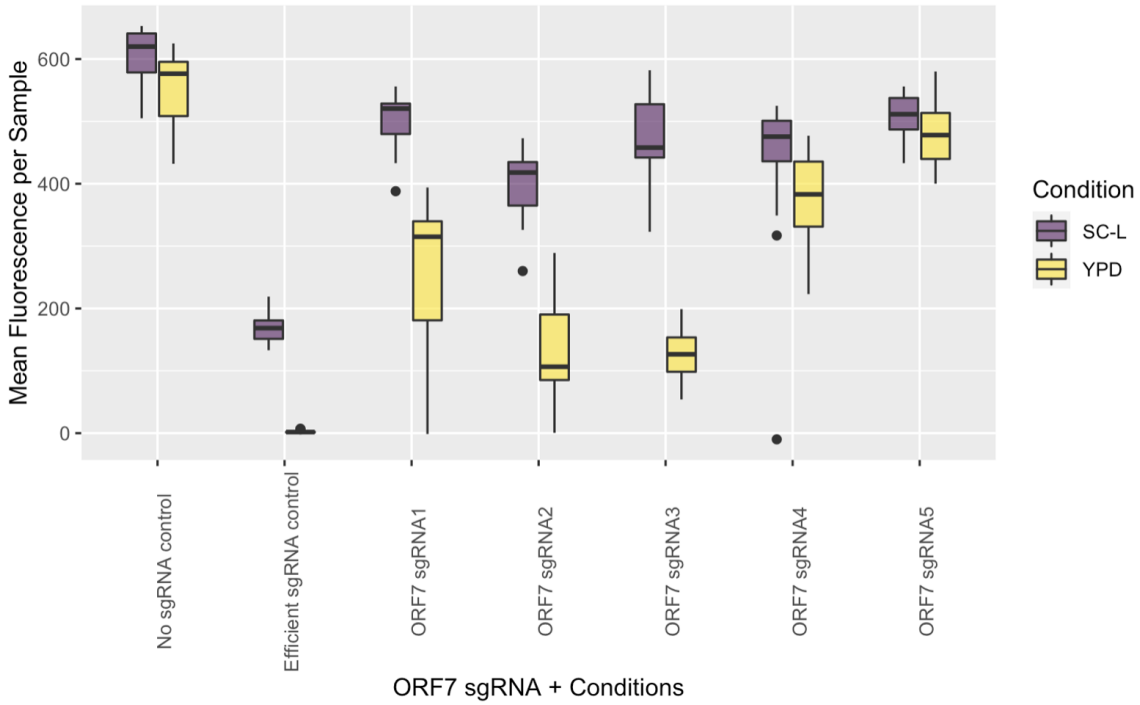
Determining most efficient sgRNA for ORF4



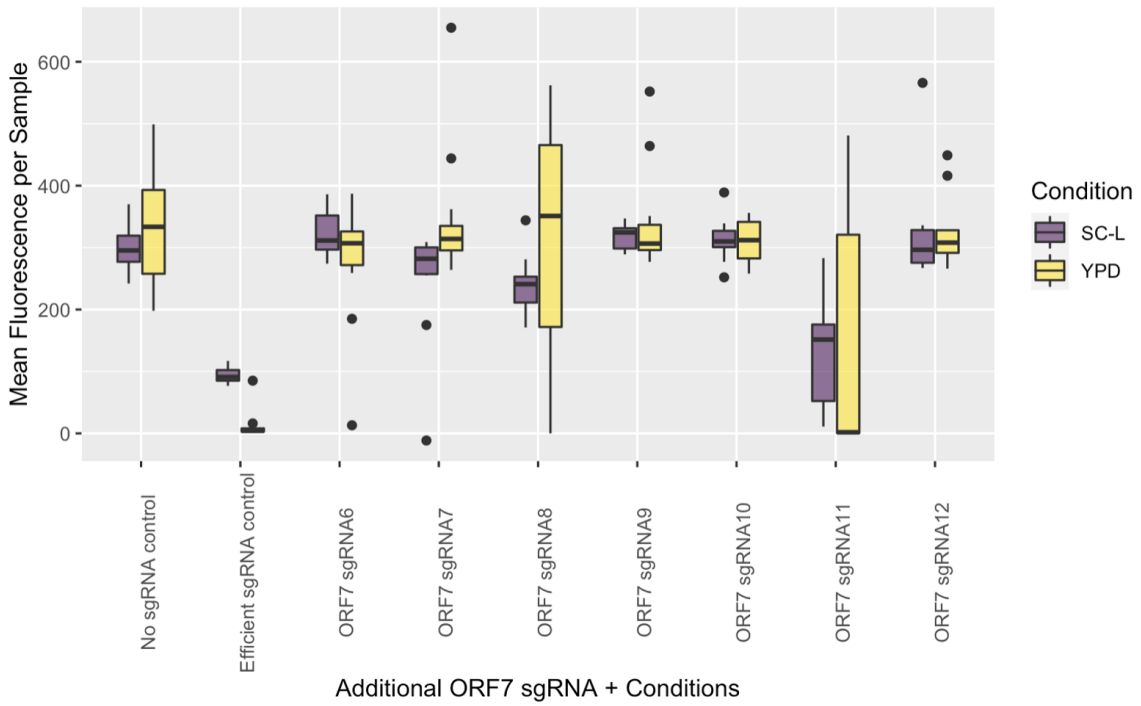
Determining most efficient sgRNA for ORF5



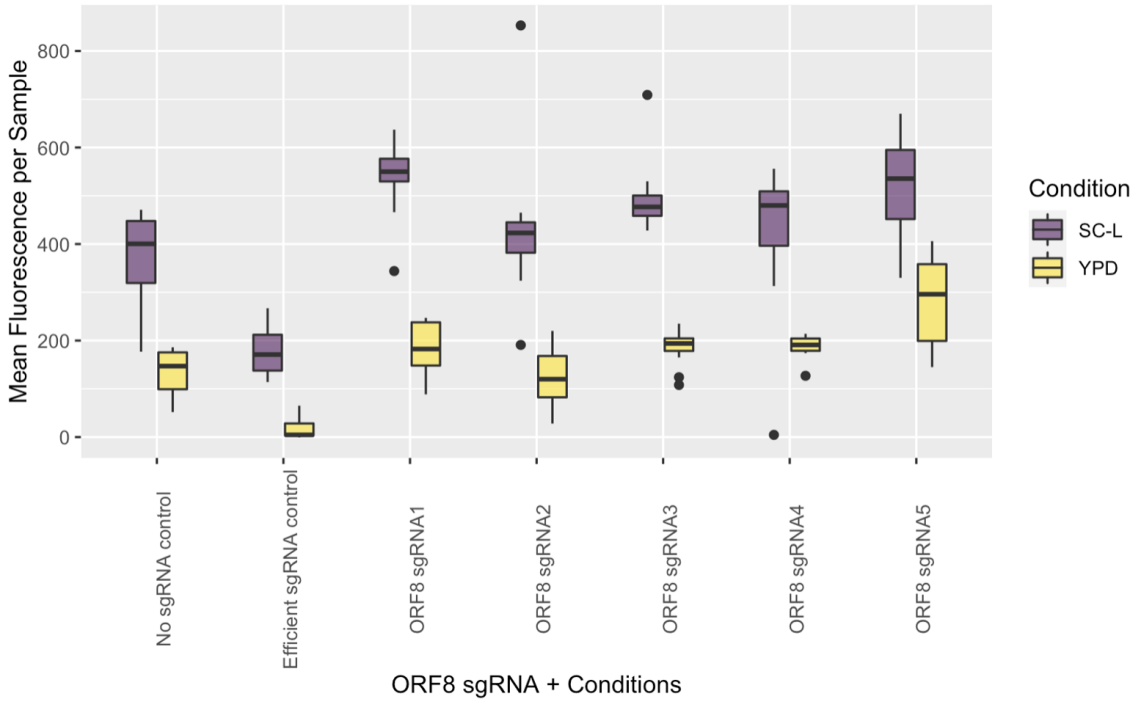
Determining most efficient sgRNA for ORF5



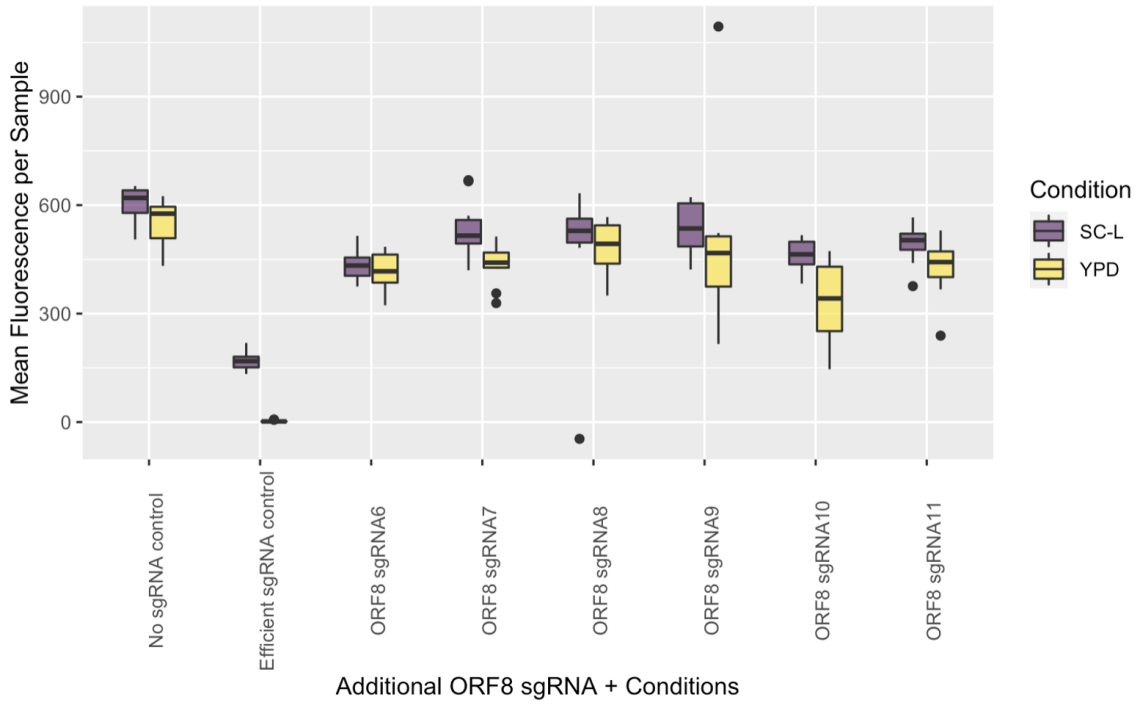
Determining most efficient sgRNA for ORF7



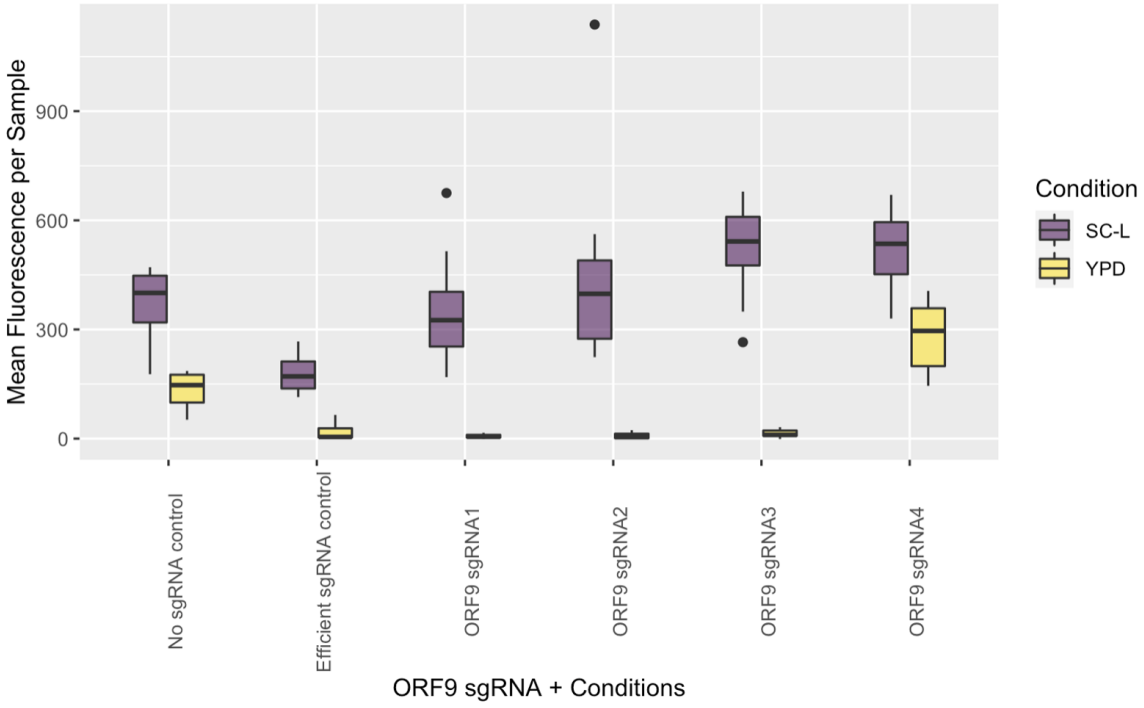
Determining most efficient sgRNA for ORF7 - additional sgRNA



