UNIVERSITY OF CALIFORNIA,
IRVINE

Cellular selection systems for continuous directed evolution using OrthoRep

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Garri Arzumanyan

Dissertation Committee:
Associate Professor Chang Liu, Chair
Associate Professor Wendy Liu
Professor Nancy Da Silva

2020

DEDICATION

To

my family and friends

for their hard work and endless support

in good times and in hard times.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to express deep gratitude and appreciation to my committee chair, Professor Chang Liu. Through setting higher expectations upon me than I would believe myself capable of, he helped me grow past my own boundaries of what I believed to be possible or comfortable. Through his contagious optimism and belief, I learned to believe in the importance of my work in a grander frame, which I struggled to see on some occasions. Without his guidance and persistent help, the work presented in this dissertation would not have been possible. In hindsight I appreciate his guidance because I will be able to look back on my work and personal growth with pride.

I would like to also thank my committee members, Professor Wendy Liu and Professor Nancy da Silva for their time and valuable insight, which helped me pay closer attention to detail when presenting my work and be more rigorous in judging the validity of my data.

Additionally, a thank you to Dr. Arjun Ravikumar for years of direct mentorship. I am glad to say that Arjun's work ethic rubbed off on me, to which I attribute many of my successes. Not only this, but the many discussions we had about research largely shaped my view of the directed evolution and protein engineering field.

I thank ACS Publications and Elsevier for permission to include copyrighted text and figures as part of my dissertation.

VITA

Garri Arzumanyan

EDUCATION

Doctor of Philosophy in Biomedical Engineering **2020**

University of California, Irvine

Masters of Science in Biomedical Engineering **2018**

University of California, Irvine

Bachelor of Science in Biomedical Engineering **2015**

University of California, San Diego

RESEARCH EXPERIENCE

Graduate Research Assistant **2015-2020**

University of California, Irvine

Undergraduate Research Assistant **2013-2015**

University of California, San Diego

TEACHING EXPERIENCE

Teaching Assistant **2017-2020**

University of California, Irvine

Cell and Molecular Engineering I *Winter 2017, Winter 2020*

CCBS Short Course *2018, 2019, 2020*

PUBLICATIONS

Zhong Z, Wong BG, Ravikumar A, Arzumanyan GA, Khalil AS, Liu CC. Automated Continuous Evolution of Proteins *in vivo*. *ACS Synthetic Biology*, **In Press** (2020).

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

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).

Du B., Zielinski DC, Kavvas ES, Dräger A, Tan J, Zhang Z, Ruggiero K, Arzumanyan GA, Palsson BO. *et al.* Evaluation of rate law approximations in bottom-up kinetic models of metabolism. *BMC Syst Biol* **10,** 40 (2016).

ABSTRACT OF THE DISSERTATION

Cellular selection systems for continuous directed evolution using OrthoRep

by

Garri Arzumanyan

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2020

Professor Chang Liu, Chair

In the past decade, newly developed continuous evolution systems have allowed biomolecules with demanding functions to be engineered with greatly reduced time and labor requirements. In this work, we expanded the range of proteins that can be evolved by our laboratory's continuous evolution system, OrthoRep, and studied nuances regarding its replication system. By leveraging the scalability of OrthoRep, we evolved drug-resistant malarial dihydrofolate reductases (DHFRs) in 90 independent replicates and uncovered a more complex fitness landscape than previously realized. This included common adaptive trajectories constrained by epistasis, rare outcomes that avoid a frequent early adaptive mutation, and a suboptimal fitness peak that occasionally traps evolving populations. Next, we developed a selection for protein-protein interactions by linking the split DHFR protein-complementation assay (PCA) with OrthoRep. The dynamic range of selection was expanded to improve binders from low (K_d ~ 10 uM) to high (K_d ~ 1 nM) affinity. Using this system, we matured the affinity of leucine zipper pairs. Lastly, we showed that the p1 and p2 linear plasmids comprising OrthoRep are replicated through mutually orthogonal mechanisms. This finding sets the ground for future applications in dual-channel recording, as well as evolution of multiple proteins at custom mutation rates.

CHAPTER 1

Introduction to continuous evolution and cellular selection systems

1.1 Continuous directed evolution

Directed evolution is a powerful approach for engineering biomolecules through repeated rounds of diversification and functional selection.[1] However, the scalability of this traditional approach is limited because each round of mutation and selection is labor-intensive and requires one to two weeks to carry out in a laboratory setting. This limits researchers' ability to evolve new or improved functions that require long mutational paths, and prevents parallelization of directed evolution experiments to study adaptive trajectories and fitness landscapes.

The limitations of traditional directed evolution can be overcome by continuous evolution systems that perform mutagenesis and selection simultaneously akin to natural evolution, but at much higher mutation rates targeted specifically to genes of interest to shorten the time scale of evolution.[2,3,4] The field of continuous evolution was pioneered by phage-assisted continuous evolution, or PACE, in which genes of interest are encoded on phage and propagated in a modified *E. coli* host with high mutation rate to generate diversity.[2] The desired function of the evolving gene is linked to phage infectivity, resulting in enrichment of gene variants that exhibit the desired protein function. PACE has made evolution through long mutational paths routine, while shortening the time scale of directed evolution from weeks to days while minimizing the labor required. However, the reliance on a specialized chemostat limits scalability and presents a difficult barrier for beginner users, hindering widespread use of PACE outside of experienced hands. The system is also limited by prokaryotic protein folding constraints and incompatibility

with selections based on cellular phenotypes. This limitation was addressed by a recently developed system, VEGAS, which leverages the high mutation rates of the Sindbis virus to perform continuous evolution in mammalian hosts.[4] However, engineering selection systems in mammalian hosts is slow and subject to a limited toolkit of synthetic biology tools. Thus, a niche remains for a scalable continuous evolution system that can evolve proteins through long mutational paths with minimal labor and expand the range of evolvable proteins beyond those that can be expressed in prokaryotic hosts.

1.2 OrthoRep: An orthogonal replication system for continuous evolution

To address this unfilled niche, our laboratory has previously developed OrthoRep, a continuous evolution system in *S. cerevisiae*.[3] OrthoRep targets mutagenesis to genes of interest encoded on the p1 linear plasmid, whose replication mechanism is completely orthogonal to genomic replication. This orthogonality allows p1 to be mutated by its dedicated, protein-primed DNA polymerase at a 100,000 fold higher mutation rate than the genome. Due to this property, the host genome remains in tact and cheater mutations which may lead to trivial solutions to a directed evolution problem are prevented. OrthoRep is easy to use and scalable because evolution happens simply through passaging of yeast cultures, which can be scaled to small volumes and ran in high-throughput in parallel. The use of a eukaryotic host expands the scope of proteins that can be evolved to those that require more nuanced post-translational modifications or folding mechanisms not possible in a prokaryotic system. In this work, we report the development of several selection systems for evolving proteins with OrthoRep. We

also engineered the replication of the auxiliary linear plasmid of p2 to discover that it forms a second, mutually orthogonal replication system in addition to p1.

1.3 Cellular selection systems for continuous evolution

In vivo continuous mutagenesis systems rely on cellular or phenotypic selections to enrich the most fit mutants from a continuously mutating population. This occurs on long time scales in natural evolution due to low mutation rates, as genes evolve individually or as parts of intricate pathways to achieve a fitness advantage for their host organism. While natural fitness landscapes are defined by environmental factors, researchers can create artificial fitness landscapes by defining their own selection pressures to drive evolution of a gene towards desired activity. The simplest selections evolve protein functions that are directly beneficial to the host, such as the ability to produce a nutrient or provide resistance against a small molecule drug. More complex protein functions can be evolved by using a genetic circuit or biosensor to link the function of interest to a selectable phenotype. Such selection systems have been designed to evolve novel or improved enzymatic activity, recognition of new protease cleavage sites, RNA polymerases with specificity to new promoter sequences, protein-protein interactions, drug resistance and more.[5,6,7,8,9] If protein activity cannot be tied to the fitness of the host, it can also be linked to facile readouts like fluorescence or colorometric that enable *ex vivo* screening, although this approach lowers throughput and nullifies the main advantages of continuous mutagenesis systems.

Selection systems designed for continuous evolution systems like PACE or OrthoRep should be designed with desirable properties like predictability, tunability of selection stringency, and

minimal potential for cheater mutants. Continuous evolution experiments are often repeated or performed in parallel with many replicates, making predictability an important feature to guide changes to the selection ramp-ups or to compare mutational outcomes between replicates. A way to dynamically tune selection stringency is also a crucial feature of well-designed selections because it allows the researcher to push the protein of interest to evolve the highest possible level desired function or activity. Ideally, stringency is tuned with simple inputs like addition or removal of a small molecule or nutrient, light induction, or adjustment of chemostat flow rate and etc. This allows the selection to be truly continuous. In some cases, however, a new host must be genetically engineered to impose stronger selection. Although this tactic may delay or pause evolution temporarily, it has been used to great effect in PACE selections for T7 RNA polymerase or Cas9 binding activity to create hosts with evolutionary "stepping stones" that allow difficult selections to get off the ground.[10] PACE is especially amenable to this approach due to the ease with which a new host can be flown into the evolving "lagoon", whereas the same approach would be more troublesome with OrthoRep due to the need to transfer linear plasmids to a new host via mating or protoplast fusion.

Without proper precautions, continuous evolution systems are prone to cheater mutants that escape the proposed selection pressure through trivial solutions. For instance, a system that selects for protein binding by linking it to the production of a reporter gene may be bypassed by trivial solutions that constitutively turn on the genetic circuit that links protein binding to reporter gene activation, rather than improving the binding activity of the protein being evolved. PACE prevents such escape mutants by using a chemostat that continuously refreshes the host *E. coli* cells before they can replicate and divide inside the lagoon. OrthoRep minimizes cheater mutants by completely targeting mutagenesis to the p1 linear plasmid where only genes of interest are

mutated. To address rare genomic cheater mutants that may occur after many generations, mating strategies are being developed for refreshing the yeast host. In cases where a protein vital for the selection must be fused to the evolving protein and therefore exposed to high mutagenesis on OrthoRep, special solutions that allow post-translational fusion are needed to avoid cheater mutants (see Chapter 3 for an example of one such solution via a split-intein approach).

1.4 Advances in OrthoRep technology presented in this work

As a proof of concept and demonstration of OrthoRep's utility, we evolved drug-resistant dihydrofolate reductases (DHFRs) in high-throughput and uncovered a more complex fitness landscape than previously realized. Malarial drug resistance evolved in large part due to mutations in the dihydrofolate reductase enzyme of *P. falciparum* (PfDHFR) that broke binding to inhibitors such as pyrimethamine. Although difficulties with culturing the parasite *in vitro* prevented parallel evolution to gather replicate data adaptive trajectories, computer simulations based on resistance phenotypes of individual mutants still revealed a potentially interesting landscape with one strongly preferred trajectory through an initial S108N mutation. We noticed a unique opportunity to study the evolutionary paths of drug resistance at unprecedented scale by linking yeast survival to PfDHFR function and using OrthoRep to mutate PfDHFR towards pyrimethamine resistance. The scalability of OrthoRep allowed us to repeat this evolution in 90 independent, replicate lines with ease. The mutational data from our parallel evolution experiment uncovered that path choice is constrained through S108N due to negative epistasis with the D54N mutation, as well as revealed a previously unknown path to resistance through a double mutant that bypasses the preferred S108N path.

After fruitful results in studying drug resistance with OrthoRep, we sought to expand the scope of protein function that can be evolved OrthoRep by engineering a selection system for intracellular protein-protein interactions. Current state of the art techniques like phage and yeast display do not directly select for binders that retain intracellular binding activity, and are still limited by throughput issues due to labor and equipment requirements. We envisioned that if protein-protein interaction could be linked to yeast survival with a selection system that is both sensitive and has a high dynamic range, binding proteins could be evolved with ease and high throughput. We based our selection system on the split DHFR protein complementation assay, which detects protein-protein interactions by binding-dependent reconstitution of murine DHFR. After improvements to the dynamic range of selection and implementation of a split-intein strategy to prevent unwanted cheater mutants, we used this new system to affinity mature intracellular anti-GFP nanobodies.

Lastly, we expanded the toolkit for engineering p2, the second linear plasmid of the OrthoRep system, and showed that its replication mechanism is mutually orthogonal to that of the p1 plasmid. To achieve this, we developed a CRISPR Cas9 method for integrating heterologous genes onto p2 and engineered p2 DNA polymerases with higher mutation rates to demonstrate mutual orthogonality. These findings lay the groundwork for applications in dual-channel DNA recording and may catalyze further studies into the role of terminal proteins in defining the specificity of p1 and p2 replication.

1.5 References

(1) Packer, M. S., and Liu, D. R. (2015) Methods for the directed evolution of proteins. *Nature Reviews Genetics 16*, 379–394.

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

(3) 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. *bioRxiv* 313338.

(4) English, J. G., Olsen, R. H. J., Lansu, K., Patel, M., White, K., Cockrell, A. S., Singh, D., Strachan, R. T., Wacker, D., and Roth, B. L. (2019) VEGAS as a Platform for Facile Directed Evolution in Mammalian Cells. *Cell 178*, 748-761.e17.

(5) Arnold, F. H. (2018) Directed Evolution: Bringing New Chemistry to Life. *Angewandte Chemie International Edition 57*, 4143–4148.

(6) Packer, M. S., Rees, H. A., and Liu, D. R. (2017) Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nature Communications 8*, 956.

(7) Negative selection and stringency modulation in phage-assisted continuous evolution | Nature Chemical Biology.

(8) Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., Turner, K. H., Wang, P., Malvar, T., and Liu, D. R. (2016) Continuous evolution of B. thuringiensis toxins overcomes insect resistance. *Nature 533*, 58–63.

(9) Brown, K. M., Costanzo, M. S., Xu, W., Roy, S., Lozovsky, E. R., and Hartl, D. L. (2010) Compensatory Mutations Restore Fitness during the Evolution of Dihydrofolate Reductase. *Mol Biol Evol 27*, 2682–2690.

(10) Continuous evolution of SpCas9 variants compatible with non-G PAMs | Nature Biotechnology.

CHAPTER 2

High-Throughput evolution of PfDHFR using OrthoRep reveals adaptive trajectories leading to drug resistance

2.1 Introduction

By subjecting genes to repeated cycles of mutation and functional selection, the field of directed evolution has yielded extraordinary successes, including numerous industrial enzymes and therapeutic proteins, expanded genetic codes, and significant insights into how RNAs and proteins evolve (Davis et al., 2017; Packer and Liu, 2015; Turner, 2009). However, existing approaches to directed evolution are difficult to scale: classical methods rely on onerous rounds of *in vitro* gene diversification followed by transformation into cells for expression and selection, and a pioneering phage-assisted continuous evolution system requires specialized setups and is largely incompatible with selections based on cellular phenotypes (Esvelt et al., 2011). These shortcomings limit the routine evolution of novel biomolecular functions that require long mutational paths to access, the rapid evolution of enzymes and metabolic pathways that fully integrate with host systems, and the extensive parallelization of directed evolution experiments to discover multiple related functions or to map the scope of adaptive trajectories leading to important properties such as drug resistance.

In principle, the most scalable and experimentally straightforward evolution systems are living cells, since populations of cells will continuously adapt when simply passaged under selective conditions. For example, microbial evolution experiments are routinely run in high-throughput (*i.e.* scores of replicate lines) to optimize strains, map adaptive landscapes, and understand evolutionary dynamics (Hegreness et al., 2008; Kim et al., 2014; Kryazhimskiy et al., 2012;

Kryazhimskiy et al., 2014; Lang et al., 2011; Tenaillon et al., 2012). However, because genomic mutation rates are low, any single gene can only evolve slowly, making the basic passaging of cells a poor approach to the directed evolution of novel biomolecules or specific genes. While it is possible to elevate genomic mutation rates with chemical mutagens or by engineering host DNA polymerases (DNAPs) and repair systems (Badran and Liu, 2015; Herr et al., 2011; Loeb et al., 1999), the large number of essential genes in a cell's genome sets both soft and hard "speed limits" on mutation rates (Biebricher et al., 2006; Bull et al., 2007; Drake 1991; Herr et al., 2011; Loeb et al., 1999; Nowak and Schuster, 1989; Wilke et al., 2001), making only modest increases sustainable. Furthermore, genome-wide mutagenesis allows adaptation to occur through mutations outside of user-defined genes, which defeats the purpose of most directed evolution experiments that aim to deliver a specific protein or enzyme with new function, and risks the fixation of trivial solutions (*e.g.* cheater mutations) that satisfy the selection conditions.

