UNIVERSITY OF CALIFORNIA,
IRVINE

Advances in Orthogonal DNA Replication and Application to Small-Molecule Biosensor Evolution

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Alex Arash Javanpour

Dissertation Committee:
Associate Professor Chang C. Liu, Chair
Assistant Professor Tim Downing
Professor Andrej Lupták
Professor Nancy da Silva

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express my gratitude and appreciation to my advisor and mentor, Dr. Chang C. Liu. I will always value his mentorship and his drive to push me towards the finish line, despite when experiments were not going my way, as well as his patience.

I also would like to express my appreciation to my lab mates. I was fortunate to work with some incredibly bright scientists and I am grateful for the advice and insight they shared with me, both within and outside science, as well as for listening to my frequent ramblings. I will miss the time we spent discussing science, cycling throughout Orange County and down to San Diego, watching basketball, and playing music loudly in lab late at night.

Lastly, I would like to thank my parents, Hamid and Zarmina Javanpour. It was through them that I learned what hard work, conviction, and dedication really meant.

Portions of this text are reprints of material as they appear in *Cell* and *ACS Synthetic Biology.* Co-authors on these works include Arjun Ravikumar, Garri A. Arzumanyan, Muaeen K.A. Obadi, and Chang C. Liu.

CURRICULUM VITAE

Alex Arash Javanpour

EDUCATION

Doctor of Philosophy in Biomedical Engineering 2020
University of California, Irvine

Bachelor of Science in Bioengineering 2012
University of California, Berkeley

PUBLICATIONS

- Jensen ED, Ambri F, Bendtsen MB, **Javanpour AA,** Liu CC, Jensen MK, Keasling JD. Integrating continuous evolution with high-throughput screening for optimization of *cis,cis*-muconic acid production in yeast. *(Submitted)*.

- **Javanpour AA**, Liu CC. Using OrthoRep to evolve a small-molecule biosensor for reprogrammed ligand specificity. *(In preparation)*.

- **Javanpour AA**, Liu CC. Facile transfer of OrthoRep via cytoduction. *(In preparation)*.

- Garcia-Garcia JD, Joshi J, Patterson JA, Trujillo-Rodriguez L, Reisch CR, **Javanpour AA**, Liu CC, Hanson AD. Potential for Applying Continuous Directed Evolution to Plant Enzymes: An Exploratory Study. *Life*, 10, 178 (2020).

- **Javanpour AA**, Liu CC. Genetic compatibility, portability, and extensibility of an orthogonal DNA replication system. *ACS Synthetic Biology*, 8, 1249-1256 (2019).

- Ravikumar A, Arzumanyan GA, Obadi MKA, **Javanpour AA**, Liu CC. Scalable, Continuous Evolution of Genes at Mutation Rates Above Genomic Error Thresholds. *Cell*, 175, 1946-1957 (2018).

- Arzumanyan GA, Gabriel KN, Ravikumar A, **Javanpour AA**, Liu CC. Mutually Orthogonal DNA Replication Systems in vivo. *ACS Synthetic Biology*, 7, 1722-1729 (2018).

- Mahboobi SH, **Javanpour AA**, Mofrad MRK. The interaction of RNA helicase DDX3 with HIV-1 Rev-CRM1-RanGTP complex during the HIV replication cycle. *PloS One*, 10(2), (2015).

AWARDS AND HONORS

- National Science Foundation Graduate Research Fellowship Program, Honorable Mention (2014)
- Graduate Dean's Recruitment Fellowship, UC Irvine (2013)
- Entering Student Fellowship, UC Irvine (2013)

RESEARCH EXPERIENCE

Laboratory for Synthetic Evolution, *UC Irvine* 2013-2020
Graduate Student Researcher, Advisor: Prof. Chang C. Liu

Laboratory for Molecular Cell Biomechanics, *UC Berkeley* 2011-2013
Undergraduate Student Researcher, Advisor: Prof. Mohammad Mofrad

Lawrence Berkeley National Lab, *Berkeley, CA* 2009
Student Assistant

TEACHING EXPERIENCE

Graduate Teaching Assistant (University of California, Irvine) *Spring 2020*
BME 50B—Cell and Molecular Engineering (*Prof. Elliot Hui*)

Graduate Teaching Assistant (University of California, Irvine) *Winter 2015*
BME 50A—Cell and Molecular Engineering (*Prof. Chang C. Liu*)

ABSTRACT OF THE DISSERTATION

Advances in Orthogonal DNA Replication and Application to Small-Molecule Biosensor Evolution

by

Alex Arash Javanpour

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2020

Professor Chang C. Liu, Chair

Directed evolution applies Darwinian evolution to engineering new and optimized proteins, pathways, or cells. Traditional directed evolution techniques rely on labor-intensive *ex vivo* mutagenesis, transformation, and screening steps. Unlike natural evolution, however, this scheme is neither continuous nor capable of substantial parallelization. Our group has developed OrthoRep, an orthogonal DNA polymerase-plasmid pair in yeast that stably mutates ~100,000-fold faster than the host genome *in vivo*. User-defined genes in OrthoRep can be continuously and rapidly evolved under selection, enabling a new paradigm of high-throughput evolution of biomolecular and cellular function. Here, I describe technological advances to OrthoRep and an application towards evolving a small-molecule biosensor for new ligand specificity. First, I cover efforts to rationally engineer the OrthoRep DNA polymerase for higher mutation rates and determine the spontaneous recombination frequency of the OrthoRep plasmid. Next, I demonstrate facile transfer of OrthoRep plasmids between yeast strains via cytoduction. Third, I show the strain generality of OrthoRep and a CRISPR/Cas9 method for expedient genetic manipulations. Finally, I demonstrate successful evolution

towards reprogramming the specificity of a transcription-based biosensor towards a non-native ligand.

CHAPTER 1

Tackling protein engineering via continuous directed evolution

1.1 Classical directed evolution

Directed evolution applies Darwinian evolution to engineering new and optimized proteins, pathways, or cells. The standard process of directed evolution utilizes *ex vivo* PCR mutagenesis or recombination on gene(s) of interest (GOI) to create libraries of variants, which are then transformed into host cells and subject to selection for the desired function. By iterating rounds of mutation, transformation, and selection, the process results in convergence onto the desired function. Unlike natural evolution, however, this scheme is neither continuous nor capable of substantial parallelization because each round of evolution requires time consuming and labor-intensive *ex vivo* mutagenesis, transformation, and screening steps. Additionally, this approach can be problematic as protein libraries with millions of members still sample only a fraction of the sequence space possible for an average protein. Screening methods that introduce biases to the library and the degeneracy of the genetic code further restrict the library design.[1]

The field of *in vivo* continuous directed evolution tackles these logistical limitations by performing both mutagenesis of the GOI and selection entirely within a host cell.[2] By performing mutagenesis of GOIs within the cell, the experimenter is not required to carry out cycles of *in vitro* mutagenesis followed by library transformation.

Instead, GOIs can be continuously evolved through serial passaging of cells under selective conditions. This approach is limited only by the generation time of the host and the number of cells that can be cultured. In this chapter, I will first discuss screening methods employed in non-continuous evolution, then approaches used to attain continuous evolution, and conclude by describing a system that we have developed to achieve *in vivo* continuous evolution.

1.2 Non-continuous directed evolution

One of the most challenging steps of directed evolution is the efficient identification of desirable mutants from large libraries. Using the appropriate screen is crucial in directed evolution, as "you get what you screen for."[3] The simplest form of non-continuous evolution, plate screening has been widely applied for the identification of desirable enzyme variants, with agar plate screening and microtiter plate screening as the two most common approaches. [4] Libraries can be plated out on many agar plates to facilitate some form of visual screening. This can involve incubation of with a substrate to create a visual signal, such as fluorescence, color, or clear/halo zone. The drawbacks of agar plate screening are that it is low throughput, limited by the number of colonies per plate, and can suffer from poor sensitivity when compounds such as enzyme substrates are needed to be supplemented into the media. Microtiter plate screening has been implemented to engineering many proteins. Typical measurements of this method are those that can be readily determined by a microplate reader, such as fluorescence and luminescence.[5] Traditional enzyme activity assays can be performed in microtiter plates by supplying reaction components and purified proteins or even

crude cell extracts. Additionally, the throughput can be significantly improved with the utilization of robotic liquid handling systems.[6] For certain enzymatic reactions, the disappearance of substrates or formation of products can be easily identified by simple visual observation or measuring UV−Vis absorbance or fluorescence using a plate reader. However, these assays are highly dependent on the chemistry and availability of suitable native substrates. Lastly, throughput is limited by the number of wells per plate and can require substantial customization of automation solutions to achieve, at best, medium throughput screening.

Another common technique is surface display. This engineering approach is often used for improving binding affinity or stability of a protein and can be typically divided into phage display and yeast display. A common application, as seen with yeast display, is developing binding proteins such as antibodies and is in some cases applied to facilitate screening of enzyme mutants. In this screening method, the target enzyme is displayed together with the fluorogenic or magnetic-labeled substrates on the cell surface, providing the critical phenotype-genotype linkage. This approach enables libraries to be screened based on cell-sorting technologies such as magnetic-activated cell-sorting and fluorescence-activated cell-sorting (FACS). Compared to plate screening, surface display screening has a much higher throughput and is therefore suitable for screening of moderately large libraries (e.g., FACS throughput can reach up to 40,000 cells per second.[7] While this approach is non-continuous due to screening being a discrete step, implementation of a robotic setup or perhaps a cyclical microfluidic design could achieve continuous evolution.

1.3 Continuous directed evolution

Continuous evolution automates all the steps of directed evolution, greatly reducing the generation time and the technical effort required from the experimenter. These advantages enable large sequence-space searches, sequential improvements over long mutational trajectories, and dramatic changes in phenotype over practical time scales. More specifically, continuous evolution implies both continuous diversification of the GOI and selection entirely within living cells, thereby not requiring manipulation to carry out the transition from genotype to phenotype, and back to genotype after selection. Critically, linkage of genotype and phenotype occurs through the host's translation and replication machinery.

In its technically simplest form, continuous evolution uses serial passaging or some kind of continuous dilution (e.g., chemostat or turbidostat).[8,9] Ideally, to achieve continuous selection, evolution of the desired phenotype should lead to an increase in replicative fitness of the corresponding GOI or the host itself. Continuous, automated cycles of diversification and selection mimic the process of biological evolution over many generations on practical time scales, permitting deep exploration of evolutionary trajectories and dramatic improvements to phenotype. The minimal reliance on experimenter intervention during continuous evolution can also make performing multiple parallel evolution experiments more accessible than using classical, non-continuous directed evolution platforms

Another advantage resulting from simplification of genetic diversification is the ability to perform facile, long neutral drift experiments. This allows non-deleterious mutations to accumulate rapidly in the gene pool under partially relaxed selection

conditions, and the increased diversity results in a starting population that has already explored a portion of sequence space, thereby facilitating subsequent adaptation.[10,11]

One implementation of continuous directed evolution is the phage assisted continuous evolution (PACE) system.[15–19] PACE utilizes rapid M13 bacteriophage replication within an *E. coli* host cell to continuously evolve a GOI. In PACE, the gene encoding the critical phage coat protein, protein III (pIII), is removed from the M13 genome, and its production is linked to selection by a GOI to be evolved via genetic circuits. Specifically, induction of pIII is genetically linked to a desired phenotype of the GOI. A turbidostat or chemostat system is used to provide a constant inflow of fresh *E. coli* hosts for the replicating phages to infect. This prevents accumulation of cheater mutations in the host genome. Phage within host cells that cannot induce pIII transcription and translation fail to assemble complete phage particles and cannot propagate and therefore are washed out of the system. A major advantage of this setup is that because phage replication completes in 10 minutes, cycles of directed evolution can be completed very rapidly. Additionally, a mutagenesis plasmid was engineered to boost mutation rates up to 7.2E-5 substitutions per base per generation of phage, allowing for rapid hypermutation. One limitation of PACE is that it requires a turbidostat or chemostat setup, which limits the parallelizability of evolution experiments to a few replicates, and this specialized setup can be difficult to establish in new labs, limiting adoption within the protein engineering community. Another limitation is that PACE requires selections to be linked to phage propagation, specifically, the production of pIII. This does not allow for selections involving techniques like FACS.

More generally, continuous directed evolution techniques are not without some

limitations. Establishing a continuous cycle of mutation, selection, and propagation that is compatible with a range of genes is challenging. So, while traditional *in vitro* directed evolution techniques enable precise control of mutation rate and selection stringency, analogous modifications are more difficult to implement *in vivo*. Because *in vivo* continuous evolution systems require continuous culturing of host cells, invasion of cultures either by cheaters (mutant alleles that bypass the intended selection) or with contaminant organisms poses additional technical. Lastly, the development of selection methods compatible with continuous evolution is typically difficult because evolved proteins with desired phenotypes in a continuous system must be linked to gene replication or host survival, rather than to simple colorimetric signals common in non-continuous screening, as described in the previous section. While establishing this genotype-phenotype linkage between desired protein function and host survival is straightforward for genes that are natively associated with organismal fitness (e.g., antibiotic resistance or temperature tolerance), designing selections compatible with continuous evolution is difficult for GOIs that are not intrinsically linked to cellular survival. In metabolic pathway engineering, protein engineers may desire optimization of the production of a small-molecule through the directed evolution of one or more enzymes within the pathway. However, if this pathway is not native to the host, as is often the case in these applications, then a way to link increased compound production to survival is required. One such approach is to utilize a biosensor that can transduce the level of sensing of a desired compound into a readable output amenable for high-throughput screening, such as the production of a fluorescent protein and subsequent FACS-based screening for higher levels of fluorescence.[20]

1.4 Orthogonal DNA Replication (OrthoRep)

Orthogonal DNA replication (OrthoRep) is a genetic platform for the rapid, continuous, and scalable *in vivo* evolution of user-specified genes of interest (GOIs) in yeast (**Figure 1.1**).[21,22] At its core, the OrthoRep system consists of a special DNA polymerase (DNAP) called TP-DNAP1 that stably replicates a cytoplasmic linear DNA plasmid called p1, without replicating genomic DNA.[22] Engineered error-prone variants of TP-DNAP1 therefore durably drive the continuous mutagenesis of p1 without elevating the mutation rate of genomic DNA, unlocking rates of extreme mutagenesis – currently up to ~10^{-5} substitutions per base (TP-DNAP1-4-2 variant; **see Chapter 2.1**) – for p1-encoded GOIs that would otherwise be unavailable *in vivo*. In addition to p1, an accessory cytoplasmic linear plasmid called pGKL2 (p2) encodes machinery responsible for the replication, transcription, and maintenance of both p1 and p2.[23] However, as p2 encodes its own dedicated DNAP (TP-DNAP2) and p1 replication is mechanistically insulated from p2 replication,[24] rapid mutation of GOIs on p1 by error-prone TP-DNAP1s occurs with complete targeting. Since p1 naturally encodes only TP-DNAP1 and a dispensable toxin/antitoxin pair and since TP-DNAP1 can be functionally encoded on nuclear DNA to sustain p1's replication,[22] p1 is free to encode only arbitrary genetic payloads for rapid mutation and evolution. Therefore, OrthoRep, specifically the orthogonal TP-DNAP1/p1 plasmid pair, forms an ideal genetic architecture for continuous *in vivo* evolution of GOIs.

Although OrthoRep has been successfully applied to evolve several proteins and enzymes, the generality of OrthoRep has not yet been systematically studied. In **Chapter 3**, we show that OrthoRep is fully compatible with all *S. cerevisiae* strains

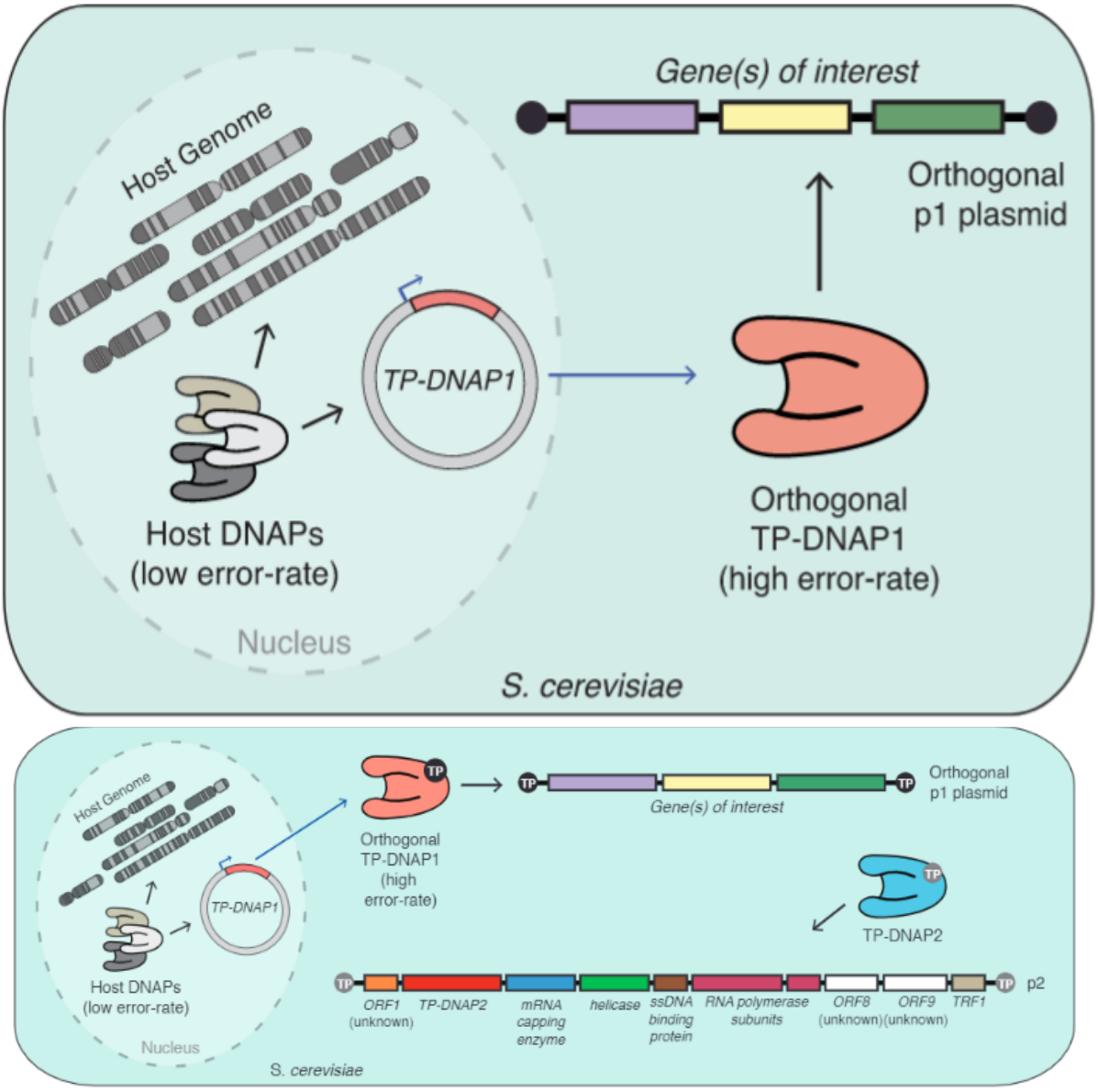

Figure 1.1—(*Top*) Conceptual illustration of OrthoRep. (*Bottom*) Expanded view of OrthoRep including p2.

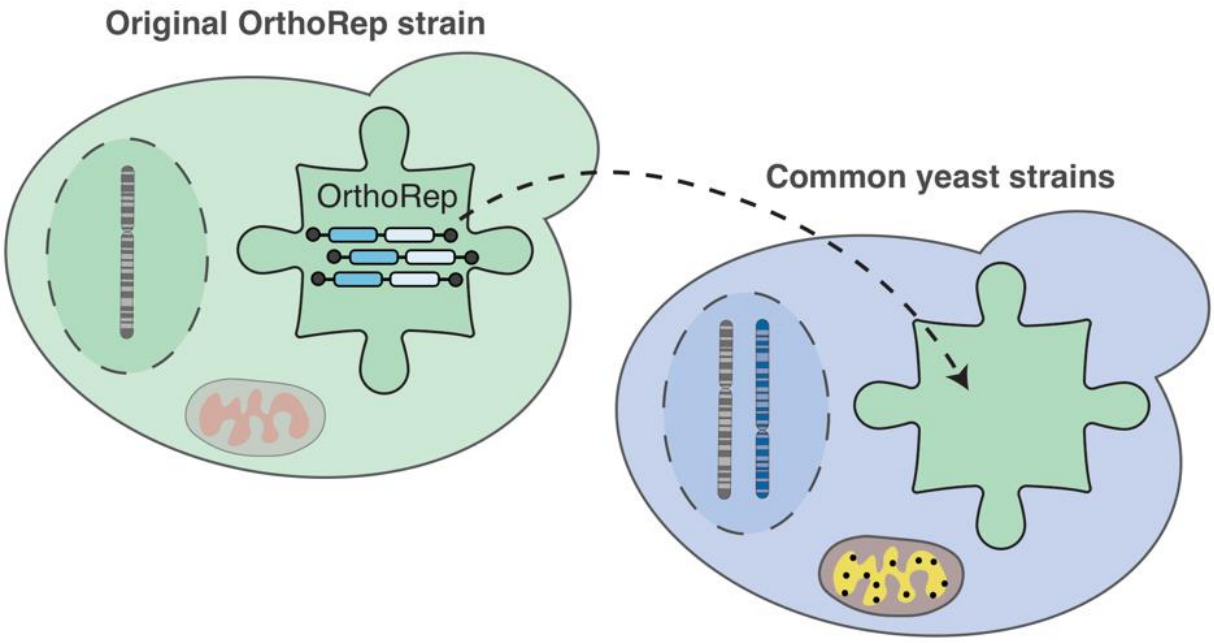

Figure 1.2—Schematic overview of transfer of OrthoRep to a wide-range of S. cerevisiae strains and further work performed in **Chapter 3** demonstrating plasmid stability.

tested, demonstrate that the orthogonal plasmid can encode genetic material of at least 22 kb, and report a CRISPR/Cas9-based method for expedient genetic manipulations of OrthoRep (**Figure 1.2**). It was previously reported that the replication system upon which OrthoRep is based is only stable in respiration-deficient *S. cerevisiae* strains that have lost their mitochondrial genome (ρ^0 strains).[25,26] However, here we trace this biological incompatibility to the activity of the dispensable toxin/antitoxin system encoded on the wild-type orthogonal plasmid. Since the toxin/antitoxin system is replaced by genes of interest in any OrthoRep application, OrthoRep is a generally compatible platform for continuous *in vivo* evolution in *S. cerevisiae*.

OrthoRep should have immediate utility as a straightforward and widely-accessible platform for continuous directed evolution, because OrthoRep realizes mutagenesis of user-defined genes entirely inside a living cell. Therefore, it does not require low-throughput DNA transformation and extraction steps or custom setups for linking selection to the propagation of successful gene variants like other systems do.[15,27] As a consequence, OrthoRep readily integrates with the existing rich ecosystem of cell-based and *in vivo* yeast genetic selections. For example, OrthoRep is already being used in our lab to evolve novel antibodies via yeast surface display and protein-protein interactions via yeast two-hybrid systems. More sophisticated selections can also readily interface with OrthoRep, including selections utilizing cell-based technologies such as FACS (see **Chapter 4**), continuous culturing devices,[9] and droplet screening systems.

In the longer term, we believe that OrthoRep has a critical architectural advantage that will make it a mainstay among the rapidly growing number of continuous

evolution systems that are becoming available.[12,13,16,28–30] In OrthoRep, the only way a user-defined gene can propagate is if it also gets mutated. This is because there is only one DNAP capable of replicating the target gene in OrthoRep and that DNAP is error-prone. Furthermore, that error-prone DNAP should remain error-prone: it is encoded on a nuclear plasmid (or the host genome) where it experiences no elevation in mutation rate, since OrthoRep is entirely orthogonal. Other fully *in vivo* continuous evolution systems achieve diversification of the target gene by recruiting mutagenesis machinery that is not essential for the target gene's replication, which is still carried out by host replication systems. Therefore, rapid evolution may eventually cease when mutations accumulate in the *cis*-elements that recruit mutagenesis machinery. Furthermore, in these systems, genomic mutation rates are elevated through off-target effects of the mutagenesis machinery, which increases the risk that the mutagenesis machinery itself will become disabled, especially since increases in genomic mutation rates are deleterious. As the field of continuous directed evolution advances to more difficult target activities that require longer and longer mutational trajectories to reach, OrthoRep's enforced continuity should become increasingly more valuable.

In addition to this critical distinction of enforced continuous mutagenesis, OrthoRep is unique in a number of additional aspects that should contribute to its long-term utility for directed evolution. First, OrthoRep supports custom and systematically engineerable mutation rates. Already, we have a series of TP-DNAP1s spanning a substitution mutation rate between ~10^{-9} to ~10^{-5}, (see **Chapter 2.1**) which should allow researchers to choose the right level of mutational accumulation for their evolution experiment. Since the supply of beneficial mutations to a gene can change evolutionary

outcomes,[31] this ability to control OrthoRep's mutation rate should be valuable in directed evolution. Ongoing engineering of TP-DNAP1, informed by *in vitro* characterization and structure determination, should also yield variants that approach the error threshold of a typical 1 kb gene (~10^{-3}), thereby maximizing the mutation rate for continuous *in vivo* directed evolution. Second, OrthoRep is a fully scalable platform, since it does not require *in vitro* library construction or specialized equipment. Therefore, it can be used to evolve genes at bioreactor-scale or in small culture volumes in a high-throughput manner with basic serial passaging. In addition to drug resistance and fitness landscape studies, large high-throughput replication of evolution experiments can be used to test and exploit the relationship between adaptive outcomes and mutational supply, gene dosage, population size, population structure, or selection dynamics. Scalability also means that genes can be evolved for many related phenotypes (*e.g.* biosensors that recognize different substrates) in parallel, expanding the throughput of directed evolution at large. Third, OrthoRep achieves continuous evolution in a eukaryotic host, whereas other well-established systems are primarily prokaryotic. The space of directed evolution problems addressable in a eukaryote is arguably more relevant to human biology and therapeutics, especially considering the sophistication of posttranslational modifications and signaling pathways available to eukaryotes. Furthermore, among eukaryotes, yeast is a particularly privileged host for directed evolution, because it can sustain large population sizes with fast generation times, and the availability of yeast mating should allow for *in vivo* recombination of genes being evolved on OrthoRep (see **Chapter 2.2**), expanding the modes of diversification available to continuous evolution. In summary, OrthoRep is a unique,

simple, and highly stable *in vivo* continuous evolution system that should enable the routine generation of new biomolecular and cellular functions.

1.5 References

(1) Lutz, S. (2010) Beyond directed evolution—semi-rational protein engineering and design. *Curr. Opin. Biotechnol. 21*, 734–743.

(2) d'Oelsnitz, S., and Ellington, A. (2018) Continuous directed evolution for strain and protein engineering. *Curr. Opin. Biotechnol. 53*, 158–163.