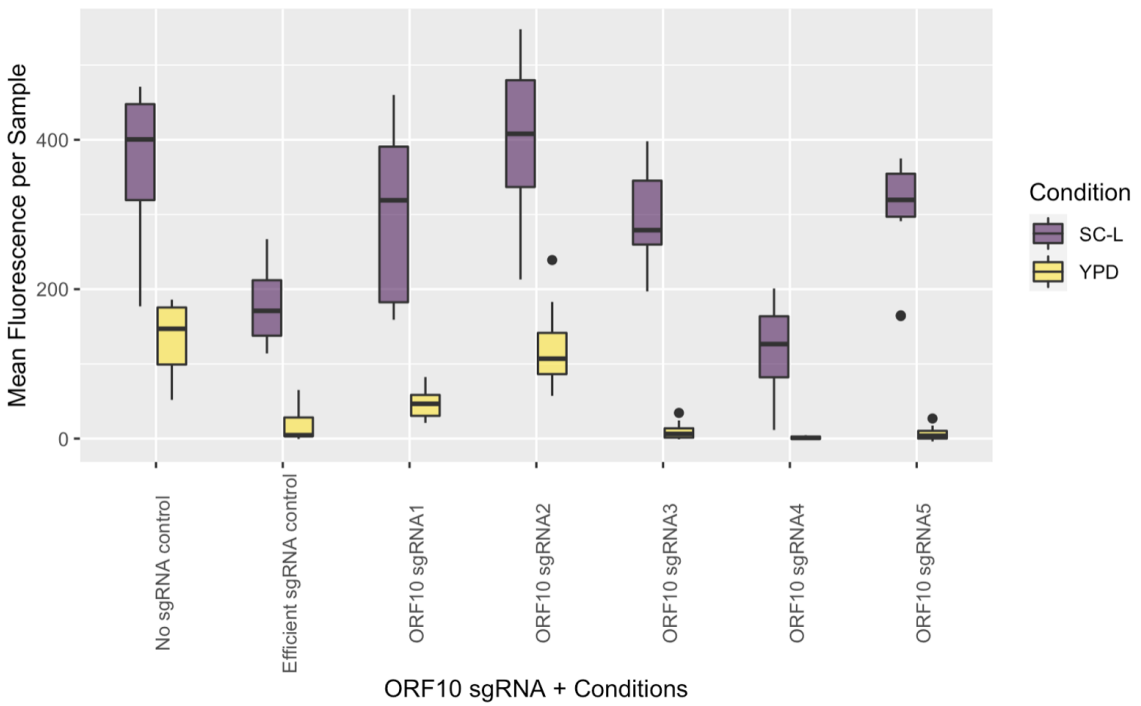
Determining most efficient sgRNA for ORF8



Determining most efficient sgRNA for ORF8 - additional sgRNA

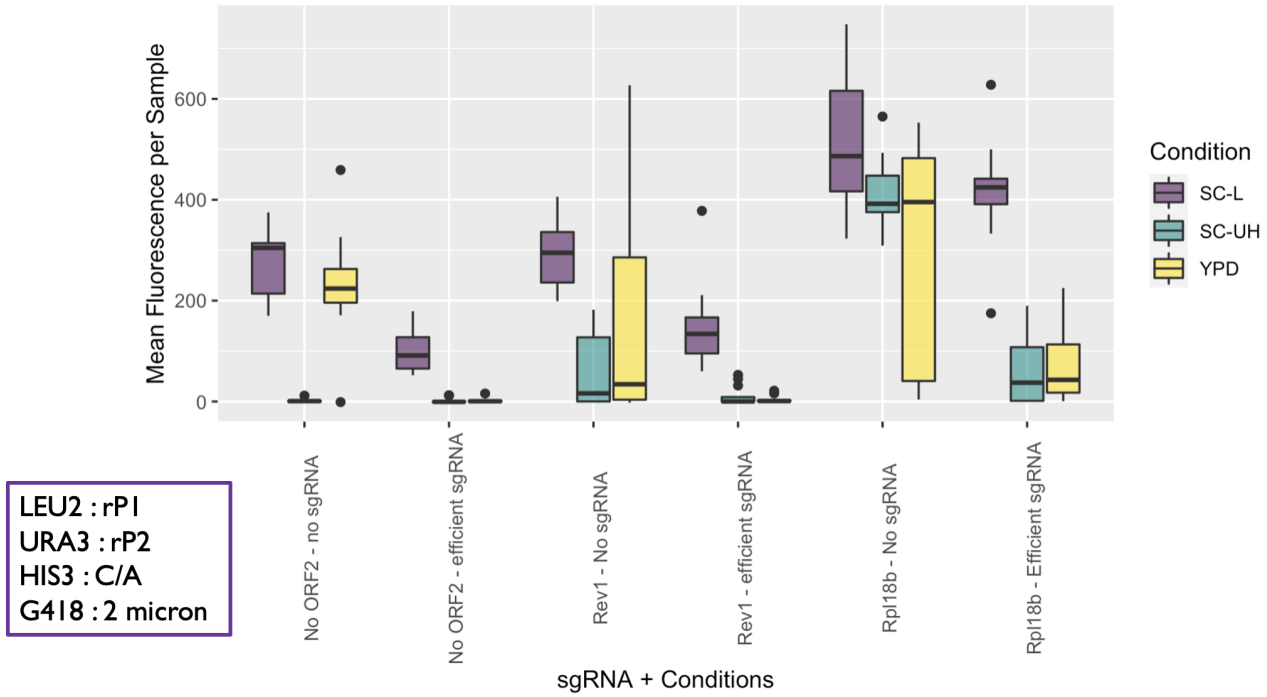


Determining most efficient sgRNA for ORF9

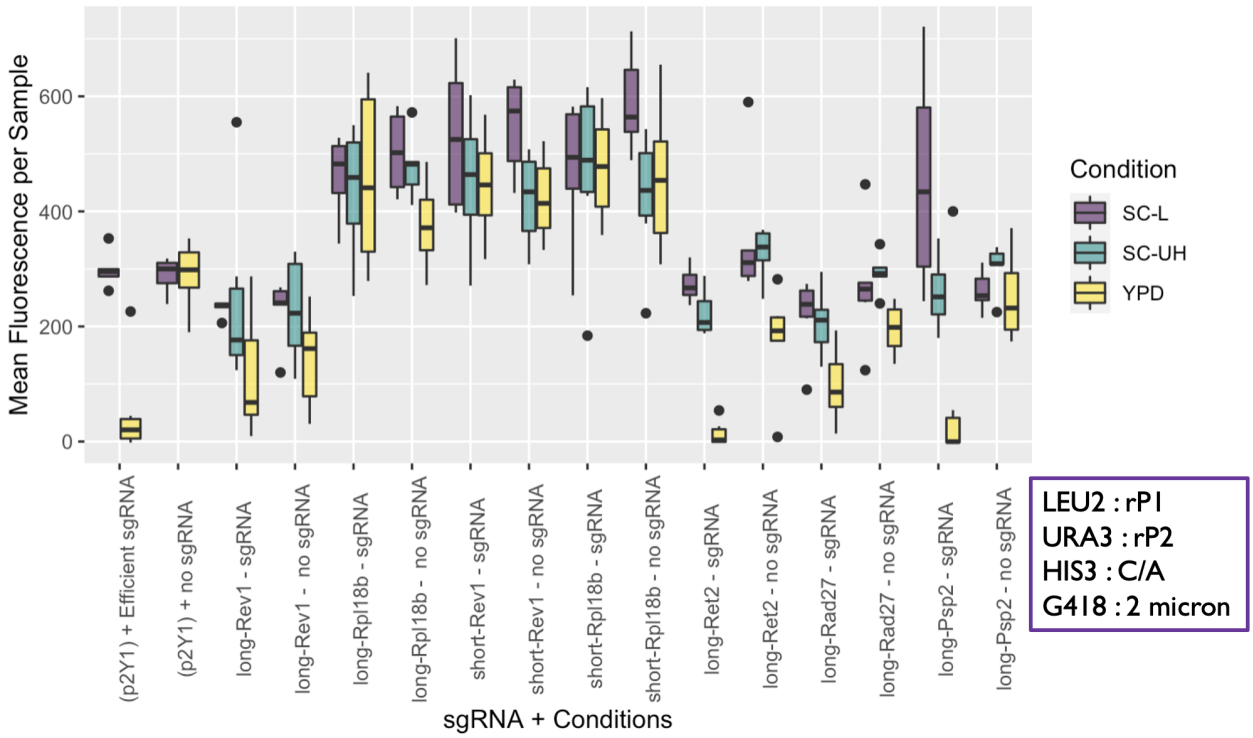


Determining most efficient sgRNA for ORF10

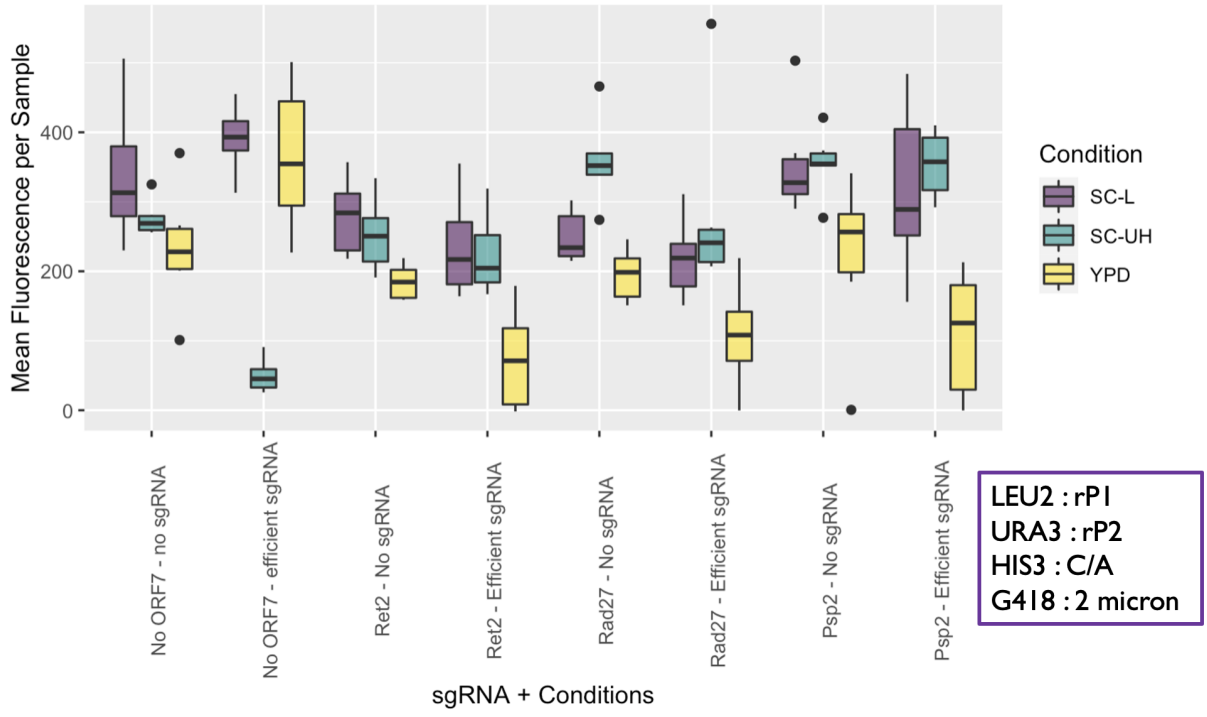
6.7.2 p2 ORF complementation data for all ORFs