We report OrthoRep, a highly-error-prone orthogonal DNAP-DNA plasmid pair in yeast (Ravikumar et al., 2014) that mutates user-defined genes at rates of ~$1x10^{-5}$ substitutions per base (s.p.b.) without increasing the genomic mutation rate (~10^{-10} s.p.b.) at all. This ~100,000-fold mutational acceleration enables the rapid, continuous evolution of genes entirely *in vivo*. We describe a substantial DNAP engineering effort that yielded highly error-prone orthogonal DNAP variants, show that OrthoRep stably replicates desired genes at error rates above the mutation-induced extinction threshold of the host genome (Herr et al., 2011), and demonstrate the utility of OrthoRep in studying drug resistance. By encoding *Plasmodium falciparum* DHFR (Wooden et al., 1997) on OrthoRep, we were able to evolve highly resistant *Pf*DHFRs in 90 independent replicates through the simple serial passaging of cells in the presence of the *Pf*DHFR inhibitor, pyrimethamine. Prevailing analyses of *Pf*DHFR resistance focus on a single fitness peak observed

widely in the field (Chusacultanachai et al., 2002; Hankins et al., 2001; Japrung et al., 2007; Lozovsky et al., 2009; Sirawaraporn et al., 1997), but our experiment reveals a more complex landscape including other genotypes of similar fitness. We find that a highly adaptive first-step mutation constrains path choice, leading to convergence, but also that rare mutations direct trajectories to alternative outcomes or a suboptimal fitness trap, illustrating the balance between fate and chance in drug resistance. By drastically scaling and simplifying directed evolution, OrthoRep should have widespread utility in the search for new biomolecular and cellular functions and the study of molecular adaptation.

2.2 Materials and Methods

DNA cloning. Plasmids used in this study are listed in Table S9. *E. coli* strain TG1 (Lucigen) was used for all of 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 individually cloned plasmids (*i.e.* excluding TP-DNAP1 libraries) were assembled by the Gibson method (Gibson et al., 2009). To clone plasmid 29, a DNA fragment encoding the open-reading frame of *PfDHFR* (819 bp) was obtained from IDT.

To clone the scanning saturation mutagenesis library of TP-DNAP1, a pool of ~19,000 oligonucleotides (130-200-mers) was obtained from Agilent Technologies and sub-cloned into plasmid 2. The oligo pool was designed as 29 sub-libraries, each covering a 25-50 variable amino acid region of the TP-DNAP1 open-reading frame and flanked by ~25 bp constant regions (Figure S2B). The variable region consisted of a replacement of each amino acid in the w.t. sequence with 19 codons representing the 19 other amino acids. The mutagenic codons were chosen from a 20-codon genetic code with a maximal codon adaptation index for the *S. cerevisiae* genome (Figure

S2C). Constant regions were chosen for efficient PCR amplification. Each sub-library was PCR amplified and assembled into corresponding PCR-amplified plasmid 2 backbones by the Gibson method (Gibson et al., 2009). Assembled sub-libraries were transformed into *E. coli* at >30-fold coverage of theoretical diversity and plated on selective LB plates. After overnight growth at 37 °C, transformants were scraped from plates and resuspended in 0.9% NaCl for plasmid extraction. Control transformations containing only the plasmid 2 backbones were similarly treated, to verify a low frequency (<5%) of full-length plasmid 2 carry-over. Plasmids were extracted from individual clones of two sub-libraries and subject to analysis via agarose gel electrophoresis and Sanger sequencing.

To clone mutant TP-DNAP1 shuffling libraries, plasmids of the 65 basis set mutants were pooled and crossed by the Gibson method (Gibson et al., 2009). Since many basis set mutations encode mutations outside of strictly conserved motifs, the TP-DNAP1 open-reading frame was segmented into four regions to define broader boundaries for shuffling: the exonuclease domain (amino acids 1-596), motif A (amino acids 597-684), motif B (amino acids 685-819) and motif C (amino acids 820-987). To cross the 7 motif B basis set mutants with the 10 motif A and 8 motif C basis set mutants, the corresponding regions were PCR amplified from individual mutant TP-DNAP1 plasmids, and PCR amplicons from each region were pooled in equimolar ratios. Pooled fragments were assembled with a PCR-amplified plasmid 2 backbone by the Gibson method (Gibson et al., 2009). Assembled libraries were transformed and extracted as described above. Shuffling libraries contained a large fraction of misassembled plasmids, as determined by agarose gel electrophoresis. The desired plasmid population was purified by gel extraction and re-transformation. Both transformation steps retained >100-fold coverage of theoretical library size. Plasmids were extracted from individual clones of the purified libraries and subject to analysis via gel

electrophoresis and Sanger sequencing. To cross round-3 mutants with exonuclease basis set mutants, a new region was defined to cover round-3 mutants (amino acids 597-987), and a similar cloning procedure was followed.

Yeast transformation. All yeast transformations (including p1 integrations) were performed as described previously (Ravikumar et al., 2014). Genomic modifications were made using a CRISPR-Cas9 system for *S. cerevisiae* (Ryan et al., 2016).

We note the following protocol modifications for library transformations: (i) 10 μg of plasmid DNA was added for each library transformation; (ii) cells were incubated at 30 °C for 45 min with rotation at ~10 r.p.m prior to heat shock; (iii) cells were resuspended in 0.9% NaCl after heat shock and a small portion was plated on selective SC medium to determine library size; (iv) the remaining resuspension was inoculated directly into 50 mL (per transformation) of selective SC media and grown to saturation at 30 °C.

Yeast DNA extraction. Whole-cell DNA extractions followed the yeast DNA miniprep procedure described previously (Ravikumar et al., 2014). Cytoplasmic plasmid extraction followed the standard whole-cell yeast DNA extraction protocol with a few modifications: (1) cells were washing in 0.9% NaCl prior to treatment with Zymolyase (US Biological); (2) 200 μg/mL proteinase K (Fisher Scientific) was supplemented during SDS treatment for degradation of TP; (3) rotation at ~10 r.p.m was used during Zymolyase treatment.

Characterization of p1-*Pf*DHFR strains. Strains GA-Y109, 149, 151 and 155 expressing *Pf*DHFR from p1 were derived from the parent strain, GA-Y102, via plasmid shuffle. GA-Y077 was constructed from AR-Y292 by concomitant deletion of genomically encoded *DFR1* and transformation of a centromeric plasmid encoding *PfDHFR*. Pilot studies shown in Figure S4 used

strains GA-Y077, 151 and 155. The results confirmed that strains dependent on *Pf*DHFR acquire sensitivity to pyrimethamine and evolve resistance exclusively by mutating p1-encoded *Pf*DHFR.

***Pf*DHFR evolution experiments.** In the large-scale experiment, GA-Y229, containing p1-encoded *Pf*DHFR replicated by TP-DNAP1-4-2, was serially passaged for evolution of pyrimethamine resistance. To start evolution, a saturated preculture of GA-Y229 was diluted 1:100 into 90 wells of a 96-well block containing 0.5 mL of selective SC supplemented with 100 μM pyrimethamine. (The remaining 6 wells served as controls, as described below.) The block was sealed and incubated at 30 °C. OD_{600} was monitored every 24 hours using a microplate reader (TECAN Infinite M200 PRO). When 70/90 experimental replicates reached an OD_{600} above 0.7, the entire block was passaged at a 1:100 dilution. If this cutoff was reached within 72 hours of the previous dilution, then the concentration of pyrimethamine was increased for all 90 replicates. Otherwise, pyrimethamine concentration remained the same. This was repeated until pyrimethamine concentration reached 3mM, the maximum concentration we were able to dissolve in SC media. The drug regimen proceeded from 100 μM to 500 μM, 1 mM, 2 mM, 2.5 mM and finally 3 mM. This regimen was guided by pilot experiments and designed to maintain strong selection throughout the experiment. Cultures were maintained at 100 μM for one passage, 500 μM for the second passage, 1mM for the third passage, 2 mM for passages 4-5, 2.5 mM for passages 6-7, and 3 mM for passages 8-13. During passage 4, the seal covering the 96-well block was punctured, so the passage was repeated from the passage 3 cultures stored at 4 °C. 50 μL volumes from each passage were stored with 25% glycerol at -80 °C.

Six randomly chosen control wells were filled with selective SC media lacking pyrimethamine, and two of these were seeded with GA-Y229. Media conditions in the control wells were kept the same throughout. No cross-contamination was detected and GA-Y229 grew robustly throughout.

After 70/90 replicates from passage 13 reached an OD_{600} of 0.7, the entire block was subject to whole-cell yeast DNA extraction. Bulk populations of p1 plasmids served as templates for PCR amplification of *PfDHFR*, and PCR amplicons were subject to Sanger sequencing. Mixed trace files were automatically annotated using Mutation Surveyor (SoftGenetics; Minton et al., 2011) and called mutations were manually verified. For focused analysis of the C50R, D54N, Y57H, C59R, C59Y and S108N mutations, trace file peak heights at bases 148, 160, 169, 175 and 323 in *PfDHFR* were converted to frequencies using QSVanalyzer (Carr et al., 2009). Insertion frequencies at base 737 in *PfDHFR* were calculated from trace files using TIDE (Brinkman et al., 2014). See Quantification and Statistical Analysis for details.

To track the dynamics of *Pf*DHFR evolution, mutational frequencies were tracked across all 13 passages of eight representative replicates. Cultures were inoculated from glycerol stocks into the same media condition they were last grown in. *PfDHFR* sequencing was performed as described for passage 13 of the large-scale evolution experiment. Mutation frequencies were calculated using QSVanalyzer (Carr et al., 2009) and TIDE (Brinkman et al., 2014). See Quantification and Statistical Analysis for details. (For mutations that did not fully fix in any of the sequenced populations, a homozygous mutant allele was constructed by PCR and the resulting amplicon was similarly subject to Sanger sequencing.)

For evolution of *Pf*DHFR encoding a synonymous initial mutation at S108, strain AR-Y470 was used (Table S9). Strain AR-Y470 was constructed identically to the OrthoRep strain used for large-scale *Pf*DHFR evolution (GA-Y229), with the exception that AR-Y470 was made to encode a mutated version of *PfDHFR* wherein codon S108 is changed from AGA to TCA. To evolve pyrimethamine resistance, AR-Y470 was serially passaged in a similar manner to the large-scale experiment. First, a saturated preculture of AR-Y470 was diluted 1:100 into 12 wells of a 96-well

block containing 0.5 mL of selective SC supplemented with 100 μM pyrimethamine. The block was sealed and incubated at 30 °C. OD_{600} was monitored every 24 hours using a microplate reader (TECAN Infinite M200 PRO). When 10/12 experimental replicates reached an OD_{600} above 0.7, the entire block was passaged at a 1:100 dilution. If this cutoff was reached within 72 hours of the previous dilution, then the concentration of pyrimethamine was increased for all 12 replicates. Otherwise, pyrimethamine concentration remained the same. This was repeated until pyrimethamine concentration reached 3mM, the maximum concentration we were able to dissolve in SC media. The drug regimen proceeded from 100 μM to 500 μM, 1 mM, 2 mM, 2.5 mM and finally 3 mM. Cultures were maintained at 100 μM for one passage, 500 μM for passages 2-3, 1mM for the fourth passage, 2 mM for the fifth passage, 2.5 mM for the sixth passage, and 3 mM for passages 7-11. 50 μL volumes from each passage were stored with 25% glycerol at -80 °C.

After 10/12 replicates from passage 11 reached an OD_{600} of 0.7, the entire block was subject to whole-cell yeast DNA extraction. Bulk populations of p1 plasmids served as templates for PCR amplification of *PfDHFR*, and PCR amplicons were subject to Sanger sequencing. Mixed trace files were automatically annotated using Mutation Surveyor (SoftGenetics; Minton et al., 2011) and called mutations were manually verified.

***Pf*DHFR MIC assay.** The MIC of pyrimethamine was measured for 50 *PfDHFR* alleles in the yeast strain, YH5 (Wooden et al., 1997). Plasmids 23-70 in Table S9, are yeast centromeric plasmids that express *Pf*DHFR variants from the weak *DFR1* promoter. Plasmids 23-70 were transformed into YH5 (Wooden et al., 1997) and transformations were plated on selective SC medium supplemented with 100 μg/mL dTMP. After 5-6 days of growth at 30 °C, three transformants representing each allele were expanded in selective SC media supplemented with 100 μg/mL dTMP. Cultures were grown for 4 days at 30 °C. Saturated cultures were washed to

remove any residual dTMP. Resuspensions were diluted 1:100 into 50 μL volumes of 14 media conditions: YPD supplemented with 100 μg/mL dTMP, YPD, and YPD supplemented with 50 nM, 100 nM, 500 nM, 5 μM, 30 μM, 100 μM, 300 μM, 600 μM, 1 mM, 1.25 mM, 1.5 mM or 2 mM pyrimethamine. Inoculums were transferred into 384-well microplate reader trays, which were then sealed thoroughly to prevent evaporation and grown at 30 °C. Trays were unsealed, subject to OD_{600} measurement using a microplate reader (TECAN Infinite M200 PRO), resealed, and returned to the 30 °C shaker at 3-6 hour intervals for 7 days. See Quantification and Statistical Analysis for details of MIC analysis.

Quantification of *PfDHFR* mutation frequencies. For focused analysis of the C50R, D54N, Y57H, C59R, C59Y and S108N mutations, trace file peak heights at bases 148, 160, 169, 175 and 323 in *PfDHFR* were converted to frequencies using QSVanalyzer (Carr et al., 2009). The *PfDHFR* trace file of GA-Y229 served as the template of the homozygous w.t. base, at all five positions. Insertion frequencies at base 737 in *PfDHFR* were calculated from trace files using TIDE (default settings; Brinkman et al., 2014). Four *PfDHFR* mutations (Y70C, V125V, V146A, and V195A) were present at high frequency in GA-Y229 prior to pyrimethamine selection, after this strain was bottlenecked through a single cell during plasmid shuffle. These mutations were commonly observed as hitchhikers in evolved cultures, and were excluded from analysis.

Mutation frequencies calculated via QSVanalyzer (Carr et al., 2009) were compared against frequencies calculated from deep sequencing. For deep sequencing analysis, *PfDHFR* was PCR amplified from DNA extract of two replicate populations. *PfDHFR* was amplified as two partially overlapping fragments. Amplicons were sent to Quintara Biosciences for library preparation and sequencing. At the sequencing vendor, the amplicons were additionally amplified to incorporate

the TruSeq HT i5 and i7 adapters. The amplified libraries were sequenced on an Illumina MiSeq with the 500-cycle v2 reagent kit (Cat #: MS-102-2003). Paired-end reads were merged using PEAR (Zhang et al., 2014). Merged reads containing insertions or deletions were removed from analysis. Mutation frequencies match closely with results obtained from QSVanalyzer (Carr et al., 2009). From one of the replicates, mutation frequencies calculated for the C50R, D54N, Y57H, C59R, C59Y and S108N mutations by QSVanalyzer (Carr et al., 2009) are 2.3%, 60.5%, 66.3%, 95.5%, <1%, and 33.2%, respectively. In comparison, the corresponding frequencies from deep sequencing analysis are <0.1%, 57.1%, 54.7%, 97.3%, <0.1%, and 39.5%, respectively. From the other replicate, mutation frequencies calculated for the C50R, D54N, Y57H, C59R, C59Y and S108N mutations by QSVanalyzer (Carr et al., 2009) are 5.9%, 35.8%, 51.7%, 97.4%, <1%, and 98.6%, respectively. In comparison, the corresponding frequencies from deep sequencing analysis are 2.5%, 34.1%, 52.9%, 99.6%, <0.1%, and 99.9%, respectively.

***Pf*DHFR MIC analysis.** MIC was defined as $\log_{10}$ of the lowest pyrimethamine concentration (in M) at which OD_{600} remained below 0.25 after 7 days of growth. MIC was individually calculated for three clones of the 50 *PfDHFR* alleles. In total, 3 clones did not grow robustly in YPD supplemented with dTMP and were omitted from subsequent analysis. Of the 147 clones included in analysis, 6 clones did not exceed the MIC threshold at a low pyrimethamine concentration, but grew robustly at several higher concentrations. In these cases, we attribute failed growth to experimental error, and determined MIC as if growth were sustained in the aberrant condition.

2.3 Results

High-throughput evolution of PfDHFR reveals adaptive trajectories leading to drug resistance

Sustainable, continuous, and targeted mutagenesis with OrthoRep can be used to understand and predict drug resistance in high-throughput evolution experiments that abundantly sample adaptive trajectories and outcomes. *Pf*DHFR resistance to the antimalarial drug, pyrimethamine, occurs in the wild primarily through four active site mutations (N51I, C59R, S108N, and I164L), but the broader resistance landscape remains largely unknown. Laboratory evolution and landscape-mapping studies have mostly been limited to the quadruple mutant fitness peak (qm-wild) and suggest that resistance reproducibly arises from the crucial S108N mutation, followed by step-wise paths to qm-wild (Chusacultanachai et al., 2002; Hankins et al., 2001; Japrung et al., 2007; Lozovsky et al., 2009; Sirawaraporn et al., 1997; Wooden et al., 1997). We asked whether high-throughput directed evolution of *Pf*DHFR resistance to pyrimethamine would reveal a more complex landscape with additional fitness peaks, including ones that forgo S108N.