(3) Schmidt-Dannert, C., and Arnold, F. H. (1999) Directed evolution of industrial enzymes. *Trends Biotechnol. 17*, 135–136.

(4) Leemhuis, H., Kelly, R. M., and Dijkhuizen, L. (2009) Directed evolution of enzymes: Library screening strategies. *IUBMB Life 61*, 222–228.

(5) He, Y.-C., Ma, C.-L., Xu, J.-H., and Zhou, L. (2011) A high-throughput screening strategy for nitrile-hydrolyzing enzymes based on ferric hydroxamate spectrophotometry. *Appl. Microbiol. Biotechnol. 89*, 817–823.

(6) Watt, A. P., Morrison, D., Locker, K. L., and Evans, D. C. (2000) Higher Throughput Bioanalysis by Automation of a Protein Precipitation Assay Using a 96-Well Format with Detection by LC−MS/MS. *Anal. Chem. 72*, 979–984.

(7) Daugherty, P. S. (2007) Protein engineering with bacterial display. *Curr. Opin. Struct. Biol. 17*, 474–480.

(8) Toprak, E., Veres, A., Yildiz, S., Pedraza, J. M., Chait, R., Paulsson, J., and Kishony, R. (2013) Building a Morbidostat: An Automated Continuous-Culture Device for Studying Bacterial Drug Resistance under Dynamically Sustained Drug Inhibition. *Nat. Protoc. 8*, 555.

(9) Zhong, Z., Wong, B. G., Ravikumar, A., Arzumanyan, G. A., Khalil, A. S., and Liu, C. C. (2020) Automated Continuous Evolution of Proteins in Vivo. *ACS Synth. Biol. 9*, 1270–1276.

(10) Dalby, P. A. (2011) Strategy and success for the directed evolution of enzymes. *Curr. Opin. Struct. Biol. 21*, 473–480.

(11) Zhong, Z., and Liu, C. C. (2019) Probing pathways of adaptation with continuous evolution. *Curr. Opin. Syst. Biol. 14*, 18–24.

(12) Fabret, C., Poncet, S., Danielsen, S., Borchert, T. V, Ehrlich, S. D., and Jannière, L. (2000) Efficient gene targeted random mutagenesis in genetically stable Escherichia coli strains. *Nucleic Acids Res. 28*, e95–e95.

(13) Camps, M., Naukkarinen, J., Johnson, B. P., and Loeb, L. A. (2003) Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci. 100*, 9727–9732.

(14) Shinkai, A., and Loeb, L. A. (2001) In vivo mutagenesis by Escherichia coli DNA polymerase I. Ile(709) in motif A functions in base selection. *J. Biol. Chem. 276*, 46759–46764.

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

(16) Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., and Turner, K. H. (2016) Continuous Evolution of Bacillus Thuringiensis Toxins Overcomes Insect Resistance. *Nature 533*, 58.

(17) Hubbard, B. P., Badran, A. H., Zuris, J. A., Guilinger, J. P., Davis, K. M., Chen, L., Tsai, S. Q., Sander, J. D., Joung, J. K., and Liu, D. R. (2015) Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nat. Methods 12*, 939–942.

(18) Dickinson, B. C., Packer, M. S., Badran, A. H., and Liu, D. R. (2014) A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nat. Commun. 5*, 5352.

(19) Pu, J., Zinkus-Boltz, J., and Dickinson, B. C. (2017) Evolution of a split RNA polymerase as a versatile biosensor platform. *Nat. Chem. Biol. 13*, 432–438.

(20) Zhang, J., Jensen, M. K., and Keasling, J. D. (2015) Development of biosensors and their application in metabolic engineering. *Curr. Opin. Chem. Biol. 28*, 1–8.

(21) Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., and Liu, C. C. (2018) Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds Resource Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell 175*, 1946.

(22) Ravikumar, A., Arrieta, A., and Liu, C. C. (2014) An orthogonal DNA replication system in yeast. *Nat. Chem. Biol. 10*, 175–177.

(23) Klassen Rolandand Meinhardt, F. (2007) Linear Protein-Primed Replicating Plasmids in Eukaryotic Microbes, in *Microbial Linear Plasmids* (Meinhardt Friedhelmand Klassen, R., Ed.), pp 187–226. Springer Berlin Heidelberg, Berlin, Heidelberg.

(24) Arzumanyan, G. A., Gabriel, K. N., Ravikumar, A., Javanpour, A. A., and Liu, C. C. (2018) Mutually Orthogonal DNA Replication Systems In Vivo. *ACS Synth. Biol. 7*, 1722–1729.

(25) Gunge, N., and Sakaguchi, K. (1981) Intergeneric transfer of deoxyribonucleic acid killer plasmids, pGKl1 and pGKl2, from Kluyveromyces lactis into Saccharomyces cerevisiae by cell fusion. *J. Bacteriol. 147*, 155.

(26) Gunge, N., and Yamane, C. (1984) Incompatibility of linear DNA killer plasmids pGKL1 and pGKL2 from Kluyveromyces lactis with mitochondrial DNA from Saccharomyces cerevisiae. *J. Bacteriol. 159*, 533–539.

(27) Simon, A. J., D'Oelsnitz, S., and Ellington, A. D. (2019) Synthetic evolution. *Nat. Biotechnol. 37*, 730–743.

(28) DeBenedictis, E. A., Chory, E. J., Gretton, D., Wang, B., and Esvelt, K. (2020) (2020) A High-Throughput Platform for Feedback-Controlled Directed Evolution. *bioRxiv*

21022.

(29) Finney-Manchester, S. P., and Maheshri, N. (2013) Harnessing Mutagenic Homologous Recombination for Targeted Mutagenesis in Vivo by TaGTEAM. *Nucleic Acids Res. 41*, 1.

(30) Halperin, S. O., Tou, C. J., Wong, E. B., Modavi, C., Schaffer, D. V, and Dueber, J. E. (2018) CRISPR-Guided DNA Polymerases Enable Diversification of All Nucleotides in a Tunable Window. *Nature 560*, 248.

(31) Desai, M. M., Fisher, D. S., and Murray, A. W. (2007) The speed of evolution and maintenance of variation in asexual populations. *Curr. Biol. 17*, 385–394.

CHAPTER 2

Technological advancement of OrthoRep—rational engineering of higher mutation rates and plasmid cytoduction

2.1 Rational engineering of TP-DNAP1 for higher error-rates

In order to enable accelerated directed evolution experiments in small culture volumes amenable to high-throughput serial passaging, we drastically increased the mutation rate of p1 replication by engineering highly error-prone variants of TP-DNAP1. First, we sought to find a collection of single amino acid mutant TP-DNAP1s with elevated mutation rates. We anticipated that single amino acid changes would yield modest mutators that could then serve as a basis set for building TP-DNAP1s with multiple mutations that act together to reduce fidelity. Indeed, in a study of *E. coli* Pol I, a combination of three moderately fidelity-reducing mutations in the exonuclease domain, active site, and O-helix resulted in an 80,000-fold increase in Pol I's error rate.[1] Initial attempts to populate the TP-DNAP1 basis set, based on homology analysis (see **2.4 Methods**) to related family B DNAPs, (**Figure 2.1**, **Appendix A**) mostly yielded low activity variants (**Appendix B**), due to the idiosyncrasies of protein-priming DNAPs. Mutators identified from this effort (referred to as Rd1 mutants within **Appendix A**) seeded our basis set, but a more comprehensive approach was pursued to find high-activity variants suitable for combination.

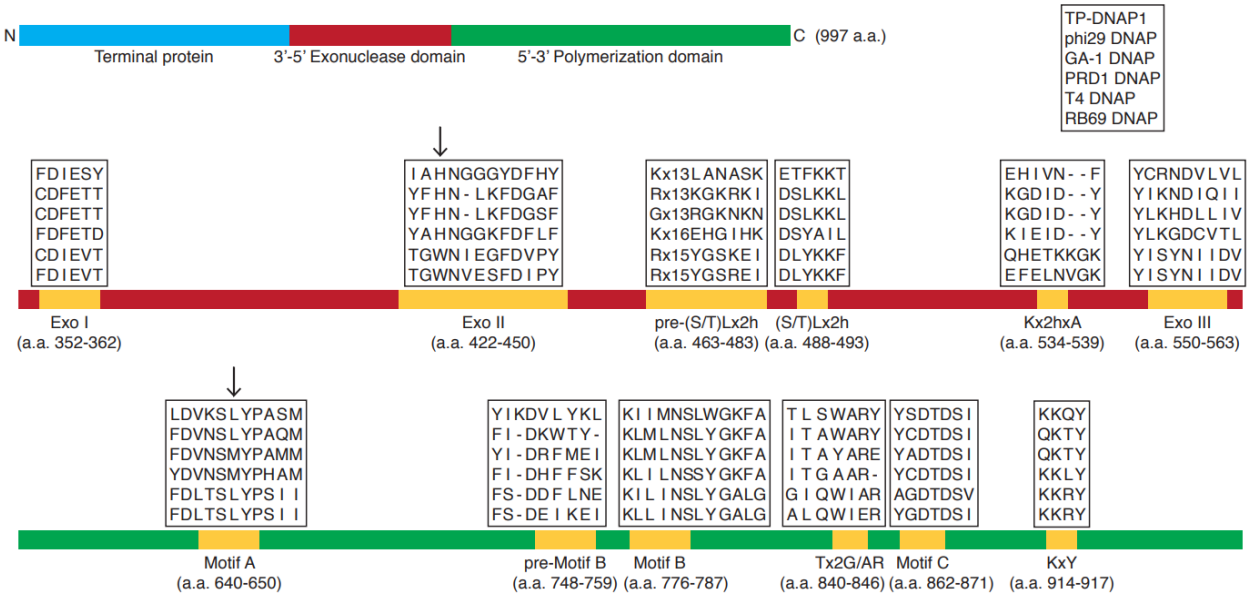

Figure 2.1—Design of TP-DNAP1 mutants by homology analysis. The architecture of TP-DNAP1 consists of a fusion between the terminal protein, a 3'-5' proofreading exonuclease domain, and a DNA polymerization domain. Motifs responsible for fidelity in the exonuclease and proofreading domains are highlighted. A multiple sequence alignment between TP-DNAP1 and five closely related family B DNAPs is shown. In the larger homology study described in the main text, multiple sequence alignment with 99 closely related DNAPs (**Appendix A**) was used to identify positions that exhibit amino acid variation and are flanked by conserved residues. Two candidate positions identified from this study are denoted with arrows. Amino acid variations found at these positions were transplanted into the corresponding location in TP-DNAP1. A total of 87 such TP-DNAP1 mutants were generated and screened. Twenty-four of the TP-DNAP1 variants displayed elevated mutation rates, but almost 60% of these suffered from low activity, judging by the copy number of p1 (**Appendix B**)

In this subsequent effort, Dr. Arjun Ravikumar and Muaeen K.A. Obadi undertook a large mutagenesis campaign to screen for further mutants.[2] To summarize, a scanning saturation mutagenesis library consisting of ~14,000 members was assayed for replicative activity and mutation rate changes. The hits from this library were crossed alongside candidates identified from the rational engineering effort to create new combinations of mutants (**Figure 2.2**). After four rounds of testing combinatorial mutants, the highest error-prone variant, TP-DNAP1-4-2 (L477V, L640Y, I777K, W814N), was discovered to achieve a substitution mutation rate of ~1E-5. Of the four mutations found within this variant, L640Y was initially discovered from the rational engineering effort. Interestingly, at position I777, two rational mutations, I777A and I777V were constructed and determined to have modestly elevated mutation rates (8.63E-9 and 4.71E-9, respectively, compared to 2.23E-9 for WT TP-DNAP1). I777A was selected as a candidate Rd1 mutant used for subsequent combinations. Finally, the high p1 mutation rates driven by error-prone TP-DNAP1s remained completely stable for the longest duration tested (90 generations; **Appendix A**), and genomic mutation rates remained unchanged in the presence of p1 replication by TP-DNAP1-4-2. Therefore, OrthoRep can durably sustain *in vivo* mutagenesis with complete orthogonality (i.e., at least ~100,000-fold mutational targeting) to enable continuous evolution experiments.

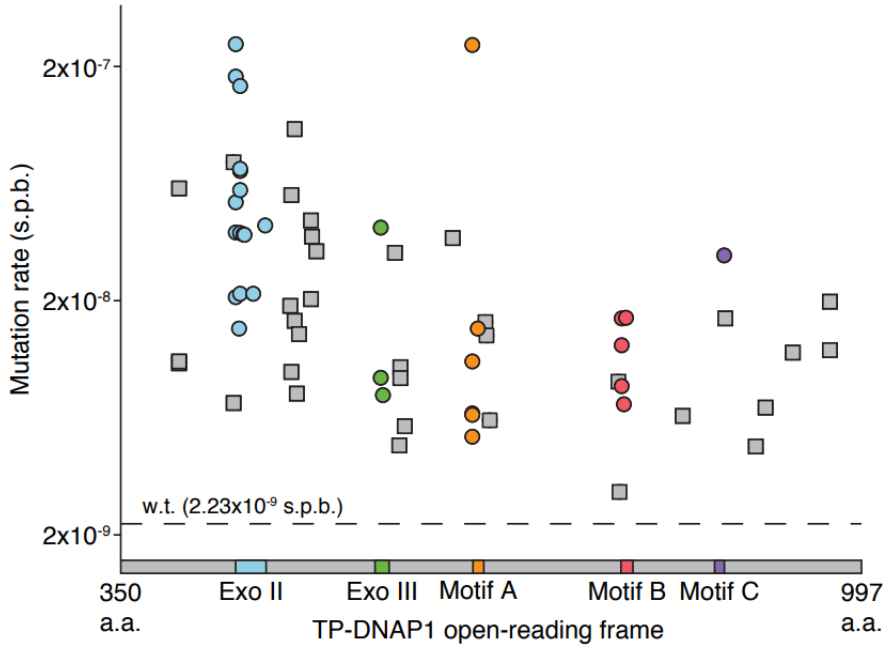

Figure 2.2—Engineering of highly error-prone orthogonal DNAPs for OrthoRep. Mutation rates of 65 basis set TP-DNAP1 variants found from the homology study and TP-DNAP1 library screen. Variants are ordered by amino acid position in the TP-DNAP1 open-reading frame. Residues 1-350, corresponding to the putative TP domain, are not shown. Color-coding indicates regions known to determine fidelity.

2.2 Sexual recombination

Sexual recombination is a critical evolutionary process that allows innovation through large steps in sequence space and that remedies the slowing effects of clonal interference and Muller's ratchet.[3–5] Through sexual recombination, advantageous mutations acquired in separate lineages can be shuffled together post-mating, and deleterious mutations can be discarded. Some genetic techniques, such as DNA shuffling, can mimic sexual recombination, but are performed *in vitro* and therefore are not tractable with an *in vivo* continuous evolution system. Currently, OrthoRep only supports asexual evolution. It would be highly valuable to incorporate sexual recombination into OrthoRep to allow for rapid combinatorial exploration through sequence and fitness space in specific genes of interest. If successful, this would make the two core processes of evolutionary innovation, mutation and sexual recombination, available at rapid rates for the continuous evolution of desired genes *in vivo*, creating a system for biomolecular evolution at maximum possible speeds.

Instead of standard yeast mating, which involves cycles of meiosis and sporulation, we can leverage the cytoplasmic localization of OrthoRep by performing karyogamy mating, or cytoduction.[6] Using yeast that contain a karyogamy mutant, cells will undergo cellular fusion and cytoplasmic mixing (cytoduction), but not nuclei fusion (karyogamy), during mating. Importantly, cytoduction will allow the mixing of OrthoRep plasmids, but not nuclear DNA. Because karyogamy will not occur, sporulation is not required post-mating, which saves 2-4 days of further growth before selecting for haploids and resuming selection. Beyond directed evolution, this pipeline would provide a facile and quick way to transfer OrthoRep plasmids between strains of yeast.

In the following sections, I discuss measurement of phenotypic recombination rates with OrthoRep, a pilot study involving standard yeast mating, and, lastly, plasmid cytoduction.

2.2.1 Measuring recombination rates

Recombination rates on p1 were measured by creating a strain carrying two distinct p1 plasmids each carrying a selectable marker (*URA3* or *KanMX*) and a split auxotrophic gene (*LEU2*) with varying amounts of overlap homology: 30bp, 100bp, or 500bp. To determine rates via a fluctuation test, 36 replicate cultures of this strain were grown under selection for the two functional markers. Then, cultures were washed and spot plated onto media that selected for recombinant products containing full-length *LEU2* (SC-L; leucine dropout media). The number of colonies observed for each replicate spot was counted and recombination rates were determined via the MSS maximum-likelihood estimator method.[7] Phenotypic recombination rates were determined by dividing the calculate rate by the average number of cells per culture. This fluctuation analysis revealed that recombination occurs at a rate that is dependent on the amount of homology and that, at 500 bp of homology, can occur frequently (**Table 2.1**) and without deliberate induction of double-strand breaks.

Amount of homology	**30 bp**	**100 bp**	**500 bp**
Recombination rate (events/cell)	1E-7	5E-6	>1E-3

Table 2.1—Recombination rates determined via fluctuation analysis

2.2.2 Standard yeast mating with OrthoRep

Yeast mating with OrthoRep strains has been performed wherein one strain contained half of the *LEU2* marker and another strain of opposite mating type contained the other half, as described above. These strains were mated and sporulated following standard procedure, with 4 days of sporulation. Selection of diploids was performed post-mating by growing in media selecting for both markers, and mating efficiencies with OrthoRep strains was high (**Table 2.2**). Cultures after sporulation were plated on solid media and colonies were picked to find haploid clones. Mating type validation was performed via a diagnostic PCR,[8] and haploids were then plated for leucine dropout plates to select for recombined plasmids repairing *LEU2* as before. All clones tested were able to grow in the absence of leucine. Recombination was verified via gel electrophoresis of miniprepped plasmids.

	30bp #1	30bp #2	100bp #1	100bp #2	500bp #1	500bp #2
α	154	88	160	67	135	139
a	123	85	105	126	88	114
Diploid	153	146	143	162	144	178
Efficiency	55.4%	63.2%	57.7%	70.7%	62.1%	61.0%
Average	**59.3 ± 5.5%**		**64.2 ± 9.2%**		**61.5 ± 0.8%**	

Table 2.2—Mating efficiency with OrthoRep plasmids. Two biological replicate mating experiments were performed.

2.2.3 Using cytoduction to transfer OrthoRep

As a proof-of-concept, cytoduction was used to simplify transfer of OrthoRep to another strain as an alternative to protoplast fusion. Using this approach, it should be possible to perform simple mating in liquid culture overnight followed by direct selection of cytoductants with inherited OrthoRep plasmids.

To perform karyogamy mating, a set of strains of opposite mating type were constructed with a set of selectable genotypes. Each mating type was constructed to uniquely contain a constitutively expressed drug-selectable marker, an auxotrophic marker driven by a promoter that is only activatable for that specific (haploid) mating type, and a recessive marker that confers additional drug resistance. This genetic setup ensures that each mating type can be robustly selected via both a positive selection (via the specific auxotrophic marker) as well as a negative selection (through counter-selection of the recessive marker). Finally, to perform karyogamy mating, these strains were further modified to contain the mutant allele, *kar1Δ15*, which has been shown to prevent karyogamy and has low diploid formation.[9] Additionally, cytoduction was performed in tandem with strains that contained WT *KAR1* to ascertain the extent to which *kar1Δ15* reduces the frequency of diploid formation.

After creating the necessary strains with the specific genetic selection system, a *CEN6/ARS4* nuclear plasmid encoding the *wt* TP-DNAP1 was introduced to all strains. Finally, the OrthoRep plasmids were introduced via protoplast fusion to one of the mating pairs to act as a “donor” strain. The p1 plasmid encodes for an auxotrophic marker, *TRP1*, allowing to directly select for cytoductants that have inherited the OrthoRep plasmids.

Karyogamy mating was performed by mixing equal amounts of a donor strain (containing OrthoRep) with a recipient strain of opposite mating types (1E7 cells, each) in liquid YPD (2 mL) and grown with shaking overnight. The following day, cells were assayed for (A) recipient (haploid) cytoductants containing OrthoRep, (B) frequency of diploids within the population, (C) presence of diploids in media that selects for haploid cytoductants (**Figure 2.3**). After overnight mating with strains containing *kar1Δ15*, 2.75% (688/25000) of the population successfully transferred the OrthoRep plasmid to desired recipient strain, while the WT *KAR1* condition had nearly no survival (3/25000, or 0.012%) (**Figure 2.3 A**). The percentage of the population containing diploids was decreased from 12.8% to 0.16% when *KAR1* was mutated, demonstrating that *kar1Δ15* is preventing karyogamy (**Figure 2.3 B**). Of the diploids that are formed in the *kar1Δ15* mating condition, none survived when selecting for the mating-specific marker and selecting for the recessive counter-selectable drug marker (**Figure 2.3 C**). This is important as diploids grow substantially faster and have the potential to overtake a population of cells, preventing iterative cycles of mating in a directed evolution campaign.

While the p1 plasmid encoded for an auxotrophic marker that is required for survival post-mating, further confirmation of the OrthoRep plasmids within the recipient haploid cytoductants was performed by DNA miniprepping 6 colonies and visualizing the purified DNA via agarose gel electrophoresis (**Figure 2.4**). All colonies (lanes 1-6) showed the presence of both p1 and p2 at the correct band size, confirming successful transfer, while the parental recipient strain (lane 7) does not contain the plasmids, as expected.

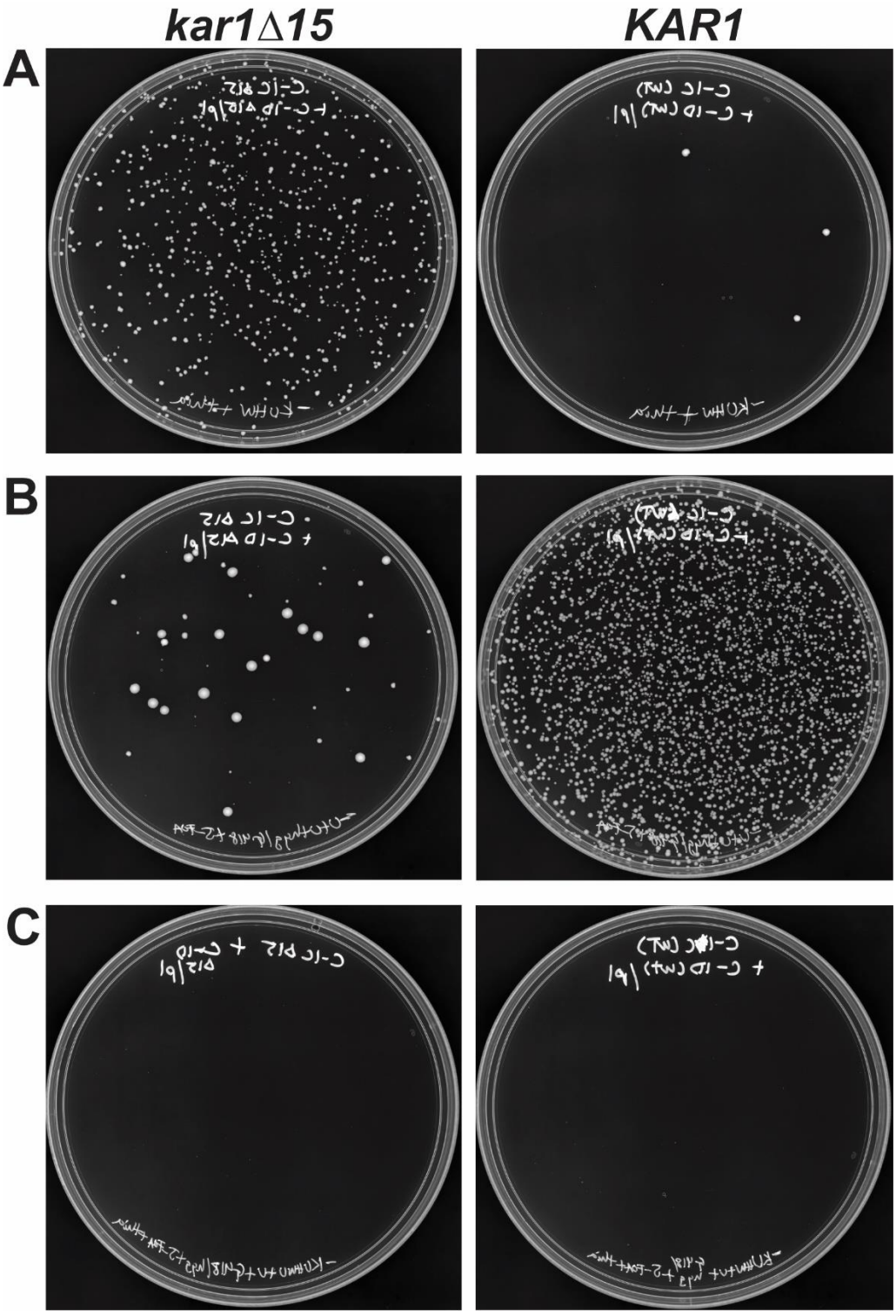

Figure 2.3—Post-mating assay testing for (A) recipient (haploid) cytoductants containing OrthoRep, (B) frequency of diploids within the population, (C) presence of diploids in media that selects for haploid cytoductants. Plates were imaged after 3.5 days of growth at 30°C.

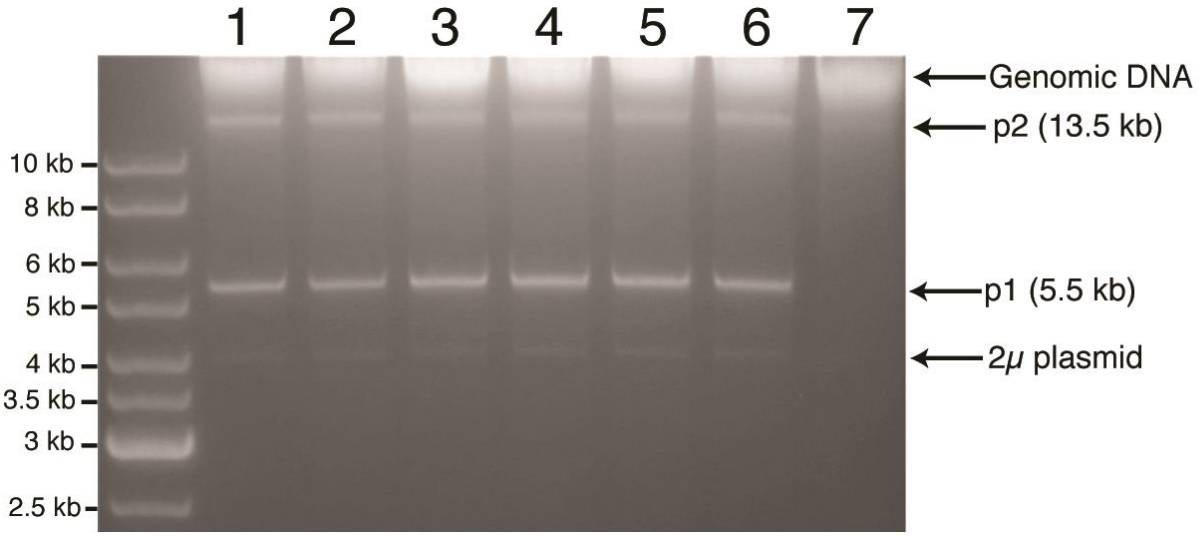

Figure 2.4—DNA minipreps of 6 colonies (lanes 1-6) picked from the selection of recipient cytoductants (left plate, Figure 2.3A) visualized via agarose gel electrophoresis. Lane 7 corresponds to the parent recipient haploid strain, which should not contain the OrthoRep plasmids. Bands corresponding to p1 and p2 are noted.