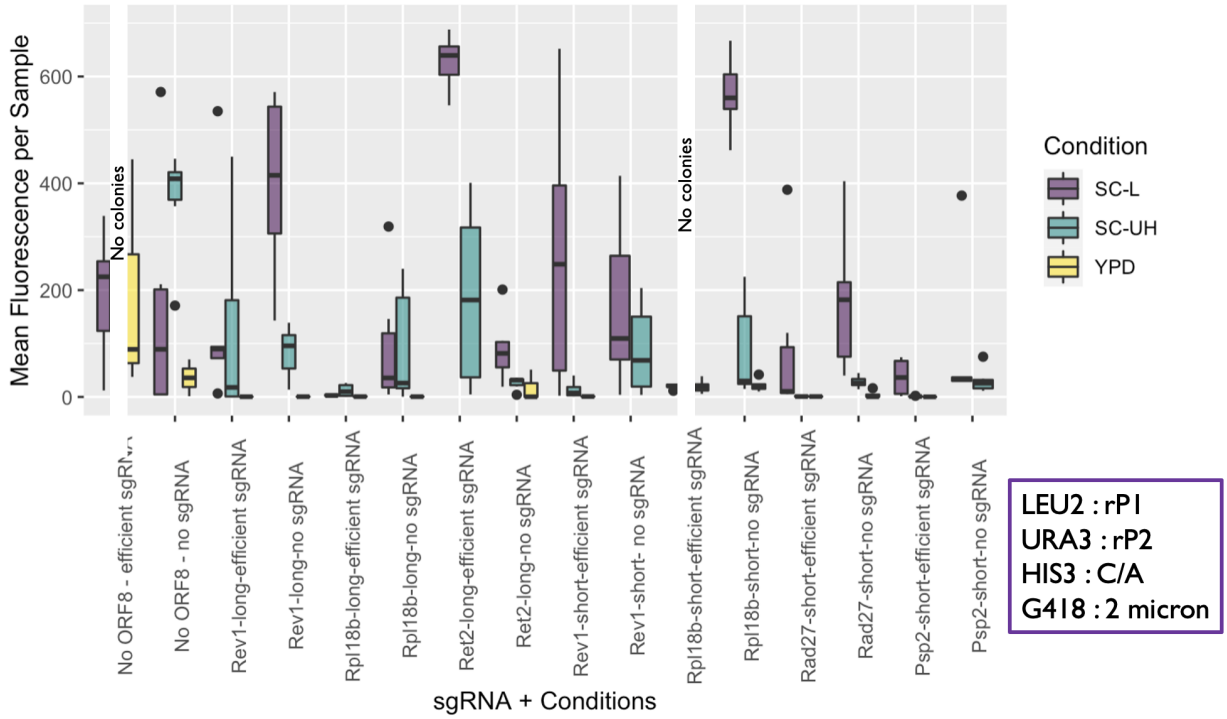
Determining best promoter for ORF2 complementation



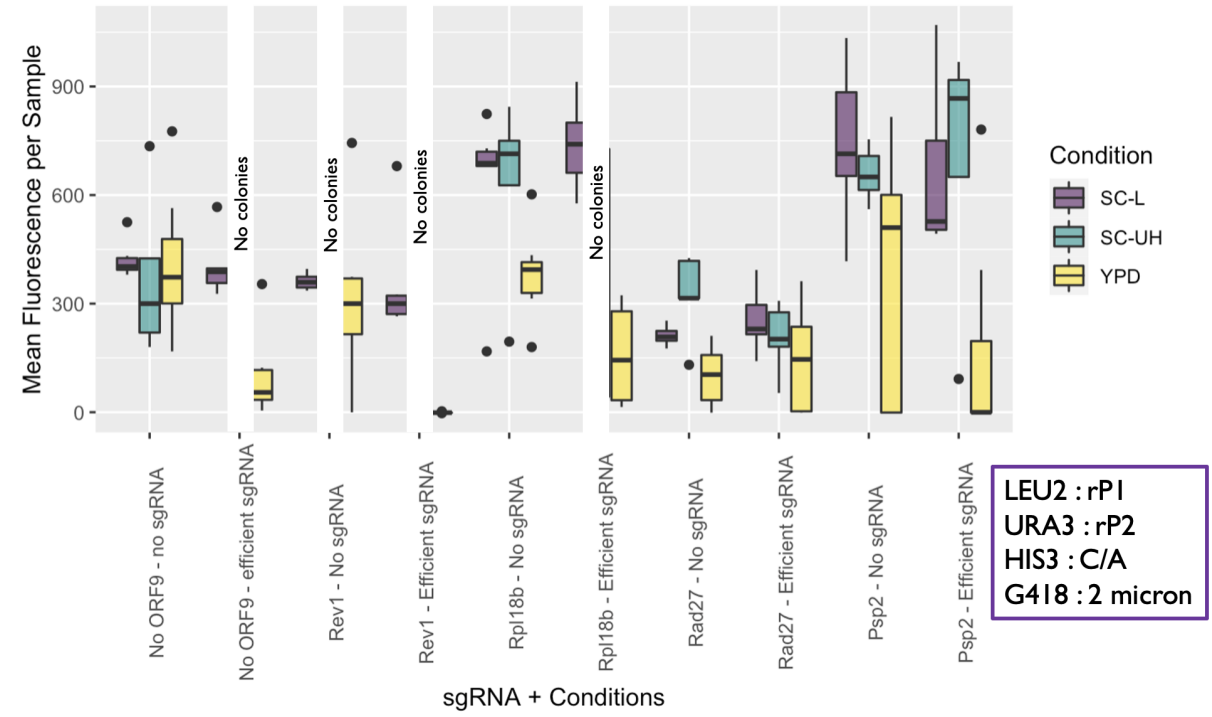
Determining best promoter for ORF6 complementation



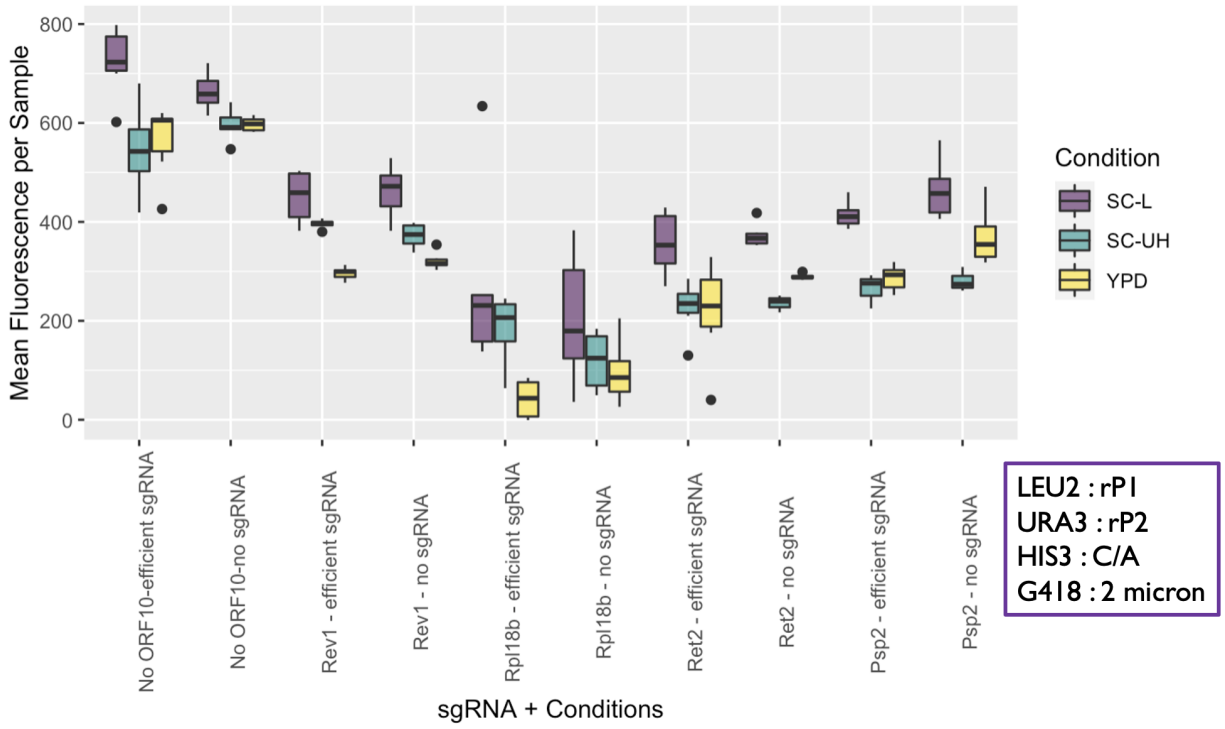
Determining best promoter for ORF7 complementation



Determining best promoter for ORF8 complementation



Determining best promoter for ORF9 complementation



Determining best promoter for ORF10 complementation

7 Cas9 Project

7.1 Introduction

Programmable editing of genes is an invaluable tool that has become irreplaceable in the quest for making lives better. Currently, the CRISPR-Cas9 system and the subsequent iterations using a variety of Cas proteins has led to editing of genes in plants for agriculture, in humans in for health-care applications and everyday in the laboratory for cloning (Engqvist & Rabe, 2019; Anzalone et al., 2019; Arzumanyan et al., 2018).

CRISPR-Cas9 is the bacterial acquired immunity against viruses. This immunity is acquired in two phases. The first phase consists of the immunization phase in which on the first infection from a new virus, the viral DNA is cleaved and a 20 bp sequence called the spacer is added into a CRISPR (Clustered Regularly Interspersed Palindromic Repeats) array (Jinek et al., 2012; Mali, Esvelt, & Church, 2013). These spacers are separated by repeats, and the array grows longer as more and more spacer get added into it (Jinek et al., 2012). The second phase is the immunity phase, and on repeat infection, the bacterial are able to produce a generalized tracrRNA and a pre-CRISPR RNA (crRNA) from the spacer located in the array as previously saved (Jinek et al., 2012; Mali et al., 2013). The tracRNA and crRNA come together with the Cas9 (CRISPR-associated) endonuclease and bind to the viral DNA (Jinek et al., 2012). Cas9 uses the spacer as a guide and looks for a specific NGG sequence known as a protospacer adjacent motif (PAM) to differentiate between the self and the non-self target leading to the double-stranded cleaving of the viral DNA (Jinek et al., 2012; Mali et al., 2013). Jinek et. al., (2012) combined the tracRNA and the crRNA into a chimeric single guide RNA (sgRNA) where a given 20bp spacer can be targeted provided there is an NGG PAM sequence adjacent to it (Mali et al., 2013). The creation of this sgRNA leads to a more programmable version of the Cas9 that can be moved to other species outside of bacteria for programmable gene editing, which is a powerful tool with several applications.