We used OrthoRep to evolve *Pf*DHFR resistance to pyrimethamine in 90 independent 0.5 mL cultures (Figure 3A). Based on a well-established yeast model of *Pf*DHFR, we constructed transgenic yeast strains that lack endogenous DHFR and depend on p1-encoded *Pf*DHFR. These strains acquired sensitivity to pyrimethamine and in pilot studies, evolved resistance by accumulating mutations in *Pf*DHFR (Figure S4). We found that resistance arose more commonly and successfully as the mutation rate of p1 was increased (Figure S4), suggesting that OrthoRep could indeed be used to drive rapid *Pf*DHFR evolution. To perform a large-scale resistance evolution experiment, strain OR-Y8, which uses the most mutagenic TP-DNAP1 (TP-DNAP1-4-2) to replicate p1-encoded *Pf*DHFR, was seeded into 90 independent 0.5 mL cultures containing

pyrimethamine. Cultures were grown to saturation and uniformly passaged at 1:100 dilutions into media containing gradually increasing pyrimethamine concentrations chosen to maintain strong selection as populations adapted (Figure 3A). After just 13 passages (*i.e.* 87 generations), 78 surviving populations adapted to media containing the maximum soluble concentration of pyrimethamine (3 mM). (Revival experiments showed that extinction of the 12 replicates was stochastic and that they could also adapt given repeated chances (Table S6).) From Sanger sequencing analysis of bulk adapted populations (see STAR Methods for details), we identified 37 unique protein-coding mutations across all replicates and as many as six amino acid changes in a single population. A large fraction of these mutations are predicted to be adaptive. For example, ten of the 37 mutations have been previously reported to yield pyrimethamine resistance (Chusacultanachai et al., 2002; Hankins et al., 2001; Japrung et al., 2007; Tanaka et al., 1990). In addition to these 37 mutations, several mutations identified in the promoter region increased gene expression (manuscript in preparation); and we hypothesize that some of the observed synonymous mutations in *PfDHFR* reduce translational suppression mediated by binding of *Pf*DHFR to its own mRNA sequence (Zhang and Rathod, 2002).

Adapted populations primarily converged on a region of the *Pf*DHFR resistance landscape that contains previously unidentified S108N-based genotypes as fit as qm-wild. Across all replicates, we observed seven pervasive coding changes (Figure 3B), including 737_738insA, which creates an adaptive C-terminal truncation (Figure 4A). The two most common mutations, C59R and S108N, occur together in 62/78 adapted populations (Figure 3B). Although these mutations are present in qm-wild, only one population accumulated a third mutation from the qm-wild peak

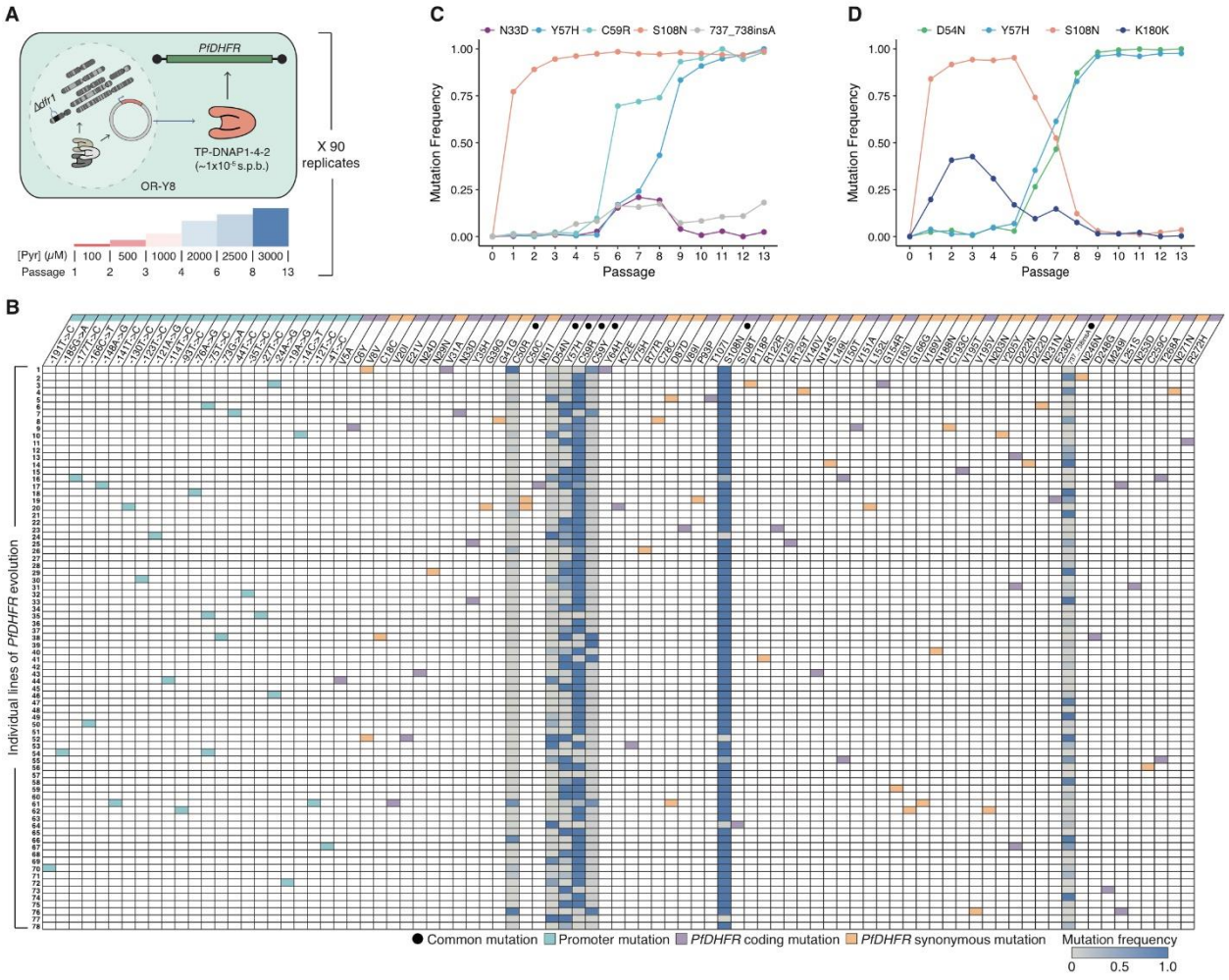


Figure 3. High-throughput evolution of *Pf*DHFR resistance to pyrimethamine

(**A**) The strain used for evolution (OR-Y8) and the drug regimen to which it was subjected. Evolving lines of OR-Y8 were monitored daily by OD_{600} measurement and passaged at a 1:100 dilution when 70/90 replicates reached an OD_{600} of 0.7. Pyrimethamine concentration was uniformly increased if diluted cultures reached the growth cutoff within 72 hours. Evolution was terminated when populations fully adapted to 3 mM pyrimethamine. (**B**) *Pf*DHFR and promoter mutations identified in 78 evolved populations from Sanger sequencing. Green, purple, and yellow shading indicates the presence of a mutation at ~20% frequency or higher. For seven commonly observed mutations, frequencies were calculated and are shown with a color on a gray-blue scale. See STAR Methods for SNP analysis details. (**C**) Mutation frequencies tracked across all 13 passages from line 60 in (**B**). (**D**) Mutation frequencies tracked across all 13 passages from line 77 in (**B**). In (**C**) and (**D**), populations from each passage were revived from glycerol stocks in the same media condition that they were initially grown in. Mutation frequencies were calculated from Sanger sequencing of revived populations, as in (**B**).

See also Figures S4 and S5 and Tables S6 and S9.

(N51I; replicate 17 in Figure 3B). Instead, most populations diverged from qm-wild and acquired combinations of C50R, D54N, or Y57H in addition to C59R and S108N, indicating a new region in the *Pf*DHFR resistance landscape with high fitness. To validate this, we fully mapped the resistance landscape of this region defined by C50R (10000), D54N (01000), Y57H (00100), C59R (00010), and S108N (00001) by constructing and measuring the MIC of all combinations of these five mutations (Figure 4A). We found that this region is indeed highly fit and contains four alleles that have similar or higher pyrimethamine MICs than qm-wild (11110, 10111, 01111 and 11111 in Figure 4A). Since these alleles are close in genotype, differing by only one or two mutations, they approximate a fitness plateau. In replicate lines 16 and 30 of our evolution experiment, this plateau is reached via 01111. Although most replicates in our experiment do not reach this particular plateau, the 00111 intermediate was frequently accessed. In these instances, additional adaptive mutations were often acquired outside of the five-mutation landscape. For example, replicate 9 contains the previously reported C6Y resistance mutation, alongside Y57H, C59R, and S108N. Since 00111 by itself is almost as resistant as genotypes on the plateau, these populations likely achieved comparable fitness atop neighboring peaks in the wider landscape. Taken together, we conclude that our evolution experiments were able to rapidly identify previously unknown solutions to *Pf*DHFR resistance.

Epistasis among mutations in S108N-based trajectories directs adaptation to 01111 and leads to the observed convergence of 00111 across replicate lines. Because S108N is a highly adaptive single mutant, 00001 rapidly and repeatedly fixed first in evolving populations (Figures 3C and S5), and blocked access to the 96/120 possible trajectories in this landscape that start with other first-step mutations. From 00001, access to the fitness plateau is constrained by negative epistasis between S108N and D54N, which is relieved and changes sign only when Y57H and C59R are

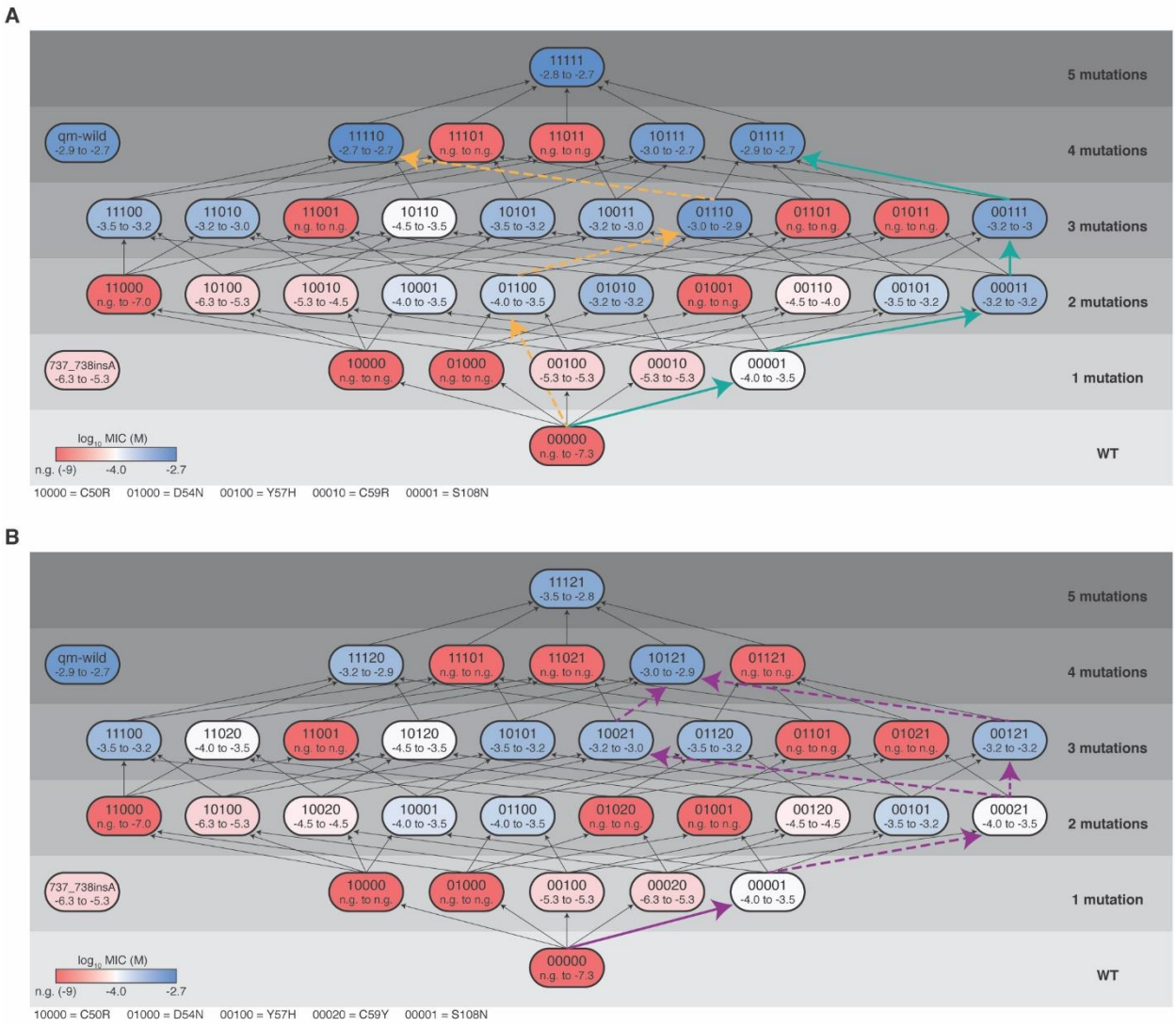


Figure 4. Five-mutation *Pf*DHFR fitness landscapes

(**A**) A fitness map of a five-mutation *Pf*DHFR landscape defined by C50R, D54N, Y57H, C59R, and S108N. Black arrows show all theoretically possible single-mutation steps in this landscape. The common S108N-dependent pathway is highlighted with a solid green arrow. The rare S108N-independent pathway is highlighted with a dashed yellow arrow. (**B**) A fitness map of a five-mutation *Pf*DHFR landscape defined by C50R, D54N, Y57H, C59Y, and S108N. Black arrows show all theoretically possible single-mutation steps in this landscape. Two commonly observed C59Y-based paths are highlighted with magenta arrows. Dashed arrows represent rare mutational steps. In (**A**) and (**B**), MIC of pyrimethamine was determined for yeast strains expressing all 32 *Pf*DHFR 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.

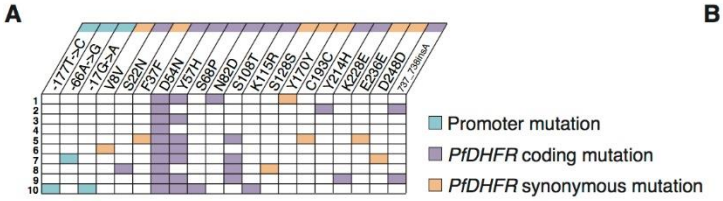
See also Figures S4 and S5 and Table S9.

both present (Figure 4A). (We note that adapted populations in our evolution experiment containing high frequencies of D54N, C59R and S108N without Y57H, typically carry other, potentially compensatory, promoter and coding mutations that take the place of Y57H.) As a result, just eight of the 24 possible paths from 00001 to the plateau avoid inactive *Pf*DHFR intermediates. Adapting populations limited to these paths likely follow the greediest one (Figure 4A). This explains why our experiment finds that evolution, particularly of 00111 and 01111, is largely repeatable.

Notably, 11110 lies on the fitness plateau without requiring S108N (Figure 4A). Three populations in our experiment avoid mutation at S108 (Figure 3B) and can access this unique quadruple mutant. We attribute this to a rare clonal interference event where the 01100 double mutant arises and displaces a population that has nearly fixed 00001 (Figure 3D). One of these replicates additionally fixed C59R to reach 01110, the triple mutant with the highest MIC (replicate 24 in Figure 3B). Stronger selection for pyrimethamine resistance, if feasible, should also fix C50R and lead to 11110 (Figure 4A).

Since 11110 is suppressed by rapid fixation of S108N, weaker early selection or greater population structure (Salverda et al., 2017; Szendro et al., 2013), should allow alternative first-step mutations (*e.g.* Y57H, C59R) to fix and increase the chance of reaching 11110. Alternatively, random initial mutations created by neutral drift have been shown to direct drug resistance evolution along new trajectories (Salverda et al., 2011). We examined this latter possibility by repeating evolution from a variant of w.t. *PfDHFR* with a synonymous codon change at S108 (AGA→TCA) that prevents mutation to N through a single substitution. Twelve populations starting from this allele were evolved under the same pyrimethamine regimen described for the large-scale experiment. In this experiment, the ten surviving populations dramatically shifted towards a new, convergent outcome

that avoids S108N and fixes D54N instead (Figure 5). Seven of these ten populations reached the 01100 double mutant that can subsequently access 11110. Since different pyrimethamine-resistant mutants should respond differently to other DHFR inhibitors, the existence of S108N-independent outcomes and the ability to steer the population towards these through weaker selection or neutral drift may have implications for drugs schedule design. In the future, we aim to leverage the scalability of OrthoRep, by starting evolution from hundreds of neutral variants of *Pf*DHFR, to capture the scope of trajectories that may be available from standing variation in natural *P. falciparum* populations and predict selection conditions that may prefer one trajectory over another. Here, we conclude that our large-scale evolution experiment is able to identify a rare path to pyrimethamine resistance that avoids the commonly observed S108N mutation that is crucial in natural *Pf*DHFR resistance.