2.3 Future Work

With karyogamy mating established, the next key step is to test sexual recombination directly by shuffling mutations from evolving lineages. Towards this end, evolution of an antibiotic resistance gene can be used as a straightforward testbed to demonstrate the utility of recombination via karyogamy mating. Our lab has previously used OrthoRep to evolve the dihydrofolate reductase gene from *Plasmodium falciparum* (*Pf*DHFR), the parasite which spreads malaria, for resistance to the anti-malarial drug, pyrimethamine, in 90 replicate cultures (**Figure 2.5**).[2] Widespread resistance to pyrimethamine spread in the wild due to a set of four mutations that accumulated within *Pf*DHFR (N51I, C59R, S108N, I164L; "qm-wild").[10] However, evolution of *Pf*DHFR for pyrimethamine resistance in 90 replicate cultures via simple serial passaging in the presence of drug revealed a complex sequence landscape comprising of genotypes that yielded full resistance to pyrimethamine that differed from the qm-wild genotype.[2]

To more closely mimic an actual directed evolution experiment, we can perform a proof-of-concept recombination experiment leveraging the mutational data collected on *Pf*DHFR pyrimethamine resistance via directed evolution. This can be performed by encoding two different mutant alleles of *Pf*DHFR onto the OrthoRep plasmid, p1, in opposite mating-type strains. For example, this can be attempted with two sets of allele pairs: (1) *abc-e* and *---d-* and (2) *ab--e* and *--cd-* (**Figure 2.5**). These two mutant allele sets contain all the necessary mutations necessary to survive on high pyrimethamine concentration but individually (i.e., without recombination) are insufficient. After encoding the alleles onto p1, the two strains containing the different alleles can be mated, and cytoductants will then be selected for with high pyrimethamine (i.e., 1-3 mM). Post-

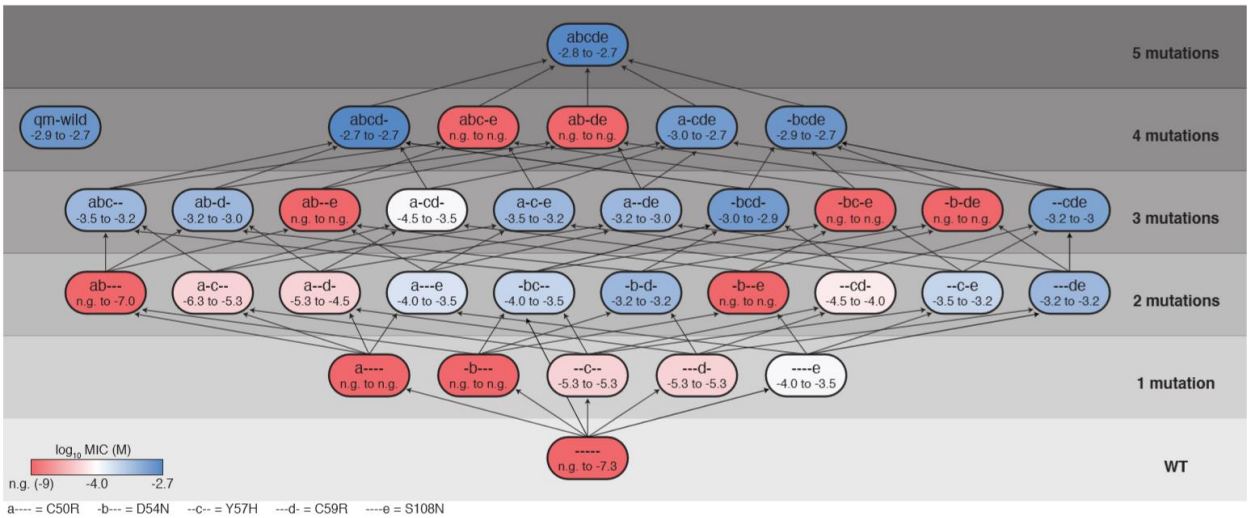

Figure 2.5— A fitness map of a five-mutation *Pf*DHFR landscape defined by (a) C50R, (b) D54N, (c) Y57H, (d) C59R, and (e) S108N. These were the five most commonly accessed mutations from the large-scale evolution experiment described in Ravikumar et al., 2018. All combinations of these five mutations (32 total) were cloned onto a nuclear plasmid to test resistance to pyrimethamine. The minimum inhibitory concentration (MIC) of pyrimethamine was determined for yeast strains expressing all 32 PfDHFR alleles from each landscape. Data shown are the range of $\log_{10}$(MIC of pyrimethamine [M]) for biological triplicates, with a color on a red-blue scale indicating the median. The mid-point of the red-blue scale is shifted to distinguish highly resistant alleles. n.g., no growth. Figure adapted from (2).

cytoduction, it is expected that the p1 plasmids encoding the two different alleles will mix together and spontaneously recombine. Recombinant products containing the correct set of mutations will provide survival during pyrimethamine selection and can be confirmed via Sanger sequencing post-selection. To control for mutations that may occur spontaneously during mating, control experiments can be performed by mating two strains that contain the same mutant allele on p1.

Beyond simply testing if recombination of genetically encoded alleles on p1 can occur through karyogamy mating, the original evolution experiment for drug resistance can be repeated to incorporate cycles of mating during evolution. Asexual evolution of *Pf*DHFR with OrthoRep was seemingly constrained by a “greedy” first-step single mutation (S108N) that biased mutational trajectories of many replicate cultures along a constrained path due to negative epistasis of this mutation with other beneficial mutations.[2] However, in some rare cases, there were replicate cultures that accessed alternate paths not relying on S108N. Moreover, of the four mutations found in qm-wild, two were accessed readily (C59R and S108N), but N51I was present in only 1 population and I164L was never detected. Thus, with mutation shuffling via recombination, evolving cultures exposed to pyrimethamine may be able to access different mutational trajectories not requiring S108N much more frequently. Finally, by incorporating sexual recombination, evolving cultures may adapt to high pyrimethamine concentrations more rapidly compared to asexual populations.[5]

2.4 Methods

TP-DNAP1 homology analysis

A list of 99 homologs to TP-DNAP1 (EMBL accession number: CAA25568.1) was generated via protein BLAST with default settings (**Appendix A**).[11] A multiple sequence alignment of TP-DNAP1 to these homologs was performed using Clustal Omega and the resulting alignment was analyzed using Jalview.[12,13] Amino acid mutations were selected based on three criteria. First, candidate positions should be flanked on both sides by residues with sequence alignment to >75% of homologs. Second, the TP-DNAP1 amino acid at a candidate position should be represented across >25% of homologs. Third, amino acids not present in TP-DNAP1 at a candidate position should be conserved across >25% of homologs. If these criteria were met, then amino acids identified from the third criterion were introduced at the candidate position in TP-DNAP1

Cytoduction and post-mating selection

Yeast strains of opposite mating type were grown to saturation as precultures. Then, 1E7 cells for each mating type were washed and then mixed in 2mL YPD in 24-well deep well blocks and grown at 30°C with shaking (250 r.p.m.) for 24 hours. Then, cells were washed with 0.9% NaCl twice and cell densities were measured via flow cytometry. Cultures were diluted and 2.5E4 cells were plated onto solid media selecting for successful cytoductant transfer (SC/MSG-KUHW + 600 mg/L thialysine), diploid survival (SC/MSG + 1 mg/mL 5-FOA + 400 mg/L G418 + 400 mg/L hygromycin + 50mg/L uracil), and diploid survival on cytoductant media (SC-KUHW + 600 mg/L thialysine + 1 mg/mL 5-FOA + 400 mg/L G418 + 400 mg/L hygromycin + 50mg/L uracil).

Linear plasmid DNA extraction. p1, p2, and recombinant p1 plasmids were extracted following the yeast DNA miniprep procedure as described previously.

2.5 References

(1) Camps, M., Naukkarinen, J., Johnson, B. P., and Loeb, L. A. (2003) Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proc. Natl. Acad. Sci. 100*, 9727–9732.

(2) Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., and Liu, C. C. (2018) Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds Resource Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell 175*, 1946.

(3) Fisher, R. A, 1930 The Genetical Theory of Natural Selection. Clarendon. Oxford.

(4) Muller, H. J. (1932) Some Genetic Aspects of Sex. *Am. Nat. 66*, 118–138.

(5) McDonald, M. J., Rice, D. P., and Desai, M. M. (2016) Sex speeds adaptation by altering the dynamics of molecular evolution. *Nature 531*, 233–236.

(6) Rose, M. D. (1996) NUCLEAR FUSION IN THE YEAST SACCHAROMYCES CEREVISIAE. *Annu. Rev. Cell Dev. Biol. 12*, 663–695.

(7) Hall, B. M., Ma, C.-X., Singh, K. K., and Liang, P. (2009) Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis. *Bioinformatics 25*, 1564–1565.

(8) Illuxley, C., Green, E. D., and Dunbam, I. (1990) Rapid assessment of S. cerevisiae mating type by PCR. *Trends Genet. 6*, 236.

(9) Georgieva, B., and Rothstein, R. (2002) kar-mediated plasmid transfer between yeast strains: Alternative to traditional transformation methods, in *Guide to Yeast Genetics and Molecular and Cell Biology - Part B* (Guthrie, C., and Fink, G. R. B. T.-M. in E., Eds.), pp 278–289. Academic Press.

(10) Lozovsky, E. R., Chookajorn, T., Brown, K. M., Imwong, M., Shaw, P. J., Kamchonwongpaisan, S., Neafsey, D. E., Weinreich, D. M., and Hartl, D. L. (2009) Stepwise acquisition of pyrimethamine resistance in the malaria parasite. *Proc. Natl. Acad. Sci. 106*, 12025 LP – 12030.

(11) Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990) Basic local alignment search tool. *J. Mol. Biol. 215*, 403–410.

(12) Sievers, F., Wilm, A., Dineen, D., Gibson, T. J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., Thompson, J. D., and Higgins, D. G. (2011)

Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol. 7*, 539.

(13) Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009) Jalview Version 2—a multiple sequence alignment editor and analysis workbench. *Bioinformatics 25*, 1189–1191.

CHAPTER 3

Genetic compatibility and extensibility of orthogonal replication

3.1 Introduction

Orthogonal DNA replication (OrthoRep) is a genetic platform for the rapid, continuous, and scalable *in vivo* evolution of user-specified genes of interest (GOIs) in yeast.[1] At its core, the OrthoRep system consists of a special DNA polymerase (DNAP) called TP-DNAP1 that stably replicates a cytoplasmic linear DNA plasmid called pGKL1 (p1) without replicating genomic DNA.[2] Engineered error-prone variants of TP-DNAP1 therefore durably drive the continuous mutagenesis of p1 without elevating the mutation rate of genomic DNA, unlocking rates of extreme mutagenesis – currently up to ~10^{-5} substitutions per base – for p1-encoded GOIs. In addition to p1, an accessory cytoplasmic linear plasmid called pGKL2 (p2) encodes machinery responsible for the replication, transcription, and maintenance of both p1 and p2.[3] Since p2 encodes its own dedicated DNAP (TP-DNAP2) and p1 replication is mechanistically insulated from p2 replication,[4] rapid mutation of GOIs on p1 by engineered error-prone TP-DNAP1s occurs with complete targeting. Since the *TP-DNAP1* gene, naturally found on p1, can instead by encoded on nuclear DNA to sustain p1's replication[2] and since the only other item naturally encoded on p1 is a dispensable toxin/antitoxin (TA) pair, p1 is free to contain only arbitrary genetic payloads for rapid mutation and evolution. Therefore, OrthoRep, specifically the orthogonal TP-DNAP1/p1 plasmid pair, forms an ideal genetic architecture for continuous *in vivo* evolution of GOIs.

We have used OrthoRep to drive the rapid *in vivo* evolution of several proteins, including a drug target to study mutational pathways leading to resistance in 90 replicates,[1] enzymes to expand their reaction scope for both basic protein science and applied biotechnology goals, and antibodies (manuscripts in preparation). In using OrthoRep, we have found that the system is functional for all GOIs we have tested on p1, operational in all *Saccharomyces cerevisiae* strains we have used, and highly durable. However, the generality and behavior of OrthoRep is so far only anecdotal.

Here, we systematically study and extend aspects of OrthoRep's generality that will impact its broad application in directed evolution and genetics. We find that OrthoRep is compatible with a wide range of *S. cerevisiae* strains, including standard lab strains (*e.g.* BY4741 and W303-1A), an industrial yeast strain (CEN.PK2-1C), and diploid yeasts (*e.g.* BY4743), despite previous reports indicating instability of p1 and p2 in diploids.[5] We also reveal that previous reports of replicative instability of p1 and p2 in ρ^+ yeast strains are due to the dispensable TA system on p1.[6] In particular, we show that the p1 orthogonal plasmid is stably replicated at expected mutation rates dictated by the wild-type (wt) or an error-prone (ep) TP-DNAP1 variant, and orthogonality to genomic replication is maintained in all strains tested. Finally, we report streamlined techniques for p1 manipulation that utilize CRISPR/Cas9 and show that genetic payloads of at least 22 kb can be readily encoded on p1, demonstrating that OrthoRep is a highly general yeast system for continuous *in vivo* evolution of GOIs that is ready for broad application.

3.2 Results and Discussion

3.2.1 Generality of OrthoRep across strains

OrthoRep is based on an autonomous replication system, so its replicative properties should be consistent across different yeast hosts. Specifically, 1) the orthogonal plasmid should stably propagate in any strain, 2) the per-base mutation rate of the orthogonal plasmid's replication by wt and ep TP-DNAP1s should be consistent across strains, matching the rates from the strain in which we have historically engineered new TP-DNAP1s[1], and 3) orthogonality against genomic replication should be complete such that no increase in the host genomic error-rate is observed when GOIs on OrthoRep are undergoing rapid mutation. We therefore characterized these three features of OrthoRep in four *S. cerevisiae* base strains beyond the F102-2 strain that our previous reported works used. Specifically, we chose to test OrthoRep in BY4741 and W303-1A, two commonly used haploid yeast strains, BY4743, a commonly used yeast diploid strain, and CEN.PK2-1C, an industrial haploid yeast strain. Unlike F102-2, all four of these new strains are ρ^+ (*i.e.*, respiration proficient).

To be viable for continuous evolution experiments, the orthogonal p1 plasmid must remain stable when replicated not only by the wt TP-DNAP1 but also with ep variants of TP-DNAP1s. Thus, each strain was constructed with either the wt TP-DNAP1 or TP-DNAP1-4-3 (L474W, L640Y, I777K, W814N), expressed *in trans* from a nuclear *CEN6/ARS4* plasmid. The endogenous copy of wt TP-DNAP1 on p1 was removed. Each strain was passaged under selection for a p1-encoded marker, *URA3*, or passaged without selection for ~100 generations (~10 generations per passage) at a dilution of 1:1000 (**Figure 3.1 A**). DNA was extracted after the first and tenth passage

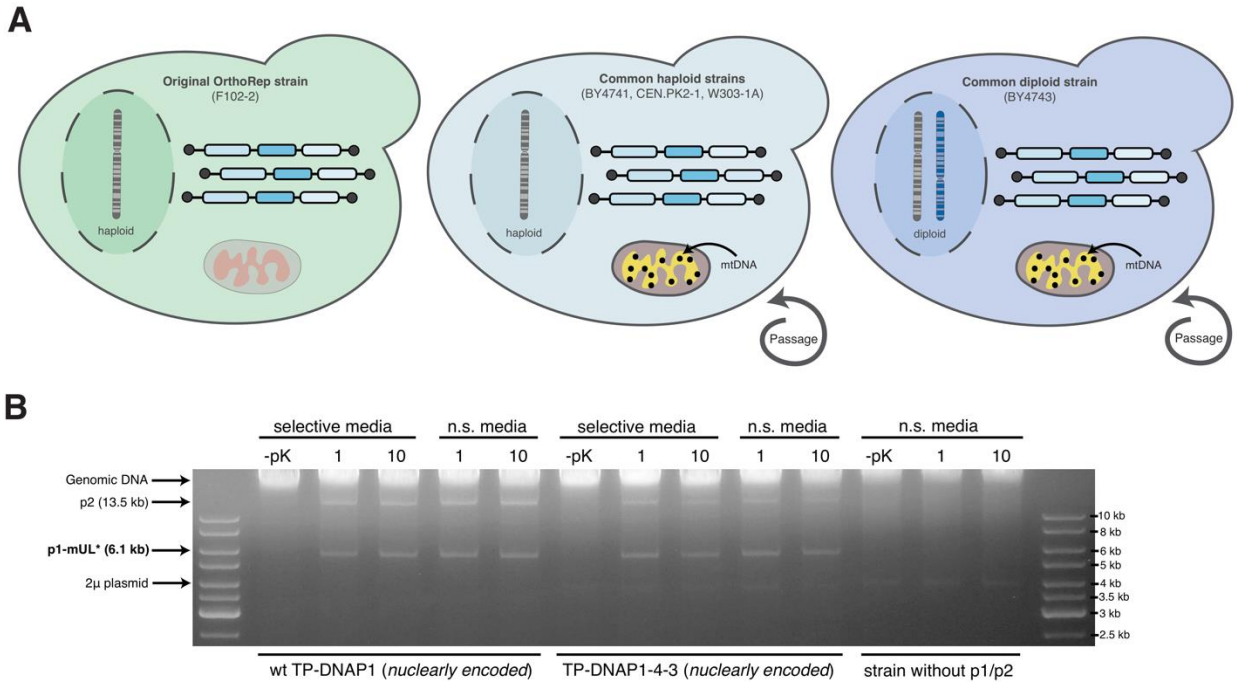

Figure 3.1— **(A)** Testing the stability of OrthoRep in common yeast strains. OrthoRep was transferred from F102-2 into BY4741, CEN.PK2-1C, W303-1A and BY4743. The ploidy and presence or absence of mitochondrial DNA (mtDNA) are indicated. After transfer, strains were passaged in media selective or non-selective (n.s.) for a p1-encoded marker (*URA3*) and a nuclear plasmid expressing a TP-DNAP1 variant (*HIS3*). **(B)** Stability of OrthoRep in BY4741 (see **Figure S1** for similar stability data on all other strains). Agarose gel electrophoresis on DNA extracted from BY4741 strains show the stability of OrthoRep (presence of p1-mUL*) over multiple cycles of passaging even without selection. Presence of p1-mUL* is observed after the first (1) and tenth (10) passage in both selective and non-selective (n.s.) media conditions in the presence of either a wt TP-DNAP1 or error-prone TP-DNAP1-4-3 expressed from a nuclear *CEN6/ARS4* plasmid. Parental BY4741 without p1/p2 is shown as a control. –pK conditions are controls referring to the lack of proteinase K treatment during DNA preparation; the lack of proteinase K treatment results in the lack of all p1/p2-derived bands because the terminal proteins on p1/p2 prevent migration into agarose gels as previously described.[2]

for all 20 cultures. Agarose gel electrophoresis (**Figure 3.1B**, Figure S1 in **Appendix C**) of the extracted DNA showed that OrthoRep is stably maintained with or without direct selection over ~100 generations in all cases. Notably, the p1/p2 plasmids are stable in the diploid strain (BY4743), contrary to previous experiments.[5] This allows the possibility to use yeast mating with OrthoRep to introduce sexual recombination during evolution experiments, which we are actively pursuing.

For OrthoRep to be useful in these *S. cerevisiae* strains, we needed to confirm that p1 is not only stable, but that its per-base substitution rate can be predictably elevated for rapid mutagenesis in the presence of ep TP-DNAP1s. Mutation rates for p1 were determined *via* fluctuation analysis for all four base strains, which were constructed to contain a p1 plasmid (p1-mUL*) encoding a fluorescent reporter (*mKate*), a selectable marker (*URA3*), and a mutation rate reporter (*leu2**), and either the wt TP-DNAP1 or ep TP-DNAP1-4-3 encoded nuclearly. Fluctuation analyses were performed via *leu2** (TAA stop codon at Q180) reversion to *LEU2* to calculate the per-base substitution rate of p1 replication. For a given strain to be measured, replicate cultures were grown in liquid media containing leucine and then spot plated on solid media without leucine. The number of colonies grown were counted for each spot and used to determine the mutation rate *via* the Ma-Sandri-Sarkar method (see Methods).[1,2] Strains encoding wt TP-DNAP1 had mutation rates comparable to our previously measured rate in F102-2 (**Table 3.1**). As expected, the mutation rate of the ep TP-DNAP1-4-3 was also comparable to that measured in F102-2 such that across all strains, the per-base mutation rate of ep TP-DNAP1-4-3 was at least 1000-fold higher than the wt TP-DNAP1 rate. Given this consistency, additional engineered TP-DNAP1s that have lower or

Base Strain	TP-DNAP1 (*nuclearly encoded*)	p1 per-base substitution rate
BY4741	wt	1.23×10^{-9} (0.96-2.10)
	TP-DNAP1-4-3	4.48×10^{-6} (3.07-5.75)
BY4743	wt	9.08×10^{-10} (7.59-20.1)
	TP-DNAP1-4-3	1.11×10^{-6} (0.65-2.08)
CEN.PK2-1C	wt	2.01×10^{-9} (1.43-3.15)
	TP-DNAP1-4-3	3.36×10^{-6} (2.18-4.24)
W303-1A	wt	1.73×10^{-9} (1.10-2.52)
	TP-DNAP1-4-3	2.71×10^{-6} (1.99-3.41)
F102-2	wt	1.95×10^{-9} (1.00-3.09)
	TP-DNAP1-4-3	2.54×10^{-6} (1.79-3.35)

Table 3.1—p1 per-base substitution rates. Mutation rates for TP-DNAP1 in different strains were measured via *leu2** fluctuation analysis. Per-base substitution rates are reported with 95% confidence intervals indicated in parenthesis, as determined by the MSS method.[17,18]

higher mutation rates are expected to behave predictably in all these strains.[1]

To ensure orthogonality against host genome replication in these new strains, we measured genomic mutation rates at the *URA3* locus *via* fluctuation analysis in the presence of replicating p1/p2 plasmids and compared the rates to those of the parental strains. BY4741, BY4743, CEN.PK2-1C, and W303-1A strains were constructed with a recombinant p1 plasmid containing a *TRP1* selectable marker, *mKate2*, and no TP-DNAP1 (p1-mW). A wt TP-DNAP1 or ep TP-DNAP1-4-3 was expressed from a nuclear plasmid in these strains, while parental controls did not contain any component of OrthoRep. Genomic per-base mutation rates are reported in **Table 3.2**. For all strains tested, the presence of OrthoRep did not increase the genomic mutation rate. This suggests that OrthoRep operates orthogonally to host replication, thereby allowing exquisitely targeted mutagenesis of p1-encoded genes.

Base strain	TP-DNAP1/p1	Per-base substitution rate ($x10^{-10}$)
BY4741	n/a (parental)	1.06 (0.22-2.59)
	wt	1.43 (0.53-3.71)
	TP-DNAP1-4-3	1.81 (0.64-2.89)
BY4743	n/a (parental)	4.28 (1.50-11.1)
	wt	5.66 (3.92-12.5)
	TP-DNAP1-4-3	3.30 (1.06-8.96)
CEN.PK2-1C	n/a (parental)	1.39 (0.55-2.72)
	wt	2.57 (1.46-4.43)
	TP-DNAP1-4-3	1.09 (0.77-3.01)
W303-1A	n/a (parental)	2.35 (1.06-2.96)
	wt	1.21 (0.74-2.61)
	TP-DNAP1-4-3	3.13 (2.41-5.55)
AH22	n/a (parental for F102-2)	1.09 (0.94-2.82)
F102-2	wt	2.16 (1.01-4.90)
	TP-DNAP1-4-3	1.71 (1.30-3.67)

Table 3.2—Orthogonality of OrthoRep. Genomic mutation rates were measured in the presence or absence of replicating p1/p2 in different base strains. Per-base substitution rates are reported with 95% confidence intervals indicated in parenthesis, as determined by the MSS method.[17,18] No statistically significant change in the genomic mutation rate is observed when OrthoRep is present.

3.2.2 Compatibility of OrthoRep with ρ⁺ strains.

Our finding that p1/p2 replication is stable in the four new base strains, all of which are respiration proficient, contradicts a previous report showing that the wt p1/p2 plasmids could not be maintained in ρ⁺ strains.[6] Specifically, it was shown that ρ⁺ *S. cerevisiae* strains containing wt p1/p2 were lost after 25 generations of growth. While early works on creating recombinant p1 plasmids in *S. cerevisiae* were assembled by removing part of the TA system, the strains used were ρ⁰.[7,8] Based on our group's accumulated experience with OrthoRep, we suspected that the p1-encoded TA system's activity was the source of incompatibility and that p1/p2 replication *per se* was in fact compatible with ρ⁺ strains. We therefore sought to test this hypothesis by assaying p1 plasmid stability in the presence and absence of TA activity in both ρ⁺ and respiration-deficient ρ⁰ strains. To do so, we constructed a p1 plasmid encoding the full TA system and the *URA3* selection marker (p1-DNAP1-TA-U) in the *S. cerevisiae* strains BY4741 (ρ⁺) and F102-2 (ρ⁰). Alongside, we constructed a p1 plasmid encoding no TA system and the *URA3* selection marker (p1-DNAP1-mUL*), also in both BY4741 and F102-2. Each of the four strains was passaged in triplicate without selection for the *URA3* gene for ~90 generations (**Figure 3.2A**). Afterwards, DNA was extracted from all strains and agarose gel electrophoresis was used to detect the plasmids, where presence of p1 and p2 would indicate their stability. We found that p1-DNAP1-TA-U was stable only in F102-2 but not in BY4741, whereas p1-DNAP1-mUL* was stable in both (**Figure 3.2B**). This suggests that the TA system is indeed the source of incompatibility for p1 replication in ρ⁺ strains.