7.1.1 Structure of SpCas9

The most commonly used Cas9 is from the bacterium Streptococcus pyogenes and is known as SpCas9 or SpyCas9. The crystal structure of the 1368 amino acid Cas9 shows presence of two nuclease domain lobes, the RuvC domain and the HNH domain (Jinek et al., 2014). The C-terminal

domain and the topoisomerase domain together make up the PAM interacting (PI) domain (Jinek et al., 2014). The two nuclease domain and the α-helix lobe are connected by an arginine-rich linker. The crystal structure shows us that the dimensions of the protein are $100\mathring{A} * 100\mathring{A} * 50\mathring{A}$ (Jinek et al., 2014). Each of the nuclease domains consists of an active site, each of which is responsible for cleaving one of the strands at the target sequence (Jinek et al., 2014). The HNH domain cleaves the complementary strand and RuVC domain cleaves the non-complementary strand. When the crystal structure is formed, the two active sites are $25\mathring{A}$ apart (Jinek et al., 2014).[Fig. 16]

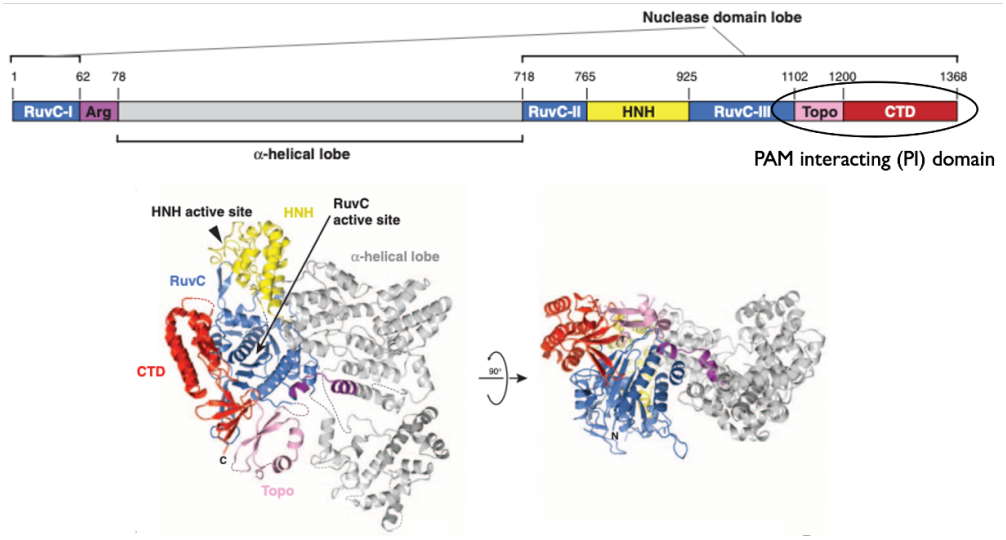


Figure 17: Crystal structure of SpCas9 shows active sites

The crystal structure of Cas9 shows two endonuclease domains: RuvC and HNH and an alpha-helical lobe that connects them. The topoisomerase domain and the C-terminal domain together make up the region that binds to the PAM sequence.

The protospacer adjacent motif (PAM) sequence is an essential component when the Cas9 protein needs to find their target (Jinek et al., 2012; Mali et al., 2013; Anders, Niewoehner, Duerst, & Jinek, 2014). In Type II-S systems, like the one that Cas9 is part of, uses the PAM to unwind the sequence upstream to determine whether the target matches the spacer sequence before cleavage (Xie et al., 2018; Anders et al., 2014). The PAM is between 2-6 bp long and is located on the non-complementary strand, but for SpCas9, the PAM is known to be 5'-NGGN-3', where N can be any nucleotide, and cleavage occurs 3 bp upstream of the PAM sequence (Jinek et al., 2014; Mali et al., 2013; Anders et al., 2014). In the absence of the PAM, the Cas9 cannot unwind the DNA even if the target sequence matches the spacer perfectly and is thus ignored (Anders et al., 2014). In the PAM sequence, the guanine in the third position is shown to play an important part in cleavage

(Xie et al., 2018). The crystal structure shows us the residues R1333 and the R1335 on Cas9 have closest direct contact with the target DNA (Anders et al., 2014). A variety of TypeII-S systems exist across various organisms but the PI domain is very conserved and several species show similar or closely related residues in those positions and they show very similar PAM requirements (Anders et al., 2014).

7.1.2 Catalytically inactive Cas9 and VP64

The catalytically active Cas9 proteins contains one active site in each of its nuclease domains. When double-stranded breaks are created, each of these catalytic sites is responsible for cleaving one strand of DNA (Jinek et al., 2012). However, there are several instances where binding of Cas9, subsequent unwinding of DNA around the PAM region and the recognition of the target sequence are necessary without the resulting cleavage. As a result, each of the active sites can be inactivated one at a time, or both at once leading to a nicking Cas9 (nCas9) or a catalytically dead Cas9 (dCas9) (Larson et al., 2013). The H840A mutation located in the HNH domain and the D10A mutation located in the RuvC domain can make dCas9 (Jinek et al., 2012; Larson et al., 2013; Anders et al., 2014). Activators and repressors can then be fused to dCas9 to affect gene regulation (Larson et al., 2013).

VP64 is a transcriptional activator made up of four-successive VP16 sequences using glycine-serine linkers (Beerli, Segal, Dreier, & III, 1998). The VP16 sequence comes from Protein 16 of the Herpes simplex virus and have been reduced to be a short sequence that can be used modularly while maintaining full transcriptional activation potential (Seipel, Georgiev, & Schaffner, 1992; Beerli et al., 1998). When VP64 is fused to a catalytically inactive Cas9, this fusion is transformed to a transciptional activator (Perez-Pinera et al., 2013). When building Cas9 assays, the cleavage activity ensures that the gene is inactivated and when repaired, there are indels that may occur leading to inactivation of the gene (Jinek et al., 2012). This cleavage and subsequent repair reaction is not reversible, and since changes in occur in DNA, the same nucleus cannot be reused to to perform the same reaction. The VP64 domain is able to recruit the machinery required for transcription, thus enabling activation of transcription at the desired location. The fusion of VP64 to dCas9 gives us control of transcription without changing the DNA sequence making designing an experiment more straightforward (Perez-Pinera et al., 2013).

This ability to fuse other modular parts that form genetic circuits has expanded the use of Cas9 as a therapeutic tool. Fusion of ssDNA-specific cytidine deaminase to dCas9 or Cas9-D10A nickase leads to a cytosine base editor that has the ability to mutate a C-G base pair to a T-A base pair in a given window on the spacer, through a series of simple reactions (Rees & Liu, 2018). Adenine base editors can be made by fusing a wild-type TadA monomer and an evolved TadA monomer to the Cas9-D10A nickase (Rees & Liu, 2018). These base editors convert an A-T base pair to a C-G base pair within the given editing window (Rees & Liu, 2018). Another application is Prime Editor, which fuses a reverse transcriptase from a Moloney Murine Leukemia Virus (M-MLV) to a Cas9-H840A nickase (Anzalone et al., 2019). The resulting Prime Editor complex can cause insertions, deletions and substitutions of longer lengths from 40 - 80 bp, using the mismatch repair pathway (Anzalone et al., 2019). This system involves three separate binding events between the guide and target, between the target binding site and the target, and between the 3' end of nicked DNA and the pegRNA leading to fewer off-targets (Anzalone et al., 2019).

7.2 The failed evolution of SpCas9

7.2.1 Engineering and Evolution Approaches for improving Cas9

There is a lot of interest in the field in improving the ability for Cas9 to bind to a variety of sequences with different PAMs to allow for more genomic editing against disease. Different labs have been involved in efforts for rational design of the enzyme as well as in evolving Cas9s to expand their PAM specificity.

Kleinstiver et.al., (2015) used rational design and directed evolution approaches to improve Cas9 binding to non-canonical PAMs. They identified two variants D1135V/R1335Q/T1337R (VQR) and D1135E/R1335Q/T1337R (EQR) which showed high cleavage efficiency towards the NGA PAM in comparison to the NGG PAM (Kleinstiver et al., 2015). Additionally, they identified a variant which showed high cleavage efficiency for NGC PAM and contained D1135V/G1218R/R1335E/T1337R (VRER) mutations. Their data showed that none of these mutation by itself or in smaller combinations of subsets was sufficient to increase efficiency of Cas9 activity, but all four mutations were necessary (Kleinstiver et al., 2015). The two previously identified residues that bind to the PAM sequence, R1335 and T1337 while two new mutations in the PAM interacting domain were also

discovered. The D1135 mutation was also discovered by other groups, but the G1218 mutation was discovered for the first time (Kleinstiver et al., 2015). This strategy of combining two approaches ensured that the mutations only showed up in the PI domain of Cas9.[Fig. 17]

Nishimasu et.al., (2018) tried to expand the PAM specificity of Cas9 but mutating the R1335A position determined to be essential for binding to the third position of the NGG PAM. They then looked for compensatory mutations that may exist in the surrounding PAM interacting domain of Cas9. A variety of mutations, some that were previously discovered including the D1135, R1335 and R1337 were found to be necessary for binding to NGN PAMs as well as the G1218 that was also seen on Cas9-VRER which is required for binding to the NGC PAM (Nishimasu et al., 2018). The new mutations discovered while still in the PAM interacting domain, were identified at unique locations and include the L1111, A1322 leading to the formation of the Cas9-VRVRFRR (Nishimasu et al., 2018). These Cas9s are more general and have the ability to detect a variety of PAM sequences. However, a large number of mutations are required for this, and more experiments will be needed to determine whether these mutations cause conformation changes leading to limitations in the targeted sequence. [Fig. 17]

Hu et. al., (2018) used the continuous directed evolution system PACE to evolve a variety of xCas9 variants that had high specificity to the target DNA while the PAM recognition specificity was relaxed. Mutation for this xCas9 almost exclusively showed up outside the PAM interacting domain, mainly in the alpha-helical lobe (Hu et al., 2018). This is interesting because mutations increasing specificity to target almost fully compensated for PAM binding mutations as seen previously. There were no mutations observed in the PAM interacting domain while the result was very similar. This shows that mutations required to relax PAM specificity need not occur only in the PAM-interacting domain. These mutations likely lead to a decrease in off-target activity since this xCas9 was more specific to the target sequence.[Fig. 17]

Kleinstiver et.al., (2016) rationally designed and then engineered a Cas9 variant named Cas9-HF that showed mutations in the alpha-helical domain and one in the RuvC domain and none in the PAM interacting domain. This Cas9-HF shows high PAM specificity as well as high target specificity thus reducing off-target activity. For this purpose they identified four mutations, N497A, R661A, Q695A and Q926A which had not been detected previously and have not been detected since (Kleinstiver et al., 2016). The paper says that the mutations they studied have contact with

	262	324	409	480	497	543	661	694	695	926	1111	1135	1218	1219	1286	1322	1331	1332	1333	1335	1337	1362	PAM	Notes
SpCas9	A	R	S	E	N	E	R	M	Q	Q	L	D	G	E	N	A	I	D	R	R	T	L	NGGN	
XCas9 3.7	T	L	I	K	N	D	R	I	Q	Q	L	E	G	V	N	A	I	D	R	R	T	L	NGAN, NGTN	
EQR	A	R	S	E	N	E	R	M	Q	Q	L	E	G	E	N	A	I	D	R	Q	R	L	NGAG	
VQR	A	R	S	E	N	E	R	M	Q	Q	L	V	G	E	N	A	I	D	R	Q	R	L	NGAN, NGCG	
VRER	A	R	S	E	N	E	R	M	Q	Q	L	V	R	E	N	A	I	D	R	E	R	L	NGCG	
VRVRFRR	A	R	S	E	N	E	R	M	Q	Q	R	V	R	F	N	R	I	D	R	V	R	L	TGNN	
QQR	A	R	S	E	N	E	R	M	Q	Q	L	D	R	E	Q	A	F	K	Q	Q	R	L	NAAN	
HF	A	R	S	E	A	E	A	M	A	A	L	D	G	E	N	A	I	D	R	R	T	L	*NGGN	> off targets

Figure 18: Mutations on Cas9 according to position

This figure shows the different mutations that have discovered on Cas9 based on experiments done around the world to allow binding to target sequences with varied PAM requirements.

the phosphate backbone of DNA, so potentially these mutations could reduce off-target activity in other scenarios as well (Kleinstiver et al., 2016). [Fig. 17]

Miller et.al., (2020) set out to evolve a Cas9 to bind the NAN PAMs using PACE. They used a combination of two target sequences to reduce target specficity and a split-intein Cas9 to be able to better isolate different mutants on the different regions of Cas9. Their efforts resulted in SpCas9-NRRH, SpCas9-NRTH and SpCas9-NRCH denoting the PAMs that each specific Cas9 could target (Miller et al., 2020). These expanded PAM binding preferences also extend to base-editors which is increasingly becoming the most common way for treating diseases called by single mutations.[Fig. 17]

There are a lot of labs invested in expanding the toolbox of CRISPR-Cas9 as shown above. However, there are still problems left to solve. The advent of Artificial Intelligence (AI) and machine learning in analyzing large amounts of sequence data and making 3D protein structures has provided opportunities to evolve a Cas9 to a variety of target and PAM sequences at very high-throughput. We would like to use OrthoRep to evolve Cas9 to bind to any given 23 bp sequence. Instead of creating more generalized Cas9 that can accrue mutations in the PI domain to relax PAM specificity or in the alpha-helix region to better bind the target, we would like to create many specialized Cas9 proteins. As more Cas9 proteins are evolved, we can use AI to detect patterns between mutations on Cas9 to their location of binding to the target. Eventually, we would be able to know what amino acid at a specific locations leads to best binding between the sgRNA, the target DNA and the Cas9 protein.