B

A.A. mutation	Codon at S108	
	AGA (w.t.)	TCA
C50R	5%	0%
D54N	14%	100%
Y57H	41%	70%
C59R	79%	0%
C59Y	9%	0%
S108N	94%	0%
S108T	1%	40%

Figure 5. Alternative evolutionary trajectories directed by a synonymous initial mutation at S108

(**A**) *Pf*DHFR and promoter mutations identified in 10 evolved populations from Sanger sequencing. Green, purple, and yellow shading indicates the presence of a mutation at ~20% frequency or higher. See STAR Methods for SNP analysis details. (**B**) Percentage of replicates encoding commonly observed mutations, from two *Pf*DHFR evolution experiments. Evolution was started from the w.t. *PfDHFR* sequence, or from a recoded allele encoding a synonymous codon at S108.

See also Figures S4 and S5 and Table S9.

Several adaptive populations in our experiment access the broader landscape beyond 11111. As described above, in some replicates, 00111 serves as a stepping-stone to neighboring fitness peaks through additional mutations like C6Y. In other replicates, we find a suboptimal peak containing C59Y (10121; Figure 4B) at which populations are occasionally trapped (Figure 3B). In replicate 64, D54N fixes with S108T and avoids negative epistasis with S108N. Future analysis will include less frequent candidate adaptive mutations that occur in multiple replicates (*e.g.* Y35H, I150T, D222N, L251S, T268A from Figure 3B) or fix independently in time (*e.g.* M249I from Figure S5F). However, our analysis of only the most common adaptive mutations and mutational paths has already uncovered new peaks in the landscape of *Pf*DHFR-mediated drug resistance and provides examples of how epistasis results in evolutionary repeatability, and how the existence of greedy mutations such as S108N can render a highly adaptive outcome (11110) rare through early fixation. In other words, high-throughput directed evolution with OrthoRep enables the discovery of new fit regions of adaptive landscapes and thorough studies of molecular evolution at the level of a single protein.

2.4 Discussion

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 (Table S7; Barbieri et al., 2017; DiCarlo et al., 2013; Ellefson et al., 2014; Esvelt et al., 2011; Jakočiūnas et al., 2018; Wang et al., 2009). 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 and others to evolve novel

antibodies via yeast surface display (Boder and Wittrup, 1997; McMahon et al., 2018), protein-protein interactions (PPIs) or PPI inhibitors via yeast two-hybrid systems (Fields and Song, 1989), and heterologous metabolic enzymes like *Pf*DHFR that can replace essential yeast functions. The last of these applications may be especially useful in efforts to create humanized yeast models (Kachroo et al., 2015; Truong and Boeke, 2018) or to improve enzymes from difficult-to-transform hosts like plants. More sophisticated selections can also readily interface with OrthoRep, including dominant negative selections for new orthogonal tRNA/aaRS pairs or sequence-specific DNA binding proteins, which may require titration of p1's copy number (Figure 6); selections utilizing cell-based technologies such as fluorescence-activated cell sorting (FACS), continuous culturing devices (Wong et al., 2018), and droplet screening systems; and other selections that have been developed for rewiring the specificities of biosensors, GPCRs, or transcription factors (Davis, 2017; Packer and Liu, 2015; Turner, 2009). To enable its immediate widespread application, we have established OrthoRep in different yeast backgrounds, including diploids and industrially relevant strains, have constructed a suite of integration vectors for customizing p1 gene expression, and have shown that p1 can stably encode at least 18 kbs of DNA, allowing for the continuous evolution of both individual genes and multi-gene pathways (manuscripts in preparation).

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 (Table S7; Badran and Liu, 2015; Camps et al., 2003; Crook et al., 2016; Esvelt et al., 2011; Fabret et al., 2000; Finney-Manchester and Maheshri, 2013; Halperin et al., 2018; Hess et al., 2016; Ma et al., 2016; Moore et al., 2018). 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 (*e.g.* promoters (Moore et al., 2018), reverse transcriptase recognition sequences (Crook et al., 2016), gRNA target sequences (Halperin et al., 2018; Hess et al., 2016; Ma et al., 2016), or target sequences for specific DNA-binding domains (Finney-Manchester and Maheshri, 2013)). 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. Indeed, we have shown here that OrthoRep stably maintains a high mutation rate for at least 90 generations, and in ongoing evolution experiments, have used OrthoRep to sustain a continuously high rate of mutagenesis for over 300 generations.

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 mutation rate between ~10^{-9} s.p.b. to ~10^{-5} s.p.b., 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

(Desai et al., 2007), this ability to control OrthoRep's mutation rate should be valuable in directed evolution. In addition, TP-DNAP1s can potentially be engineered with custom mutational spectra (Table S8) or with high in/del rates for specialized experiments, such as the evolutionary optimization of loop regions in protein scaffolds. 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, as demonstrated here, in small culture volumes in a high-throughput manner with basic serial passaging. To the best of our knowledge, no study has performed directed evolution of a protein in more than 30 replicate experiments, but here, one experimenter easily performed 90 replicates of *Pf*DHFR evolution. 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 devolution, 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, 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.

2.5 References

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.

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.

Badran, A.H., and Liu, D.R. (2015). Development of potent *in vivo* mutagenesis plasmids with broad mutational spectra. Nat. Commun. *6*, 8425.

Barbieri, E.M., Muir, P., Akhuetie-Oni, B.O., Yellman, C.M., Isaacs, and Isaacs, F.J. (2017). Precise editing at DNA replication forks enables multiplex genome engineering in eukaryotes. Cell *171*, 1453-1467.

Bebenek, A., Dressman, H.K., Carver, G.T., Ng, S., Petrov, V., Yang, G., Konigsberg, W.H., Karam, J.D., and Drake, J.W. (2001). Interacting fidelity defects in the replicative DNA polymerase of bacteriophage RB69. J. Biol. Chem. *276*, 10387–10397.

Biebricher, C.K., and Eigen, M. (2006). What Is a Quasispecies? In Quasispecies: Concept and Implications for Virology, E. Domingo, ed. (Springer), pp. 1–31.

Boder, E.T., and Wittrup, K.D. (1997). Yeast surface display for screening combinatorial polypeptide libraries. Nat. Biotechnol. *15*, 553-557.

Boeke, J.D., La Croute, F., and Fink, G.R. (1984). A positive selection for mutants lacking orotidine-5'-phosphate decarboxylase activity in yeast: 5-fluoro-orotic acid resistance. Mol. Gen. Genet. *197*, 345–346.

Brinkman, E.K., Chen, T., Amendola, M., and van Steensel, B. (2014). Easy quantitative assessment of genome editing by sequence trace decomposition. Nucleic Acids Res. *42*, e168.

Bull, J.J., Sanjuán, R., and Wilke, C.O. (2007). Theory of lethal mutagenesis for viruses. J. Virol. *81*, 2930-2939.

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. USA *100*, 9727–9732.

Carr, I.M., Robinson, J.I., Dimitriou, R., Markham, A.F., Morgan, A.W., and Bonthron, D.T. (2009). Inferring relative proportions of DNA variants from sequencing electropherograms. Bioinformatics *25*, 3244–3250.

Chusacultanachai, S., Thiensathit, P., Tarnchompoo, B., Sirawaraporn, W., and Yuthavong, Y. (2002). Novel antifolate resistant mutations of *Plasmodium falciparum* dihydrofolate reductase selected in *Escherichia coli*. Mol. Biochem. Parasit. *120*, 61–72.

Crook, N., Abatemarco, J., Sun, J., Wagner, J.M., Schmitz, A., and Alper, H.S. (2016). *In vivo* continuous evolution of genes and pathways in yeast. Nat. Commun. *7*, 13051.

Davis, A.M., Plowright, A.T., and Valeur, E. (2017). Directing evolution: the next revolution in drug discovery? Nat. Rev. Drug Discov. *16*, 681-698.

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.

DiCarlo, J.E., Norville, J.E., Mali, P., Rios, X., Aach, J., and Church, G.M. (2013). Genome engineering in *Saccharomyces cerevisiae* using CRISPR-Cas systems. Nucleic Acids Res. *41*, 4336-4343.

Drake, J.W. (1991). A constant rate of spontaneous mutation in DNA-based microbes. Proc. Natl. Acad. Sci. USA *88*, 7160–7164.

Ellefson, J.W., Meyer, A.J., Hughes, R.A., Cannon, J.R., Brodbelt, J.S., and Ellington, A.D. (2014). Directed evolution of genetic parts and circuits by compartmentalized partnered replication. Nat. Biotechnol, *32*, 97-101.

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

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.

Fenton, L. (1960). The Sum of Lognormal Probability Distributions in Scatter Transmission Systems. IEEE T. Commun. Syst. *8*, 57-67.

Fields, S., and Song, O. (1989). A novel genetic system to detect protein-protein interactions. Nature *340*, 245-246.

Finney-Manchester, S.P., and Maheshri, N. (2013). Harnessing mutagenic homologous recombination for targeted mutagenesis *in vivo* by TaGTEAM. Nucleic Acids Res. *41*, e99.

Foster, P.L. (2006). Methods for determining spontaneous mutation rates. Method. Enzymol. *409*, 195–213.

Frishman, F. (1975). On the Arithmetic Means and Variances of Products and Ratios of Random Variables. In A Modern Course on Statistical Distributions in Scientific Work, G.P. Patil, S. Kotz, J.K. Ord, eds. (Springer), pp. 401-406.

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

Gunge, N., and Sakaguchi, K. (1981). Intergeneric transfer of deoxyribonucleic acid killer plasmids, pGKLl and pGKL2, from Kluyveromyces lactis into Saccharomyces cerevisiae by cell fusion. J. Bacteriol. *147*, 155-160.

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*, 248-252.

Hankins, E.G., Warhurst, D.C., and Sibley, C.H. (2001). Novel alleles of the *Plasmodium falciparum dhfr* highly resistant to pyrimethamine and chlorcycloguanil, but not WR99210. Mol. Biochem. Parasit. *117*, 91–102.

Hegreness, M., Shoresh, N., Damian, D., Hartl, D., and Kishony, R. (2008). Accelerated evolution of resistance in multidrug environments. Proc. Natl. Acad. Sci. USA *105*, 13977-13981.

Herr, A.J., Ogawa, M., Lawrence, N.A., Williams, L.N., Eggington, J.M., Singh, M., Smith, R. A., and Preston, B.D. (2011). Mutator suppression and escape from replication error–induced extinction in yeast. PLoS Genet. *7*, e1002282.

Hess, G.T., Frešard, L., Han, K., Lee, C.H., Li, A., Cimprich, K.A., Montgomery, S.B., and Bassik, M.C. (2016). Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. Nat. Methods *13*, 1036-1042.

Jakočiūnas, T., Pedersen, L.E., Lis, A.V., Jensen, M.K., and Keasling, J.D. (2018). CasPER, a method for directed evolution in genomic contexts using mutagenesis and CRISPR/Cas9. Metab. Eng. *48*, 288-296.

Japrung, D., Leartsakulpanich, U., Chusacultanachai, S., and Yuthavong, Y. (2007). Conflicting requirements of *Plasmodium falciparum* dihydrofolate reductase mutations conferring resistance to pyrimethamine-WR99210 combination. Antimicrob. Agents Ch. *51*, 4356–4360.

Joyce, C.M., and Steitz, T.A. (1994). Function and structure relationships in DNA polymerases. Annu. Rev. Biochem. *63*, 777–822.

Kachroo, A.H., Laurent, J.M., Yellman, C.M., Meyer A.G., Wilke, C.O., and Marcotte, E.M. (2015). Systematic humanization of yeast genes reveals conserved functions and genetic modularity. Science *348*, 921-925.

Kim, S., Lieberman, T.D., and Kishony, R. (2014). Alternating antibiotic treatments constrain evolutionary paths to multidrug resistance. Proc. Natl. Acad. Sci. USA *111*, 14494-14499.

Kryazhimskiy, S., Rice, D.P., and Desai, M.M. (2012). Population subdivision and adaptation in asexual populations of Saccharomyces cerevisiae. Evolution *66*, 1931-1941.

Kryazhimskiy, S., Rice, D.P., Jerison, E.R., and Desai, M.M. (2014). Global epistasis makes adaptation predictable despite sequence-level stochasticity. Science *344*, 1519-1522.

Lang, G.I., Botstein, D., and Desai, M.M. (2011). Genetic variation and the fate of beneficial mutations in asexual populations. Genetics *188*, 647-661.

Lang, G.I., and Murray, A.W. (2008). Estimating the per-base-pair mutation rate in the yeast *Saccharomyces cerevisiae*. Genetics *178*, 67–82.

Loeb, L.A., Essigmann, J.M., Kazazi, F., Zhang, J., Rose, K.D., and Mullins, J.I. (1999). Lethal mutagenesis of HIV with mutagenic nucleoside analogs. Proc. Natl. Acad. Sci. USA *96*, 1492-1497.

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. USA *106*, 12025–12030.

Ma, Y., Zhang, J., Yin, W., Zhang, Z., Song, Y., and Chang, X. (2016). Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. Nat. Methods *13*, 1029-1035.

McMahon, C., Baier, A.S., Pascolutti, R., Wegrecki, M., Zheng, S., Ong, J.X., Erlandson, S.C., Hilger, D., Rasmussen, S.G.F., Ring, A.M., Manglik, A., and Kruse, A.C. (2018). Yeast surface display platform for rapid discovery of conformationally selective nanobodies. Nat. Struct. Mol. Biol. *25*, 289-296.

Minton, J.A.L., Flanagan, S.E., and Ellard, S. (2011). Mutation Surveyor: Software for DNA Sequence Analysis. In PCR Mutation Detection Protocols, B.D.M. Theophilus and R. Rapley, eds. (Humana Press), pp. 143–153.

Moore, C.L., Papa, L.J. III, and Shoulders, M.D. (2018). A processive protein chimera introduces mutations across defined DNA regions *in vivo*. J. Am. Chem. Soc. *140*, 11560-11564.

Nowak, M., and Schuster, P. (1989). Error thresholds of replication in finite populations mutation frequencies and the onset of Muller's ratchet. J. Theor. Biol. *137*, 375-395.

Packer, M.S., and Liu, D.R. (2015). Methods for the directed evolution of proteins. Nat. Rev. Genet. *16*, 379–394.

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

Ryan, O.W., Poddar, S., and Cate, J.H.D. (2016). CRISPR–Cas9 genome engineering in *Saccharomyces cerevisiae* cells. Cold Spring Harbor Protocols 2016, pdb.prot086827.

Salverda, M.L.M., Dellus, E., Gorter, F.A., Debets, A.J.M., van der Oost, J., Hoekstra, R.F., Tawfik, D.S., and de Visser, J.A.G.M. (2011). Initial mutations direct alternative pathways of protein evolution. PLoS Genet. *7*, e1001321.

Salverda, M.L.M., Koomen, J., Koopmanschap, B., Zwart, M.P., and de Visser, J.A.G.M. (2017). Adaptive benefits from small mutation supplies in an antibiotic resistance enzyme. Proc. Natl. Acad. Sci. USA *114*, 12773–12778.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al. (2014). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539–539.

Sirawaraporn, W., Sathitkul, T., Sirawaraporn, R., Yuthavong, Y., and Santi, D.V. Antifolate-resistant mutants of Plasmodium falciparum dihydrofolate reductase. (1997). Proc. Natl. Acad. Sci. USA *94*, 1124–1129.

Tanaka, M., Gu, H.M., Bzik, D.J., Li, W.B., and Inselburg, J. (1990). Mutant dihydrofolate reductase-thymidylate synthase genes in pyrimethamine-resistant *Plasmodium falciparum* with polymorphic chromosome duplications. Mol. Biochem. Parasit. *42*, 83–91.

Tenaillon, O., Rodríguez-Verdugo, A., Gaut, R.L., McDonald, P., Bennett, A.F., Long, A.D., and Gaut, B. S. (2012). The molecular diversity of adaptive convergence. Science *335*, 457-461.

Truong, D.M., and Boeke, J.D. (2018). Resetting the yeast epigenome with human nucleosomes. Cell *171*, 1508-1519.

Turner, N.J. (2009). Directed evolution drives the next generation of biocatalysts. Nat. Chem. Biol. *5*, 567-573.

Wang, H.H., Isaacs, F.J., Carr, P.A., Sun, Z.Z., Xu, G., Forest, C.R., and Church, G.M. (2009). Programming cells by multiplex genome engineering and accelerated evolution. Nature *460*, 894-898.

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.

Wilke, C.O., Wang, J.L., Ofria, C., Lenski, R.E., and Adami, C. (2001). Evolution of digital organisms at high mutation rates leads to survival of the flattest. Nature *412*, 331-333.

Wong, B.G., Mancuso, C.P., Kiriakov, S., Bashor, C.J., and Khalil, A.S. (2018). Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. Nat. Biotechnol. *36*, 614-623.

Wooden, J.M., Hartwell, L.H., Vasquez, B., and Sibley, C.H. (1997). Analysis in yeast of antimalarial drugs that target the dihydrofolate reductase of *Plasmodium falciparum*. Mol. Biochem. Parasit. *85*, 25–40.