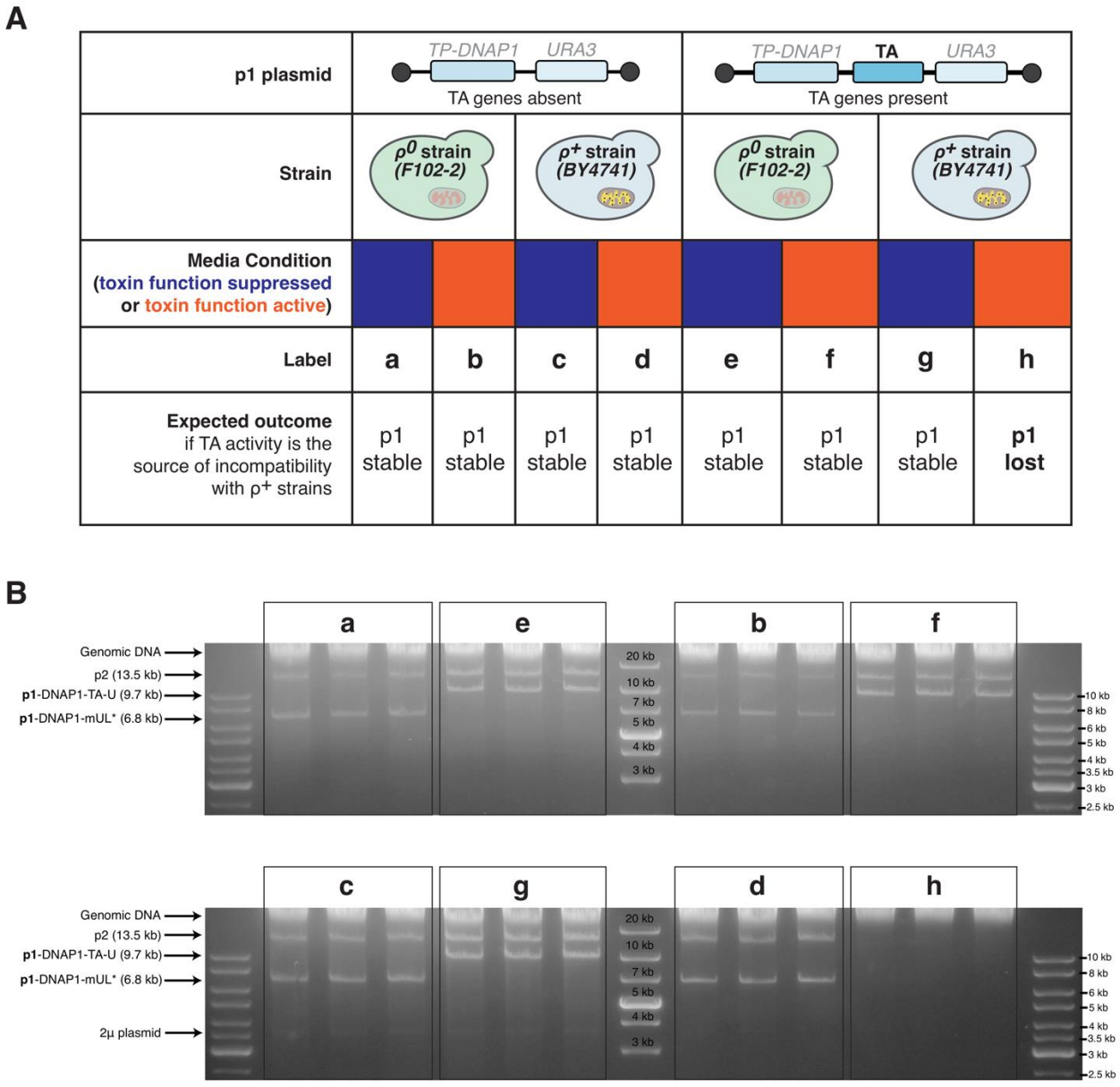

Figure 3.2— **(A)** Design of experiments tracing source of incompatibility between p1 plasmid replication and ρ^+ strains containing mtDNA and full mitochondrial function. A p1 plasmid with or without TA genes encoded were transferred to ρ^0 (F102-2) or ρ^+ (BY4741) strains and passaged in media that either suppressed (blue) or activated (orange) the toxin phenotype. **(B)** Stability of p1/p2 in experimental conditions described in (**A**) after passaging. Agarose gel electrophoresis on DNA extracted from technical triplicates of experiments outlined in (**A**) show the presence of the expected p1 plasmid in all conditions except for the condition where TA is encoded and activated in a ρ^+ strain.

To further test our hypothesis, we carried out a second experiment that exploits the inactivation of the p1-encoded TA system at low pH. Previous studies on the TA system revealed that its encoded killer phenotype is active in *S. cerevisiae* at pH ~4.5-5, but not outside this range.[6,9] Therefore, we also passaged our four strains (BY4741 (ρ^+) or F102-2 (ρ^0) with or without a p1-encoded TA system) at a low pH condition (~4), which should suppress toxin activity. We found that when passaged at low pH, p1-DNAP1-TA-U was indeed stable in BY4741 whereas at the pH that activated the toxin phenotype, it was unstable (**Figure 3.2B**, g *versus* h). Assays on synthetic solid media plates indicated that all BY4741 strains from these experiments retained respiratory competence (growth on glycerol plates) (Figure S2 in **Appendix C**). Taken together, this demonstrates that p1-DNAP1-TA-U is only stable in ρ^+ when TA activity is inactive. As the TA system is dispensable and replaced with GOIs in applications of OrthoRep, we conclude that OrthoRep's generality extends across all common *S. cerevisiae* strains.

3.2.3 Expedient CRISPR/Cas9 manipulation of p1

Recombinant p1 or p2 plasmids are currently created *via* transformation of an integration cassette with 5' and 3' homology arms targeting the parental p1 or p2 plasmids in a recipient cell. Typically, DNA analysis of cells shortly after transformation reveals a mixed population of parental and recombinant plasmid. This is because integration is not quantitative and the wt copy number of p1 and p2 is high. To generate a pure population of cells that only encode the recombinant plasmid, the user must passage the initial transformants for 1-2 weeks under selection for the recombinant

plasmid. While straightforward, the added time is undesirable, and in the case where the recombinant plasmid is larger than the parental plasmid, the parental plasmid has a replicative advantage that slows down this process even further. We reasoned that co-transformation of our integration cassette in conjunction with a CRISPR/Cas9 system that targets the parental p1 plasmid would result in a pure population of recombinant p1 immediately after transformation, as unintegrated parental p1 plasmid would be destroyed (**Figure 3.3A**).

To test this approach, we modified a yeast CRISPR/Cas9 vector to encode an sgRNA targeting *MET15* or *mKate2* and a galactose-inducible Cas9, localized to the cytoplasm.[10] These constructs were co-transformed with various p1 integration cassettes of different sizes containing a unique selectable marker into strains of BY4741 containing OrthoRep components and a parental p1 plasmid encoding *MET15* or *mKate2*. Induction of Cas9 and selection for the desired p1 plasmid resulting from integration onto the parental p1 plasmid were directly imposed on solid media after transformants were plated. Control transformations were performed in parallel but without induction of Cas9. Transformed colonies were then grown to saturation in liquid culture and DNA was extracted and analyzed by gel electrophoresis to determine if a pure population of the desired p1 plasmid was obtained. Indeed, transformants that were induced for Cas9 expression displayed only desired p1 plasmid bands, whereas transformants that were not induced for Cas9 contained a mixed pool of parental and desired p1 plasmids (**Figure 3.3B**, Figure S3 in **Appendix C**). As the CRISPR/Cas9 vector used is readily lost when not selecting for the plasmid,[11] this process allows us to quickly generate a strain that only contains the desired recombinant p1 plasmid.

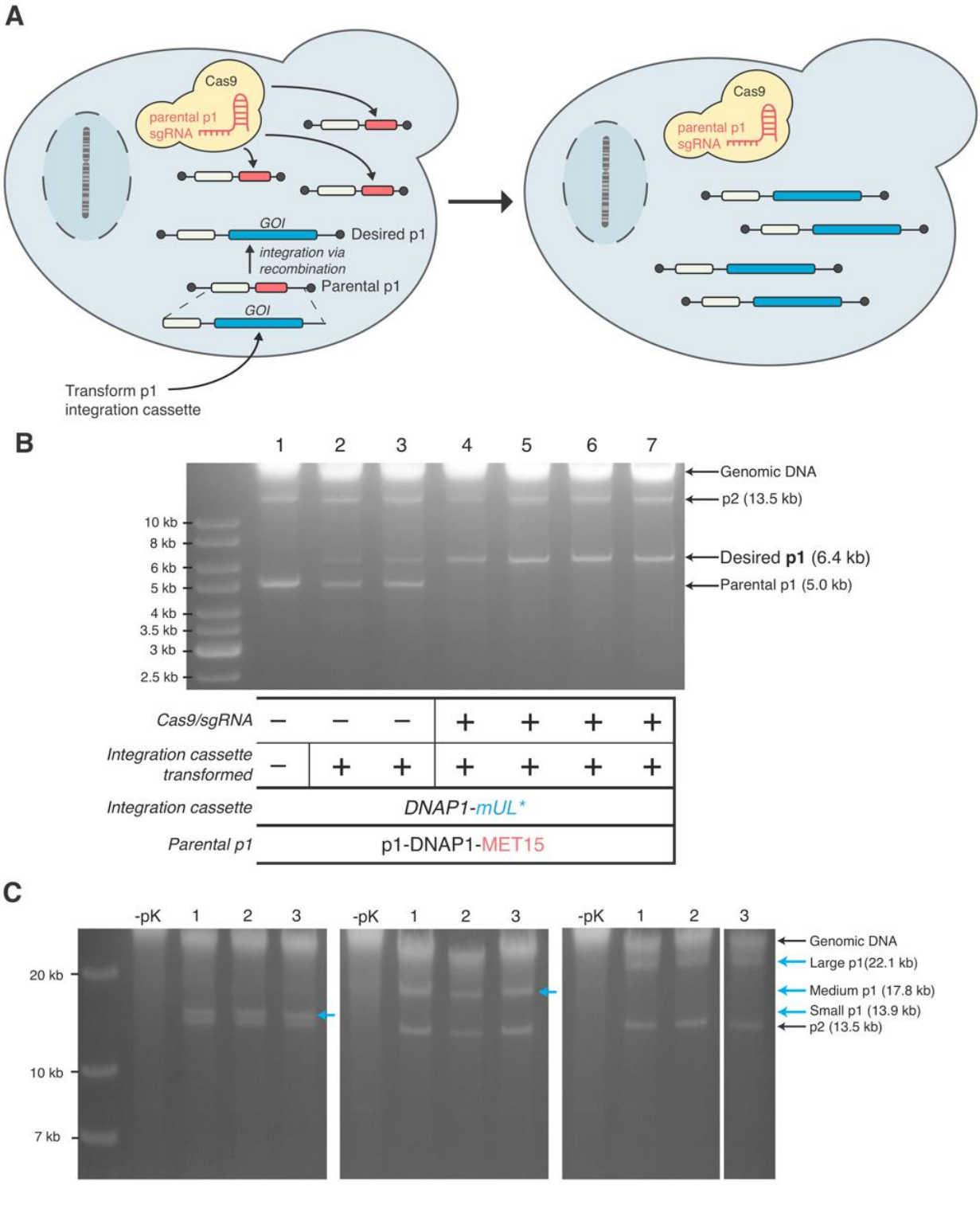
A
Cas9
parental p1
sgRNA
GOI
Desired p1
integration via
recombination
Parental p1
GOI
Transform p1
integration cassette
Cas9
parental p1
sgRNA
B
1 2 3 4 5 6 7
Genomic DNA
p2 (13.5 kb)
Desired p1 (6.4 kb)
Parental p1 (5.0 kb)
10 kb
8 kb
6 kb
5 kb
4 kb
3.5 kb
3 kb
2.5 kb
Cas9/sgRNA – – – + + + +
Integration cassette transformed – + + + + + +
Integration cassette DNAP1-mUL*
Parental p1 p1-DNAP1-MET15
C
-pK 1 2 3 -pK 1 2 3 -pK 1 2 3
Genomic DNA
Large p1(22.1 kb)
Medium p1 (17.8 kb)
Small p1 (13.9 kb)
p2 (13.5 kb)
20 kb
10 kb
7 kb

Figure 3.3—(A) Accelerating genetic manipulation of p1 with CRISPR/Cas9. After transformation of an integration cassette containing a GOI and a selectable marker, transformants are plated on media selecting for the cassette and inducing CRISPR/Cas9. Parental p1 plasmids that did not receive the GOI are cut by CRISPR/Cas9, resulting in a cell containing just the desired p1 plasmid. **(B)** Integration of a model cassette (*DNAP1-mUL**) onto p1 in strains with p1-DNAP1-MET15 as the parental plasmid. Agarose gel electrophoresis on DNA extracted from cells after transformation and plating is shown (lanes 2-7). Lane 1 is the parental strain. Biological replicates from transformation without Cas9 induction (lanes 2 and 3) show both parental and desired p1 bands whereas biological replicates from a transformation with the integration cassette and concomitant induction of Cas9 (lanes 4-7) shows only the desired p1 band. **(C)** Exploration of cassette sizes that can be integrated on p1. Agarose gel electrophoresis on DNA extracted from cells (BY4741 (AJ-Y92)) after transformation of three integration cassettes generating three sizes of p1 – small (14 kb), medium (18 kb), or large (22 kb). Three biological replicates are shown for each size. –pK conditions are controls referring to the lack of proteinase K treatment during DNA preparation; the lack of proteinase K treatment results in the lack of all p1/p2-derived bands because the terminal proteins on p1/p2 prevent migration into agarose gels as previously described.[2]

3.2.4 Exploring the possible size of recombinant orthogonal p1 plasmids

In OrthoRep, GOIs that are subjected to rapid evolution are encoded on p1. Thus far, we have encoded up to 4 genes simultaneously on p1, but it may be desirable in future applications to encode larger genetic payloads, such as entire metabolic pathways or gene clusters. While the minimal size of p1 is known – only ~400 bp of DNA are responsible for initiation of p1 replication, as most of p1's replication origin are the terminal protein components – p1 plasmids beyond its wt size (8.9 kb) have not been constructed.[12] We sought to determine if we could create pure populations of recombinant p1s that far exceed the size of wt p1 while simultaneously removing the parental p1 plasmid. To explore this possibility, we created three integration cassettes containing *URA3*, *mKate2*, and varying lengths of stuffer DNA, taken from *Kluveromyces lactis* genomic DNA. These cassettes would generate p1 plasmid sizes of 14, 18, and 22 kb. Using the aforementioned CRISPR/Cas9 technique, transformation of these integration cassettes into a BY4741 strain containing p1/p2 yielded strains containing only the desired recombinant p1 for all three sizes, as determined via DNA miniprep and gel electrophoresis (**Figure 3.3C**). Therefore, it is possible to encode at least 22 kb of DNA on p1, which should allow the continuous evolution of gene clusters and pathways using OrthoRep. While the lack of a dedicated primase may limit the efficient replication of p1 beyond a maximum size,[13] we can now conclude that this size limit must exceed 22 kb.

3.3 Conclusion

We have demonstrated that OrthoRep is stable and functional in all standard yeast strains tested and have resolved conflicting reports of p1 replicative instability in ρ^+ and diploid strains. We have shown that the mutational properties of OrthoRep hold across all strains tested and developed a streamlined technique to install GOIs on p1, including cargos of at least 22 kb. These developments solidifying the generality of OrthoRep should make it a highly applicable system for continuous *in vivo* evolution.

3.4 Methods

General DNA Cloning. Plasmids used in this study are listed in **Table S1**. *E. coli* strain TG1 (Lucigen) was used for all the DNA cloning steps. All primers used in this study were purchased from IDT. All enzymes for PCR and cloning were obtained from NEB. All plasmids were cloned via Gibson assembly.[14] Cloning of sgRNAs was performed as described by Ryan et al.[11]

Yeast strains and Media. All yeast strains used in this study are listed in **Table S2**. Auxotrophic selection markers used on p1 were first fully deleted from the genome via CRISPR/Cas9[11] before integrating genetic cassettes onto p1 or p1/p2 transfer through protoplast fusion.

Yeast strains were grown in standard media including YPD (10 g/L Bacto Yeast Extract; 20 g/L Bacto Peptone; 20 g/L Dextrose) and appropriate synthetic drop-out media (6.7 g/L Yeast Nitrogen Base w/o Amino Acids (US Biological); 2 g/L Drop-out Mix Synthetic minus the appropriate nutrients w/o Yeast Nitrogen Base (US Biological); 20 g/L Dextrose). For the TA-induced instability experiments, SC media was buffered

with citrate phosphate buffer (0.05 M citric acid and 0.1 M sodium dibasic heptahydrate) and pH was adjusted with 5 M NaOH. For all fluctuation test experiments, all liquid and solid media was adjusted to pH 5.8 with 5 M NaOH.

Yeast transformation and integration. All transformations were performed via the high efficiency Gietz method.[15] For p1 integrations, 2-4 µg of plasmid was linearized by digestion with ScaI, which generated blunt ends containing homologous regions to p1. For *CEN6/ARS4* nuclear plasmid transformations, roughly 100-500 ng of plasmid was transformed. Transformants were selected on the appropriate selective solid SC media. Plates were grown at 30°C for 2 days for nuclear transformations and 4-5 days for p1 integrations.

CRISPR/Cas9-based p1 manipulation. The pCas plasmid, containing a KanMX selection marker and expressing both sgRNA and Cas9, was used for all CRISPR/Cas9 editing.[11] Spacer sequences were designed using Yeastriction v0.1.[16] Genomic knockouts were performed as described previously.[11] For all p1 integrations, a version of pCas with the 8xHIS and SV-40 nuclear localization tags removed was used.

For integration of genes onto p1 in F102-2, a pCas plasmid (~3-5 µg) containing an sgRNA spacer targeting the WT p1 plasmid (*GCTGATTATACATATACAGA*) was co-transformed with a ScaI-linearized integration cassette (2-4µg). Transformants were recovered overnight and plated on SC dropout solid media containing 1 g/L monosodium glutamate (MSG) as the nitrogen source and 400 µg/mL G418. Transformation plates were incubated at 30°C for 4-5 days and colonies were picked into liquid media and 200 µg/mL G418. To confirm integration and loss of parental p1, DNA was extracted (see **linear plasmid DNA extraction**) once cultures were saturated

(~2 days) and analyzed by agarose gel electrophoresis.

For the integration of genes onto p1 in BY4741 seen in Figure 3, a base strain was constructed containing a galactose-inducible Cas9 and a *MET15* marker integrated onto p1. To create this strain, a pCas plasmid was cloned to contain a spacer targeting *mKate2* (*TCTTCAAGTTGCAGATCAGG*) and Cas9 driven by the *GAL2* promoter. This was subsequently transformed into a BY4741 strain containing p1-DNAP1-mUL*. Afterwards, a p1 integration cassette containing *MET15* was transformed into this strain, and transformants were recovered in YPD for 1 hr and then plated on SC-MC dropout containing MSG (1 g/L), G418 (400 µg/mL), and 2% raffinose and 2% galactose as the carbon sources. A clone containing p1-DNAP1-*MET15* and no parental p1-DNAP1-mUL* was re-streaked on solid media (SC-MC) to lose the pCas plasmid. Afterwards, another pCas plasmid with a galactose-inducible Cas9 was cloned containing a *MET15* spacer (*GCTAAGAAGTATCTATCTAA*) and transformed into this BY4741 strain containing p1-DNAP1-*MET15*. Transformants were grown in SC-MC containing MSG (1 g/L) + G418 (400 µg/mL). All subsequent p1 integrations into this strain, shown in Figure 3, were recovered for 1 hr in YPD and then plated on appropriate SC dropout media containing MSG (1g/L), G418 (400 µg/mL), and 2% raffinose and 2% galactose.

Yeast protoplast fusion. Transfer of p1/p2 from donor strains into recipient strains was performed via protoplast fusion. Recipient strains express a genomically-encoded auxotrophic marker not found in the donor strain. The donor strain contains p1/p2 with a p1-encoded auxotrophic marker not found in the recipient strain.

Protoplast fusion was performed as described before.[2] Briefly, small precultures (~3-4 mL) of donor and recipient strains were grown in appropriate SC dropout media.

Strains were then diluted (1:50) into YPD (80 mL) and grown overnight. A portion (35 mL) of both cultures was spun down at 3000g for 5 min and washed with ddH_2O (10 mL). After spinning down again, the mass of the cell pellets was measured to roughly ~0.3 g. Pellets were then resuspended in β-mercaptoethanol (0.2%, 1.8 mL) and EDTA (0.06 M). Cells were incubated at 30°C for 30 min with occasional rotation. Cells were then washed with KCl (0.6M, 3 mL), spun (3000g, 5 min), and then resuspended in a solution (4.8 mL) of citrate phosphate buffer (0.1 M, pH 6.1), KCl (0.6 M), EDTA (0.01 M), and Zymolyase (29 units, US Biological). Cells were incubated at 30°C for 60 min with occasional rotation. Afterwards, cells were spun down at 2000 rpm (700 x g) for 10 min at 4°C, and then washed twice with 3 mL KCl (0.6M). Cells were then gently resuspended in a solution (3 mL) of citrate phosphate buffer (0.1 M, pH 6.1), KCl (0.6 M), and EDTA (0.01 M). Half of the donor strain (1.5 mL) and half of the recipient strain (1.5 mL) were mixed together and centrifuged at 2000 rpm (700 x g) for 10 min at 4°C. The cell pellet was resuspended gently in a solution (5 mL) containing PEG 3350 (33%), KCl (0.6 M), and $CaCl_2$ (50 mM), followed by incubation at 30°C for 30 min with occasional rotation. Cells were then spun down at 2000 rpm (700 x g) for 10 min at 4°C and then resuspended gently in a solution (5 mL) containing KCl (0.6M) and $CaCl_2$ (50 mM). Roughly 20-150 µL of cell suspension was mixed with 10 mL of warm, liquid SC dropout media (selecting for the p1 auxotrophic marker and the auxotrophic marker genomically-encoded only in the recipient strain) containing KCl (0.6M) and Bacto-Agar (3%), and the entire mixture was plated. The appearance of colonies following protoplast fusion typically took 2-3 days of incubation at 30°C. p1/p2 plasmid DNA extraction was first performed to assess for their presence in the recipient strain. Strains

that contained both plasmids were tested for respiration proficiency by plating on glycerol with appropriate SC dropout media.

Typically, six colonies were inoculated from plates into liquid SC dropout media (same dropout used for plating of the protoplast fusion) for subsequent verification. Recipient strains were verified to contain p1/p2 by agarose gel electrophoresis of extracted linear plasmid DNA. Strains that had p1 and p2 plasmids at the correct band sizes were subsequently tested for respiration proficiency by plating on glycerol media containing SC dropout media selecting for the p1-auxtrophic marker and the nuclear marker on a *CEN6/ARS4* plasmid also containing a TP-DNAP1 variant.

Linear plasmid DNA extraction. p1, p2, and recombinant p1 plasmids were extracted following the yeast DNA miniprep procedure as described previously.[2]

p1 fluctuation test. Fluctuation analyses on p1-encoded *leu2** were performed to determine per-base substitution rates of p1 replication, as described previously.[1] *leu2** contains a C->T mutation at base 538 *LEU2* that results in an ochre nonsense mutation at a site permissive to all single point mutants that generate missense mutations.

To perform fluctuation analyses, yeast strains containing p1-mUL* and TP-DNAP1 expressed *in trans* from a *CEN6/ARS4* plasmid were grown in SC media lacking uracil, histidine, and tryptophan (SC-UHW, pH 5.8). Once grown to saturation, the strains were diluted 1:10,000 into the same media and split into 48 100 µl replicate cultures in 96-well trays. Trays were grown with aluminum seal covers for 2-2.5 days until saturation. Cultures were washed with 0.9% NaCl and resuspended in 0.9% NaCl (35 µl). Four replicates were pooled and titered on SC-UHW (pH 5.8) solid plates, while 10 µl of the remaining 44 replicates were spot plated on SC-UHLW (pH 5.8) solid plates.

Spots were allowed to dry before incubation. Plates were incubated at 30°C. Colonies were counted from SC-UHW titer plates after 2 days and from spot plates after 4-6 days.

The number of *LEU2* revertant colonies was used to calculate the number of mutants (*m*) using the Ma-Sandri-Sarkar (MSS) maximum likelihood estimator as implemented by the FALCOR tool.[17,18] After correcting for partial plating, *m* was normalized by the titers, copy number (see **Determination of copy number *via* quantitative PCR**), and the number of ways to revert *leu2** to *LEU2* (2.33) to obtain the per-base mutation rate. 95% confidence intervals were obtained from the MSS method and were similarly scaled.[1,2,4]

Genomic orthogonality measurements. Fluctuation analyses on the genomically-encoded *URA3* gene were performed to determine genomic per-base substitution rates, as previously described.[1,2] *URA3* was integrated into yeast strains its native locus. For BY4743, a hemizygous strains was constructed (*URA3*::*kanMX/ura3Δ0*) with a geneticin-resistance gene (*kanMX*) inserted immediately downstream of a hemizygous *URA3* gene, in order to select against 5-FOA-resistant mitotic recombinants.[19] These *URA3* strains also contained p1-mW, a *CEN6/ARS4* plasmid containing TP-DNAP1 (wt or TP-DNAP1-4-3), and a *HIS3* marker. Parental control strains did not contain p1/p2 or a *CEN6/ARS4* plasmid. p1/p2-containing strains were inoculated into liquid SC media lacking uracil, histidine, and tryptophan (SC-UHW, pH 5.8), while parental strains were inoculated into liquid SC media lacking only uracil (SC-U, pH 5.8). Saturated cultures were diluted 1:5,000 into the same SC media with uracil added back and split into replicate 200 µl cultures in 96-well trays. Trays were grown with sealed covers for 2-2.5

days until saturation. Cultures were washed with 0.9% NaCl and resuspended in 0.9% NaCl (200 µl). Four cultures from each strain were pooled, diluted and titered on solid SC-HW or SC medium. 190 µl of all the remaining cultures were spot plated onto solid SC (SC-HW for strains containing p1/p2) medium supplemented with 5-FOA (1 g/L) and uracil (50 mg/L), and spots were allowed to dry before incubation. Plates were incubated at 30°C. Colonies were counted from titer plates after 2 days and from spot plates after 5 days. Per-base mutation rates were determined as above (see **p1 fluctuation test**) but normalized by the target size of *URA3* for 5-FOA resistance via substitution mutation (104 bp). 95% confidence intervals were similarly obtained as before.[2]

Determination of copy number *via* quantitative PCR. The copy numbers of p1-mUL* strains were determined by quantitative PCR (qPCR), as described previously.[2,4] Whole-cell DNA extracts from 40 mL cultures were prepared by the large-scale cytoplasmic plasmid extraction protocol specific for qPCR, as described previously.[2,4] All of the extracts were diluted 4,000-fold for qPCR reactions. As before, qPCR was performed using primers specific for *leu2** and *LEU3*.

qPCR reactions were performed in 20 µL mixtures containing template DNA (5 µL), forward primer (2 µl, 5 µM), reverse primer (2 µl, 5 µM), ddH_2O (1 µl), and 2X Maxima SYBR Green/Fluorescein qPCR master mix (10 µl, Thermo Scientific).