7.2.2 Pilot: Confirming that the assay works

The Cas9 evolution experiment was designed such that dCas9-VP64 would be expressed from the mutating p1. It would then use the sgRNA expressed from the HO locus to find the target also located in the HO locus. Once this binding occurred the VP64 would bind and activate the downstream HIS3 gene (Perez-Pinera et al., 2013). The HIS3 gene encodes for Imidazoleglycerol-phosphate dehydratase which catalyzes the sixth-step in the histidine biosynthetic pathway (Caufield, Sakhawalkar, & Uetz, 2012). The organic compound 3-Amino-1,2,4-triazole, (3-AT), is a competitive inhibitor of the product of the HIS3 gene, Imidazoleglycerol-phosphate dehydratase (Caufield et al., 2012). Thus, in presence of high concentrations of 3AT, increased expression of HIS3 is required for the survival of the gene.[Fig. 18C]

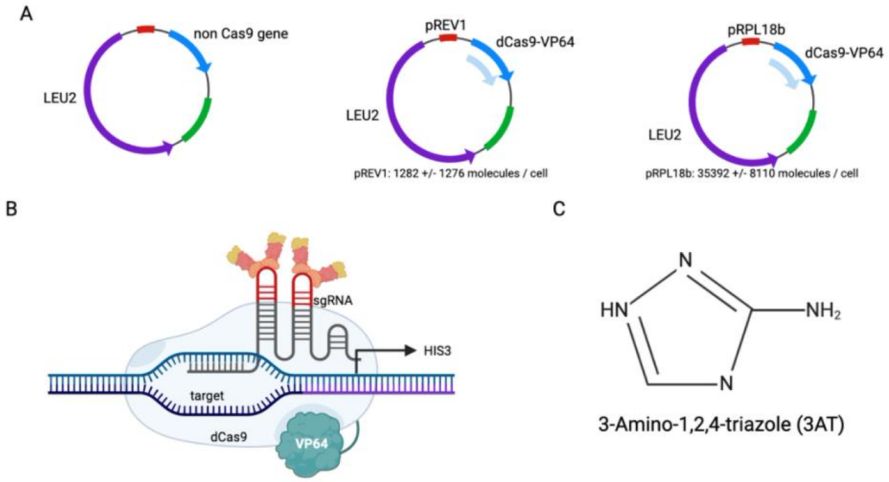


Figure 19: Setup for Cas9 pilot experiment

A: The three plasmids used for the pilot growth rate experiment. B: Schematic of how the experiment will work. C: The structure of 3-Amino-1,2,4-triazole (3AT).

The OrthoRep error-prone polymerase will replicate the dCas9-VP64 located on p1 creating random mutations on the length of the gene. This dCas9-VP64 with the mutations will use the sgRNA expressed from the HO locus to bind to the target that contains a non-NGG PAM. The Cas9's that contain mutations that allow for this binding to non-canonical PAM will be able to produce enough HIS3 to survive in increasing concentrations of 3-AT. With each successive passage, we will be able to identify the Cas9 containing mutations that allow binding to the non-canonical PAMs. We can thus, evolve different Cas9's with a variety of non-canonical PAMs to determine if

OrthoRep is able to identify any similar mutations to those that have been previously identified by other engineering and evolution efforts.[Fig.18B]

To determine whether this experiment will work, we first need to determine the level of expression needed by Cas9 to be expressed at a level where enough HIS3 would be produced. We encoded the dCas9-VP64 construct on a CEN/ARS plasmid and used two different promoter, pREV1 which has low strength and pRPL18b which is a medium strength promoter. We also integrated two different PAMs, a canonical TGGG PAM and a non-canonical TGTG PAM at the target locus upstream of the HIS3 gene. We then transformed each of the CEN/ARS plasmids into each of the yeast strains and performed a growth curve such that OD600 measurements were taken every 30 minutes for a period of 24 hours The 96-well plate consisted of four yeast strains and eight concentrations of 3-AT. Each of the strains was in replicates of 3 and the per minute growth rate was calculated for each sampled and then averaged for the replicates. The data is showed in the figure below. The data show that at low concentrations of 3-AT, there is no difference in growth between the four strains suggesting that the strains are able to grow irrespective of the amount of Cas9 and the efficiency of binding to a non-canonical PAM. However, at higher concentrations of 3-AT, and particularly at the highest concentration of 3-AT measured for this assay, both concentration and efficiency of Cas9 play an important part in determining whether cells can grow. In presence of 2 mM 3-AT, the highest expression of Cas9 (pRPL18b), and the NGG PAM show most growth, while lower expression of Cas9 in presence of NGG PAM shows higher growth when compared to the non-canonical PAMs. While the growth rate is higher for the non-canonical PAMs with higher expression of Cas9, it is still significantly lower (3-fold) than the growth rate with canonical PAMs.[Fig. 18A, Fig.19]

The data show that higher expression of Cas9 is required for activity, and that the Cas9 is substantially worse at binding non-canonical PAMs as compared to canonical ones. There is scope for substantial improvement in the ability of Cas9 to bind non-canonical PAMs, and increasing concentrations of 3-AT can retard growth rate when binding is inefficient.

7.2.3 V1: Designing a scalable experiment

Evolving Cas9 to better bind the non-canonical PAM, required Cas9 to attempt to bind to a variety of PAMs. To that extent, we designed a scalable experiment that utilized yeast mating behavior.

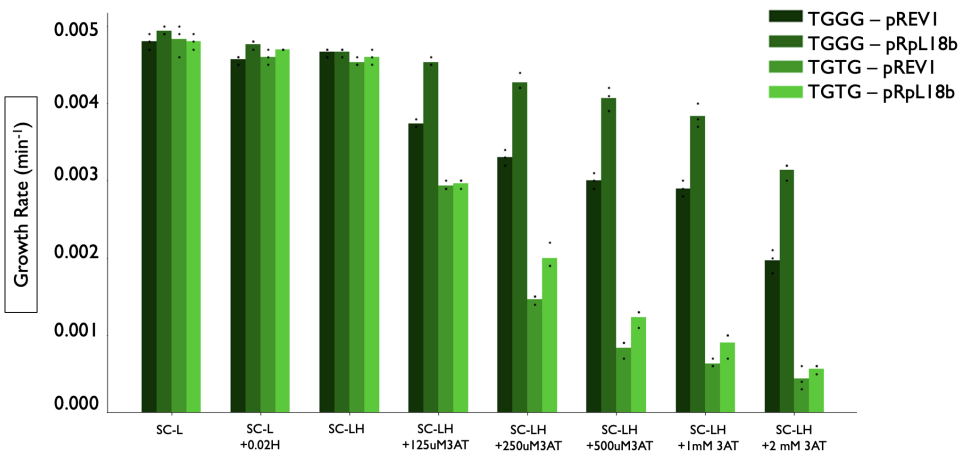


Figure 20: Growth of Cas9 depends on the PAM and level of Cas9 expression

The figure shows changing concentrations of Leucine, Histidine and 3AT in the media and the growth rate per minute for each strain comparing two different PAMs and with two promoters with different expression strength

Commonly used laboratory yeast strains are haploid in nature and may have mating type either a or α. Mating occurs when a cells are in close physical proximity to α cells (Merlini, Dudin, & Martin, 2013). The a cells release a-factor pheromones that signal to α cells that a cells are near. These cells of the opposite mating type then create a projection called a shmoo towards each other leading to the formation of a diploid cell by first fusing the cell walls, and cytoplasm and ultimately the nuclear envelopes (Merlini et al., 2013). Typically, mating results in the scrambling of genes. these diploid cells can then be sporulated to form haploids again (Merlini et al., 2013). However, scrambling of the genome will result in the loss of the nucleus required for evolution, and a kar1Δ13 mutation can prevent nuclear fusion (Vallen, Hiller, Scherson, & Rose, 1992).

For our first go around of this experiment, we designed it such that the BY4742 with mating type α would contain the appropriate nucleus and the a-type BY4741 cells would contain the Cas9 encoded p1. We integrated the error-prone polymerase with the pPSP2 promoter using the NatMX cassette as the yeast selectable marker at the can1 locus. After confirming the integration, we transformed the sgRNA with the promoter and terminator, as well as the target and downstream HIS3 with a minimal promoter and KanMX as the selectable marker at the HO locus. The HO integration cassette was made as a library to ensure a variety of PAMs. Once these plasmids were sequenced, we picked 5 different PAM, TGGG, TGAG, TGCG, TGTG, TAAA. The TGGG PAM is

the canonical PAM, while the TAAA PAM is the furthest away. The other PAMs included vary the third position, where a G is typically critical for binding and recognition. Thus, after integration of the HO cassette we now have 5 yeast strains. [Fig. 20]

The a-type BY4741 strain contained the kar1Δ13 mutation introduced by Cas9. This strain previously contained the MET15 landing pad onto which we integrated the dCas9-VP64 with a LEU2 selectable marker. While this transformation was not very efficient, it yielded some colonies which when miniprepped had the correct size p1. This strain had a wild-type polymerase in the genome and no error-prone polymerase. Once this strain was confirmed, it was mated with each of the five BY4742 strains on YPD plates overnight and then plated onto SC-LR+Nat+G418+Canavanine (200mg/L) to select for the correct nucleus and the correct cytoplasm containing the transformed p1. The colonies took about 3 days to be fully grown up. They were then picked into SC-L media and passaged twice in 2 ml of media to stabilize the copy number until ready for evolution. These strains were then stocked in 50% glycerol and then also passaged into SC-L + 0.02H. After successively passaging into SC-LH followed by SC-LH + 125uM 3AT, then SC-LH+150uM 3AT, then SC-LH+175uM 3AT etc. The data from this experiment showed that the cells did grow faster over time, but when plated onto YPD+Hyg plates, there was growth observed showing that the cells were likely diploid. PCR to confirm mating type also showed that there were cells that were diploid, or showed both a and α bands on the gel from the PCR suggesting that the cells were either diploid or that they contained both a and α nuclei.

To repeat this experiment, we needed to ensure that the other nucleus that we were selecting against was not able to survive. To that end, we increased the concentration of canavanine used in the plates from 200mg/L to 600mg/L. Additionally, after mating, we streaked the matings out on SC-LR+Nat+G418+canavanine plates and then patched single colonies onto the same kind of plate. We then replica plated each plate onto individual plates to check for markers both for the correct nucleus and against the wrong nucleus. The details of that are shown below. Only colonies that grew on the correct plates and did not grow on the ones they shouldn't have grown were picked for evolution.