Zhang, J., Kobert, K., Flouri, T. and Stamatakis, A. (2014). PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics *30*, 614–620.

Zhang, K., and Rathod, P.K. (2002). Divergent regulation of dihydrofolate reductase between malaria parasite and human host. Science *296*, 545–547.

Zheng, Q. (2017). rSalvador: An R Package for the Fluctuation Experiment. G3-Genes Genom. Genet. *7*, 3849–3856.

Author contributions

A.R. and G.A.A. designed and performed *Pf*DHFR evolution experiments. A.R., G.A.A., and C.C.L. analyzed results and wrote the manuscript. C.C.L. supervised the research.

CHAPTER 3

Continuous selection for intracellular protein-protein interactions using OrthoRep and split DHFR

3.1 Introduction

Natural interactions between proteins in metabolic and signaling pathways, multi-enzyme complexes, and cytoskeletal structure have inspired researchers to develop methods for creating custom protein-protein interactions (PPIs).[1,2] Aside from rational and computational approaches, discovery of binding proteins relies heavily on traditional directed evolution where binding activity is improved through cycles of diversification followed by screening or selection techniques such as phage and yeast display.[3,4] However, the limitations of traditional, iterative directed evolution methods are becoming apparent as continuous evolution techniques continue to increase scalability of directed evolution campaigns, while decreasing the time, labor and expensive equipment requirements.[5,6,7] For example, Badran et al. used phage-assisted continuous evolution (PACE) to rapidly evolve the Bt toxin Cry1Ac to bind to a novel receptor, TnCAD, in just 528 hours of chemostat-controlled continuous culture.[8] The impressive speed of PACE selections is balanced, however, by limitations in scalability as well as the scope of proteins which can be successfully expressed in *E. coli.* Recent developments adaptations of PACE into a non-continuous format (PANCE) may improve scalability with a tradeoff in speed.[9] We envisioned that the scalability and scope of binding proteins that can be evolved by continuous evolution could be improved by linking OrthoRep, a eukaryotic continuous evolution system, with an intracellular selection for PPIs.

Taking note of the extensive work required to expand the dynamic range of the PACE two-hybrid selection, we searched carefully for a PPI selection method which was amenable to engineering and functions robustly in *S. cerevisiae*. Variations of the protein-fragment complementation assay (PCA) presented several such options. In PCA, the fitness of a cell is tied to the interaction between "bait" and "prey" proteins by covalently linking them to a variety of available split reporters.[10] Yeast two-hybrid is a widely used PCA that has been instrumental in protein interactome studies because of its ability to quickly screen complex libraries. However, high false positive rates force researchers to conduct multiple sequential rounds of positive and negative screening, often with multiple reporter genes.[11] The potential for constitutively active mutants in the synthetic promoter driving expression of the reporter in yeast two-hybrid make this flaw especially prohibitive in the context of continuous directed evolution, where these mutants would quickly sweep a population and outcompete real binding mutants. Cheater mutations in the reporter or activating domain are less fatal, but provide an even larger target size for weakening the selection throughout a directed evolution experiment. This made us favor the split DHFR PCA as a more robust starting point for developing our selection, where interaction between bait and prey refolds mDHFRI and mDHFRII fragments via proximity.[12] The simplicity afforded by directly restoring the enzymatic activity of an essential enzyme instead of a transcription factor bypasses the need for a secondary genetic circuit with reporter genes which make two-hybrid methods so susceptible to false positives. Though cheater mutants in mDHFRI and mDHFRII could still weaken the dynamic range of selection, the available target size for these mutants is much smaller than in two-hybrid methods and more importantly, no mutant has the potential to fully disable the selection. Additionally, we saw potentially clear paths for tuning

the dynamic range of selection of split DHFR through modulation of expression levels, and activity-altering mutations in DHFR.

Here, we linked the split DHFR system with OrthoRep to perform continuous directed evolution of intracellular PPIs. We first tuned the dynamic range of selection to distinguish binding interactions within a K_d range of 10 uM-1 nM. We then applied this improved selection to improve the binding affinity of leucine zipper pairs. These initial mock selections with leucine zipper pairs revealed the susceptibility of the system to mutations that did not improve the PPI under selection, but rather increased the activity of DHFR through resistance to methotrexate. This undesired route of adaptation was minimized by encoding mDHFRI in the genome while maintaining post-translational fusion to the prey via a split-intein, thereby sparing mDHFRI from high mutagenesis.

3.2 Materials and Methods

DNA cloning. Plasmids cloned and used in this study are listed in Table S1. All oligonucleotide primers and synthesized gene fragments (gBlocks) were purchased from IDT. Enzymes for PCR and cloning were obtained from NEB. Plasmids were cloned using Gibson assembly with overlap regions of 20–30 bp[13] or restriction cloning.

p1 integration cassettes were cloned by swapping out the gene contents of the established "fulldelpol" integration cassette used in our lab, via Gibson assembly. Bait and prey plasmids were constructed by utilizing the MluI/BamHI restriction sites flanking both bait and prey. Prey plasmids were constructed to encode prey as an N-terminal fusion to mDHFRI in a CEN6/ARS4 vector with a LEU2 selection marker. Bait plasmids were constructed to encode bait as an N-terminal fusion to mDHFRII (I114A) in a CEN6/ARS4 vector with a URA3 selection marker.

For OrthoRep evolution experiments, bait plasmids also encoded a mutagenic p1 polymerase (either “633” or “611”). In version 2 of the system, preys were instead encoded as N-terminal fusions to the CfaN split intein and encoded on the p1 plasmid.

Yeast strains and media. All yeast strains used in this study are listed in Table S2. YH5 was used on it’s own for double CEN/ARS experiments to verify protein-protein interactions. To perform p1 evolution experiments in a strain that can uptake dTMP, Y299 was created by protoplast fusion of an F102-derived strain gifted by Alex Javanpour harboring a landing pad p1-TRP1-mTurq2 with the YH5 *tup1* mutant that can uptake dTMP (from Tun Huang and Carol Sibley). The protocol for protoplast fusion is described in detail elsewhere.[14] GA-Y321 was made by integrating CfaC_mDHFRII_KanMX into the HO locus of GA-Y299. This strain was used for experiments requiring split-intein mediated fusion of prey and mDHFRI. Note: all strains derived from YH5 were propagated in media with supplemented dTMP (50 ug/mL) and pH in the range of 5.5-6.0 for slightly increased growth rate.

Yeast transformation. All yeast transformations were performed using the high-efficiency LiAc/SS carrier DNA/PEG method.[15] For p1 integrations, 2−5 μg of plasmid containing the appropriate integration cassette was digested with ScaI, yielding a linearized cassette with blunt ends containing the genes of interest flanked by regions of homology to p1. The products of the digestion reaction were directly transformed into GA-Y299 or GA-Y321 harboring harboring the p1-TRP1-mTurq2 landing pad and plated on selective solid SC medium. For transformation into YH5 derived strains, colonies appeared after 5-7 d of growth at 30 °C. CEN6/ARS4 plasmids were also transformed with the LiAc/SS carrier DNA/ PEG method, but with only 500−3000 ng of DNA for individual vectors, and at least 1 ug of each plasmid for co-transformations.

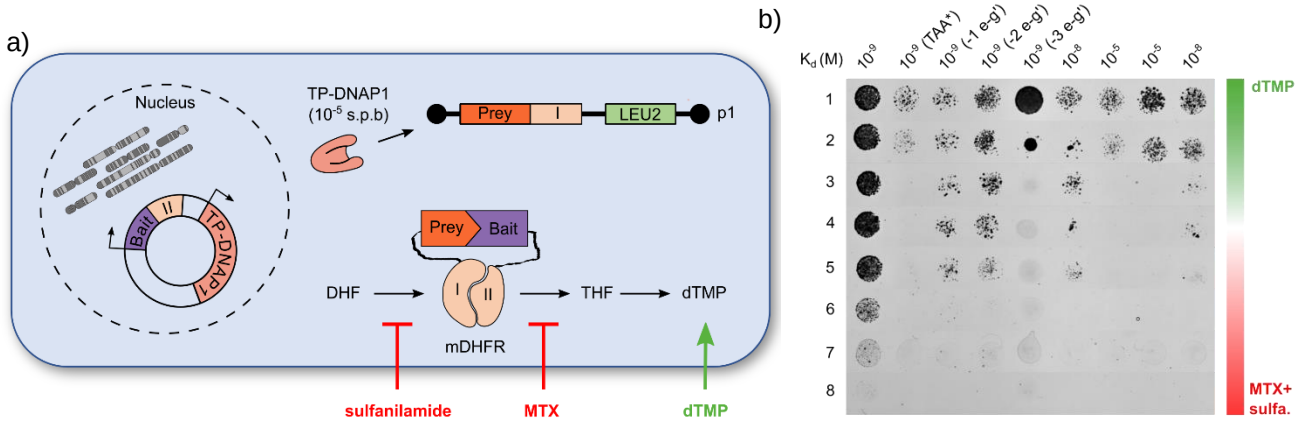
Linear plasmid DNA extraction. p1, p2, and recombinant p1 plasmids were extracted following the yeast DNA miniprep procedure as described previously.[16]

Spot plating assays. Spot plating assays were performed by plating 10 uL of saturated or 10-fold diluted cultures of yeast strains expressing bait and prey as fusions to the split DHFR fragments. Plates with varying selection stringencies were prepared by tuning the concentration of dTMP, methotrexate and sulfanilamide in agar plates based on synthetic complete media. The following stock solutions were prepared and stored at 4C: 100 mg/mL dTMP in water, 10 mM methotrexate in DMSO, and 4 g/L sulfanilamide in DMSO.

3.3 Results and Discussion

We first assessed the ability of the split DHFR system to serve as a growth selection in the context of OrthoRep. The linear plasmids p1 and p2 comprising OrthoRep were protoplast fused into YH5, a dTMP-permeable mutant of *S. cerevisiae* lacking the native yeast *DFR1* gene.[17] YH5 can sustain growth without DHFR activity by uptaking exogenous dTMP, a vital feature for weakening selection strength when starting evolution from low-affinity binders (Fig. 1a). To test for successful reconstitution of split DHFR fragments in YH5, we fused heterodimerizing leucine zippers (K_d ~ 1 nM) as "bait" and "prey" to the N-termini of the murine DHFR fragments mDHFRI (residues 0-105) and mDHFRII (residues 106-187). This strongly interacting pair of leucine zippers fully restored DHFR activity in YH5, as witnessed by complete growth complementation on plates lacking dTMP (Fig. 1b).

Next, we wanted to ensure that refolding of mDHFR is completely dependent on bait and prey interaction, rather than any latent affinity between mDHFRI and mDHFRII which has been previously reported.[12] We indeed observed concentration-dependent growth in strains expressing



c)

			p10B2 promoter					Prey						mDHFRI				
Bait	Prey	Starting K_d	-60	-40	-37	-34	-5	7	24	25	28	34	36	7	10	32	45	104
EE12RR345L	RR12EE345L	10^-9	-	-	-	-	-	-	-	-	-	-	-	-	-	F31S	-	-
EE12RR345L	RR12EE345L (TAA*)	no binding	-	-	-	-	-	-	-	*25Q	-	-	-	-	-	F31S	-	-
EE12RR345L	RR12EE345L (-1 e-g')	n.d. (<10^-9)	-	-	-	G→A	-	A7T	-	-	-	-	-	-	-	F31S	-	-
EE12RR345L	RR12EE345L (-2 e-g')	n.d. (<10^-9)	insA	-	A→G	-	-	-	-	-	-	-	-	-	A9V	F31S	E44K	P103L
EE12RR345L	RR12EE345L (-3 e-g')	n.d. (<10^-9)	-	insA	-	-	-	-	E24T	-	-	-	-	I6F	-	-	-	P103Q
EE34	RR34	10^-8	-	-	-	-	G→A	-	-	-	-	V34I	-	-	-	F31S	-	-
RR34	EE34	10^-8	-	-	-	-	-	-	-	-	-	-	-	-	-	F31S	-	P103L
EE34	EE34	10^-5	-	-	-	-	-	-	-	K25Q	G28H	-	K36Q	-	-	-	-	-
RR34	RR34	10^-5	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-

Figure 1. Split DHFR selection using OrthoRep. (A) Schematic of split DHFR selection using OrthoRep. Prey-mDHFRI fusion is encoded on p1 and replicated by a highly mutagenic DNA polymerase (TP-DNAP1-4-2). Bait-mDHFRII fusion is encoded on a nuclear plasmid to spare it from high mutagenesis. Interaction between the prey and bait leads to proximity-dependent refolding of DHFR restoration of its function in the folate pathway (conversion of DHF into THF). Selection stringency for DHFR activity, and therefore prey/bait affinity, can be modulated by tuning the concentrations of exogenous dTMP, methotrexate and sulfanilamide. (B) Spot plating assay to determine the dynamic range of split DHFR selection linked with OrthoRep. Selection stringency was increased from weak (green, top row) to strong (red, bottom row) by tuning concentrations of dTMP,methotrexate and sulfanilamide in the plates. From top to bottom: 50 ug/mL dTMP, 10 ug/mL dTMP, 1 ug/mL dTMP, 0 ug/mL dTMP, 10 uM MTX, 100 uM MTX, 100 uM MTX + 4 mg/mL sulfanilamide, 100 uM MTX + 8 mg/mL sulfanilamide. Leucine zippers spanning a ranging from no binding to K_d ~ 1 nM were introduced into YH5 as bait and prey following the scheme in 1a, plated and incubated for 6 days. Modified versions of the K_d ~ 1 nM pair with a premature stop codon (TAA*) and 1, 2 or 3 disrupted electrostatic interactions (-1 e-g', -2 e-g', -3 e-g'). (C) Mutational outcomes of mock selections for affinity maturation of leucine zipper pairs. Mutations shown are from sequenced populations that have reached the maximal selection or a plateau in the selection* (see supplement for selection ramp-up explanation). Green: mutations at electrostatic interaction sites that suggest an increase an affinity. Red: methotrexate resistant cheater mutants.

mDHFRI and mDHFRII without bait and prey (pRNR2: weak growth, pRpl18b: strong growth). To weaken this spontaneous refolding, we introduced the I114A mutation into mDHFRII, which breaks hydrophobic contacts to residues I51 and L93 in mDHFRI.[12] As expected, background growth was no longer observed at low expression levels (pRNR2), but was not completely eliminated at mid-level expression of mDHFRI and mDHFRII (pRpl18b) (Supplemental Figure 1). The I114A mutation did not affect the ability of K_d ~ 1 nM leucine zippers to refold mDHFR and restore growth (Fig 1b).

An ideal binding-dependent growth selection should have a large dynamic range, where selection strength can be adjusted to sustain growth at weak binding affinities at the beginning of an experiment and later increased to force maturation to high affinity. Because split DHFR selection already has an innate sensitivity for weak PPIs, we first focused our efforts on expanding the upper ceiling of selection. We reasoned that maintaining low cytoplasmic expression of prey-mDHFRI and bait-mDHFRII would maintain a high selection ceiling by requiring high affinity to reconstitute the sparse fragments. The bait-mDHFRII(I114A) fusion was encoded on a nuclear plasmid driven by the weak pRNR2 promoter, with closely matched expression of prey-mDHFRI on p1 using the p10B2 promoter.[18] Following this format, three leucine zipper pairs spanning a range of affinities (K_d 10 uM, 10 nM, and 1 nM) were encoded as bait and prey to gauge the system's ability to distinguish affinities by growth rate. Selection imposed by full dropout of dTMP was strong enough to hinder growth with the 10 uM zipper pair, but not with the 10 nM and 1 nM pairs. To expand dynamic range further, we used methotrexate and sulfanilamide to indirectly impose stronger selection for higher affinity by inhibiting DHFR activity. This resulted in full inhibition of strains with the 10 nM and 1 nM zipper pairs, as well as a clear increase in fitness in response to increase in affinity (Fig. 1b). This served as evidence

that selection with this system could be maintained by tuning dTMP, methotrexate and sulfanilamide in a manner that could mature binding proteins from K_d ~ 10 uM to 1 nM.

We then attempted affinity maturation of these leucine zipper pairs by exposing them to continuous mutagenesis using OrthoRep. A YH5-derivative encoding each of the three leucine zipper pairs was passaged in 50 mL liquid cultures with 1:100 dilutions after reaching saturation (usually 2-5 days per passage). Evolution from the K_d ~ 10 uM pair yielded mutations at key electrostatic interaction sites between leucine zippers, suggesting selection for higher affinity was working (Fig. 1c). However, evolution from stronger binding leucine zipper pairs (K_d ~ 10 nM and 1 nM) favored mutations in mDHFRI, in particular the F31S methotrexate-resistance mutation, suggesting there was an accessible path to gain higher DHFR activity without improving PPI affinity (Fig. 1c).