Standard curves for both primer pairs were prepared using six serial fivefold dilutions of DNA extracted from the respective parental strain, which contain *LEU2* and *LEU3*, and a non-template control with only ddH_2O, in triplicates. All samples were measured in triplicate with both primer pairs. qPCR was performed using a Roche

LightCycler 480 System using the following protocol:

1. 95°C for 10min
2. 95°C for 15s, 60°C for 1min; 40x
3. 95°C for 1min
4. 55°C for 1min

Cycle threshold (C_t) values were determined by the *LightCycler* 480 software (fit-points method, threshold = 1.75). C_t values from both standard curves were fit to a semi-log regression line plot of C_t versus log([DNA]). From this, a slope and *y* intercept were calculated, and relative copy numbers of samples were determined by the equation: copy number = $10^{((\text{sample } C_t - y \text{ intercept})/\text{slope})}$.[2] Triplicate samples were averaged, and p1-derived plasmid copy numbers were normalized to those of the genome by dividing the average *leu2** copy number by the average *LEU3* copy number.

3.6 References

(1) Ravikumar, A., Arzumanyan, G. A., Obadi, M. K. A., Javanpour, A. A., and Liu, C. C. (2018) Scalable, Continuous Evolution of Genes at Mutation Rates above Genomic Error Thresholds. *Cell 175*, 1946–1957.e13.

(2) Ravikumar, A., Arrieta, A., and Liu, C. C. (2014) An orthogonal DNA replication system in yeast. *Nat. Chem. Biol. 10*, 175–177.

(3) Klassen Rolandand Meinhardt, F. (2007) Linear Protein-Primed Replicating Plasmids in Eukaryotic Microbes, in *Microbial Linear Plasmids* (Meinhardt Friedhelmand Klassen, R., Ed.), pp 187–226. Springer Berlin Heidelberg, Berlin, Heidelberg.

(4) Arzumanyan, G. A., Gabriel, K. N., Ravikumar, A., Javanpour, A. A., and Liu, C. C. (2018) Mutually Orthogonal DNA Replication Systems In Vivo. *ACS Synth. Biol. 7*, 1722–1729.

(5) Gunge, N., Murakami, K., Takesako, T., and Moriyama, H. (1990) Mating type locus-dependent stability of the Kluyveromyces linear pGKL plasmids in Saccharomyces cerevisiae. *Yeast 6*, 417–427.

(6) Gunge, N., and Yamane, C. (1984) Incompatibility of linear DNA killer plasmids pGKL1 and pGKL2 from Kluyveromyces lactis with mitochondrial DNA from Saccharomyces cerevisiae. *J. Bacteriol. 159*, 533–539.

(7) Kämper, J., Esser, K., Gunge, N., and Meinhardt, F. (1991) Heterologous gene expression on the linear DNA killer plasmid from Kluyveromyces lactis. *Curr. Genet. 19*, 109–118.

(8) Schickel, J., Helmig, C., and Meinhardt, F. (1996) Kluyveromyces lactis Killer System: Analysis of Cytoplasmic Promoters of the Linear Plasmids. *Nucleic Acids Res. 24*, 1879–1886.

(9) Gunge, N., and Sakaguchi, K. (1981) Intergeneric transfer of deoxyribonucleic acid killer plasmids, pGKl1 and pGKl2, from Kluyveromyces lactis into Saccharomyces cerevisiae by cell fusion. *J. Bacteriol. 147*, 155.

(10) Ryan, O. W., Skerker, J. M., Maurer, M. J., Li, X., Tsai, J. C., Poddar, S., Lee, M. E., DeLoache, W., Dueber, J. E., Arkin, A. P., and Cate, J. H. D. (2014) Selection of chromosomal DNA libraries using a multiplex CRISPR system. *Elife 3*, e03703.

(11) Ryan, O. W., and Cate, J. H. D. (2014) Chapter Twenty-Two - Multiplex Engineering of Industrial Yeast Genomes Using CRISPRm, in *Methods in Enzymology* (Doudna, J. A., and Sontheimer, E. J., Eds.), pp 473–489. Academic Press.

(12) McNeel, D. G., and Tamanoi, F. (1991) Terminal region recognition factor 1, a DNA-binding protein recognizing the inverted terminal repeats of the pGKl linear DNA plasmids. *Proc. Natl. Acad. Sci. 88*, 11398.

(13) Krupovic, M., and Koonin, E. V. (2014) Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nat. Rev. Microbiol. 13*, 105.

(14) Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., and Smith, H. O. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods 6*, 343–345.

(15) Gietz, R. D., and Schiestl, R. H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat. Protoc. 2*, 31.

(16) van Maris, A. J. A., Backx, A., van Rossum, H. M., Pronk, J. T., van den Broek, M., Wijsman, M., Kuijpers, N. G. A., Daran-Lapujade, P., Mans, R., and Daran, J.-M. G. (2015) CRISPR/Cas9: a molecular Swiss army knife for simultaneous introduction of multiple genetic modifications in Saccharomyces cerevisiae. *FEMS Yeast Res. 15*.

(17) Sarkar, S., Ma, W. T., and Sandri, G. v. H. (1992) On fluctuation analysis: a new, simple and efficient method for computing the expected number of mutants. *Genetica 85*, 173–179.

(18) Hall, B. M., Ma, C.-X., Singh, K. K., and Liang, P. (2009) Fluctuation AnaLysis CalculatOR: a web tool for the determination of mutation rate using Luria–Delbrück fluctuation analysis. *Bioinformatics 25*, 1564–1565.

(19) Herr, A. J., Kennedy, S. R., Knowels, G. M., Schultz, E. M., and Preston, B. D. (2014) DNA Replication Error-Induced Extinction of Diploid Yeast. *Genetics 196*, 677.

CHAPTER 4

Applying OrthoRep towards biosensor evolution

4.1 Introduction

Microbial metabolic production is an attractive alternative to traditional chemical synthesis for a wide array of commercially relevant molecules. Engineering microbes to produce a target chemical efficiently often requires substantial modification of host cell metabolism, which necessitates searching a vast genetic space of enzyme genes and expression levels. Millions of pathway designs can now be built but identifying the most productive cells remains low throughput.[1,2] The ability to detect and report on the presence of any arbitrary target molecule within individual cells would greatly aid the field of metabolic engineering.

A genetically-encoded biosensor can transduce molecular recognition of a target molecule into biological output within a cell, such as reporter gene expression. By expressing fluorescence when a chemical is produced at high titer, for example, the biosensor allows bright, high-producing cells to be collected from a diverse population in a high-throughput manner. Regulatory proteins such as transcription factors have proven highly effective devices for sensitive and specific detection of small molecules.[3,4] These proteins bind a chemical effector which triggers an allosteric response that controls the transcription of one or more genes.[5] Transcription factors have been described for sensing various small molecules and have enabled high-throughput

screening to improve product titers, including dicarboxylic acids,[2,6] alcohols,[6] lactone,[7] and aromatic products of type II PKSs.[8]

Recent work from Jensen and colleagues demonstrated that ligand-inducible transcriptional activators from the largest family of transcriptional regulators found in prokaryotes (LysR-type transcriptional regulators, LTTR) can be ported to the eukaryotic chassis and used to measure the level of a small molecule inside the cell and activate transcription.[9] Specifically, they showed that direct transplantation of a prokaryotic transcriptional activator, BenM, behaves as a biosensor for *cis,cis*-muconic acid (CCM), its cognate ligand, in *S. cerevisiae*.

Based on a multi-parametric engineering strategy, they identified a functional design for the biosensor. They established this platform in yeast by engineering a common yeast promoter, CYC1p, to contain BenM's DNA operator sequence, *BenO* (here termed as the 209bp_CYC1p_*BenO_T1* reporter). This hybrid promoter drove the expression of GFP. In the presence of ligand, CCM, BenM showed robust production of GFP, whereas the absence of CCM exhibited strong inhibition of GFP transcription. Next, they performed one round of directed evolution on BenM to screen for variants via FACS on GFP fluorescence that had higher dynamic range for CCM. The best hit they discovered, a triple mutant ("TM"; H110R, F211V, Y286N), had two-fold higher induction of GFP compared to WT BenM. Finally, their design was shown to be applicable to a range of other biosensors founded on small-molecule-induced transcriptional activators from the LTTR family (e.g., *FdeR* and its ligand, naringenin). These biosensors form a crucial interface that transmits information about the presence of a target molecule

within a cell into actuation of genes that can report this information outside the cell for screening.

We seek to evolve BenM using OrthoRep for higher dynamic range for its cognate ligand, CCM, as well as for detection of a non-cognate ligand, adipic acid (AA). In this chapter, I will discuss preliminary work towards establishing an OrthoRep strain compatible with this biosensor, setting up a survival-based selection, and FACS-based evolution of this biosensor towards better CCM detection and reprogramming specificity towards AA. I will conclude by discussing remaining experiments as well as a recent result from a collaborator showing improved production of CCM through OrthoRep-based evolution of the rate-limiting enzymes from the pathway.

4.2 Preliminary work—Establishing an OrthoRep biosensor strain

The OrthoRep plasmids were moved into strains containing the 209bp_CYC1p_*BenO_T1*::GFP reporter via protoplast fusion. The p1 plasmid contained a *URA3* auxotrophic selection marker and either the WT or TM BenM gene. An additional variant for each p1-encoded BenM was constructed containing a genetically-encoded polyA tail (i.e., string of 75 adenines appended 3' of the stop codon). This has been shown to boost protein expression for p1-encoded genes.[10] Thus, in total, 4 different recombinant p1 plasmids were made. These plasmids were then transferred into a strain containing the 209bp_CYC1p_*BenO_T1*::GFP reporter encoded in the genome as well as a nuclear plasmid expressing either the WT TP-DNAP1 or the mutagenic TP-DNAP1-4-2. Thus, 8 total strains were constructed (**Table 4.1**).

Strain number	TP-DNAP variant (WT or 4-2)	BenM variant (WT or TM)	polyA tail?
1	WT	WT	No
2	WT	TM	No
3	WT	WT	**Yes**
4	WT	TM	**Yes**
5	4-2	WT	No
6	4-2	TM	No
7	4-2	WT	**Yes**
8	4-2	TM	**Yes**

Table 4.1—Strains constructed containing BenM on p1. Each strain has either the WT TP-DNAP1 or TP-DNAP1-4-2, the WT or TM variant of BenM, and encodes or does not encode the polyA tail on the 3' end of BenM to increase expression.

These strains underwent a set of experiments to verify that they were correct both genetically and phenotypically. DNA minipreps off all strains showed the correct p1 band size at ~2.4 kb. Next, DNA sequencing of the p1-encoded *BenM* gene showed no additional mutations. Finally, a set of fluorescence-based experiments were performed to test the behavior of the GFP reporter.

First, induction of the reporter was performed with and without the ligand, CCM, for each of the 8 strains. Precultures for three biological replicates of each strain were grown to saturation and then passaged 1:100 in the same media with or without CCM. After 24 hrs, the cultures were washed and GFP fluorescence was measured via flow cytometry (**Figure 4.1**). Strong fold-activation was seen across all biological replicates for all 8 strains. Additional, negative control strains, which had no BenM and either with or without GFP, showed weak background fluorescence, and the positive control strain, which contained GFP and the TM BenM, had high fluorescence as expected.

Next, a titration curve of fluorescence across a range of CCM concentrations was measured. One biological replicate for each strain was grown as before and then passaged into media containing a range of CCM (0, 100, 200, 400 mg/L). As expected, GFP fluorescence increased with increasing amounts of CCM and began to plateau at 200-400 mg/L, consistent with published results (**Figure 4.2**).[9]

Lastly, as we expect rounds of positive and negative sorting will be required to evolve BenM for improved dynamic range for small-molecules of interest, it was necessary to demonstrate that the reporter can turn “off” once the ligand is removed and that the system should turn on if CCM is added again. To test if BenM-containing strains turn off, cultures of these strains were grown in 0 or 200 mg/L CCM in the

titration curve experiment were then passaged 1:100 in 0 mg/L CCM. After 24 hrs, GFP fluorescence was measured again on flow cytometry (**Figure 4.3**). Almost all strains had returned to background fluorescence, while some strains containing the TM BenM variant had some fluorescence at ~1.1-1.2-fold above background. One additional passage was sufficient to completely turn off these cultures that were slightly still fluorescent. These “OFF” cultures were then passaged 1:100 into 200 mg/L CCM and then GFP fluorescence was re-measured. In all cases, mean GFP fluorescence returned to the same level as the initial induction.

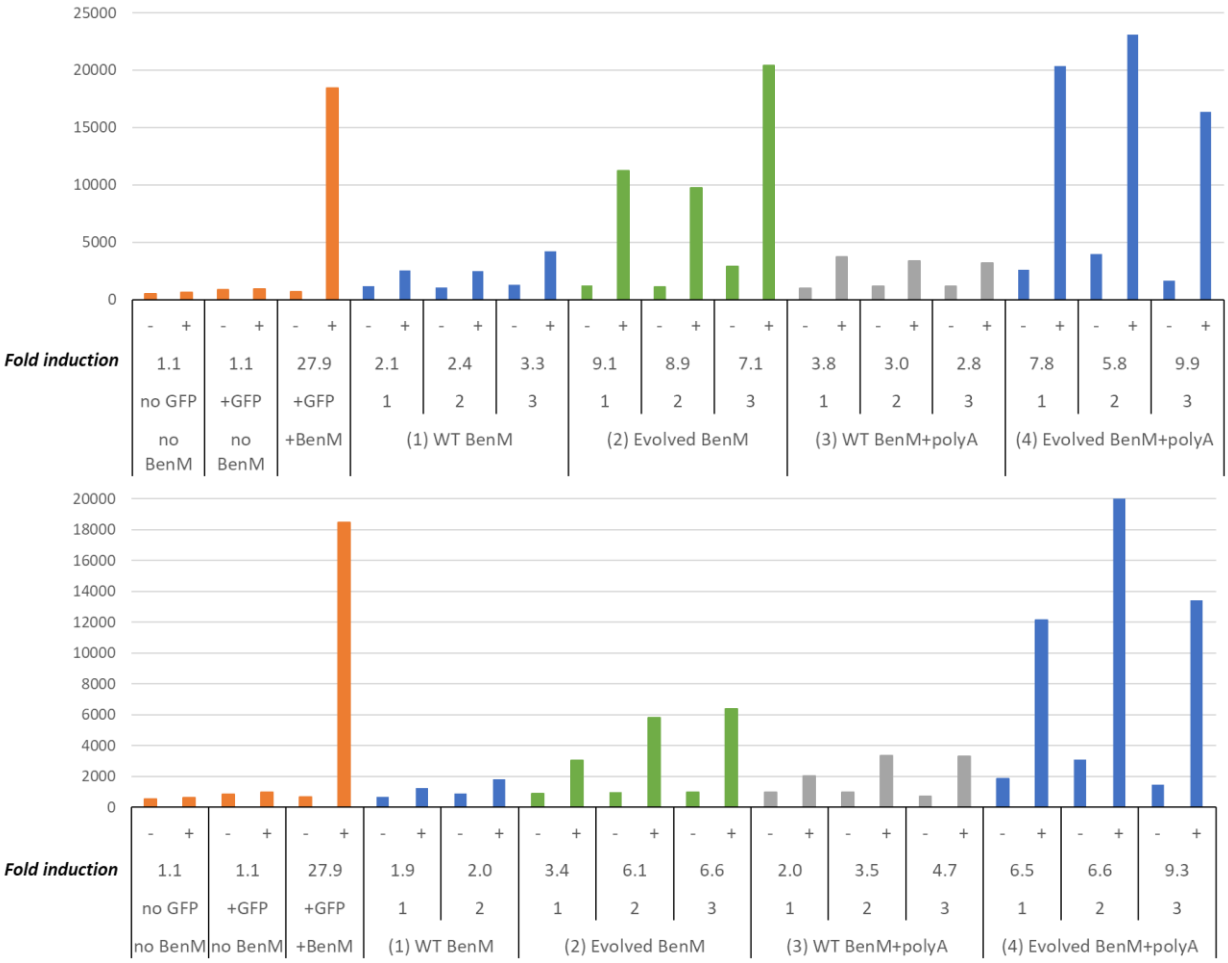

Figure 4.1—Flow cytometry measurements of mean GFP fluorescence. Roughly 100,000 events were collected for each strain. Below each graph, -/+ refer to presence of CCM. Beneath this row shows the fold induction. Three biological replicates were tested for each strain, except for controls (one replicate). (Top) Strains with WT TP-DNAP1 and (Bottom) strains with TP-DNAP1-4-2.

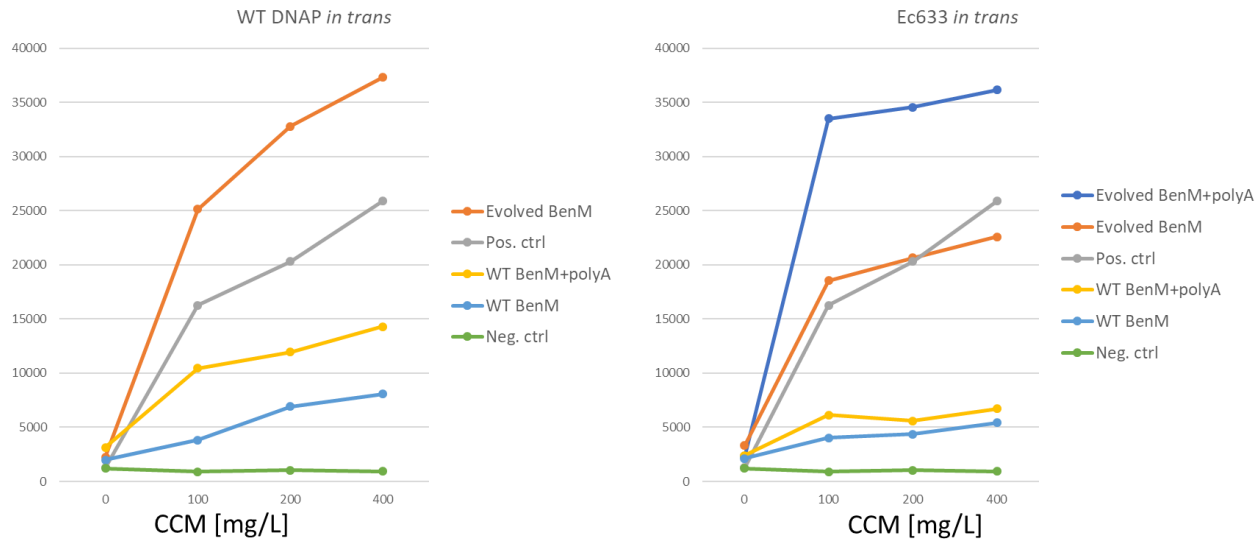

Figure 4.2—Flow cytometry measurements of mean GFP fluorescence. For each strain, a range of CCM concentration was used. (Left) shows strains with WT TP-DNAP1 while (right) shows strains with TP-DNAP1-4-2 (Ec633).

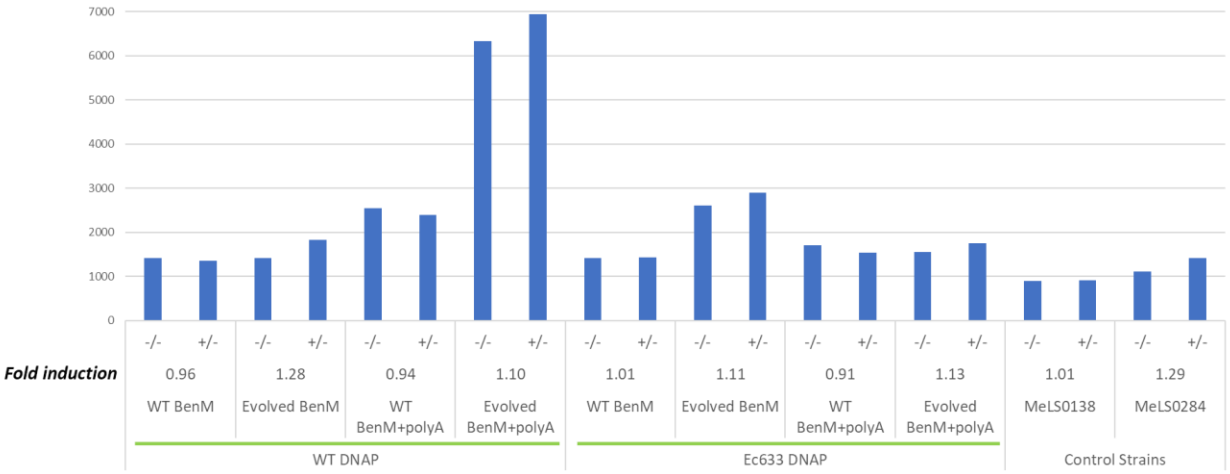

Figure 4.3—Flow cytometry measurements of mean GFP fluorescence. Here, -/- refers to cultures that were grown in 0 mg/L CCM and then passaged 1:100 in 0 mg/L CCM before measurement, whereas +/- refers to 200 mg/L CCM and then passaging 1:100 into 0 mg/L CCM. The fold induction is the ratio between these two conditions. Ec633 DNAP refers to TP-DNAP1-4-2 strains. MeLS0138 is a negative control strain that has the GFP reporter but no BenM present. MeLS0284 is a control strain that has the GFP reporter and a genomically expressed TM BenM.

4.3 Preliminary work—Establishing a survival-based selection for biosensor evolution with OrthoRep

While evolution of BenM for improved dynamic range for CCM has been demonstrated via FACS by Jensen and colleagues,[9] this process is laborious and is limited by the number of samples that can be feasibly sorted. Therefore, a simpler and high-throughput approach utilizing survival-based selection is desirable. Previous work has shown that GFP at the 209bp_CYC1p_*BenO_T1* reporter locus can be swapped with KanMX, a drug-selectable marker.[11] In doing so, it was demonstrated that the growth rate of strains harboring 209bp_CYC1p_*BenO_T1::KanMX* in the presence of drug, G418, was dependent on the concentration of ligand, CCM, in the media. This recapitulated the dose-dependent response observed when using GFP as the output reporter. Thus, to evolve BenM via positive selection, GFP was swapped out for KanMX. Because there is a risk of BenM evolving the ability to induce expression of the reporter gene without the presence of the ligand (i.e., auto-activation), it is important to have a negative selection mechanism. To achieve this, the negatively-selectable marker *CAN1* is genetically fused to *KanMX* at the reporter locus. By encoding *CAN1* first and then fusing *KanMX* to the 3' end of *CAN1*, this protects directed evolution experiments against cheater mutations that inactivate *CAN1* via nonsense and insertion/deletion (indel) mutations. Any cell that acquires a nonsense or indel mutation during a negative selection step will subsequently die during the following positive selection step as it will not be able to produce KanMX protein to provide resistant to G418.

Strains containing 209bp_CYC1p_*BenO_T1::GFP* reporter, WT TP-DNAP1 (nuclear plasmid), and either WT or TM BenM on p1 (no polyA tail) were transformed with integration constructs with either *CAN1*, *KanMX*, *CAN1-P2A-KanMX,* or *CAN1-*

T2A-KanMX to swap out GPF at the reporter using CRISPR/Cas9 markerless integration.[12] Afterwards, integrations were verified by PCRing and then sequencing the reporter locus. In total, two base strains (differing based on which BenM variant was encoded on p1) with 4 different reporter variants for each base strain were constructed.

The positive selection capability of these strains was tested by assaying for G418 resistance. Cultures were grown in normal selective yeast media and then passaged 1:100 into media that contained CCM (2.8 mM) or did not. After ~28 hrs, the cultures were washed and then roughly 1000 cells were spot plated onto solid plates containing a range of G418 concentrations (0, 50, 100, 200, 400 mg/L). Cultures that were grown with CCM were plated on the same concentration of CCM, while the uninduced cultures had no ligand; plates were grown for 3.5 days. For strains containing WT BenM, they were able to survive until ~50-100 mg/L G418 without induction and until ~100-200 mg/L G418 with induction. TM BenM strains grew until 400 mg/L robustly with induction, however there was substantial growth without induction, indicating leak expression of the reporter. Negative controls were not able to grow.

This assay was repeated with canavanine added during liquid culture growth prior to plating in order to kill off cells exhibiting leak expression. This was achieved by expression of *CAN1* at the reporter locus. Plating again on solid plates with a range of G418 resulting in far fewer surviving colonies for cultures that were grown in the presence of canavanine prior to plating compared to cultures grown without canavanine (**Figure 4.4**).

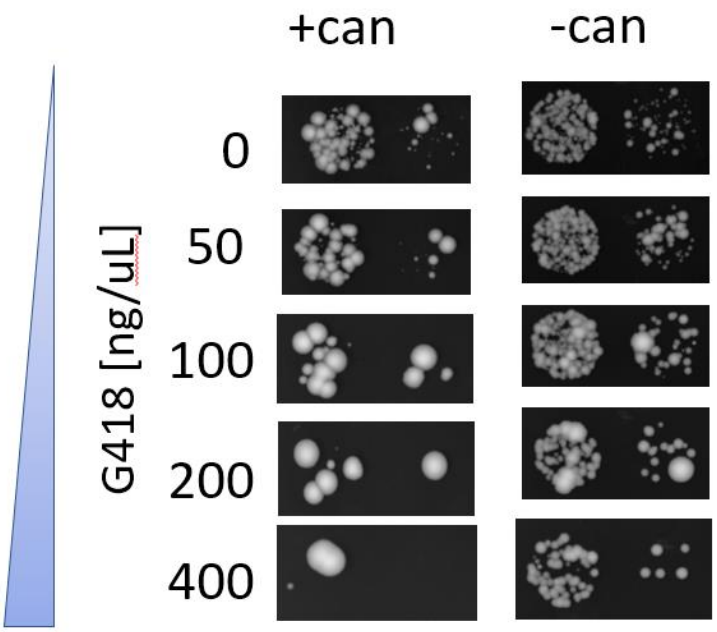

Figure 4.4—Spot plating for strain containing TM BenM and *CAN1*-T2A-*KanMX* reporter. Cultures was grown in +/- can (canavanine), then washed and plated onto solid media containing a range of G418 and 0 mg/L CCM. Each spot of colonies represents 1E4 and 1E5 dilution of same culture. Pre-culture in +*can* resulted in marked decrease in colony survival, indicating cells with spurious activation of the reporter during culturing were killed off.

4.4 FACS-based evolution of a biosensor for improved *cis,cis*-muconic acid detection and reprogramming specificity to adipic acid

The strain consisting of WT BenM encoded on p1 with no polyA tail and the mutagenic TP-DNAP1-4-2 (strain 5 in **Table 4.1**) was used to perform directed evolution of BenM. We sought to (1) increase the dynamic range for sensing of CCM and (2) reprogram BenM to detect adipic acid (AA). This strain underwent random drift to accumulate mutations prior to FACS-based sorting. Random drift was performed through daily passages ($1:2^7$) in 150mL of media for ~150 generations (21 passages). Afterwards, FACS-based directed evolution was initiated. Each round of FACS consisted of a positive sort (top 0.5% of GFP fluorescence cells sorted) followed by an outgrowth in 50mL and a subsequent negative sort to purify the population of cheater mutants capable of auto-inducing GFP fluorescence without the presence of the ligand. Separate parallel evolution campaigns were carried out for AA and CCM (one replicate culture, each).