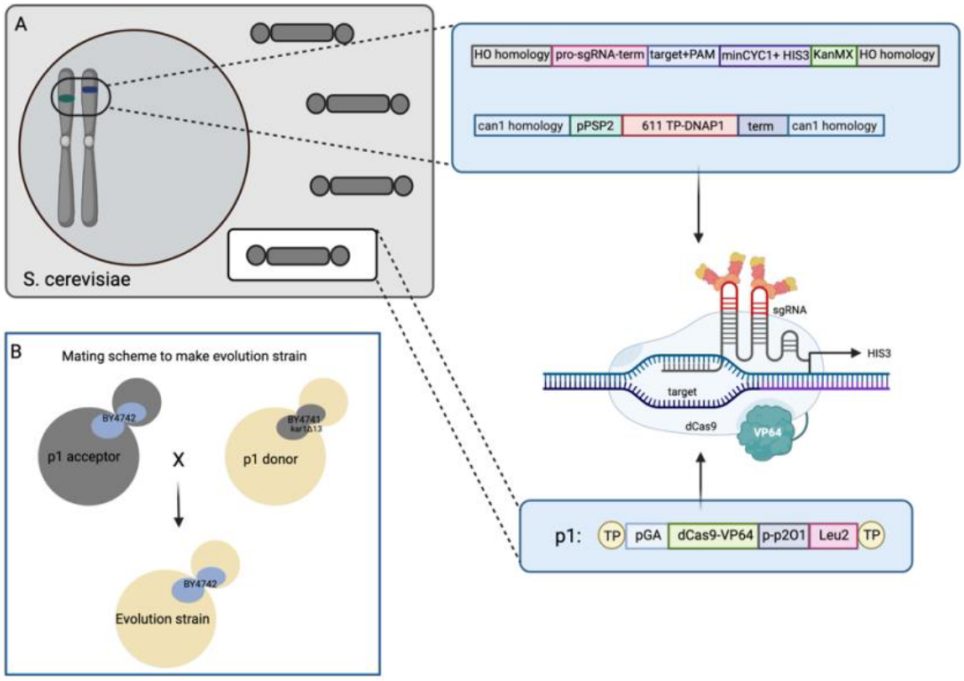


Figure 21: V1 of the Cas9 Evolution

A: Shows schematic of yeast cell with genomic integration and p1 integration of dCas9-VP64 with Leu2. Shows all the elements interacting at the location of the target and HIS3 being produced. B: Shows mating technique used to transfer p1 from the p1 donor to the p1 acceptor resulting in the final evolution.

Selection	BY4742 cell	BY4741 cell	Mated cell (BY4742 + p1)
SC-L	Should not grow	Should grow	Should grow
Nat	Should grow	Should not grow	Should grow
G418	Should grow	Should not grow	Should grow
Hyg	Should not grow	Should grow	Should now grow
SC-R + Canavanine	Should grow	Should not grow	Should grow
SC-H	Should grow	Should not grow	Should grow

A set of control experiments were performed to determine validity of the data. We showed that in absence of the dCas9-VP64 located on p1, the PAM and the HIS3 gene, no growth was observed in SC-LH, SC-LH+100uM 3AT, SC-LH+150uM 3AT and SC-LH+200uM 3AT showing that the parent strain did not have HIS3 or LEU2 and that any growth observed was due to the transformation that had been performed. This evolution scheme followed similar to the first one. However, after 7 passages, a long miniprep was done and the HO locus was amplified and sequenced to confirm that the PAMs had not been mutated to NGGN PAMs. Sequencing revealed recombination at the HO

locus between the 20bp target sequence and the upstream located sgRNA sequence that contained the identical spacer. This led to a loss of the sgRNA required for Cas9 binding and part of the p1 was lost since Cas9 was no longer necessary for survival of the cell.[Fig. 20B]

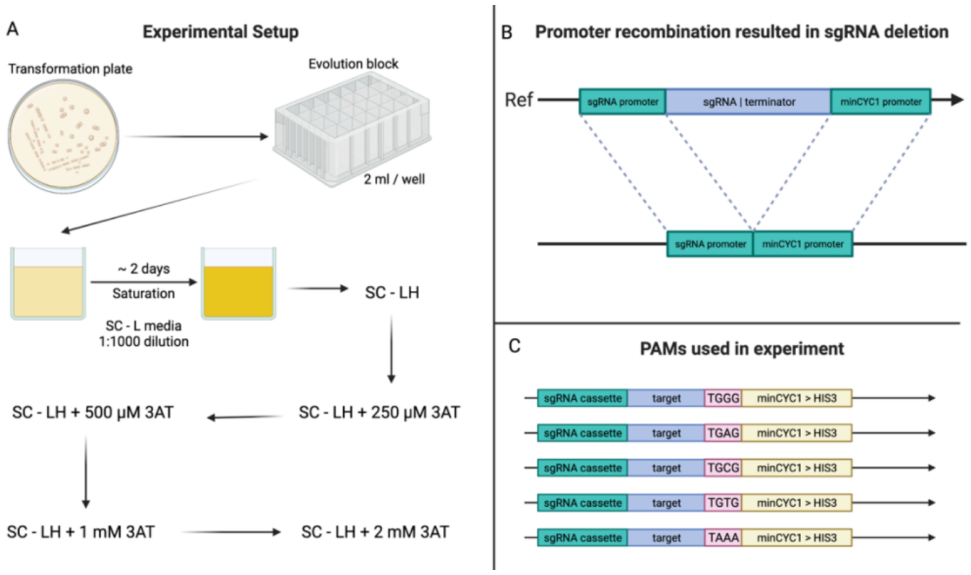


Figure 22: Results for V1 of the Cas9 Evolution

A: Shows passaging scheme for initial evolution. B: Shows the result of the evoution experiment - recombination leading to loss of sgRNA. C: shows the 5 PAMs used in the initial evolution experiment

7.2.4 V2: Changing architecture to prevent recombination

Since the 20bp in the target recombined with the 20bp spacer in the sgRNA sequence leading to the loss of the guide RNA, we needed to redesign the strains to prevent recombination. To make evolution more simple, and to test whether one of several new options of architectures would work, we decided to evolve the Cas9 to bind and recognize the NGAG PAM which showed lower activity when compared to the NGGG PAM. The added bonus was that mutations from previously engineered Cas9 proteins to show activity towards NGAG PAMs have been identified and their locations confirmed. This information gives us a sense of what is possible and what to expect when performing these experiments. To prevent recombination, we integrated the target, PAM and HIS3 gene with the Kan selectable marker at HO and integrated the sgRNA with its promoter and terminator at the ade2 locus with a HygMX cassette. This time we also used a variety of combinations of polymerases to affect error-rate in an effort to speed up the number of mutations that could occur on Cas9. Some back of the envelope calculations as shown below were performed

to determine the evolution volumes and the bottlenecks required to be able to detect and sample the evolved variants. This math was done incorrectly in the beginning, but the correct calculations are shown below. In addition to the BY4742 strains, we also used an F102-2 strain since that one has shown more stability when maintaining p1 over long periods of time. We also performed a spot assay to ensure that 3AT does not interfere with the functioning of the media and other processes of yeast and that growth on plates or liquid containing 3AT. For this, we used all the strains that we made for the evolution and made serial dilutions of 1:10, 1:100, 1:1000 and 1:10,000 on plates made with a variety of media. The data is shown in Fig. 22D.

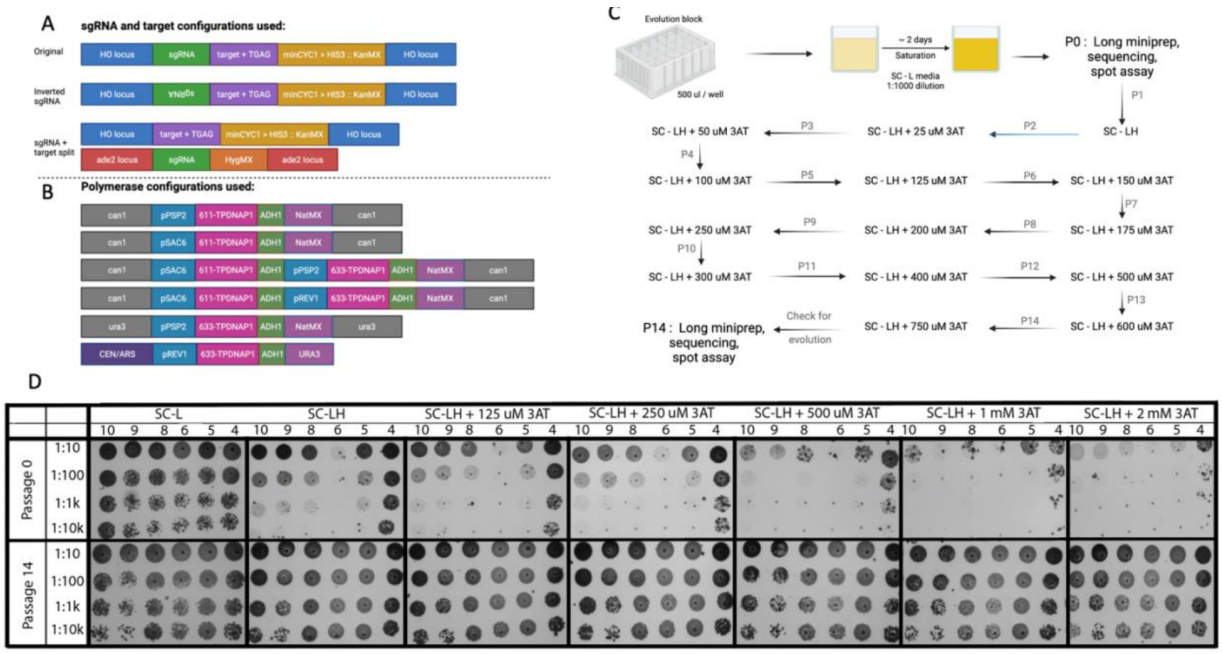


Figure 23: Even after 14 passages, there was no evidence of evolution

[A] shows the sgRNA and target configurations used in the experiments. [B] the many options for error-prone polymerase combinations. [C] Passaging scheme for the evolution experiment. [D] Spot assay data comparing six strains, at different concentrations of 3AT before evolution (Passage 0) and at the end of evolution (Passage 14). The six strains are in table..

$$\text{Mutation rate of 611} = 6.32 * 10^{-7}$$

$$\text{Copy number of of 611} \approx 20$$

$$\text{Size of dCas9-VP64} = 4362 \text{ bp}$$

$$\text{Possible mutations per base pair} = 3$$

$$\text{No. of possible mutants} = \text{length of gene x possible mutations per base}$$

$$\text{No. of possible mutants} = 4362 * 3 = 1.3 * 10^4$$

No. of bases replicated per cell division = length of gene x copy number

No. of bases replicated per cell division = 4362 x 20 = 8.7 x 10^4

No. of bases to be replicated for 1 mutation = $6.32 * 10^7$

No. of mutations per cell = $\frac{8.7*10^4}{6.32*10^7} = 1.3 * 10^{-3}$

No. of cells to cover all single mutants = $\frac{No. of single mutants}{No. of mutations per cell}$

No. of cells to cover all single mutants = $\frac{1.3*10^4}{1.3*10^{-3}} = 1 * 10^7$

The average number of yeast cells in 1 ml of OD600 = $1.5 * 10^7$

1 ml of OD600 at 1.0 should be used as a bottleneck to cover all single mutants.