Next, we aimed to prevent unwanted DHFR mutants by targeting mutagenesis to the evolving prey and not mDHFRI via a split-intein approach. Split-inteins can carry out a protein trans-splicing reaction between two proteins of interest while leaving behind a minimal scar sequence. Post-translational fusion of prey and mDHFRI by a split-intein would lower the rate of cheater mutants by allowing mDHFRI to be encoded in the genome, where the rate of mutation is ~10,000 fold lower than on p1. To attempt this, we chose the synthetic Cfa intein for its fast reaction time of $t_{1/2}$ ~ 20s at 30C, which is also optimal for *S. cerevisiae* growth.[19] Because the dynamic range of split DHFR selection is highly dependent on predictable cytoplasmic concentration levels of the prey-mDHFRI fusion, we wanted to gauge the effect of split-intein mediated fusion on cytoplasmic protein levels. Split GFP fragments were fused to CfaN and CfaC, the N- and C- terminal fragments of the Cfa split-intein, and GFP production levels were compared to a control strain expressing GFP driven by the same promoter. A slight drop in GFP

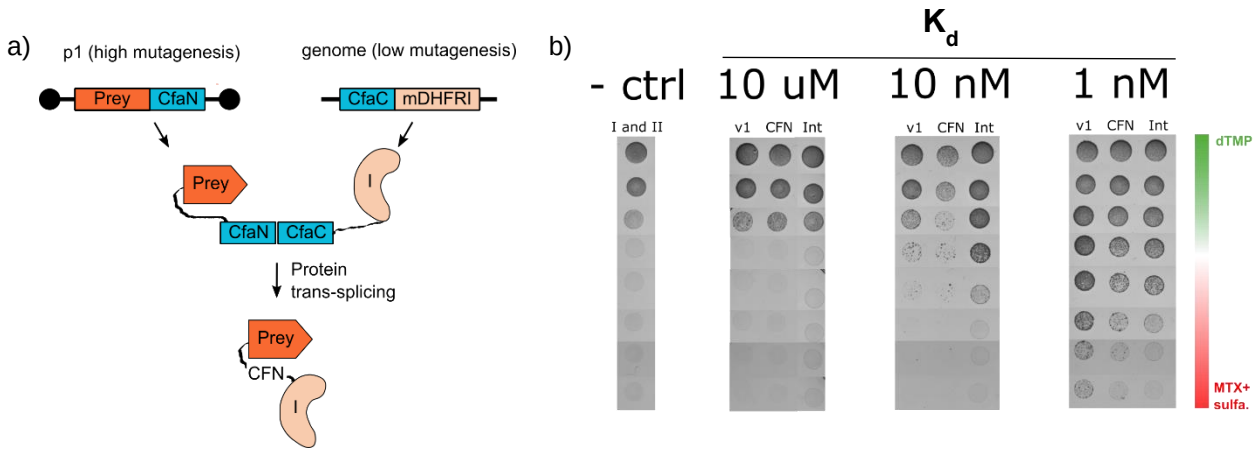


Figure 2. Split-intein approach to mitigate cheater mutants in mDHFRI. a) Post-translational fusion of prey and mDHFRI performed by the trans-splicing Cfa split-intein. This approach allows the mDHFRI fragment to be encoded separately at a genomic locus with low mutation rate, significantly lowering the rate of cheater mutations in mDHFRI, while maintaining high mutagenesis of the prey on p1. b) Spot plating assay to determine the effect of the split-intein on dynamic range of split DHFR selection. Negative control strain (I and II): expressing split DHFR fragments with no bait and prey. Leucine zippers with K_d ~10 uM, 10 nM and 1 nM were expressed as bait and prey in GA-Y321; prey was encoded on p1 as a fusion to the CfaN split intein. Selection stringency was increased from weak (green, top row) to strong (red, bottom row) by tuning concentrations of dTMP,methotrexate and sulfanilamide in the plates. From top to bottom: 50 ug/mL dTMP, 10 ug/mL dTMP, 1 ug/mL dTMP, 0 ug/mL dTMP, 10 uM MTX, 100 uM MTX, 100 uM MTX + 4 mg/mL sulfanilamide, 100 uM MTX + 8

production was caused by Cfa-mediated post-translational fusion at low and medium expression levels (pRNR2 and pRpl18b), indicating efficient post-translational fusion by Cfa in *S. cerevisiae* (Supplemental Figure 2). We then moved mDHFRI from p1 into the genome, now encoded as a fusion to CfaC (CfaC-mDHFRI). The prey protein was fused to CfaN on p1, still driven by the p10B2 promoter to minimize changes to the dynamic range of selection. We reevaluated the growth response of the selection with the previously used leucine zippers (K_d ~ 10 uM, 10 nM and 1 nM) and observed similar growth rates on plates with varying selection stringency, suggesting that the split-intein was minimally invasive to split DHFR selection dynamics.

3.4 References

(1) Pawson, T., and Nash, P. (2000) Protein–protein interactions define specificity in signal transduction. *Genes Dev. 14*, 1027–1047.

(2) Protein–Protein interactions, cytoskeletal regulation and neuronal migration | Nature Reviews Neuroscience.

(3) McCafferty, J., Griffiths, A. D., Winter, G., and Chiswell, D. J. (1990) Phage antibodies: filamentous phage displaying antibody variable domains. *Nature 348*, 552–554.

(4) Boder, E. T., and Wittrup, K. D. (1997) Yeast surface display for screening combinatorial polypeptide libraries. *Nature Biotechnology 15*, 553–557.

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

(6) 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.

(7) English, J. G., Olsen, R. H. J., Lansu, K., Patel, M., White, K., Cockrell, A. S., Singh, D., Strachan, R. T., Wacker, D., and Roth, B. L. (2019) VEGAS as a Platform for Facile Directed Evolution in Mammalian Cells. *Cell 178*, 748-761.e17.

(8) Badran, A. H., Guzov, V. M., Huai, Q., Kemp, M. M., Vishwanath, P., Kain, W., Nance, A. M., Evdokimov, A., Moshiri, F., Turner, K. H., Wang, P., Malvar, T., and Liu, D. R. (2016) Continuous evolution of B. thuringiensis toxins overcomes insect resistance. *Nature 533*, 58–63.

(9) Miller, S. M., Wang, T., Randolph, P. B., Arbab, M., Shen, M. W., Huang, T. P., Matuszek, Z., Newby, G. A., Rees, H. A., and Liu, D. R. (2020) Continuous evolution of SpCas9 variants compatible with non-G PAMs. *Nature Biotechnology 38*, 471–481.

(10) Stynen, B., Tournu, H., Tavernier, J., and Dijck, P. V. (2012) Diversity in Genetic In Vivo Methods for Protein-Protein Interaction Studies: from the Yeast Two-Hybrid System to the Mammalian Split-Luciferase System. *Microbiol. Mol. Biol. Rev. 76*, 331–382.

(11) Serebriiskii, I. G., and Golemis, E. A. (2001) Two-Hybrid System and False Positives, in *Two-Hybrid Systems: Methods and Protocols* (MacDonald, P. N., Ed.), pp 123–134. Humana Press, Totowa, NJ.

(12) Pelletier, J. N., Campbell-Valois, F.-X., and Michnick, S. W. (1998) Oligomerization domain-directed reassembly of active dihydrofolate reductase from rationally designed fragments. *Proc Natl Acad Sci U S A 95*, 12141–12146.

(13) Gibson DG (2011) Enzymatic assembly of overlapping DNA fragments. Meth. Enzymol 498, 349–361.

(14) Javanpour, A. A., and Liu, C. C. (2019) Genetic Compatibility and Extensibility of Orthogonal Replication. *ACS Synth Biol 8*, 1249–1256.

(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– 34, DOI: 10.1038/nprot.2007.13

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

(17) Huang, T., Barclay, B. J., Kalman, T. I., von Borstel, R. C., and Hastings, P. J. (1992) The phenotype of a dihydrofolate reductase mutant of Saccharomyces cerevisiae. *Gene 121*, 167–171.

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

(19) Stevens, A. J., Brown, Z. Z., Shah, N. H., Sekar, G., Cowburn, D., and Muir, T. W. (2016) Design of a Split Intein with Exceptional Protein Splicing Activity. *J. Am. Chem. Soc. 138,* 2162–2165.

Author Contributions:

G.A.A. and C.C.L. conceived of the project and experimental strategy. G.A.A. designed and performed all experiments. G.A.A. and J.S. designed and performed split-intein experiments. G.A.A. and C.C.L. wrote the manuscript. C.C.L. supervised the research.

G.A.A. – Garri A Arzumanyan

J.S. – Justin Sadownick

C.C.L – Chang C. Liu

CHAPTER 4
Mutually Orthogonal DNA replication systems *in vivo*

4.1 Introduction

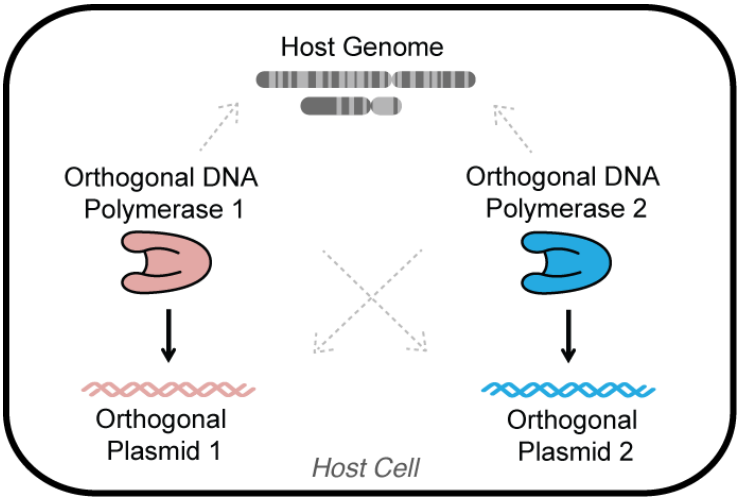
Expanding the capabilities of DNA polymerases (DNAPs) to change how they copy DNA *in vitro* has resulted in staple biotechnologies including PCR mutagenesis, DNA sequencing, nucleic acid diagnostics, and the evolution of synthetic genetic polymers.[1-5] Engineering DNAPs with new properties *in vivo* could lead to similarly important advances, including cells that continuously evolve target genes, record lineage or exposure to custom stimuli through mutating barcodes, or use new genetic alphabets for biocontainment.[6-10] However, the operation of highly engineered DNAPs in a living cell is challenging because they can easily harm genomic replication.[11] We recently developed a special DNA replication system in *Saccharomyces cerevisiae* consisting of a DNAP/plasmid pair that is orthogonal to genomic replication and showed that engineered orthogonal DNAPs only replicate the orthogonal plasmid.[6,7] Here, we demonstrate that an additional DNAP/plasmid pair is orthogonal to genomic replication in *S. cerevisiae* and that the two DNAP/plasmid pairs are mutually orthogonal to each other. This solidifies two platforms for independently expanding the properties of DNA replication *in vivo*.

Our two orthogonal replication systems are based on the cytoplasmically-localized pGKL1/pGKL2 (p1/p2) plasmids originating from *Kluyveromyces lactis*.[11,12] Both p1 and p2 encode their own DNAPs, TP-DNAP1 and TP-DNAP2, respectively. We previously showed that engineered error-prone variants of TP-DNAP1 increase the mutation rate of p1 to ~10^{-5} substitutions per base (s.p.b.) without affecting the genomic mutation rate of ~10^{-10} s.p.b. in *S. cerevisiae*.[6] We also found that p1 replication strictly requires TP-DNAP1.[7] This allowed us to

conclude that TP-DNAP1 and p1 constitute an orthogonal DNAP/plasmid pair such that engineered changes to TP-DNAP1 only act on p1 but not on the genome. However, it was not known whether orthogonality to genomic replication holds true for TP-DNAP2 and p2 nor whether the TP-DNAP1/p1 and TP-DNAP2/p2 pairs are mutually orthogonal. Here, we develop error-prone TP-DNAP2s, report the associated genetic techniques necessary to engineer the p2 plasmid and TP-DNAP2, and ultimately prove that the TP-DNAP2/p2 pair is both orthogonal to genomic replication and to p1 replication.

The existence of mutually orthogonal genetically tractable replication systems is significant for three main reasons. First, our finding of two mutually orthogonal DNA replication systems should lay the foundation for novel applications in synthetic biology. For example, *in vivo* accelerated evolution[6] of different genes or sets of genes can now be carried out at two distinct custom mutation rates, which could be useful for evolving components in hierarchically organized signaling pathways. Another possibility includes using inducible error-prone orthogonal DNAPs to record multiple cellular events or external stimuli, where the number of mutational events in p1 and p2 would be independent readouts of the amount of exposure to two signals experienced by cells. In addition, the freedom to engineer two orthogonal DNAPs *in vivo* may enable propagation of different XNA's in living cells, whereas current efforts are limited to either using novel base-pairs recognized by host DNAPs or engineering DNAPs to synthesize XNA with novel backbones *in vitro*.[4,5,9] Second is a practical consideration. Our main motivation for developing an orthogonal replication (OrthoRep) system was to achieve continuous rapid evolution of target genes *in vivo* at extreme mutation rates that the genome cannot withstand.[6] This was achieved by making TP-DNAP1 highly error-prone so that it rapidly mutates the p1 plasmid but spares the genome. The mutual orthogonality demonstrated here ensures that the essential accessory genes encoded on p2

are also spared by error-prone TP-DNAP1s during directed evolution experiments of genes on p1 in OrthoRep. Third, our result provides *in vivo* evidence that DNA initiation of p1 and p2 use independent components. p1 and p2 both contain terminal proteins (TPs) linked to their 5' termini, which act as origins of replication, akin to other protein-primed DNA replication systems like those found in bacteriophage Φ29 and adenovirus.[14,15] The lack of homology between the TPs of p1 and p2 suggested that TP-DNAP1 and TP-DNAP2 may use distinct molecular interactions for plasmid initiation. In addition, preliminary *in vitro* biochemical data from our lab shows that TP-DNAP1 can initiate replication from p1's inverted terminal repeat (ITR), hypothesized to act in concert with p1's TP to form an original of replication, but cannot initiate replication from p2's ITR (unpublished). Our observation of mutual orthogonality between p1 and p2 replication all but proves that highly specific TP-DNAP interactions with cognate TPs and ITRs govern plasmid initiation, encouraging future studies on the mechanisms of protein-primed DNA replication and suggesting a potential approach for engineering additional orthogonal replication systems that operate concurrently in the same cell.



4.2 Materials and methods

DNA cloning

Plasmids cloned and used in this study are listed in Table S4. All oligonucleotide primers and synthesized gene fragments (gBlocks®) were purchased from IDT. Enzymes for PCR and cloning were obtained from NEB. All plasmids, including TP-DNAP2 libraries were cloned using Gibson assembly with overlap regions of 20-30 bp.[22] Vectors harboring homologous recombination cassettes for p2 integrations were cloned as previously described for p1 integration cassettes.[6,7]

To clone pGA55, three gene fragments constituting recoded *TP-DNAP2* were assembled with the vector backbone of pAR318, a yeast shuttle vector containing *CEN6/ARS4* and *HIS3* for propagation in yeast, and *ColE1* and *KanR* for propagation in *E. coli.* The resulting vector was used for TP-DNAP2 complementation and generation of TP-DNAP2 libraries (see "TP-DNAP2 library generation" section).

Yeast strains

All yeast strains used in this work are listed in Table S5. *S. cerevisiae* strain AR-Y292 served as the parent for all strains used in this study and contains the wild type pGKL1 and pGKL2 (or p1 and p2) linear plasmids. GA-Y021 and GA-Y069 were created from AR-Y292 by p2 integration methods described below. AR-Y436 is a derivative of AR-Y292 encoding a functional copy of *URA3* at the endogenous genomic locus, for 5-FOA-based fluctuation tests of genomic mutation rates in presence of mutagenic TP-DNAP2 variants.

Strains for testing mutual orthogonality were generated by transforming a panel of *CEN6/ARS4* vectors encoding *TP-DNAP1* or *TP-DNAP2* variants into two base strains, AR-Y304 and GA-Y021. AR-Y304 (described previously[7]) contains recombinant p1 encoding *mKate2*, *URA3* and *leu2** without disturbing the native TP-DNAP1 ORF. Similarly, GA-Y021 encodes recombinant

p2 that replaces *ORF1* with *mKate2*, *URA3* and *leu2** without disturbing the native TP-DNAP2 ORF.

Yeast transformation

All yeast transformations were performed using the high-efficiency LiAc/SS carrier DNA/PEG method.[23] For integrations into p2, 2-5 μg of plasmid containing the appropriate integration cassette was digested with ScaI, yielding a linearized cassette with blunt ends containing the genes of interest flanked by regions of homology to p2. The products of the digestion reaction were directly transformed into appropriate AR-Y292-derived strains harboring wild type p1 and p2, and plated on selective solid SC medium. Colonies appeared after 4-5 d of growth at 30 °C. *CEN6/ARS4* plasmids were also transformed with the LiAc/SS carrier DNA/PEG method, but with only 500-3000 ng of DNA for individual vectors and with at least 10 μg of plasmid DNA for TP-DNAP2 library transformations, to maintain 6-fold coverage.