Alternating rounds of positive sorting followed by negative sorting was performed for 11 rounds with the concentration of ligand increasing from 7 mM to 14 mM at round 6. After round 11, clones from the AA and CCM evolution experiments were isolated to characterize the performance of the evolved biosensors and determine the associated mutations. BenM alleles from the AA and CCM evolved populations from round 11 were obtained through DNA miniprep and PCR. Subsequently, these BenM alleles were transformed as a library and integrated as a single copy genomically (*ura3Δ0::REV1p-*

BenM-tTDH1). Colonies from this transformation were grown and then tested for dynamic range.

The top 7 hits with greatest dynamic range were sequenced and selected for further characterization (**Tables 4.2** and **4.3**). For both AA and CCM evolution experiments, a core set of consensus mutations was found among the top hits in addition to auxiliary mutations. Additionally, mutations were found throughout BenM (304 amino acids in length). Titration curves for the top AA (**Figure 4.5**) and CCM (**Figure 4.6**) hits were measured. Mutants were grown in a range of concentrations in biological triplicates for the evolved ligand as well as the opposite ligand to assess specificity. The fold-induction was calculated for all concentrations relative to the negative control condition (no ligand). In addition to the WT BenM variant, a set of additional mutant variants were measured alongside. TiSNO120 (A130D, A153G, P201S, E287V) was previously evolved for AA.[13] TM (H110R, F211V, Y286N) and MP17_D08 (A230V, F253S, Y286N, Y293H) were previously evolved for improved CCM dynamic and operational range, respectively.[9,13] Nearly all alleles evolved in this experiment outperform these three mutants for dynamic range and have comparable or improved operational ranges. Additionally, while some mutants exhibit elevated promiscuity at high ligand concentrations, several alleles were quite specific for their evolved ligand. AA-5 and AA-6 show comparable, low detection of CCM compared to TiSNO120 (**Figure 4.5**, *bottom*), and CCM-4 has no detectable activity towards AA, while CCM-7 shows very weak activity, comparable to MP17_D08 (**Figure 4.6**, *bottom*). At no point throughout the evolution campaign was negative selection performed against the opposite ligand to directly evolve specificity.

	AA-1	AA-2	AA-3	AA-4	AA-5	AA-6	AA-7
G44D (GGT-> G**A**T)							
D76G (GAT-> G**G**T)							
T82I (ACC-> A**T**C)							
Y112C (TAC-> T**G**C)							
A115V (GCT-> G**T**T)							
N118D (AAC-> **G**AC)							
I188N (ATC-> AA**C**)							
P199T (CCA->**A**CA)							
P201S (CCA-> **T**CA)							
S284P (TCA->**C**CA)							
Y293C (TAC-> T**G**C)							
A294V (GCC-> G**T**C)							

Table 4.2—Mutations sequenced from the top 7 hits from the AA evolution experiment. Amino acid substitutions and their corresponding nucleotide substitution (bolded and underlined) are listed. Boxes shaded green correspond to consensus mutations (4 total) while purple boxes correspond to auxiliary mutations.

	CCM-1	CCM-2	CCM-3	CCM-4	CCM-5	CCM-6	CCM-7
D23G (GAC-> G**G**C)							
A28T (GCT-> **A**CT)							
E41G (GAA-> G**G**A)							
Y67C (TAC-> T**G**C)							
I69V (ATC-> **G**TC)							
V88A (GTT-> G**C**T)							
A115V (GCT-> G**T**T)							
L174S (TTG-> T**C**G)							
I192V (ATC-> **G**TC)							
D213G (GAC-> G**G**C)							
E223G (GAA-> G**G**A)							
Y257H (TAC-> **C**AC)							

Table 4.3—Mutations sequenced from the top 7 hits from the CCM evolution experiment. Amino acid substitutions and their corresponding nucleotide substitution (bolded and underlined) are listed. Boxes shaded blue correspond to consensus mutations (6 total) while yellow boxes correspond to auxiliary mutations.

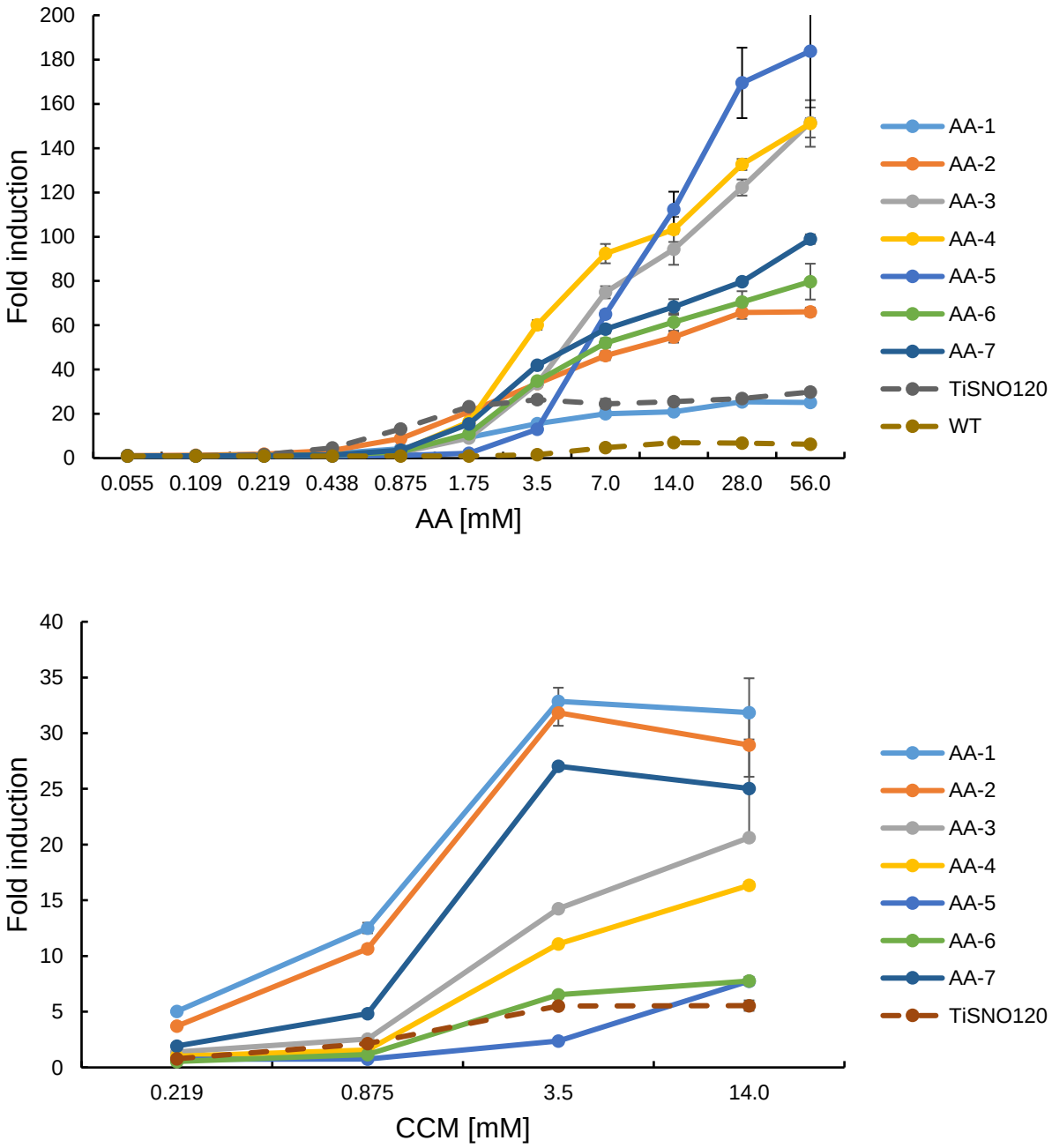

Figure 4.5—Titration curves of the top 7 AA hits, TiSNO120, and WT BenM. (*Top*) Fold induction measured for a range of AA concentrations (operational and dynamic range). (*Bottom*) Fold induction measured for a range of CCM concentrations (specificity). Each concentration was measured in biological triplicates, and the mean ± one standard deviation is shown.

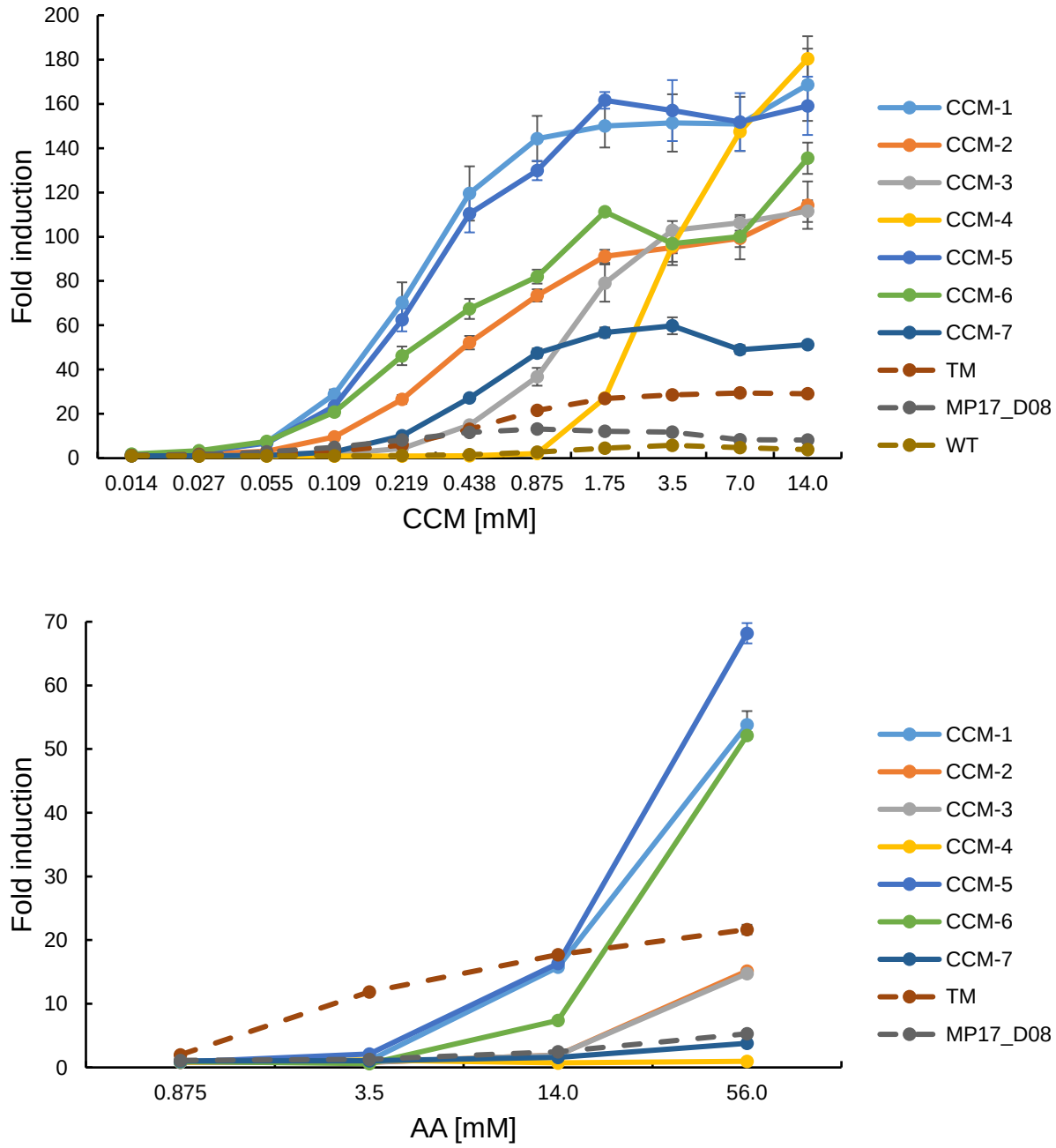

Figure 4.6—Titration curves of the top 7 CCM hits, TM, MP17_D08, and WT BenM. (*Top*) Fold induction measured for a range of CCM concentrations (operational and dynamic range). (*Bottom*) Fold induction measured for a range of AA concentrations (specificity). Each concentration was measured in biological triplicates, and the mean ± one standard deviation is shown.

4.5 Future Work

Further mutational analysis is underway to assess the functionality of all evolved mutants. Specifically, all combinations of consensus mutations from the AA evolution campaign (**Table 4.2**) have been cloned and are being tested for AA and CCM sensitivity. Additionally, all auxiliary mutations have been cloned in isolation to determine AA and CCM sensitivity. This information along with structure mapping could elucidate the mutations responsible for reprogramming specificity of BenM towards AA. Mutating these positions critical for AA sensing may be useful in reprogramming BenM towards other novel small-molecule ligands. We will also be able to determine which mutations, if any, were simply neutral hitchhikers and whether there exists epistatic interactions among mutations. Using this information, we may be able to add certain mutations on top of evolved alleles to test if we can further improve biosensor performance (e.g., higher dynamic range or altered operational range).

Beyond AA, there are other small-molecules that are targets for bio-based production. A report from the DOE identified a set of high-value bio-based chemicals that can be derived from simple biomass.[14] Of these top value-added chemicals, 1,4-diacids (succinic acid, fumaric acid, and malic acid) are attractive novel ligands because they are structurally and chemically similar to CCM, which is a 1,6-diacid. Another high value-added chemical is glucaric acid, for which a biosynthetic pathway has been engineered in yeast.[15] Evolving BenM for these small-molecules would be of great utility towards optimizing engineered biosynthetic pathways.

Lastly, recent work performed in collaboration with Jensen and colleagues resulted in successful evolution of the biosynthetic pathway that produces CCM for

higher titers used OrthoRep (*manuscript in preparation*). Specifically, we encoded the rate-limiting step of the pathway onto OrthoRep, consisting of the enzymes *KpAroY.B* and *KpAroY.Ciso*,[16] and performed FACS-based evolution using a previously evolved BenM variant (MP17_D08) over a two-week period. Afterwards, screening of best-performing variants identified mutants that improved CCM titers >13-fold compared to reference enzymes. Utilizing some of the variants evolved in this work (**Table 4.3** and **Figure 4.6**) that are superior to MP17_D08 could result in even better titer improvements.

4.6 References

(1) Rogers, J. K., Taylor, N. D., and Church, G. M. (2016) Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol. 42*, 84–91.

(2) Raman, S., Rogers, J. K., Taylor, N. D., and Church, G. M. (2014) Evolution-guided optimization of biosynthetic pathways. *Proc. Natl. Acad. Sci. 111*, 17803–17808.

(3) Rogers, J. K., Taylor, N. D., and Church, G. M. (2016) Biosensor-based engineering of biosynthetic pathways. *Curr. Opin. Biotechnol. 42*, 84–91.

(4) Zhang, J., Jensen, M. K., and Keasling, J. D. (2015) Development of biosensors and their application in metabolic engineering. *Curr. Opin. Chem. Biol. 28*, 1–8.

(5) Cuthbertson, L., and Nodwell, J. R. (2013) The TetR Family of Regulators. *Microbiol. Mol. Biol. Rev. 77*, 440 LP – 475.

(6) Dietrich, J. A., Shis, D. L., Alikhani, A., and Keasling, J. D. (2013) Transcription Factor-Based Screens and Synthetic Selections for Microbial Small-Molecule Biosynthesis. *ACS Synth. Biol. 2*, 47–58.

(7) Tang, S.-Y., Qian, S., Akinterinwa, O., Frei, C. S., Gredell, J. A., and Cirino, P. C. (2013) Screening for Enhanced Triacetic Acid Lactone Production by Recombinant Escherichia coli Expressing a Designed Triacetic Acid Lactone Reporter. *J. Am. Chem. Soc. 135*, 10099–10103.

(8) Tahlan, K., Ahn, S. K., Sing, A., Bodnaruk, T. D., Willems, A. R., Davidson, A. R., and Nodwell, J. R. (2007) Initiation of actinorhodin export in Streptomyces coelicolor. *Mol. Microbiol. 63*, 951–961.

(9) Skjoedt, M. L., Snoek, T., Kildegaard, K. R., Arsovska, D., Eichenberger, M., Goedecke, T. J., Rajkumar, A. S., Zhang, J., Kristensen, M., Lehka, B. J., Siedler, S., Borodina, I., Jensen, M. K., and Keasling, J. D. (2016) Engineering prokaryotic transcriptional activators as metabolite biosensors in yeast. *Nat. Chem. Biol. 12*, 951–958.

(10) Zhong, Z., Ravikumar, A., and Liu, C. C. (2018) Tunable Expression Systems for Orthogonal DNA Replication. *ACS Synth. Biol. 7*, 2930–2934.

(11) Snoek, T., Romero-Suarez, D., Zhang, J., Ambri, F., Skjoedt, M. L., Sudarsan, S., Jensen, M. K., and Keasling, J. D. (2018) An Orthogonal and pH-Tunable Sensor-Selector for Muconic Acid Biosynthesis in Yeast. *ACS Synth. Biol. 7*, 995–1003.

(12) Ryan, O. W., and Cate, J. H. D. (2014) Chapter Twenty-Two - Multiplex Engineering of Industrial Yeast Genomes Using CRISPRm, in *Methods in Enzymology* (Doudna, J. A., and Sontheimer, E. J., Eds.), pp 473–489. Academic Press.

(13) Snoek, T., Chaberski, E. K., Ambri, F., Kol, S., Bjørn, S. P., Pang, B., Barajas, J. F., Welner, D. H., Jensen, M. K., and Keasling, J. D. (2020) Evolution-guided engineering of small-molecule biosensors. *Nucleic Acids Res. 48*, e3–e3.

(14) Werpy, T., Holladay, J., and White, J. (2004) Top Value Added Chemicals From Biomass: I. Results of Screening for Potential Candidates from Sugars and Synthesis Gas.

(15) Gupta, A., Hicks, M. A., Manchester, S. P., and Prather, K. L. J. (2016) Porting the synthetic D-glucaric acid pathway from Escherichia coli to Saccharomyces cerevisiae. *Biotechnol. J. 11*, 1201–1208.

(16) Weber, C., Brückner, C., Weinreb, S., Lehr, C., Essl, C., and Boles, E. (2012) Biosynthesis of cis,cis-muconic acid and its aromatic precursors, catechol and protocatechuic acid, from renewable feedstocks by Saccharomyces cerevisiae. *Appl. Environ. Microbiol. 78*, 8421–8430.

Appendix A

99 TP-DNAP1 homologs generated via protein BLAST

GenBank accession number		
YP_001648053.1	AAP92340.1	AGT77289.1
CAA30136.1	ABF00945.1	AAW33516.1
AET14262.1	ABF00940.1	YP_002213831.1
CAA09497.1	AGT75584.1	AGT75888.1
CAD91889.1	AP_000576.1	AAZ15249.2
ACY41087.1	AGT76272.1	AAZ13824.2
ABO48392.1	AAW33476.1	AGT76147.1
NP_044849.1	ADQ38371.1	AGT77111.1
EPY75987.1	AGT77732.1	AGT75671.1
YP_001648062.1	AET87303.1	AFQ34339.1
XP_004185906.1	AAW33432.1	AET87262.1
EMH76195.1	AFH58030.1	AET87221.1
CAA38621.1	ACV41281.1	AFQ34378.1
XP_001735501.1	ACU57037.1	AGV32761.1
XP_654477.1	ACO81791.1	AP_000266.1
EMD48635.1	AAW33115.1	CAC08221.2
EKE43021.1	AGF90825.1	AAW33337.2
CAA37450.1	AFQ34496.1	AAW33246.2
YP_002004529.1	AFQ34457.1	AAT97530.2
P03680.1	NP_073685.1	AAS16276.1
CAA37451.1	AGT76234.1	YP_006272954.1
P06950.1	AAR89955.1	YP_068023.1
NP_690635.1	AGT76890.1	P05664.1
ACH57069.1	AGT77245.1	AP_000304.1
P19894.1	AGT76978.1	P87503.1
CAJ57275.1	AFV96276.1	0905196A
P33538.1	AFQ34417.1	AAS10360.1
1XHX_A	AET87180.1	AAS10432.1
CAA36327.1	AET87139.1	AAS10396.1
EPR79322.1	AAW33386.2	AAW33204.1
AAN62492.1	AP_000539.1	AGT76666.1
YP_002213842.1	ABB17778.1	ABK35035.1
P22374.1	AFQ34300.1	AFA46720.1

Appendix B

TP-DNAP1 variants characterized for Rd1, Rd2, and Rd4 by fluctuation tests in Chapter 2.1

All independent measurements of mutation rate are shown, with corresponding 95% confidence intervals and the number of replicates performed for each fluctuation test listed. The number of replicates assayed for determination of p1 copy number is shown as (n).

TP-DNAP1	p1 copy number (n)	Mutation rate (s.p.b.)	Lower 95% C.I. (s.p.b.)	Upper 95% C.I. (s.p.b.)	Number of replicates	Notes
K302G	87.8 (1)	2.27×10^{-9}	1.57×10^{-9}	3.15×10^{-9}	36	From saturation mutagenesis library
T352E	31.8 (1)	4.58×10^{-9}	3.12×10^{-9}	3.12×10^{-9}	34	From homology study
C354F	25.1 (1)	3.86×10^{-9}	2.46×10^{-9}	2.46×10^{-9}	36	From homology study
F355M	88.8 (1)	4.63×10^{-9}	3.22×10^{-9}	6.35×10^{-9}	36	From saturation mutagenesis library
Y360F	47 (1)	3.76×10^{-9}	2.65×10^{-9}	2.65×10^{-9}	36	From homology study
N371A	16.6 (1)	6.03×10^{-8}	4.69×10^{-8}	7.54×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
N371C	26.9 (1)	1.08×10^{-8}	7.85×10^{-9}	1.42×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
N371M	23.3 (1)	1.56×10^{-8}	1.14×10^{-8}	2.06×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
	27 (1)	1.10×10^{-8}	7.53×10^{-9}	1.53×10^{-8}	36	
C376I	58.1 (1)	1.73×10^{-9}	1.13×10^{-9}	1.13×10^{-9}	33	From homology study
K384D	91.1 (1)	2.84×10^{-9}	1.88×10^{-9}	4.06×10^{-9}	36	From saturation mutagenesis library
V405I	68.3 (1)	2.34×10^{-9}	1.63×10^{-9}	1.63×10^{-9}	36	From homology study
D407Q	176 (1)	1.50×10^{-9}	8.05×10^{-10}	2.50×10^{-9}	48	From saturation mutagenesis library
	160 (1)	1.31×10^{-9}	6.33×10^{-10}	2.35×10^{-9}	48	
G410H, N423Q	55.4 (1)	4.77×10^{-8}	3.94×10^{-8}	5.62×10^{-8}	24	From saturation mutagenesis library; Rd2 mutant
	60.3 (1)	3.91×10^{-8}	3.21×10^{-8}	4.63×10^{-8}	24	
	61.7 (1)	3.50×10^{-8}	2.97×10^{-8}	4.05×10^{-8}	36	
G410Y	138 (1)	1.06×10^{-9}	6.42×10^{-10}	1.61×10^{-9}	24	From saturation mutagenesis library
	132 (1)	3.08×10^{-10}	1.23×10^{-10}	6.20×10^{-10}	24	
E411T, G426C	31.7 (1)	3.42×10^{-9}	2.05×10^{-9}	5.28×10^{-9}	24	From saturation mutagenesis library
	42.2 (1)	3.10×10^{-9}	1.82×10^{-9}	4.81×10^{-9}	24	

	46 (1)	$1.84x10^{-9}$	$9.68x10^{-10}$	$3.10x10^{-9}$	24	
	39.8 (1)	$1.58x10^{-9}$	$8.04x10^{-10}$	$2.72x10^{-9}$	24	
N413I	25.6 (1)	$7.89x10^{-9}$	$3.62x10^{-9}$	$1.46x10^{-8}$	48	From saturation mutagenesis library
	32.3 (1)	$4.47x10^{-9}$	$1.78x10^{-9}$	$9.04x10^{-9}$	48	
	41.4 (1)	$4.14x10^{-9}$	$1.78x10^{-9}$	$7.99x10^{-9}$	48	
	33.4 (1)	$2.30x10^{-9}$	$5.73x10^{-10}$	$5.96x10^{-9}$	45	
I414H	81.4 (1)	$1.87x10^{-9}$	$1.16x10^{-9}$	$2.79x10^{-9}$	24	From saturation mutagenesis library
	94 (1)	$8.43x10^{-10}$	$4.34x10^{-10}$	$1.44x10^{-9}$	24	
I420Y	3.52 (1)	$2.45x10^{-9}$	$4.08x10^{-10}$	$4.08x10^{-10}$	35	From homology study
A421N	5.73 (1)	$9.76x10^{-8}$	$5.94x10^{-8}$	$1.49x10^{-7}$	46	From saturation mutagenesis library; Rd2 mutant
	7.37 (1)	$7.80x10^{-8}$	$4.62x10^{-8}$	$1.21x10^{-7}$	45	
A421S	33.6 (1)	$7.31x10^{-9}$	$5.34x10^{-9}$	$9.65x10^{-9}$	36	From saturation mutagenesis library; Rd2 mutant
N423R	36.5 (1)	$3.67x10^{-8}$	$3.02x10^{-8}$	$4.36x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	37.1 (1)	$3.27x10^{-8}$	$2.67x10^{-8}$	$3.92x10^{-8}$	36	
	37.6 (1)	$2.57x10^{-8}$	$1.99x10^{-8}$	$3.19x10^{-8}$	24	
	43.3 (1)	$2.07x10^{-8}$	$1.59x10^{-8}$	$2.60x10^{-8}$	24	
	43.9 (1)	$1.97x10^{-8}$	$1.52x10^{-8}$	$2.47x10^{-8}$	24	
	69.5 (1)	$1.55x10^{-8}$	$1.20x10^{-8}$	$1.93x10^{-8}$	23	
N423D	4.61 (1)	$5.14x10^{-7}$	$3.89x10^{-7}$	$6.58x10^{-7}$	48	From Ravikumar et al., 2014; Rd1 mutant
	4.84 (1)	$4.03x10^{-7}$	$2.99x10^{-7}$	$5.24x10^{-7}$	48	
	10.6 (3)	$1.74x10^{-7}$	$1.49x10^{-7}$	$2.01x10^{-7}$	48	
	9.48 (3)	$1.82x10^{-7}$	$1.52x10^{-7}$	$2.14x10^{-7}$	44	
	7.46 (3)	$2.49x10^{-7}$	$2.07x10^{-7}$	$2.94x10^{-7}$	45	
N423E	12.6 (1)	$2.67x10^{-7}$	$2.04x10^{-7}$	$3.39x10^{-7}$	48	From saturation mutagenesis library; Rd2 mutant
	17.5 (1)	$1.81x10^{-7}$	$1.27x10^{-7}$	$2.48x10^{-7}$	42	
N423Q	36.2 (1)	$6.43x10^{-8}$	$5.34x10^{-8}$	$7.59x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	20.3 (1)	$6.19x10^{-8}$	$4.94x10^{-8}$	$7.52x10^{-8}$	24	
	25.8 (1)	$5.89x10^{-8}$	$4.78x10^{-8}$	$7.07x10^{-8}$	36	
	18.1 (1)	$5.26x10^{-8}$	$4.11x10^{-8}$	$6.52x10^{-8}$	24	
	23.7 (1)	$4.86x10^{-8}$	$3.82x10^{-8}$	$5.98x10^{-8}$	24	
	46.8 (1)	$3.10x10^{-8}$	$2.49x10^{-8}$	$3.74x10^{-8}$	24	
N423W	22.5 (1)	$4.67x10^{-9}$	$2.96x10^{-9}$	$2.96x10^{-9}$	35	From homology study
N423Y	4.65 (1)	$1.29x10^{-9}$	$7.36x10^{-11}$	$7.36x10^{-11}$	34	From homology study
G425S	42.9 (1)	$3.04x10^{-9}$	$2.05x10^{-9}$	$2.05x10^{-9}$	36	From homology study
G426A	52.2 (1)	$3.35x10^{-9}$	$2.35x10^{-9}$	$2.35x10^{-9}$	36	From homology study
G426K	106 (1)	$1.31x10^{-9}$	$7.67x10^{-10}$	$2.04x10^{-9}$	24	