After 14 successive passages, there seemed to be an improvement in growth of the strains over time, however the control strain was also growing better after successive passaging. Spot assay data showed no evidence of evolution and the long minipreps showed that all of the p1 were at the expected size, there were no mutations that had accumulated on p1. We suspected that the strain had adapted to the higher concentrations of 3AT in the media and had adapted to survive without the need for Cas9. To confirm our suspicions, we used the control (containing Cas9 on a CENARS plasmid) and grew it in SC-LH media and then did a 1:500 stringent passage into 125uM 3AT and followed the selection scheme showed in figure ... Since there was no evolution possible, the only way for the strain to survive at higher concentrations was by utilizing other pathways. Our best guess was that the a histidine permease that would have transported 3AT into the cell was altered in some way to prevent this transport making the cells fully immune to the concentration of 3AT in the media. And we did observe this when after 3 passages, the strains were able to grow with no problem at even 20 mM of 3AT. We realized that this selection scheme could not work to evolve Cas9. [Fig.23]

7.2.5 V3: Using FACS for non-continuous evolution

We changed our strategy a third time, to evolve Cas9. For this attempt, we decided to not rely on titrating a competitive inhibitor and instead chose to make the selection less continuous by introducing the need for fluorescence based cell-sorting. The dCas9-VP64 is still encoded on p1, while the sgRNA is encoded at ade2 and the target and GFP are encoded at the HO locus. The error-prone polymerases continue to be encoded at the can1 locus. The same BY4741 strain was

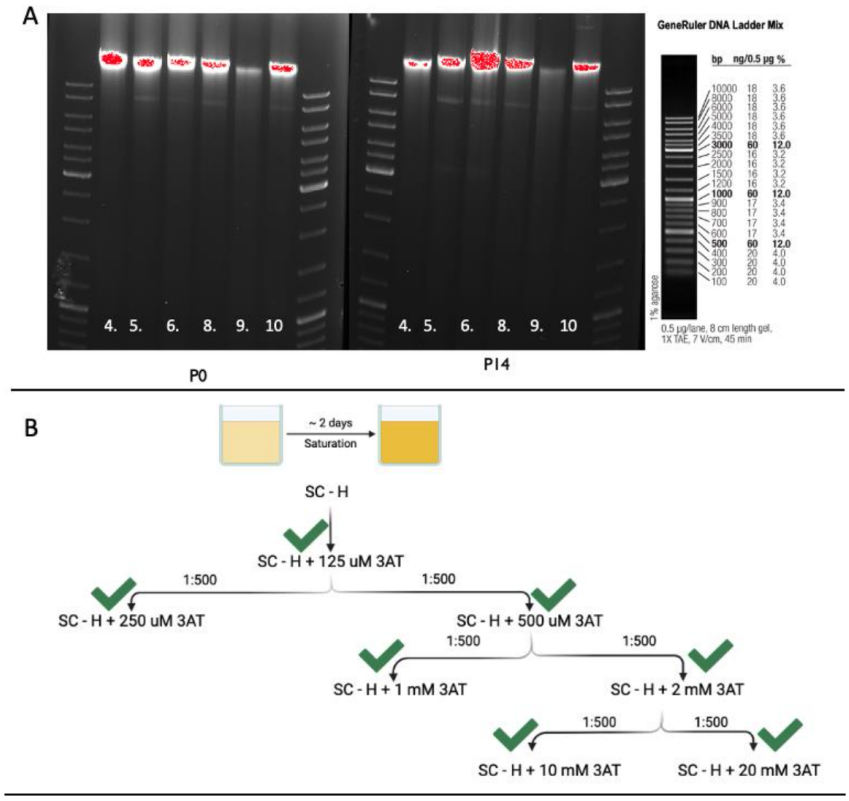


Figure 24: Adaptation of strain led to no evolution of Cas9

[A] Shows gel images of proper p1 integration. The wild-type p2 band is not clearly visible, it is of size 13.4 kB and should be closest to the brightest band at the top which is genomic DNA. The band visible between the 6kB and 8kB markers on the ladder is the expected p1 size after integration. [B] Evolution scheme for test evolution to see if he background strain can grow at increasing concentrations of 3AT.

mated into the BY4742 strain once the correct nucleus was created. The only difference between this evolution and the previous one was the replacement of HIS3 by GFP and the subsequent use of FACS based sorting to select for the highest expressing cells instead of the use of 3AT. [Fig.24]

In the absence of the GFP cassette, there is minimal fluorescence of the cells, showing that there is no autofluorescence that is interfering with the measurements of fluorescence. When a GGG PAM is used, and the no Cas9, low expression Cas9 and medium expression of Cas9 from a CEN/ARS plasmid are compared, highest fluorescence is observed at the higher expression of Cas9. However, when a GAG PAM is used, the levels of fluorescence are very similar showing that there is scope for improvement for the recognition of GAG PAM by evolving Cas9. When we did the evolution, we observed that even after 5 successive passages, there was no evidence of evolution when comparing increased fluorescence of GFP in presence of GAG PAM. Moreover, bulk

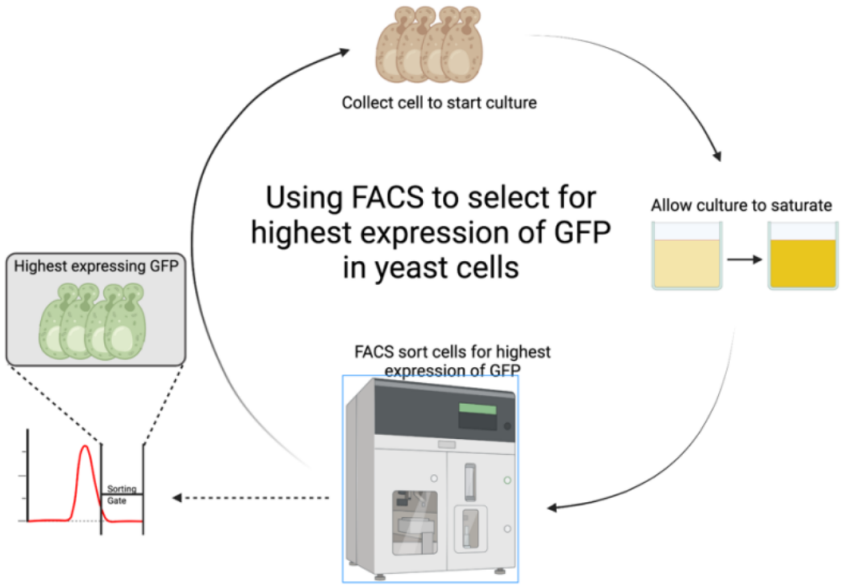


Figure 25: Schematic showing the process for evolution using FACS

sequencing of p1 showed no mutations.

The mutation spectrum of the 611 and 633 polymerases is such that the mutations they make are 95.1% transitions and 4.9% transversions. To allow Cas9 to recognize and bind to GAG PAMs, the EQR and VQR mutations as described in ... are required. The majority of these mutations are tranversion mtations which are currently very hard to access using OrthoRep. The transversion mutations are detailed below. Transition mutations are easily accessible to OrthoRep, but they are not very useful for the Cas9 evolution. Thus, engineering approached would be necessary to allow such transition mutations to occur more easily on Cas9. However, making these mutations would only be a temporary fix in this evolution process. To be able to evolve more complex targets, more work is needed to upgrade OrthoRep, which will be useful to the community at large as well.

7.2.6 Learnings and Conclusions

There were several learnings from this project that are applicable to improving how OrthoRep works and to get a better sense of why certain evolutions may fail. One of the important lessons is that sustained exposure to chemicals like 3-AT, that may be toxic at small quantities when not seen often by the cells, may lose their potency over time rendering them ineffective. Yeast cells have survived many millions of years and have the ability to adapt to their surroundings in a way that

promotes their survival very easily. These changes may be due to mutations that may be detected using sequencing but they may also be harder to detect. Some of these may include epigenetic tags that may suppress the expression of certain genes, or may increase promoter strength in the genes that cause higher survival or faster growth.

Second, since OrthoRep is dependent on the error-prone polymerases replicating the genes of interest on p1, it is essential that the polymerases have balanced transition and transversion mutation rates. Transition mutations are able to mutate A ¡-¿ G and G ¡-¿ A as well as C ¡-¿ T and T ¡-¿ C. However these only comprise one-third of the possible mutations when assuming that A, T, G and C can all be mutated to all other residues at any given location. The older generation of polymerases 611 and 633, used for evolving Cas9 are able to make 95.1% transitions and only 4.9% transversion mutations. For evolutions requiring transversion mutations, the newer polymerases with more balanced transition and transversion rates are essential for improving the range of proteins that can be evolved and the range of function that is available. Transition mutations also allow us to sample more sequence space and will allow us to reach newer fitness peaks.

Third, when evolving proteins of larger size, the size of p1 increases and in this case it was close to 6 kb. The increase in p1 size allows for replication of fewer p1 copies being replicated by the error-prone polymerase in the time taken for each cell division. As a result, the number of copies within each cell will be reduced. If we assume that the number of p1 copies do not change, even with a larger p1, for evolution to be as effective, bottleneck sizes which would depend on the number of all single mutations, which would increase leading to increased evolution volumes. An increase in evolution and thus passaging volumes complicated the number of replicates that can be performed and processes easily during the course of an evolution campaign. For labs getting started with OrthoRep or unable to passage large volumes of culture, using this technology might be a hindrance.

7.2.7 Possible future directions

I believe that there is still some value in evolving Cas9 to bind any given 20-23 bp sequence in the human genome. As new technology like base editors and prime editors get invented and becomes mainstream for treating human disease, it is essential to be able to expand the targeting range and

improve the specificity of Cas9. While SaCas9 given its smaller size is preferred since the protein and all of its accompaniments can fit into an AAV, it would be prudent to evolve that version of Cas9 instead. The learnings from these Cas9 experiments have illustrated the ways in which evolutions can fail and also highlight the current limitations of OrthoRep. These are great opportunities to develop new techniques like the recently developed polymerases that have both higher error-rate and higher copy-number as well as a more balanced mutation spectrum as compared to the previous polymerases. These newer polymerases will enable faster editing, and will also allow all kinds of mutations to accrue on p1.

References

Anders, C., Niewoehner, O., Duerst, A., & Jinek, M. (2014). Structural basis of pam-dependent target dna recognition by the cas9 endonuclease. *Nature*, *513*, 569-573.

Anzalone, A. V., Randolph, P. B., Davis, J. R., Sousa, A. A., Koblan, L. W., Levy, J. M., ... Liu, D. R. (2019). Search-and-replace genome editing without double-strand breaks or donor dna. *Nature*, *576*, 149–157.

Arnold, F. (1998). Design by directed evolution. *Accounts of Chemical Research*, *31*(3).

Arzumanyan, G. A., Gabriel, K. N., Ravikumar, A., Javanpour, A. A., & Liu, C. C. (2018). Mutually orthogonal dna replication systems in vivo. *ACS Synthetic Biology*, *7*, 1722-1729.

Badran, A. H., & Liu, D. R. (2015). In vivo continuous directed evolution. *Current Opinion in Chemical Biology*, *24*, 1-10.

Beerli, R. R., Segal, D. J., Dreier, B., & III, C. F. B. (1998). Toward controlling gene expression at will: Specific regulation of the erbb-2/her-2 promoter by using polydactyl zinc finger proteins constructed from modular building blocks. *PNAS*, *95*(25), 14628-14633.

Cadwell, R. C., & Joyce, G. F. (1992). Randomization of genes by pcr mutagenesis. *Genome Research*, *2*(1), 28-33.

Caufield, J., Sakhawalkar, N., & Uetz, P. (2012). A comparison and optimization of yeast two-hybrid systems. *Methods*, *58*, 317-324.

Chen, R. (2001). Enzyme engineering: rational redesign versus directed evolution. *Trends in Biotechnology*, *19*(1), 13-14.

Chevallereau, A., Meaden, S., van Houte, S., Westra, E. R., & Rollie, C. (2019). The effect of bacterial mutation rate on the evolution of crispr-cas adaptive immunity. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *374*(1772).

Cobb, R. E., Chao, R., & Zhao, H. (2013, May). Directed evolution: Past, present and future. *AICHe Journal*, *59*(5), 1432-1440.

Denes, C. E., Cole, A. J., Tran, M. T. N., Khalid, M. K. N. M., Hewitt, A. W., Hesselson, D., & Neely, G. G. (2022). "cheater" particles render the vegas platform unsuitable for mammalian directed evolution. *bioRxiv*.

Engler, C., Kandzia, R., & Marillonnet, S. (2008). A one pot, one step, precision cloning method

with high throughput capability. *PLOS One*, *3*(11), e3647.

English, J. G., Olsen, R. H. J., Lansu, K., Patel, M., White, K., Cockrell, A. S., ... Roth, B. L. (2019). Vegas as a platform for facile directed evolution in mammalian cells. *Cell*, *178*, 748-761.