Linear plasmid DNA extraction

p1, p2, and all derived linear plasmids were extracted using a modified version of the yeast DNA extraction protocol detailed by Amberg et al..[24] The modifications were as follows: (i) cells spun down from 40 mL of saturated culture were washed in 0.9% NaCl before treatment with Zymolyase (US Biological) to break up flocculated cells; (ii) 200 μg/mL proteinase K (Sigma) was supplemented during SDS treatment for degradation of TP; (iii) Rotation at ~10 r.p.m. was used during Zymolyase and proteinase K treatments. This large-scale extraction protocol was used for preparing DNA for absolute quantification by qPCR. For qualitative analysis by agarose gel electrophoresis, this extraction protocol was scaled down to extract DNA from only 1.5 mL of saturated yeast culture.

Curing of cytoplasmic plasmids using Cas9

To achieve active curing of the cytoplasmic p2 plasmid, the yeast Cas9 genomic modification system developed by Cate and coworkers was repurposed for cytoplasmic targeting.[25] The SV-40 nuclear localization signal and 8x HIS tag were removed from the pCAS plasmid (Addgene plasmid # 60847) to localize Cas9 to the cytoplasm where p2 (and p1) plasmids propagate. Appropriate 20 nt spacers were cloned into this vector to target different sites in p2 (Table S1). These modified pCAS vectors were transformed into the strains harboring p2 plasmids to be cut, and plated on solid selective SC medium containing 1 g/L monosodium glutamate (MSG) as the nitrogen source and supplemented with G418 (400 μg/mL). Colonies that appeared after incubation at 30 °C for 2 d were inoculated into liquid selective SC medium with 1 g/L MSG and G418 (200 μg/mL) and passaged once at a 1:1000 dilution to cure the targeted p2 plasmid. The resulting cultures were then subjected to DNA extraction and analysis by gel electrophoresis to verify loss of the targeted p2 plasmid. To minimize potential toxicity due to Cas9 expression, final strains lacking the pCAS vector were isolated by passaging without G418 selection, and replica plating clones on solid medium with and without G418 to screen for loss of the pCAS vector.

TP-DNAP2 library generation

TP-DNAP1 and TP-DNAP2 peptide sequences were aligned using protein BLAST and four candidates residues for library generation were chosen by two criteria.[26] First, candidate TP-DNAP2 residues must match a residue in TP-DNAP1 known to affect fidelity, based on prior studies.[6,7] Second, at least 25% of the 20 neighboring residues must align. This analysis yielded positions S370, Y424, L474 and F882.

To clone the expression vector for wild type TP-DNAP2 (pGA55), *TP-DNAP2* was codon optimized for expression in *S. cerevisiae* with GenScript's OptimumGene™ tool, and the recoded ORF was synthesized as three gene fragments, which were assembled downstream of the *REV1* promoter in a *CEN6/ARS4* vector containing selection markers *HIS3* and *KanR*. The four TP-DNAP2 NNK libraries were cloned from pGA55 *via* Gibson assembly. Gibson fragments containing the NNK codon at each library site were prepared with two sequential PCR's, in order to limit bias in the NNK incorporation that may result from annealing between the degenerate codon and the plasmid template. The first PCR generated a linear fragment from pGA55 with a 5' end that terminates immediately 3' of the library codons, and a 3' end in *KanR*. These linear amplicons were purified, diluted to 40 ng/μL, and re-amplified in a second PCR reaction with a forward primer containing Gibson overlap regions and NNK overhangs at the corresponding library site in TP-DNAP2, and the same reverse primer used in the initial PCR. These PCR products were then purified and treated with DpnI for 6 h at 37 ^{o}C to digest any pGA55 plasmid carry-through. The second Gibson fragment was PCR amplified from pGA55 to include the vector backbone starting 5' in *KanR* and 3' leading up to, but not including the library codon. For each library, 100 ng of corresponding PCR amplicons were combined in a 20 μL Gibson assembly reaction, and incubated at 50 ^{o}C for 1 h. The assemblies were purified and concentrated in 12 μL of ddH_2O. 5 μL of the purified assembly products were then transformed into electrocompetent Top10 cells and recovered at 37 ^{o}C for 1 h. Each transformation was plated at 1x, 10x and 100x dilutions on solid LB medium supplemented with kanamycin (50 μg/mL). After overnight incubation at 37 ^{o}C, colony counts on the 10x an 100x plates were used to calculate the transformation efficiency. All transformations yielded more than 3200 transformants, corresponding to >100-fold coverage of each library in *E. coli* (each NNK library has a theoretical

diversity of 32). The transformants from the 1x plates were then harvested by resuspension in 5 mL of sterile ddH_2O. Library plasmid DNA was extracted from *E. coli* using the Zyppy™ Plasmid Miniprep Kit. Library quality was determined by verifying plasmid library sizes by agarose gel and Sanger sequencing of bulk library populations as well as 8 individual clones from each library.

p2 fluctuation tests using *leu2

For fluctuation tests used to measure p2 mutation rates, *leu2** served as a marker for detecting mutational events. *leu2** is a disabled version of *LEU2*, where Q180 is replaced with a TAA stop codon. Q180 is a permissive site where mutation to any other codon other than TAG and TGA results in functional reversion to *LEU2*. These mutational events can be detected by plating scores of parallel cultures on medium lacking leucine and counting the number distribution of functional *LEU2* mutants.

For TP-DNAP2 library screening, each library member was subjected to small-scale *leu2** fluctuation tests with six replicates. 190 library members from each yeast library transformation were arrayed and inoculated into liquid SC medium lacking uracil and histidine, and passaged three times at 1:10,000 dilutions. *mKate2* fluorescence was measured at every passage on a microplate reader (TECAN Infinite® 200 PRO, settings: λ_{ex} = 588, λ_{em} = 633) to track p2 copy number stabilization. To perform fluctuation tests, each library member was diluted 1:10,000 into liquid SC medium buffered to pH 5.8 and lacking uracil, histidine, tryptophan, and dilutions were split into six 100 µL replicates. Cultures were grown for 48 h at 30 ºC to reach saturation. Saturated cultures were washed with 200 µL 0.9% NaCl to remove residual leucine and resuspended in 35 µL of 0.9% NaCl. 10 µL of this resuspension was spot plated onto solid SC medium buffered to pH 5.8 and lacking uracil, histidine, tryptophan and leucine. Spot plates

were allowed to dry and incubated at 30 ºC. Colonies were counted after 5 d. Colony counts were used to calculate the expected number of *LEU2* functional mutants (*m*), using the p0 method.[17]

To precisely measure p1 and p2 per-base substitution mutation rates, several modifications were made to the small-scale protocol. First, 36 replicates were performed for reconfirmation of candidate TP-DNAP2 mutators, and 48 replicates for genomic and mutual orthogonality experiments. Titers were also determined for each strain after spot plating by pooling the residual volume from 4 replicates and plating dilutions on YPD. The expected number of *LEU2* functional mutants (*m*) was determined by the Ma-Sandri-Sarkar maximum likelihood estimator (as calculated by the FALCOR tool and corrected for partial plating.[17-19] The mean *mKate2* fluorescence was determined from 50,000 event counts on Attune™ NxT Flow Cytometer and converted to p2 copy number by using a calibration curve (see below). To determine per-base substitution rates, the corrected *m* was normalized to the average cell titer, the p2 copy number, and the target size for functional *leu2** reversion (2.33 bp). 95% confidence intervals were similarly scaled.

Genomic fluctuation tests using *URA3*

Fluctuation tests using genomically encoded *URA3* were performed in the presence of TP-DNAP2 variants to determine the genomic per-base substitution rates, similarly to previously described protocols.[20] AR-Y292 derived strains harboring the appropriate TP-DNAP2 variant encoded on a *CEN6/ARS4* vector (with a *HIS3* marker) were grown in liquid SC medium lacking uracil and histidine until saturation. Each strain was diluted 1:5,000 into SC medium lacking histidine and aliquoted into 48 replicates of 200 μL each. Cultures were grown for 48 hours at 30 ºC to reach saturation. Saturated cultures were washed with 400 μL 0.9% NaCl and resuspended in 420 μL

0.9% NaCl. 400 μL from each replicate was spot plated on pre-dried solid SC medium lacking histidine and supplemented with 5-FOA (1 g/L). The residual 20 μL from six replicates were pooled, diluted, and plated on solid YPD medium to determine cell titers. Plates were allowed to dry before incubation at 30 ^{o}C. Colonies were counted on titer plates after 2 days, and on spot plates after 5 days of growth. The expected number of mutants (*m*) was calculated using the MSS maximum likelihood estimator method *via* the FALCOR tool, and corrected for partial plating, as described above. To determine per-base substitution rates, the corrected *m* was normalized to the average cell titer, the *URA3* copy number (1 in haploid yeast), and the target size for 5-FOA resistance *via* substitutions in *URA3* (104 bp). 95% confidence intervals were similarly scaled.

Calibration curve relating p2 copy number to *mKate2* fluorescence

A standard curve relating p2 copy number to *mKate2* fluorescence was prepared by combining quantitative PCR with flow cytometry. During the 1:10,000 back dilution step of the *leu2** fluctuation tests for the mutual orthogonality experiment, six strains with mKate2 encoded on p2 were diluted into liquid SC medium buffered to pH 5.8 and lacking uracil, histidine and tryptophan to yield 50 mL of saturated culture. After 48 h of growth at 30 ^{o}C, a small portion of each culture was diluted 1:100 in 0.9% NaCl and analyzed on a flow cytometer (Attune™ NxT Flow Cytometer, settings: λ_{ex} = 561, λ_{em} = 620; gain = 550) to determine the mean red fluorescence from 50,000 counts. Genomic DNA and linear plasmids were extracted from the remaining 40 mL of each culture using the large-scale DNA extraction protocol detailed in 'Linear plasmid DNA extraction' to ensure complete and unbiased extraction of linear plasmids relative to genomic DNA. All extracts were diluted 4000-fold for use in two distinct qPCR reactions, one to quantify p2-encoded *leu2** and the other to quantify the genomic copy of LEU3 (see Table S3 for primers). Each 20 uL qPCR reaction consisted of 5 uL of template DNA, 2 uL

forward primer (5 uM), 2 uL reverse primer (5 uM), 1 uL ddH_2O, and 10 uL of Thermo Scientific™ Maxima SYBR Green/Fluorescein qPCR Master Mix (2X).

A standard curve for each primer set was prepared by performing qPCR on a dilution series of DNA extracted from F102-2 (25x, 125x, 625x, 3125x). Non-template controls with only ddH_2O were included for each primer set to detect contamination. All qPCR's were performed in triplicate on the Roche *LightCycler® 480* System using the following protocol:

qPCR:

1) 95°C for 10 min
2) 40x:
   95 °C for 15 s
   60 °C for 1 min
   measurement
3) 95 °C for 1 min
4) 55 °C for 1 min

Primer melting curve:

Ramp up to 95 °C at 0.11°C/s, with 5 measurements per °C.

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 plotted against log ([DNA]). The slope and y-intercept were calculated using linear regression (Table S3). Each sample's average C_t values were converted into copy number values by using the following equation: copy number = $10^{((\text{sample Ct} - y\text{intercept})/\text{slope})}$. The calculated *leu2** copy number was divided by the LEU3 copy number to normalize to genomes extracted and account for variance in DNA extraction efficiency across samples. For detailed information on qPCR data analysis, see Figure S5.

4.3 Results and Discussion

Strategy for characterizing orthogonality

Our strategy for probing orthogonality of the TP-DNAP2/p2 DNAP/plasmid pair is based on engineering and using error-prone TP-DNAP1 and TP-DNAP2 variants to measure whether they increase mutation rates of genes on p1, p2, and/or the host genome. If error-prone TP-DNAP1s only increase the mutation rate of p1 (but not of p2 and the host genome) and error-prone TP-DNAP2s only increase the mutation rate of p2 (but not of p1 and the host genome), then we may conclude that TP-DNAP1/p1 and TP-DNAP2/p2 are mutually orthogonal DNA replication systems that are both orthogonal to genomic replication. This is indeed what we found.

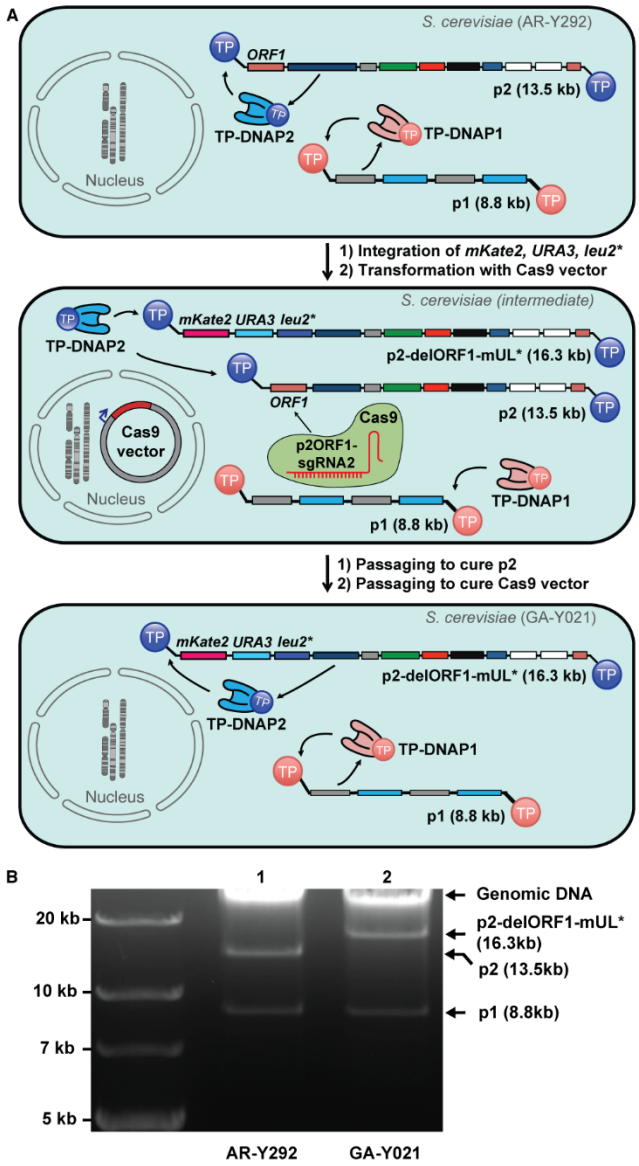


Figure 1. Engineering recombinant p2. (A) Top: AR-Y292 strain harboring p1 and p2, replicated by self-encoded TP-DNAP1 and TP-DNAP2, respectively. Middle: Intermediate strain derived from AR-Y292 by integration of *mKate2*, *URA3*, and *leu2** over *ORF1* of p2 and expression of Cas9 targeting *ORF1* of p2, which is missing from the recombinant p2-delORF1-mUL*. Bottom: Complete curing of the parental p2 plasmid induced by Cas9 results in GA-Y021, which contains only p1 and recombinant p2-delORF1-mUL*. GA-Y021 is used as the base strain for subsequent measurements of p2 mutation rate. **(B)** Gel analysis (0.9% agarose) verifying integration of *mKate2*, *URA3*, *leu2** and loss of parental p2 after Cas9 treatment. Lane 1: DNA extracted from AR-Y292, containing linear plasmids p1 and p2. Lane 2: DNA extracted from GA-Y021, confirming loss of p2 and its replacement by p2-delORF1-mUL* at high copy.

Integration of heterologous genes onto p2 using CRISPR/Cas9

We first developed a reliable method for encoding and expressing user-defined genes on p2. To measure mutation rates of p2 replication, we needed to encode *mKate2*, *URA3*, and *leu2** (mUL*) on p2. *mKate2* would serve as a fluorescent reporter for copy number, *URA3* as a selection marker, and *leu2**, which contains a stop codon at a permissive site in *LEU2* (Q180*), would serve as a reporter for substitution mutation rates in fluctuation tests that measure reversion to functional *LEU2*. We expected that mUL* could be integrated onto p2 via *in vivo* homologous recombination, following our procedures for manipulating p1.[6,7] We constructed a DNA cassette encoding mUL* flanked by regions homologous to p2 (Figure S1) such that successful recombination would result in the replacement of the non-essential ORF1 found on wildtype (wt) p2.[16] After transformation of this cassette into *S. cerevisiae* strain AR-Y292 containing wt p1 and p2 (Figure 1A, top), we isolated several clones exhibiting uracil prototrophy and detectable fluorescence from mKate2. Extraction of cytoplasmic plasmids from these clones confirmed presence of the recombinant p2-delORF1-mUL*, but only at low copy as confirmed by DNA gel electrophoresis and PCR with primers specific to p2-delORF1-mUL*. In contrast to similar cassette integrations into p1, passaging under selection for *URA3* expression failed to cure the parental wt p2 plasmid and increase the copy number of p2-delORF1-mUL* to levels easily detectable by gel electrophoresis. Although p2-delORF1-mUL* encodes all the necessary genes for its own replication and was selected for through *URA3*, it is likely that the shorter size of wt p2 provided it with enough of a replicative advantage to be maintained.