	153 (1)	$1.25x10^{-9}$	$7.87x10^{-10}$	$1.85x10^{-9}$	24	From saturation mutagenesis library
G426P	20.2 (1)	$1.52x10^{-8}$	$1.09x10^{-8}$	$2.02x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
Y427A	6.36 (3)	$2.14x10^{-8}$	$5.36x10^{-9}$	$5.51x10^{-8}$	24	From Ravikumar et al., 2014; Rd1 mutant
Y427N	2.28 (1)	$1.65x10^{-7}$	$1.25x10^{-7}$	$1.25x10^{-7}$	36	From homology study; Rd1 mutant
Y427D	1.49 (1)	$7.16x10^{-8}$	$4.41x10^{-8}$	$4.41x10^{-8}$	36	From homology study; Rd1 mutant
Y427G	2.92 (1)	$7.31x10^{-8}$	$5.20x10^{-8}$	$5.20x10^{-8}$	36	From homology study; Rd1 mutant
Y427H	15.4 (1)	$5.93x10^{-8}$	$4.14x10^{-8}$	$8.11x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
Y427K	7.09 (1)	$3.89x10^{-8}$	$2.88x10^{-8}$	$2.88x10^{-8}$	36	From homology study; Rd1 mutant
H430Q	1.85 (1)	$3.84x10^{-8}$	$2.07x10^{-8}$	$2.07x10^{-8}$	36	From homology study; Rd1 mutant
Y431H	31.7 (1)	$3.82x10^{-8}$	$3.10x10^{-8}$	$4.59x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
P439A	1.74 (1)	$2.14x10^{-8}$	$9.21x10^{-9}$	$9.21x10^{-9}$	36	From homology study; Rd1 mutant
N450P	19.2 (1)	$4.19x10^{-8}$	$3.29x10^{-8}$	$5.17x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
L473Y	19.1 (1)	$1.90x10^{-8}$	$1.32x10^{-8}$	$2.61x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
L474F	17.2 (1)	$5.65x10^{-8}$	$4.42x10^{-8}$	$7.00x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
L474W	33.4 (1)	$1.02x10^{-8}$	$7.51x10^{-9}$	$1.34x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	34.3 (1)	$9.92x10^{-9}$	$7.16x10^{-9}$	$1.32x10^{-8}$	36	
L477I	63.1 (1)	$4.33x10^{-9}$	$2.83x10^{-9}$	$6.24x10^{-9}$	36	From saturation mutagenesis library
L477T	16 (1)	$1.08x10^{-7}$	$8.86x10^{-8}$	$1.29x10^{-7}$	36	From saturation mutagenesis library; Rd2 mutant
L477V	36.4 (1)	$2.06x10^{-8}$	$1.62x10^{-8}$	$2.54x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	46.2 (1)	$1.64x10^{-8}$	$1.22x10^{-8}$	$2.14x10^{-8}$	36	

N479P	23.3 (1)	8.02×10^{-9}	5.37×10^{-9}	1.13×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
S481F	25.5 (1)	1.44×10^{-8}	1.04×10^{-8}	1.91×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
K492T	18.3 (1)	4.94×10^{-8}	3.94×10^{-8}	6.04×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
	20 (1)	4.40×10^{-8}	3.48×10^{-8}	5.41×10^{-8}	36	
K492V	21.2 (1)	2.69×10^{-8}	2.05×10^{-8}	3.43×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
	27.3 (1)	2.03×10^{-8}	1.56×10^{-8}	2.57×10^{-8}	36	
T493P	16.2 (1)	4.95×10^{-8}	3.89×10^{-8}	6.13×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
	16.8 (1)	3.75×10^{-8}	2.39×10^{-8}	5.50×10^{-8}	36	
H497N	15.7 (1)	5.37×10^{-8}	4.03×10^{-8}	6.91×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
	15 (1)	3.25×10^{-8}	2.43×10^{-8}	4.20×10^{-8}	36	
K511H	60.6 (1)	1.49×10^{-9}	8.81×10^{-10}	2.32×10^{-9}	36	From saturation mutagenesis library
S514R	64.3 (1)	3.57×10^{-9}	2.57×10^{-9}	4.77×10^{-9}	36	From saturation mutagenesis library
S533I	48.9 (1)	1.86×10^{-9}	1.10×10^{-9}	2.89×10^{-9}	36	From saturation mutagenesis library
I549L	52.7 (1)	3.75×10^{-9}	2.63×10^{-9}	2.63×10^{-9}	35	From homology study
E550D	56.3 (1)	2.36×10^{-9}	1.60×10^{-9}	1.60×10^{-9}	36	From homology study
C556G	32.6 (1)	1.36×10^{-8}	9.85×10^{-9}	1.81×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
	33.3 (1)	9.37×10^{-9}	6.41×10^{-9}	1.30×10^{-8}	36	
	35.4 (1)	7.27×10^{-9}	5.21×10^{-9}	9.72×10^{-9}	36	
C556T	16.9 (1)	4.10×10^{-8}	3.04×10^{-8}	5.34×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
R557A	41.4 (1)	1.90×10^{-9}	1.06×10^{-9}	3.06×10^{-9}	36	From saturation mutagenesis library
R557C	43.8 (1)	2.65×10^{-9}	1.73×10^{-9}	1.73×10^{-9}	35	From homology study
N558S	14.1 (1)	7.90×10^{-9}	4.39×10^{-9}	1.29×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
L561I	54.1 (1)	2.84×10^{-9}	1.94×10^{-9}	1.94×10^{-9}	36	From homology study
S564A	53.7 (1)	2.67×10^{-9}	1.84×10^{-9}	1.84×10^{-9}	36	From homology study
E569A	52 (1)	2.51×10^{-9}	1.69×10^{-9}	1.69×10^{-9}	36	From homology study
E569K	48.7 (1)	2.72×10^{-9}	1.82×10^{-9}	1.82×10^{-9}	36	From homology study

E569P	3.72 (1)	3.20×10^{-8}	1.90×10^{-8}	1.90×10^{-8}	36	From homology study; Rd1 mutant
A573T	32.9 (1)	5.24×10^{-9}	3.37×10^{-9}	7.65×10^{-9}	36	From saturation mutagenesis library; Rd2 mutant
	34.8 (1)	4.82×10^{-9}	3.25×10^{-9}	6.77×10^{-9}	36	
V574F	20.9 (1)	1.04×10^{-8}	7.17×10^{-9}	1.43×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
V574L	31.2 (1)	9.34×10^{-9}	6.35×10^{-9}	1.30×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
E575L	35.5 (1)	3.72×10^{-9}	2.45×10^{-9}	5.32×10^{-9}	36	From saturation mutagenesis library
F578I	22.8 (1)	6.37×10^{-9}	4.12×10^{-9}	9.26×10^{-9}	36	From saturation mutagenesis library; Rd2 mutant
	22.7 (1)	5.82×10^{-9}	3.54×10^{-9}	8.85×10^{-9}	36	
	26.6 (1)	4.38×10^{-9}	2.75×10^{-9}	6.51×10^{-9}	36	
L592M	51.3 (1)	1.28×10^{-9}	7.43×10^{-10}	7.43×10^{-10}	36	From homology study
A599S, C639T	49.4 (3)	3.33×10^{-9}	2.47×10^{-9}	4.35×10^{-9}	48	From saturation mutagenesis library
	54.7 (1)	2.53×10^{-9}	1.61×10^{-9}	3.71×10^{-9}	36	
D614L	50.5 (3)	3.22×10^{-9}	1.48×10^{-9}	5.98×10^{-9}	45	From homology study
M618V	99.8 (1)	2.20×10^{-9}	1.58×10^{-9}	1.58×10^{-9}	35	From homology study
E620D	20.8 (1)	3.68×10^{-9}	2.11×10^{-9}	2.11×10^{-9}	36	From homology study
A621K	6.3 (1)	2.35×10^{-9}	5.85×10^{-10}	5.85×10^{-10}	36	From homology study
A621S	63.5 (1)	4.02×10^{-9}	2.94×10^{-9}	2.94×10^{-9}	35	From homology study
L622I, C639I	31.3 (3)	4.78×10^{-8}	4.12×10^{-8}	5.46×10^{-8}	48	From saturation mutagenesis library; Rd2 mutant
	42.9 (1)	3.70×10^{-8}	3.06×10^{-8}	4.38×10^{-8}	36	
C627V	42.7 (1)	4.73×10^{-9}	3.35×10^{-9}	3.35×10^{-9}	36	From homology study
V630K	57.3 (1)	3.26×10^{-9}	2.28×10^{-9}	2.28×10^{-9}	36	From homology study
V630Y	50.7 (1)	2.48×10^{-9}	1.59×10^{-9}	1.59×10^{-9}	35	From homology study
N631A	42.3 (1)	1.87×10^{-9}	1.12×10^{-9}	2.90×10^{-9}	36	From saturation mutagenesis library
C639V	52.8 (1)	2.54×10^{-9}	1.73×10^{-9}	1.73×10^{-9}	36	From homology study
C639Y	67.6 (3)	2.90×10^{-9}	1.16×10^{-9}	5.88×10^{-9}	48	From homology study
L640A	4.26 (3)	2.47×10^{-7}	1.54×10^{-7}	3.71×10^{-7}	45	From homology study; Rd1 mutant
L640N	5.62 (1)	6.60×10^{-9}	2.84×10^{-9}	2.84×10^{-9}	36	From homology study; Rd1 mutant
L640D	4.51 (1)	1.05×10^{-9}	6.00×10^{-11}	6.00×10^{-11}	36	From homology study

L640G	30 (1)	$7.47x10^{-9}$	$5.27x10^{-9}$	$5.27x10^{-9}$	36	From homology study; Rd1 mutant
	38.8 (1)	$5.25x10^{-9}$	$3.67x10^{-9}$	$7.16x10^{-9}$	36	From saturation mutagenesis library; Rd1 mutant
L640K	5.98 (1)	$1.75x10^{-9}$	$2.92x10^{-10}$	$2.92x10^{-10}$	35	From homology study
L640F	19.7 (3)	$1.10x10^{-8}$	$5.52x10^{-9}$	$1.91x10^{-8}$	47	From homology study; Rd1 mutant
L640W	51.6 (1)	$2.25x10^{-9}$	$1.51x10^{-9}$	$1.51x10^{-9}$	36	From homology study
L640Y	59.3 (3)	$6.51x10^{-9}$	$3.85x10^{-9}$	$1.01x10^{-8}$	46	From homology study; Rd1 mutant
K643N	37.4 (1)	$2.69x10^{-9}$	$1.68x10^{-9}$	$1.68x10^{-9}$	36	From homology study
K643S	3.34 (1)	$2.42x10^{-9}$	$8.52x10^{-10}$	$8.52x10^{-10}$	35	From homology study
K643V	7.31 (1)	$5.92x10^{-10}$	$1.47x10^{-10}$	$1.47x10^{-10}$	36	From homology study
L645N	12.1 (3)	$6.21x10^{-9}$	$2.36x10^{-9}$	$1.13x10^{-8}$	48	From homology study
L645M	69.1 (3)	$1.52x10^{-8}$	$8.75x10^{-9}$	$2.39x10^{-8}$	21	From homology study; Rd1 mutant
Y646A	3.46 (1)	$2.25x10^{-9}$	$3.74x10^{-10}$	$3.74x10^{-10}$	36	From homology study
Y646F	22.3 (1)	$4.42x10^{-9}$	$2.78x10^{-9}$	$2.78x10^{-9}$	36	From homology study
A648S	30.2 (1)	$1.80x10^{-9}$	$9.99x10^{-10}$	$9.99x10^{-10}$	35	From homology study
S649A	46.1 (1)	$3.88x10^{-9}$	$2.73x10^{-9}$	$2.73x10^{-9}$	35	From homology study
F652G	58.5 (1)	$1.62x10^{-9}$	$9.06x10^{-10}$	$2.64x10^{-9}$	24	From saturation mutagenesis library
	56 (1)	$1.29x10^{-9}$	$6.33x10^{-10}$	$2.27x10^{-9}$	23	
F652L	2.45 (1)	$1.62x10^{-8}$	$7.00x10^{-9}$	$7.00x10^{-9}$	36	From homology study; Rd1 mutant
Y653R	26.1 (1)	$3.08x10^{-9}$	$1.83x10^{-9}$	$1.83x10^{-9}$	35	From homology study
Y653H	40.6 (1)	$2.79x10^{-9}$	$1.76x10^{-9}$	$1.76x10^{-9}$	36	From homology study
Y653L	2.81 (1)	$1.42x10^{-8}$	$6.13x10^{-9}$	$6.13x10^{-9}$	36	From homology study; Rd1 mutant
Q655H	37.9 (1)	$2.08x10^{-9}$	$1.26x10^{-9}$	$1.26x10^{-9}$	36	From homology study
P656A	24.5 (1)	$6.17x10^{-9}$	$4.18x10^{-9}$	$4.18x10^{-9}$	36	From homology study; Rd1 mutant
R662T	106 (1)	$1.83x10^{-9}$	$1.21x10^{-9}$	$2.61x10^{-9}$	36	From saturation mutagenesis library
	109 (1)	$1.26x10^{-9}$	$8.34x10^{-10}$	$1.81x10^{-9}$	36	
S664H	56 (1)	$1.57x10^{-9}$	$8.58x10^{-10}$	$2.58x10^{-9}$	36	From saturation mutagenesis library
	63.7 (1)	$1.27x10^{-9}$	$6.69x10^{-10}$	$2.12x10^{-9}$	36	
I708Q	122 (1)	$9.31x10^{-10}$	$6.05x10^{-10}$	$1.35x10^{-9}$	36	From saturation mutagenesis library
N725R	54.6 (1)	$2.70x10^{-9}$	$1.80x10^{-9}$	$3.83x10^{-9}$	36	From saturation mutagenesis library
K769G	49.9 (1)	$3.74x10^{-9}$	$2.57x10^{-9}$	$5.17x10^{-9}$	36	From saturation mutagenesis library

N773Q	46 (1)	$4.80x10^{-9}$	$3.46x10^{-9}$	$3.46x10^{-9}$	36	From homology study
V774A	52.8 (1)	$2.41x10^{-9}$	$1.59x10^{-9}$	$1.59x10^{-9}$	36	From homology study
V774I	22.7 (1)	$9.02x10^{-9}$	$6.46x10^{-9}$	$6.46x10^{-9}$	35	From homology study; Rd1 mutant
I775A	16.1 (3)	$8.66x10^{-9}$	$4.09x10^{-9}$	$1.44x10^{-8}$	47	From homology study; Rd1 mutant
	45.7 (1)	$3.05x10^{-9}$	$2.06x10^{-9}$	$2.06x10^{-9}$	36	
I777A	28.5 (1)	$8.63x10^{-9}$	$6.40x10^{-9}$	$6.40x10^{-9}$	36	From homology study; Rd1 mutant
I777K	19.6 (3)	$2.39x10^{-8}$	$1.95x10^{-8}$	$2.88x10^{-8}$	48	From saturation mutagenesis library; Rd2 mutant
	46.6 (1)	$8.06x10^{-9}$	$6.15x10^{-9}$	$1.02x10^{-8}$	36	
	49.8 (1)	$7.12x10^{-9}$	$5.21x10^{-9}$	$9.36x10^{-9}$	36	
	28.9 (3)	$1.68x10^{-8}$	$1.35x10^{-8}$	$2.05x10^{-8}$	45	
	24.8 (3)	$2.05x10^{-8}$	$1.65x10^{-8}$	$2.49x10^{-8}$	45	
I777K, W814N	31.5 (3)	$1.86x10^{-8}$	$1.53x10^{-8}$	$2.23x10^{-8}$	48	From saturation mutagenesis library; Rd2 mutant
	39.6 (1)	$1.29x10^{-8}$	$9.96x10^{-9}$	$1.63x10^{-8}$	36	
I777V	46.3 (1)	$4.71x10^{-9}$	$3.39x10^{-9}$	$3.39x10^{-9}$	36	From homology study
I778A	3.41 (3)	$5.13x10^{-9}$	$7.22x10^{-10}$	$1.22x10^{-8}$	48	From homology study
M779L	43.2 (3)	$7.22x10^{-9}$	$4.34x10^{-9}$	$1.11x10^{-8}$	46	From homology study; Rd1 mutant
M779S	3.58 (1)	$4.04x10^{-9}$	$1.01x10^{-9}$	$1.01x10^{-9}$	36	From homology study
S781G, L782G, W783Y	23.7 (3)	$1.69x10^{-8}$	$8.83x10^{-9}$	$2.89x10^{-8}$	48	From homology study; Rd1 mutant
K785R	5.93 (1)	$5.04x10^{-9}$	$2.17x10^{-9}$	$2.17x10^{-9}$	36	From homology study
K785S	3.6 (1)	$2.23x10^{-9}$	$3.70x10^{-10}$	$3.70x10^{-10}$	36	From homology study
A787V	49.9 (3)	$3.99x10^{-9}$	$1.72x10^{-9}$	$7.69x10^{-9}$	48	From homology study
W790P	4.81 (3)	$1.03x10^{-8}$	$1.72x10^{-9}$	$3.19x10^{-8}$	48	From homology study
V791H	50.9 (1)	$1.46x10^{-9}$	$8.37x10^{-10}$	$2.33x10^{-9}$	36	From saturation mutagenesis library
D804M	86.1 (1)	$1.25x10^{-9}$	$7.69x10^{-10}$	$1.88x10^{-9}$	36	From saturation mutagenesis library
I824V	75 (1)	$1.59x10^{-9}$	$1.02x10^{-9}$	$2.32x10^{-9}$	36	From saturation mutagenesis library
P833C	22 (1)	$6.44x10^{-9}$	$4.11x10^{-9}$	$9.47x10^{-9}$	36	From saturation mutagenesis library; Rd2 mutant
M848N	25.5 (1)	$3.31x10^{-9}$	$2.07x10^{-9}$	$2.07x10^{-9}$	35	From homology study
M848D	6.65 (1)	$2.51x10^{-10}$	$3.38x10^{-11}$	$3.38x10^{-11}$	36	From homology study
M848H	4.95 (1)	$3.48x10^{-9}$	$1.08x10^{-9}$	$1.08x10^{-9}$	36	From homology study
M848P	7.73 (1)	$2.45x10^{-9}$	$7.62x10^{-10}$	$7.62x10^{-10}$	35	From homology study
	25.9 (3)	$2.79x10^{-9}$	$1.80x10^{-9}$	$4.07x10^{-9}$	48	

K849H, K857S	44 (1)	$2.12x10^{-9}$	$1.23x10^{-9}$	$3.35x10^{-9}$	36	From saturation mutagenesis library
I851L	63.1 (1)	$4.68x10^{-9}$	$3.43x10^{-9}$	$6.14x10^{-9}$	36	From saturation mutagenesis library
E858H	63.4 (1)	$1.90x10^{-9}$	$1.23x10^{-9}$	$2.75x10^{-9}$	36	From saturation mutagenesis library
E861R	66 (1)	$1.69x10^{-9}$	$1.08x10^{-9}$	$1.08x10^{-9}$	36	From homology study
C862I	4.18 (1)	$4.82x10^{-10}$	$6.23x10^{-11}$	$6.23x10^{-11}$	35	From homology study
S865G	46.3 (1)	$2.03x10^{-9}$	$1.27x10^{-9}$	$1.27x10^{-9}$	35	From homology study
D866N	3.46 (1)	$2.54x10^{-9}$	$4.23x10^{-10}$	$4.23x10^{-10}$	36	From homology study
S869T	4.28 (1)	$1.74x10^{-9}$	$6.09x10^{-10}$	$6.09x10^{-10}$	36	From homology study
F871I	48.3 (1)	$1.24x10^{-9}$	$7.03x10^{-10}$	$2.00x10^{-9}$	36	From saturation mutagenesis library
F871Y	28 (3)	$3.92x10^{-8}$	$3.27x10^{-8}$	$4.61x10^{-8}$	42	From saturation mutagenesis library; Rd2 mutant
	29.3 (3)	$4.29x10^{-8}$	$3.62x10^{-8}$	$5.00x10^{-8}$	43	
	35.5 (1)	$3.12x10^{-8}$	$2.50x10^{-8}$	$3.81x10^{-8}$	36	
V872I	5.18 (3)	$1.68x10^{-8}$	$5.21x10^{-9}$	$3.89x10^{-8}$	48	From homology study; Rd1 mutant
V872L	60.4 (1)	$2.17x10^{-9}$	$1.46x10^{-9}$	$1.46x10^{-9}$	36	From homology study
H873R	52.9 (1)	$2.75x10^{-9}$	$1.81x10^{-9}$	$1.81x10^{-9}$	35	From homology study
H873T	49.5 (1)	$3.16x10^{-9}$	$2.16x10^{-9}$	$2.16x10^{-9}$	36	From homology study
K874P	87.5 (1)	$3.01x10^{-9}$	$2.19x10^{-9}$	$3.98x10^{-9}$	36	From saturation mutagenesis library
L900S	55.9 (1)	$5.24x10^{-9}$	$3.72x10^{-9}$	$7.05x10^{-9}$	36	From saturation mutagenesis library; Rd2 mutant
	68.8 (1)	$4.77x10^{-9}$	$3.50x10^{-9}$	$6.26x10^{-9}$	36	
L909F	58.2 (1)	$6.99x10^{-9}$	$5.34x10^{-9}$	$8.87x10^{-9}$	36	From saturation mutagenesis library; Rd2 mutant
K934F	44.8 (1)	$2.23x10^{-9}$	$1.36x10^{-9}$	$3.41x10^{-9}$	36	From saturation mutagenesis library
K934W	34 (1)	$1.20x10^{-8}$	$9.01x10^{-9}$	$1.55x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
S955W, K967C	12.6 (3)	$5.69x10^{-9}$	$3.61x10^{-9}$	$8.41x10^{-9}$	48	From saturation mutagenesis library
	21.9 (1)	$4.55x10^{-9}$	$2.46x10^{-9}$	$7.55x10^{-9}$	36	
D958A	88 (1)	$1.98x10^{-9}$	$1.34x10^{-9}$	$2.78x10^{-9}$	36	From saturation mutagenesis library
F968C	14.4 (1)	$1.23x10^{-8}$	$7.42x10^{-9}$	$1.88x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
F968T	12.8 (1)	$1.98x10^{-8}$	$1.24x10^{-8}$	$2.96x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant

Y431H, L640Y, I777K, W814N	9.69 (3)	$3.65x10^{-6}$	$2.65x10^{-6}$	$4.71x10^{-6}$	11	Rd4 mutant; **Entries 4-6 for this TP-DNAP1 variant are measurements taken after 90 generations.**
	17.3 (3)	$1.65x10^{-6}$	$1.30x10^{-6}$	$1.99x10^{-6}$	11	
	7.97 (3)	$2.67x10^{-6}$	$1.48x10^{-6}$	$3.89x10^{-6}$	3	
	12.3 (3)	**$2.33x10^{-6}$**	**$1.70x10^{-6}$**	**$2.90x10^{-6}$**	**5**	
	7.65 (3)	**$4.29x10^{-6}$**	**$2.93x10^{-6}$**	**$5.66x10^{-6}$**	**6**	
	4.83 (3)	**$3.24x10^{-6}$**	**$2.25x10^{-6}$**	**$4.17x10^{-6}$**	**5**	
L474W, L640Y, I777K, W814N	9.56 (3)	$4.03x10^{-6}$	$2.85x10^{-6}$	$5.13x10^{-6}$	5	Rd4 mutant; **Entries 3-4 for this TP-DNAP1 variant are measurements taken after 90 generations.**
	3.55 (3)	$2.74x10^{-6}$	$1.84x10^{-6}$	$3.64x10^{-6}$	5	
	3.82 (3)	**$2.20x10^{-6}$**	**$1.49x10^{-6}$**	**$2.89x10^{-6}$**	**5**	
	2.38 (3)	**$1.04x10^{-6}$**	**$6.08x10^{-7}$**	**$1.54x10^{-6}$**	**6**	
V574F, I777K, L900S	7.8 (3)	$4.86x10^{-6}$	$3.77x10^{-6}$	$5.94x10^{-6}$	11	Rd4 mutant; TP-DNAP1-4-1 **Entries 4-6 for this TP-DNAP1 variant are measurements taken after 90 generations.**
	5.95 (3)	$8.53x10^{-6}$	$6.26x10^{-6}$	$1.09x10^{-5}$	10	
	5.09 (3)	$3.78x10^{-6}$	$2.58x10^{-6}$	$4.86x10^{-6}$	4	
	6.02 (3)	**$7.58x10^{-6}$**	**$5.50x10^{-6}$**	**$9.47x10^{-6}$**	**5**	
	4.08 (3)	**$6.87x10^{-6}$**	**$4.76x10^{-6}$**	**$8.89x10^{-6}$**	**5**	
	3.39 (3)	**$7.71x10^{-6}$**	**$5.20x10^{-6}$**	**$1.00x10^{-5}$**	**4**	
L477V, L640Y, I777K, W814N	11.8 (3)	$7.19x10^{-6}$	$5.45x10^{-6}$	$8.68x10^{-6}$	5	Rd4 mutant; TP-DNAP1-4-2; **Entries 4-6 for this TP-DNAP1 variant are measurements taken after 90 generations.**
	6.09 (3)	$1.88x10^{-5}$	$1.40x10^{-5}$	$2.30x10^{-5}$	5	
	9.67 (3)	$9.41x10^{-6}$	$6.53x10^{-6}$	$1.17x10^{-5}$	3	
	5.05 (3)	**$1.15x10^{-5}$**	**$8.14x10^{-6}$**	**$1.45x10^{-5}$**	**4**	
	7.45 (3)	**$1.19x10^{-5}$**	**$8.70x10^{-6}$**	**$1.48x10^{-5}$**	**5**	
	7.82 (3)	**$9.22x10^{-6}$**	**$6.48x10^{-6}$**	**$1.16x10^{-5}$**	**4**	
WT	59.8 (1)	$2.49x10^{-9}$	$1.77x10^{-9}$	$3.37x10^{-9}$	48	**Entries 24-27 for this TP-DNAP1 variant are measurements taken after 90 generations.**
	47.8 (3)	$4.21x10^{-9}$	$1.51x10^{-9}$	$9.03x10^{-9}$	23	
	59.8 (1)	$2.63x10^{-9}$	$1.89x10^{-9}$	$1.89x10^{-9}$	36	
	72.6 (1)	$4.37x10^{-9}$	$2.42x10^{-9}$	$7.16x10^{-9}$	48	
	46.9 (1)	$3.83x10^{-9}$	$1.65x10^{-9}$	$7.39x10^{-9}$	48	
	53.9 (1)	$2.69x10^{-9}$	$1.79x10^{-9}$	$3.83x10^{-9}$	36	
	55.8 (1)	$2.38x10^{-9}$	$1.39x10^{-9}$	$3.73x10^{-9}$	24	
	57.3 (1)	$2.36x10^{-9}$	$1.46x10^{-9}$	$3.55x10^{-9}$	36	
	78.6 (1)	$2.26x10^{-9}$	$1.44x10^{-9}$	$3.30x10^{-9}$	24	
	55.3 (1)	$2.23x10^{-9}$	$1.41x10^{-9}$	$3.28x10^{-9}$	36	
	58.9 (1)	$2.06x10^{-9}$	$1.36x10^{-9}$	$2.95x10^{-9}$	36	
	54.3 (1)	$1.96x10^{-9}$	$1.14x10^{-9}$	$3.09x10^{-9}$	36	
	52.7 (1)	$1.95x10^{-9}$	$1.10x10^{-9}$	$3.14x10^{-9}$	36	
	59.8 (1)	$1.84x10^{-9}$	$9.75x10^{-10}$	$3.07x10^{-9}$	19	
	68.3 (1)	$1.73x10^{-9}$	$1.11x10^{-9}$	$2.52x10^{-9}$	36	
	71.5 (1)	$1.65x10^{-9}$	$1.08x10^{-9}$	$2.39x10^{-9}$	36	
	59.5 (1)	$1.62x10^{-9}$	$1.02x10^{-9}$	$2.39x10^{-9}$	36	