Engqvist, M. K. M., & Rabe, K. S. (2019). Applications of protein engineering and directed evolution in plant research. *Plant Physiology*, *179*, 907-917.

Esvelt, K. M., Carlson, J. C., & Liu, D. R. (2011). A system for the continuous directed evolution of biomolecules. *Nature*, *472*, 499-503.

Fujimura, H., Hishinuma, F., & Gunge, N. (1987). Terminal segment of kluyveromyces lactis linear dna plasmid pgkl2 supports autonomous replication of hybrid plasmids in saccharomyces cerevisiae. *Current Genetics*, *12*(2), 99-104.

Fujimura, H., Yamada, T., Hishinuma, F., & Gunge, N. (1987). Dna replication in vivo of linear dna killer plasmids pgkl1 and pgkl2 in saccharomyces cerevisiae. *FEMS Microbiology Letters*, *49*(3), 441-444.

Garcia-Garcia, J. D., Gelder, K. V., Joshi, J., Bathe, U., Leong, B. J., Bruner, S. D., ... Hanson, A. D. (2022). Using continuous directed evolution to improve enzymes for plant applications. *Journal of Plant Physiology*, *188*, 971-983.

Gunge, N. (1986). Linear dna killer plasmids from the yeast kluyveromyces. *Yeast*, *2*(3), 153-162.

Gunge, N., Fukuda, K., Takahashi, S., & Meinhardt, F. (1995). Migration of the yeast linear dna plasmid from the cytoplasm into the nucleus in saccharomyces cerevisiae. *Current Genetics*, *28*(3), 280-288.

Gunge, N., Murata, K., & Sakaguchi, K. (1982, July). Transformation of saccharomyces cerevisiae with linear dna killer plasmids from kluyveromyces lactis. *Journal of Bacteriology*, *151*(1), 462-464.

Gunge, N., & Sakaguchi, K. (1981, July). Intergeneric transfer of deoxyribonucleic acid killer plasmids, pgkl1 and pgkl2, from kluyveromyces lactis into saccharomyces cerevisiae by cell fusion. *Journal of Bacteriology*, *147*(1), 55-160.

Gunge, N., Tamaru, A., Ozawa, F., & Sakaguchi, K. (1981, January). Isolation and characterization of linear deoxyribonucleic acid plasmids from kluyveromyces lactis and the plasmid-associated killer character. *Journal of Bacteriology*, *145*(1), 382-390.

Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V., & Dueber, J. E. (2018). Crispr-guided dna polymerases enable diversification of all nucleotides in a tunable window. *Nature*, *560*, 248-252.

Hu, J. H., Miller, S. M., Guerts, M. H., Tang, W., Chen, L., Sun, N., ... Liu, D. R. (2018). Evolved cas9 variants with broad pam compatibility and high dna specificity. *Nature*, *556*, 57-63.

Javanpour, A. A., & Liu, C. C. (2019, May). Genetic compatibility and extensibility of orthogonal replication. *ACS Synthetic Biology*, *8*, 1249-1256.

Javanpour, A. A., & Liu, C. C. (2021). Evolving small-molecule biosensors with improved performance and reprogrammed ligand preference using orthorep. *ACS Synthetic Biology*, *10*(10), 2705–2714.

Jeske, S., Tiggeman, M., & Meinhardt, F. (2006). Yeast autonomous linear plasmid pgkl2: Orf9 is an actively transcribed essential gene with multiple transcription start points. *FEMS Microbiology Letters*, *255*(2), 321-327.

Jinek, M., Chylinski, K., Fonfara, K., Hauer, I., Doudna, J., & Charpentier, E. (2012). A programmable dual-rna-guided dna endonuclease in adaptive bacterial immunity. *Science*, *337*(6096), 816-821.

Jinek, M., Jiang, F., Taylor, D. W., Sternberg, S. H., Kaya, E., Ma, E., ... Doudna, J. A. (2014). Structures of cas9 endonucleases reveal rna-mediated conformational activation. *Science*, *343*(6176), 1247997-1 - 1247997-11.

Kamper, J., Esser, K., Gunge, N., & Meinhardt, F. (1991). Heterologous gene expression on the linear dna killer plasmid from kluyveromyces lactis. *Current Genetics*, *19*(2), 109-118.

Kikuchi, Y., Hirai, K., & Hishinuma, F. (1984, July). The yeast linear dna killer plasmids, pgkl1 and pgkl2, possess terminally attached proteins. *Nucleic Acids Research*, *12*(14), 5685-5692.

Kleinstiver, B. P., Pattanayak, V., Prew, M. S., Tsai, S. Q., Nguyen, N. T., Zheng, Z., & Joung, J. K. (2016). High-fidelity crispr–cas9 nucleases with no detectable genome-wide off-target effects. *Nature*, *529*, 490-495.

Kleinstiver, B. P., Prew, M. S., Tsai, S. Q., Topkar, V. V., Nguyen, N. T., Zheng, Z., ... Joung, J. K. (2015). Engineered crispr-cas9 nucleases with altered pam specificities. *Nature Biotechnology*, *523*(7561), 481-485.

Labun, K., Montague, T. G., Krause, M., Cleuren, Y. N. T., Tjeldnes, H., & Valen, E. (2019).

Chopchop v3: expanding the crispr web toolbox beyond genome editing. *Nucleic Acids Research*, *47*, W171–W174.

Larsen, M., Gunge, N., & Meinhardt, F. (1998). Kluyveromyces lactis killer plasmid pgkl2: evidence for a viral-like capping enzyme encoded by orf3. *Plasmid*, *40*(3), 243-246.

Larson, M. H., Gilbert, L. A., Wang, X., Lim, W. A., Weissman, J. S., & Qi, L. S. (2013). Crispr interference (crispri) for sequence-specific control of gene expression. *Nature Protocols*, *8*, 2180–2196.

Lee, M. E., DeLoache, W. C., Cervantes, B., & Dueber, J. E. (2015). A highly characterized yeast toolkit for modular, multipart assembly. *ACS Synthetic Biology*, *4*(9), 975-986.

Mali, P., Esvelt, K. M., & Church, G. M. (2013). Cas9 as a versatile tool for engineering biology. *Nature Methods*, *10*(10), 957-963.

McNeel, D., & Tamanoi, F. (1991). Terminal region recognition factor 1, a dna-binding protein recognizing the inverted terminal repeats of the pgkl linear dna plasmids. *Proceedings of the National Academy of Sciences*, *88*(24), 11398-11402.

Merlini, L., Dudin, O., & Martin, S. G. (2013). Mate and fuse: How yeast cells do it. *Open Biology*, *3*(130008).

Miller, S. M., Wang, T., Randolph, P. B., Arbab, M., Shen, M. W., Huang, T. P., ... Liu, D. R. (2020). Continuous evolution of spcas9 variants compatible with non-g pams. *Nature Biotechnology*, *38*, 471-481.

Molina, R. S., Rix, G., A, A. M., Álvarez, B., Seo, D., Chen, H., ... Liu, C. C. (2022, May). In vivo hypermutation and continuous evolution. *Nature Methods*, *2*(36), 1-22.

Moore, C. L., Papa, L. J., & Shoulders, M. D. (2018). A processive protein chimera introduces mutations across defined dna regions in vivo. *Journal of the American Chemical Society*, *140*, 11560-11564.

Nishimasu, H., Shi, X., Ishiguro, S., Gao, L., Hirano, S., Okazaki, S., ... Nureki, O. (2018). Engineered crispr-cas9 nuclease with expanded targeting space. *Science*, *361*, 1259–1262.

Perez-Pinera, P., Kocak, D. D., Vockley, C. M., Adler, A. F., Kabadi, A. M., Polstein, L. R., ... Gersbach, C. A. (2013). Rna-guided gene activation by crispr- cas9–based transcription factors. *Nature Methods*, *10*(10), 973-976.

R, S., & Meacock, P. (1995). Kluyveromyces lactis killer plasmid pgkl2: Molecular analysis of an

essential gene, orf5. *Yeast*, *11*(7), 615-628.

Ravikumar, A., Arrieta, A., & Liu, C. C. (2014). An orthogonal dna replication system in yeast. *Nature Chemical Biology*, *10*.

Ravikumar, A., Arzumanyan, G. A., Obadi, M. K., Javanpour, A. A., & Liu, C. C. (2018). Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell*, *175*, 1946–1957.

Rees, H., & Liu, D. R. (2018). Base editing: precision chemistry on the genome and transcriptome of living cells. *Nature Reviews*, *19*, 770-788.

Rix, G., & et.al. (2023). In preparation.

Rix, G., Watkins-Dulaney, E. J., Almhjell, P. J., Boville, C. E., Arnold, F. H., & Liu, C. C. (2020). Scalable continuous evolution for the generation of diverse enzyme variants encompassing promiscuous activities. *Nature Communications*, *11*.

Romanos, M., & Boyd, A. (1988). A transcriptional barrier to expression of cloned toxin genes of the linear plasmid k1 of kluyveromyces lactis: evidence that native k1 has novel promoters. *Nucleic Acids Research*, *16*(15), 7333-7350.

Schaffrath, R., Meinhardt, F., & Meacock, P. (1997). Orf7 of yeast plasmid pgkl2: analysis of gene expression in vivo. *Current Genetics*, *31*(2), 190-192.

Seipel, K., Georgiev, O., & Schaffner, W. (1992). Different activation domains stimulate transcription from remote ('enhancer') and proximal ('promoter') positions. *The EMBO Journal*, *11*(13), 4961-4968.

Siloto, R. M. P., & Weselake, R. J. (2012). Site saturation mutagenesis: Methods and applications in protein engineering. *Biocatalysts and Agricultural Biotechnology*, *1*(3), 181-189.

Stark, M., Boyd, A., Mileham, A., & Romanos, M. A. (1990). The plasmid-encoded killer system of kluyveromyces lactis: a review. *Yeast*, *6*(1).

Stemmer, W. P. C. (1994). Rapid evolution of a protein in vitro by dna shuffling. *Nature*, *370*, 389-391.

Sykora, M., Pospisek, M., Novak, J., Mrvova, S., Krasny, L., & Vopalensky, V. (2018). Transcription apparatus of the yeast virus-like elements: Architecture, function, and evolutionary origin. *PLOS Pathogens*, *14*(10).

Taylor, M. R., Simon, E. J., Dickey, J. L., & Hogan, K. (2022). *Campbell biology.* Pearson

Education.

Tokunaga, M., Wada, N., & Hishinuma, F. (1987). Expression and identification of immunity determinants on linear dna killer plasmids pgkl1 and pgkl2 in kluyveromyces lactis. *Nucleic Acids Research*, *15*(3), 1031–1046.

Tommasino, M., Ricci, S., & Galeotti, C. (1988). Genome organization of the killer plasmid pgkl2 from kluyveromyces lactis. *Nucleic Acids Research*, *16*(13), 5863-5878.

Vallen, E. A., Hiller, M. A., Scherson, T. Y., & Rose, M. D. (1992). Separate domains of kar1 mediate distinct functions in mitosis and nuclear fusion. *Journal of Cell Biology*, *117*(6), 1277-1287.

Wellner, A., McMahon, C., Gilman, M. S. A., Clements, J. R., Clark, S., Nguyen, K. M., ... Liu, C. C. (2021). Rapid generation of potent antibodies by autonomous hypermutation in yeast. *Nature Chemical Biology*, *17*, 1057–1064.

Wesolowski, M., Algeri, A., Goffrini, P., & Fukuhara, H. (1982, August). Killer dna plasmids of the yeast kluyveromyces lactis. *Current Genetics*, *5*(3), 191-197.

Xie, H., Tang, L., He, X., Liu, X., Zhou, C., Liu, J., ... Song, Z. (2018). Sacas9 requires 5-nngrrt-3 pam for sufficient cleavage and possesses higher cleavage activity than spcas9 or fncpf1 in human cells. *Biotechnology Journal*, *13*(3).

Zhong, Z., Ravikumar, A., & Liu, C. C. (2018). Tunable expression systems for orthogonal dna replication. *ACS Synthetic Biology*, *7*(12), 2930–2934.