	71.8 (1)	$1.37x10^{-9}$	$8.77x10^{-10}$	$2.01x10^{-9}$	36	
	65.9 (1)	$1.22x10^{-9}$	$7.30x10^{-10}$	$1.89x10^{-9}$	36	
	58.8 (1)	$1.16x10^{-9}$	$6.05x10^{-10}$	$1.96x10^{-9}$	24	
	53.1 (3)	$3.10x10^{-9}$	$2.18x10^{-9}$	$4.24x10^{-9}$	48	
	51.8 (3)	$2.23x10^{-9}$	$1.50x10^{-9}$	$3.15x10^{-9}$	45	
	47.7 (3)	$2.49x10^{-9}$	$1.74x10^{-9}$	$3.42x10^{-9}$	45	
	54.2 (3)	**$1.85x10^{-9}$**	**$1.22x10^{-9}$**	**$2.65x10^{-9}$**	**45**	
	46.9 (3)	**$2.95x10^{-9}$**	**$2.05x10^{-9}$**	**$4.06x10^{-9}$**	**45**	
	45.2 (3)	**$1.84x10^{-9}$**	**$1.15x10^{-9}$**	**$2.74x10^{-9}$**	**45**	
	51.4 (3)	**$4.28x10^{-9}$**	**$3.05x10^{-9}$**	**$5.77x10^{-9}$**	**45**	

Appendix C

Supplementary information for Chapter 3

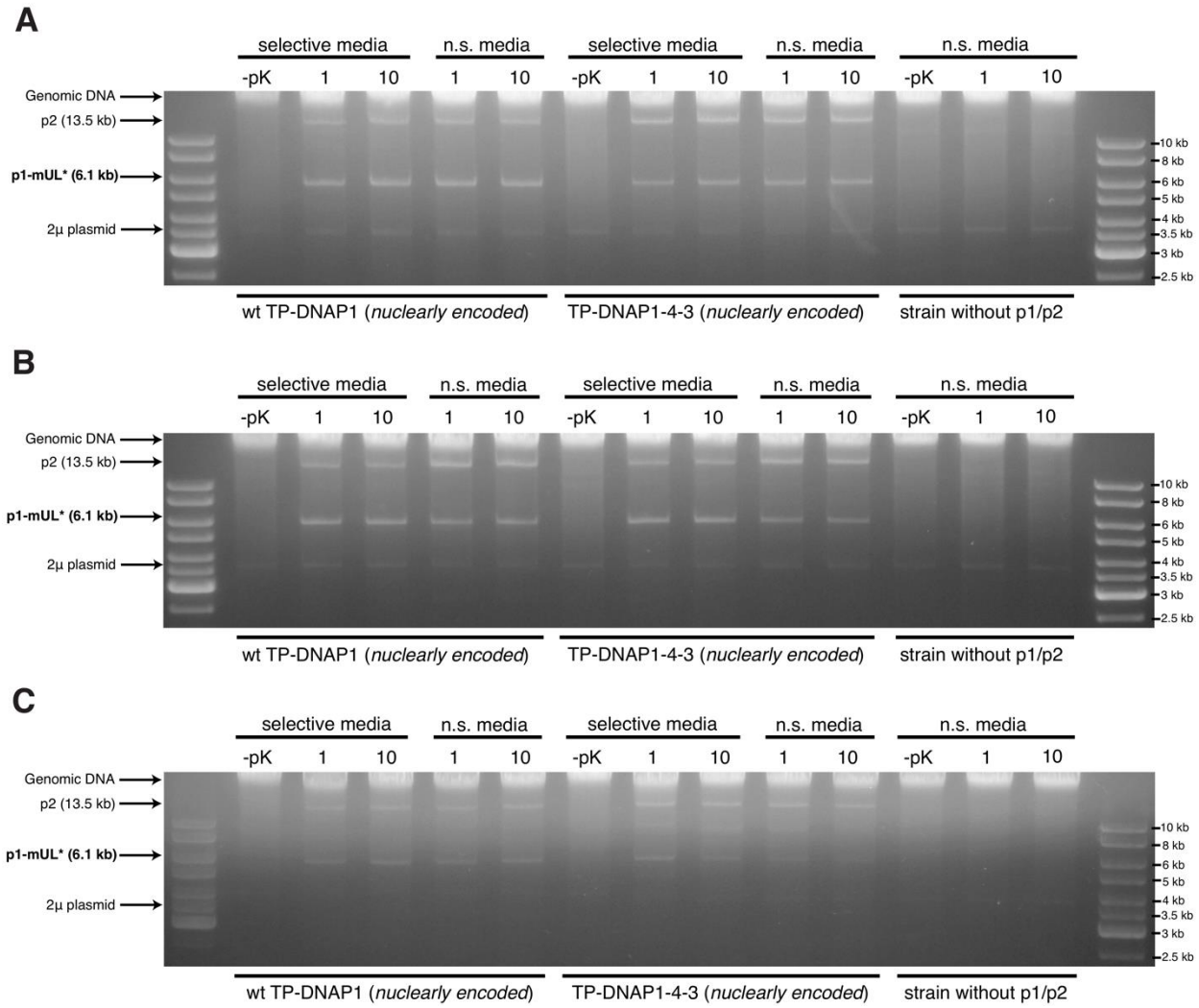

Figure S1— Stability of OrthoRep in CEN.PK2-1C (**A**), W303-1A (**B**), and BY4743 (**C**) strains. Agarose gel electrophoresis on DNA extracted from strains show the stability of OrthoRep (presence of p1-mUL*) over multiple cycles of passaging even without selection. Presence of p1-mUL* is observed after the first (1) and tenth (10) passage in both selective and non-selective (n.s.) media conditions in the presence of either a wt TP-DNAP1 or error-prone TP-DNAP1-4-3 expressed from a nuclear *CEN6/ARS4* plasmid. Parental strains without p1/p2 are shown as controls. –pK conditions are controls referring to the lack of proteinase K treatment during DNA preparation; the lack of proteinase K treatment results in the lack of all p1/p2-derived bands because the terminal proteins on p1/p2 prevent migration into agarose gels as previously described.[2]

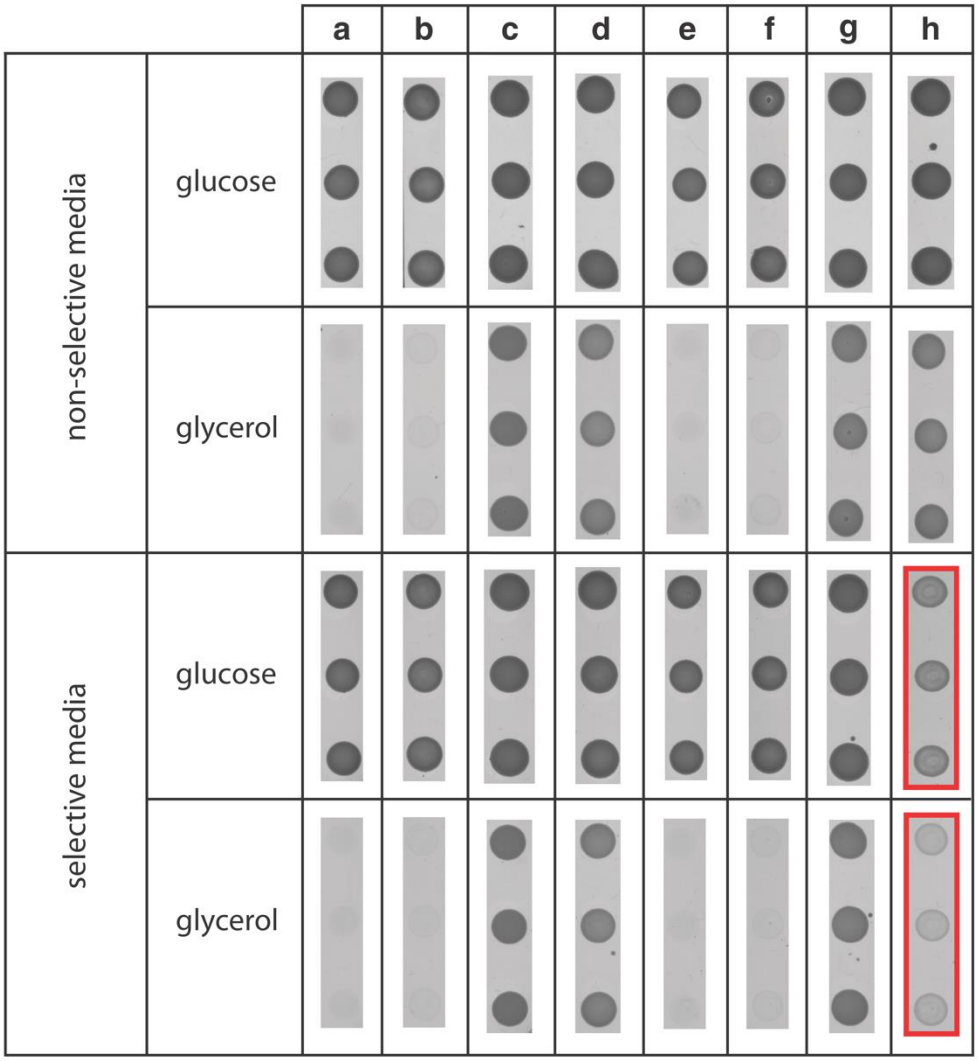

Figure S2—Ensuring that mitochondrial function hasn't changed in experiments outlined in **Figure 3.2**. Triplicates cultures from each condition in **Figure 3.2** (a-h) were spot plated (~10^6 cells) on media non-selective or selective for p1. Plates also either contained glucose or glycerol to assay for any respiration deficiency. Plates were incubated for ~4 days. For a, b, e, and f, lack of growth on glycerol shows the expected respiration deficiency of ρ^0 strains. For c, d, g, and h, growth on glycerol shows the expected respiration activity of ρ^+ strains. The red boxes highlight cells where the toxin is active in a ρ^+, which leads to instability of p1 and results in slower growth when selecting for p1-encoded genes.

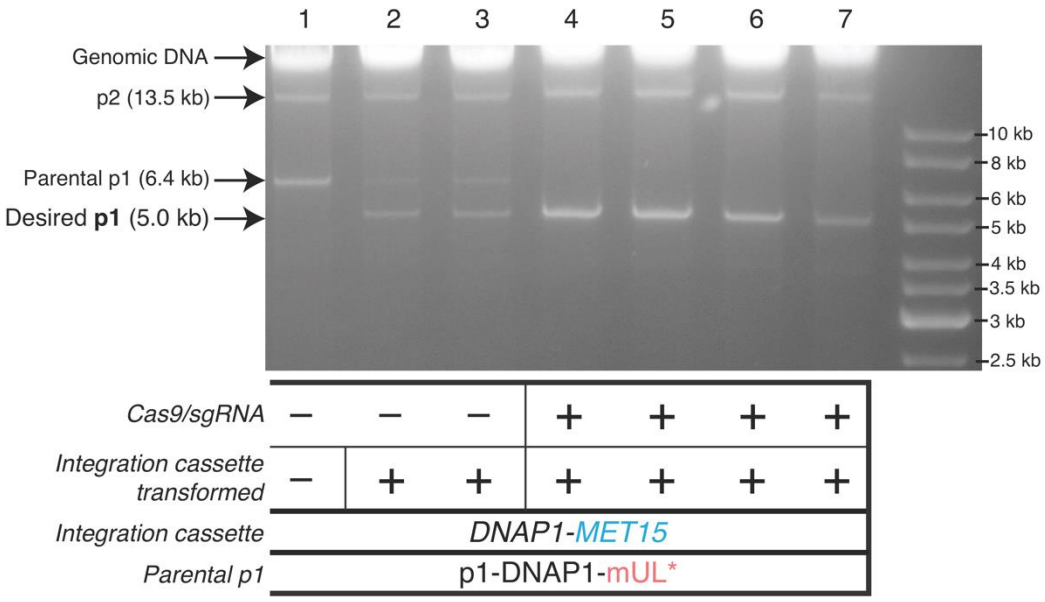

Figure S3— Integration of a model cassette (DNAP1-MET15) onto p1 in strains with p1-DNAP1-mUL* as the parental plasmid. This figure is exactly the same as Figure 3.3B, except it shows an experiment where the integration cassette genes and the parental p1 genes are swapped. Agarose gel electrophoresis on DNA extracted from cells after transformation and plating is shown (lanes 2-7). Lane 1 is the parental strain. Biological replicates from transformation without Cas9 induction (lanes 2 and 3) show both parental and desired p1 bands whereas biological replicates from a transformation with the integration cassette and concomitant induction of Cas9 (lanes 4-7) shows only the desired p1 band.

Table S1— Key plasmids used in this study.

#	Name	Source	Parent Plasmid	Origin of replication (yeast, bacterial)	Selection Marker (yeast, bacterial)	Notes
1	AR-Ec318	Previous work (Ravikumar, 2018)	See previous work	*CEN6/ARS4, ColE1*	*HIS3, KanR*	*REV1* promoter > Recoded wt *TP-DNAP1*
2	AR-Ec632	Previous work (Ravikumar, 2018)	See previous work	*CEN6/ARS4, ColE1*	*HIS3, KanR*	*REV1* promoter > Recoded *TP-DNAP1 (L474W, L640Y, I777K, W814N)* with *HIS3* (TP-DNAP1-4-3)
3	AI-Ec133	Addgene (#60847)	N/A	*2μ, ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + sgRNA > *SNR52* terminator and RNR2 promoter > *SpCAS9-SV40NLS-8XHIS* > *CYC1* terminator
4	AI-Ec145	This Work	AI-Ec133	*2μ, ColE1*	*KanMX, KanR*	contains 2 sgRNAs targetting *LEU2* locus (*GTAATAATGAAGATTATACA* and *TCAAGAATTTTACTCTGTCA*)
5	AI-Ec170	This Work	AI-Ec133	*2μ, ColE1*	*KanMX, KanR*	contains sgRNA targetting *URA3* (*GGGTCAACAGTATAGAACCG*)
6	AI-Ec200	This Work	AI-Ec133	*2μ, ColE1*	*KanMX, KanR*	contains sgRNA targetting *TRP1* (*GTTCATTGGTGAAAGTTTG*)
7	AI-Ec137	This Work	AI-Ec133	*2μ, ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + sgRNA > *SNR52* terminator and RNR2 promoter > *SpCAS9* > *CYC1* terminator
8	AI-Ec138	This Work	AI-Ec137	*2μ, ColE1*	*KanMX, KanR*	contains sgRNA targetting p1.ORF2 (*GCTGATTATAGTATATACAGA*)
9	AI-Ec147	This Work	AI-Ec137	*2μ, ColE1*	*KanMX, KanR*	contains sgRNA targetting *mKate2* (*TCTTCAAGTTGTCGCAGATCAGG*)
10	AI-Ec167	This Work	AI-Ec147	*2μ, ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + *mKate2* sgRNA > *SNR52* terminator and *GAL2* promoter > *SpCAS9* > *CYC1* terminator
11	AI-Ec168	This Work	AI-Ec167	*2μ, ColE1*	*KanMX, KanR*	*tRNA TYR* promoter > *HDV* ribozyme + *MET15* sgRNA (*GCTAAGAAGTATCTATCTATA*) > *SNR52* terminator and *GAL2* promoter > *SpCAS9* > *CYC1* terminator
12	AI-Ec180	This Work	N/A	*N/A, ColE1*	*URA3* (p1), *AmpR*	p1 recombination cassette that integrates *(p2ORF10 promoter > URA3)* to the 3' of *ORF4* to make p1-DNAP1-T/A-U
13	AR-Ec278	Previous Work (Ravikumar, 2014)	See previous work	*N/A, ColE1*	*URA3* (p1), *AmpR*	p1 recombination cassette that integrates *URA3* and *leu2* (538C>T, 539?>G)* in place of ORFs 2-4 to create p1-DNAP1-U(*TGA)
14	AI-Ec154	This work	AR-Ec278	*N/A, ColE1*	*MET15* (p1), *AmpR*	p1 recombination cassette that integrates *(p1ORF2 promoter > MET15)* in place of ORFs 2-4 to make p1-DNAP1-MET15
15	AI-Ec196	This work	AR-Ec505 (unpublished work)	*N/A, ColE1*	*URA3* (p1), *AmpR*	p1 recombination cassette that integrates *mKate2, URA3*, and *leu2* (538 ?>G)* in place of ORFs 2-4 to make p1-DNAP1-mU*(TAA)
16	AI-Ec290	This work	AI-Ec196	*N/A, ColE1*	*URA3* (p1), *AmpR*	p1 recombination cassette that integrates *mKate2, URA3*, and *K. lactis* genomic DNA in place of ORFs 2-4 to make p1-DNAP1-mU-kl14
17	AI-Ec291	This work	AI-Ec196	*N/A, ColE1*	*URA3* (p1), *AmpR*	p1 recombination cassette that integrates *mKate2, URA3*, and *K. lactis* genomic DNA in place of ORFs 2-4 to make p1-DNAP1-mU-kl18
18	AI-Ec292	This work	AI-Ec196	*N/A, ColE1*	*URA3* (p1), *AmpR*	p1 recombination cassette that integrates *mKate2, URA3*, and *K. lactis* genomic DNA in place of ORFs 2-4 to make p1-DNAP1-mU-kl22

Table S2— Key strains used in this study.

Strain	Genotype	Parent Strain	Source	
AH22	*MAT* **a** *can1 his4-519 leu2-3, 112*	N/A	ATCC (Catalog #38626)	
F102-2	*MAT* **a** *can1 his4-519 leu2-3, 112* ρ^0 + p1 + p2	N/A	ATCC (Catalog#200585)	
AJ-Y86	*MAT* **a** *can1 leu2Δ0 ura3Δ0 his4-519* ρ^0 + p1 + p2	F102-2	This work	
AR-Y292	*MAT* **a** *can1 his3 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1 + p2	See previous work	Previous work (Ravikumar, 2018)	
AJ-Y119	*MAT* **a** can1 *his3 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1-DNAP1-TA-U [*p1ORF4::URA3*] + p2	AR-Y292	This work	
AJ-Y95	*MAT* **a** *can1 leu2Δ0 ura3Δ0 his4-519* ρ^0 + p1-DNAP1-mUL*(TAA) [*p1orf2Δ::URA3/mKate2/leu2(538C>T*)] + p2	AJ-Y86	This work	
AR-Y	*MAT* **a** *can1 his3 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec318	See previous work	Previous work (Ravikumar, 2018)	
AR-Y	*MAT* **a** *can1 his3 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec632	See previous work	Previous work (Ravikumar, 2018)	
AR-Y	*MAT* **a** *can1 his3 LEU2 URA3 HIS4* ρ^0 + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec318	See previous work	Previous work (Ravikumar, 2018)	
AR-Y	*MAT* **a** *can1 his3 LEU2 URA3 HIS4* ρ^0 + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec632	See previous work	Previous work (Ravikumar, 2018)	
BY4741	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+	N/A	Gift from J. Dueber	
AJ-Y90	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-DNAP1-mUL*(TAA) [*p1orf2Δ::URA3/mKate2/leu2(538C>T*)] + p2	BY4741	This work	
AJ-Y91	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-DNAP1-mUL*(TAA) *[p1orf2Δ::URA3/mKate2/leu2(538C>T)]* + p2 + AJ-Ec167	AJ-Y90	This work	
AJ-Y92	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-DNAP1-MET15 [*p1orf2Δ::MET15*] + p2 + AJ-Ec168	AJ-Y91	This work	
AJ-Y170	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec318	BY4741	This work	
AJ-Y171	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec632	BY4741	This work	
AJ-Y190	*MAT* **a** *his3Δ1 LEU2 met15Δ0 URA3* ρ^+	BY4741	This work	
AJ-Y191	*MAT* **a** *his3Δ1 LEU2 met15Δ0 URA3* ρ^+ + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec318	AJ-Y190	This work	
AJ-Y192	*MAT* **a** *his3Δ1 LEU2 met15Δ0 URA3* ρ^+ + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec632	AJ-Y190	This work	
AJ-Y120	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-DNAP1-TA-U [*p1met15Δ::URA3*] + p2	AJ-Y92	This work	
AJ-Y231	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-DNAP1-mU-kl14 [*p1met15Δ::mKate2/URA3/* kl14] + p2	AJ-Y92	This work	
AJ-Y232	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-DNAP1-mU-kl18 [*p1met15Δ::mKate2/URA3/* kl18] + p2	AJ-Y92	This work	
AJ-Y233	*MAT* **a** *his3Δ1 leu2Δ0 met15Δ0 ura3Δ0* ρ^+ + p1-DNAP1-mU-kl22 [*p1met15Δ::mKate2/URA3/* kl22] + p2	AJ-Y92	This work	
BY4743	*MAT* **a**/α *his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0* ρ^+	N/A	Gift from S. Sandmeyer	
AJ-Y172	*MAT* **a**/α *his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0* ρ^+ + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec318	BY4743	This work	
AJ-Y173	*MAT* **a**/α *his3Δ1/his3Δ1 leu2Δ0/leu2Δ0 LYS2/lys2Δ0 met15Δ0/MET15 ura3Δ0/ura3Δ0* ρ^+ + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec632	BY4743	This work	
AJ-Y193	*MAT* **a**/α *his3Δ1/his3Δ1 LEU2/LEU2 LYS2/lys2Δ0 met15Δ0/MET15 URA3::kanMX/ura3Δ0* ρ^+	BY4743	This work	
AJ-Y194	*MAT* **a**/α *his3Δ1/his3Δ1 LEU2/LEU2 LYS2/lys2Δ0 met15Δ0/MET15 URA3::kanMX/ura3Δ0* ρ^+ + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec318	AJ-Y193	This work	
AJ-Y195	*MAT* **a**/α *his3Δ1/his3Δ1 LEU2/LEU2 LYS2/lys2Δ0 met15Δ0/MET15 URA3::kanMX/ura3Δ0* ρ^+ + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec632	AJ-Y193	This work	
CEN.PK2-1C	*MAT* **a** *ura3-52 trp1-289 leu2-3,112 his3Δ1* ρ^+	N/A	Gift from J. Dueber	
AJ-Y073	*MAT* **a** *ura3Δ0 trp1-289 leu2Δ0 his3Δ1* ρ^+	CEN.PK2-1C	This work	
AJ-Y078	*MAT* **a** *ura3Δ0 TRP1 leu2Δ0 his3Δ1* ρ^+	AJ-Y073	This work	
AJ-Y174	*MAT* **a** *ura3Δ0 TRP1 leu2Δ0 his3Δ1* ρ^+ + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec318	AJ-Y073	This work	
AJ-Y175	*MAT* **a** *ura3Δ0 TRP1 leu2Δ0 his3Δ1* ρ^+ + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec632	AJ-Y073	This work	
AJ-Y196	*MAT* **a** *URA3 trp1Δ0 LEU2 his3Δ1* ρ^+	CEN.PK2-1C	This work	
AJ-Y197	*MAT* **a** *URA3 trp1Δ0 LEU2 his3Δ1* ρ^+ + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec318	AJ-Y196	This work	
AJ-Y198	*MAT* **a** *URA3 trp1Δ0 LEU2 his3Δ1* ρ^+ + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec632	AJ-Y196	This work	
W303-1A	*MAT* **a** *ade2-1 his3-11,15 leu2-3,112 trp1-1 ura3-1 can1-100* ρ^+	N/A	Gift from J. Dueber	
AJ-Y074	*MAT* **a** *ade2-1 his3-11,15 leu2Δ0 trp1-1 ura3Δ0 can1-100* ρ^+	W303-1A	This work	
AJ-Y079	*MAT* **a** *ade2-1 his3-11,15 leu2Δ0 TRP1 ura3Δ0 can1-100* ρ^+	AJ-Y074	This work	
AJ-Y176	*MAT* **a** *ade2-1 his3-11,15 leu2Δ0 TRP1 ura3Δ0 can1-100* ρ^+ + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec318	AJ-Y074	This work	
AJ-Y177	*MAT* **a** *ade2-1 his3-11,15 leu2Δ0 TRP1 ura3Δ0 can1-100* ρ^+ + p1-mUL*(TAA) [*tpdnap1Δ::URA3/mKate2/leu2(538C>T)*] + p2 + AR-Ec632	AJ-Y074	This work	
AJ-Y199	*MAT* **a** *ade2-1 his3-11,15 LEU2 trp1-1 URA3 can1-100* ρ^+	W303-1A	This work	
AJ-Y200	*MAT* **a** *ade2-1 his3-11,15 LEU2 trp1-1 URA3 can1-100* ρ^+ + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec318	AJ-Y199	This work	
AJ-Y201	*MAT* **a** *ade2-1 his3-11,15 LEU2 trp1-1 URA3 can1-100* ρ^+ + p1-mW [*tpdnap1Δ::TRP1/mKate2*] + p2 + AR-Ec632	AJ-Y199	This work	