To cure the parental wt p2 plasmid, we attempted more active methods. We used a yeast CRISPR/Cas9 vector[25] and three candidate sgRNAs to target ORF1 (Table S1), which is present in wt p2 but not in p2-delORF1-mUL* (Figure 1A, middle). One of the three sgRNA's expressed

in conjunction with a cytoplasmically-localized Cas9 achieved complete curing of p2 and a concomitant increase in the copy number of p2-delORF1-mUL*. This was evidenced by an increase in mKate2 fluorescent signal and a brighter p2-delORF1-mUL* DNA gel electrophoresis band (Figure 1B). Curing of parental p2 to undetectable levels was confirmed by lack of PCR amplification with primers specific to p2, yielding strain GA-Y021 (Figure 1A, bottom).

Using GA-Y021, we measured the mutation rate of p2 replication by wt TP-DNAP2 still encoded on p2-delORF1-mUL*, with a previously described fluctuation test where the number distribution of functional *LEU2* mutants is used to calculate mutation rate by the MSS method.[17-19] The mutation rate of p2-delORF1-mUL* replication was 5.96×10^{-10} s.p.b. (95% C.I.: 3.57×10^{-10}-8.77×10^{-10}) with a copy number of 128 per cell (Table 1, Entry 1; Table S3). This is similar to the wild type p1 mutation rate and copy number, which are 1.39×10^{-9} and 124, respectively.[7]

Replication of p2 by TP-DNAP2 *in trans*

To facilitate the straightforward testing of error-prone TP-DNAP2s, we showed that p2 replication could be fully sustained by TP-DNAP2 encoded *in trans*, on a standard yeast nuclear plasmid, rather than *in cis*, on p2. First, TP-DNAP2, which is *ORF2* of p2, was deleted by homologous recombination of a synthetic cassette encoding *URA3* (Figure S1). Since p2 is a multi-copy plasmid, the resulting strain (GA-Y069) harbored a mixture of the parental wt p2 and the recombinant p2 with *ORF2* deleted (p2-delORF2-URA3), along with unaltered p1 (Figure 2A). In this strain, both the wt p2 and recombinant p2-delORF2-URA3 plasmids rely on TP-DNAP2 encoded on the parental wt p2 plasmid for replication. Thus, loss of wt p2 should disable replication of p2-delORF2-URA3. Indeed, when we cured the parental p2 plasmid by targeting *ORF2* of wt p2 with Cas9 (Table S1), we found that all p2-delORF2-URA3 was also lost and that the strain could no longer grow in the absence of uracil (Figure 2B). Next, we repeated the same

experiment but now in the presence of a codon-optimized TP-DNAP2 expressed *in trans* from a standard yeast *CEN6/ARS4* nuclear plasmid (pGA55-reTP-DNAP2). After p2 was fully cured by Cas9, we found that p2 was not present, p2-delORF2-URA3 remained, and this strain could grow in the absence of uracil (Figure 2B). In addition, p1 was also maintained, indicating that the accessory genes encoded on p2-delORF2-URA3 necessary for replication of both p1 and p2 and transcription of TP-DNAP1 on p1 were still functional. Therefore, p2-derived plasmids can be replicated by TP-DNAP2 encoded on a standard nuclear plasmid, simplifying the characterization of p2 replication by error-prone TP-DNAP2 variants.

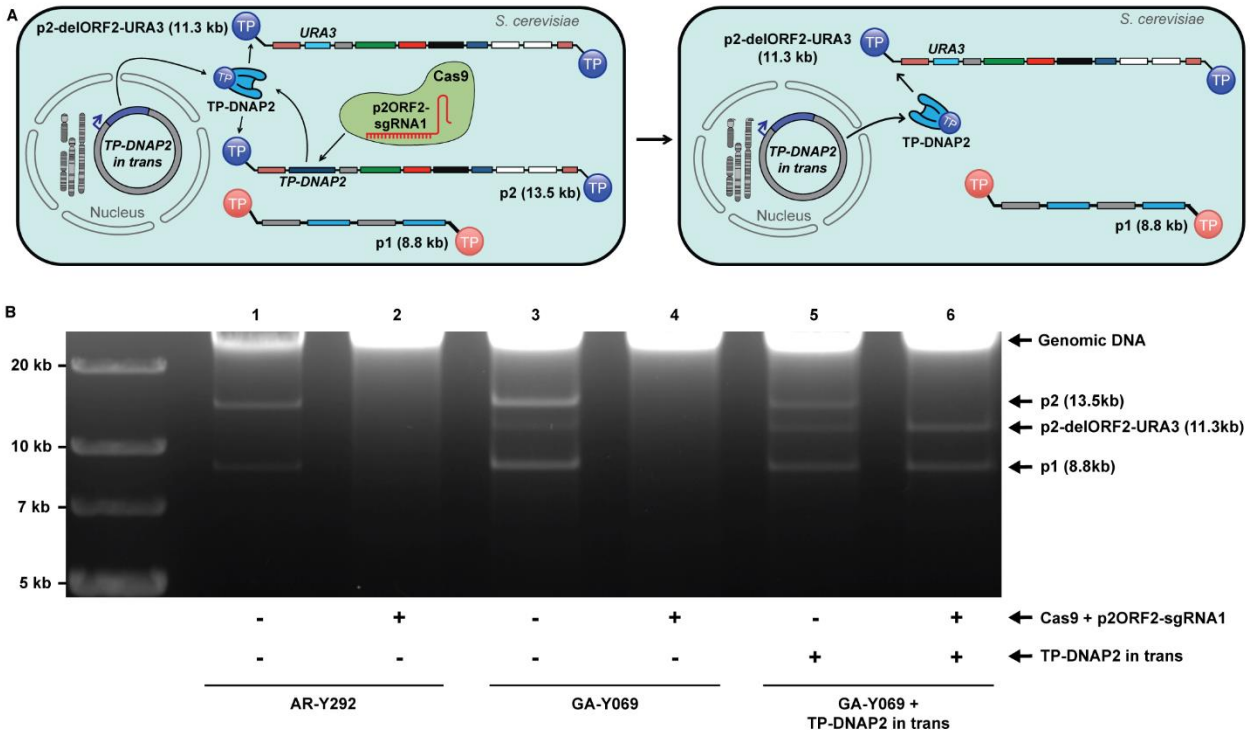


Figure 2. Functional complementation by recoded TP-DNAP2 *in trans*. (A) Left: Linear plasmid system consisting of p1, p2, and recombinant p2-delORF2-URA3. Cas9 co-expressed with p2ORF2-sgRNA1 specifically targets double strand breaks to *ORF2* of p2. Right: Complete curing of p2 by Cas9 results in loss of p2's self-encoded TP-DNAP2. This deletion is functionally complemented by TP-DNAP2 encoded *in trans*, which can stably replicate p2. **(B)** Gel analysis (0.9% agarose) testing for stability of linear plasmid combinations in the absence (odd-numbered lanes) or the presence (even-numbered lanes) of Cas9 activity targeted to *ORF2* of p2. Lanes 1 and 2: DNA extracted from AR-Y292, containing only wt p1 and p2, which are both lost in presence of Cas9. Lanes 3 and 4: DNA extracted from GA-Y069 containing p1, p2, and recombinant p2-delORF-URA3. Elimination of wt p2 by Cas9 curing eliminates the sole source of TP-DNAP2, causing subsequent loss of p2-delORF1-URA3 and p1. Following loss of the linear plasmids, the strain in lane 4 was no longer prototrophic for uracil. Lanes 5 and 6: DNA extracted from a GA-Y069 derivative, which contains the linear plasmids p1, p2 and p2-delORF2-URA3 and encodes TP-DNAP2 on a nuclear *CEN6/ARS4* plasmid *in trans*. Treatment with Cas9 results in a stable linear plasmid system lacking wt p2, instead consisting only of p1 and p2-delORF2-URA3 replicated entirely by TP-DNAP2 encoded *in trans*.

Discovery of error-prone TP-DNAP2 variants through library screening

To discover error-prone TP-DNAP2s, we screened a small library of TP-DNAP2s diversified at locations we hypothesized would be responsible for DNAP fidelity. An alignment between TP-DNAP1 and TP-DNAP2 revealed that S370, Y424, L474, and F882 in TP-DNAP2 were homologous to residues in TP-DNAP1 that we previously found could be mutated to yield error-prone TP-DNAP1s (Figure S2).[6,7] We therefore generated four distinct site-saturation mutagenesis libraries, each diversifying S370, Y424, L474, or F882 in TP-DNAP2 encoded on pGA55-reTP-DNAP2. Due to unsuccessful attempts to generate a strain with a full deletion of p2-encoded TP-DNAP2 and simultaneous integration of *mKate, URA3,* and *leu2**, each library was transformed into GA-Y021 for convenient screening. However, since GA-Y021 still encodes wt TP-DNAP2 on p2-delORF1-mUL*, p2 mutation rates measured in this format are the result of in tandem replication of p2 by wt TP-DNAP2 and each TP-DNAP2 variant encoded *in trans*. We sampled each NNK library (theoretical diversity = 32 codon variants) at 6-fold coverage by picking 190 yeast colonies and passaging under selection for *URA3* to stabilize the copy number of p2-delORF1-mUL* in the presence of the newly introduced TP-DNAP2 variants. To screen for TP-DNAP2 variants with increased p2 mutation rate, each library member was subjected to a preliminary, small scale *leu2** fluctuation test with six replicates. Seventeen candidate mutators with the highest expected number of mutants *m* calculated by the p0 method were then chosen for reconfirmation. Reconfirmation consisted of extracting the *CEN6/ARS4* plasmids encoding TP-DNAP2 variants, retransforming into a fresh GA-Y021 background and repeating *leu2** fluctuation tests with 36 replicates to determine p2 mutation rate with higher precision (Table 1). Two error-prone TP-DNAP2 variants, S370Q and Y424Q, increased p2's substitution mutation rate by ~14- and ~13-fold, respectively (Table 1, Entries 2 and 3). These variants were not active enough to fully complement a deletion of the native TP-DNAP2 and sustain p2-delORF2--URA3

replication on their own, making our measured mutation rates an underestimate of their true per-base substitution rate. We suspect that even slight decreases in p2 copy number when replicated by mutant polymerases may cause instability in the system due to disruption of expression levels of the nine essential genes encoded on p2. Despite this, these two error-prone TP-DNAP2 variants elevate p2 mutation rate to high enough levels for measuring orthogonality.

Entry	TP-DNAP2 variant *in trans*	Per-base substitution rate of p2 ($x10^{-10}$)	Fold increase
1	None	5.96 (3.57-8.77)	-
2	S370Q	83.0 (62.8-105)	13.9
3	Y424Q	77.5 (56.9-100)	13.0
4	Y424E	73.3 (55.2-93.3)	12.3
5	S370P	50.1 (32.4-70.4)	8.4
6	Y424K	44.4 (32.2-58.0)	7.5
7	S370R	31.9 (22.4-42.5)	5.3
8	Y424G	28.8 (20.0-38.7)	4.8
9	S370E	27.9 (19.7-37.2)	4.7
10	S370K	26.2 (18.0-35.4)	4.4
11	Y424R	19.6 (13.0-27.1)	3.3
12	S370L	14.8 (9.32-21.1)	2.5
13	L474D	12.2 (7.33-17.9)	2.0
14	F882A	11.7 (4.92-20.4)	2.0
15	F882V	10.8 (6.30-16.2)	1.8
16	L474A	8.37 (4.80-12.6)	1.4
17	F882R	7.32 (3.91-11.5)	1.2
18	L474V	5.76 (2.88-9.33)	1.0

Table 1. TP-DNAP2 variants with elevated mutation rates. Entry 1 is the native mutation rate of p2 replication measured in GA-Y021, where p2-delORF1-mUL* is replicated only by the self-encoded wt TP-DNAP2. Entries 2-18 are mutation rate measurements of p2 replication in presence of 17 error-prone TP-DNAP2 hits from library screening. All per-base substitution rates were measured using *leu2** fluctuation analyses in GA-Y021, where p2-delORF1-mUL* is replicated in tandem by wt TP-DNAP2 and a TP-DNAP2 variant encoded *in trans* on a nuclear *CEN6/ARS4* plasmid. Ranges in parentheses correspond to 95% confidence intervals, determined according to the MSS method[17-19].

Orthogonality between p2 replication and genomic replication

To show that p2 replication is orthogonal to genomic replication, we measured the mutation rate of the host genome in the presence of error-prone TP-DNAP2 variants. Like p1 replication by TP-DNAP1, p2 replication by TP-DNAP2 occurs in the cytoplasm *via* a protein-primed mechanism, making it likely that p2 replication is orthogonal to host genome replication. To test this, we transformed *CEN6/ARS4* vectors lacking TP-DNAP2, or encoding codon optimized versions of TP-DNAP2^{WT}, TP-DNAP2^{S370Q}, or TP-DNAP2^{Y424Q} into AR-Y436 (Figure S3). AR-Y436 contains wt p1 and p2, as well as an intact *URA3* locus in the host genome. Genomic per-base substitution rates were determined via fluctuation tests based on the frequency of 5-FOA resistant clones arising from mutations in the genomic *URA3* locus, as previously described,[7,20] and are reported in Table 2. The substitution rich spectrum of *URA3*-disabling mutations makes this assay ideal for detecting whether the elevated substitution rate of TP-DNAP2^{S370Q} or TP-DNAP2^{Y424Q} contributes to genomic mutation. We observed no increase in the host genomic mutation rate when error-prone variants of TP-DNAP2^{S370Q} and TP-DNAP2^{Y424Q} were present.

Entry	TP-DNAP2 *in trans*	Per-base substitution rate at genomic *URA3* locus (x10^{-10})
1	None	1.6 (0.8-2.4)
2	WT	2.2 (1.2-3.4)
3	Y424Q	1.8 (1.0-2.9)
4	S370Q	1.8 (1.0-2.8)

Table 2. Genomic mutation rate is unaltered by error-prone TP-DNAP2 variants. Entries 1-4 are per-base substitution rates of genomic replication in presence of TP-DNAP2 variants, measured using the genomic URA3 locus. Mutagenic TP-DNAP2 variants Y424Q and S370Q cause no statistically significant change in genomic mutation rate, relative to no TP-DNAP2 variant *in trans* (84% CI overlap method*)*.[21] Ranges in parentheses correspond to 95% confidence intervals, determined according to the MSS method.[17-19]

Mutual orthogonality between p1 and p2 replication

Next, we wanted to test whether the replication mechanisms of p1 and p2 are mutually orthogonal *in vivo* by measuring p1 and p2 mutation rates in the presence of a panel of TP-DNAP1 and TP-DNAP2 variants with varying mutation rates. Changes in p1 or p2 mutation rate induced by TP-DNAP variants would therefore signal a degree of cross-replication between TP-DNAP1/p1 and TP-DNAP2/p2, if any. A panel of six polymerases was introduced on *CEN6/ARS4* vectors into two separate base strains, AR-Y304 (p1 mutation rate reporter strain) and GA-Y021 (p2 mutation rate reporter strain), encoding *mKate2*, *URA3*, and *leu2** on either p1 or p2, respectively (Figure 3A). Included in this panel were TP-DNAP1^{WT} and two error-prone TP-DNAP1 variants found in previous screens: TP-DNAP1$^{I777K,\ L900S}$ and TP-DNAP1$^{L477V,\ L640Y,\ I777K,\ W814N}$ (Figure S4). Also included were TP-DNAP2^{WT} and the error-prone TP-DNAP2^{S370Q} and TP-DNAP2^{Y424Q} variants found in this work. In AR-Y304 and GA-Y021, both p1 and p2 still encode their native wt TP-DNAP's. Any contribution to replication by a third TP-DNAP encoded *in trans* is monitored by detecting changes in linear plasmid mutation rate. Importantly, these experiments afford each DNAP an opportunity to replicate its native plasmid and lose its attached TP, becoming "spent" and perhaps more likely to replicate its noncognate linear plasmid through DNAP exchange.[13] The presence of TP-DNAP1^{WT} or TP-DNAP2^{WT} *in trans* had no effect on their noncognate linear plasmids' mutation rate (Figures 3B and 3C, Table S2). Mutagenic variants TP-DNAP1$^{I777K,\ L900S}$ and TP-DNAP1$^{L477V,\ L640Y,\ I777K,\ W814N}$ increased p1's mutation rate by 380- and 870-fold, respectively, but caused no statistically significant change in p2's mutation rate (Figure 3B, 84% CI overlap method).[21] Likewise, p1 mutation rate was unaltered in the presence of mutagenic TP-DNAP2^{S370Q} and TP-DNAP2^{Y424Q}, which increased p2 mutation rate by 29- and 16-fold, respectively (Figure 3C). Thus, TP-DNAP1 replicates p1 with at least 870-fold specificity over

p2, while TP-DNAP2 targets p2 with at least 29-fold specificity over p1. The level of mutual orthogonality measured here is limited by the error-rates of DNAPs used, especially that of TP-DNAP2 variants. Future discovery of more mutagenic TP-DNAP1 or TP-DNAP2 variants may prove an even greater orthogonality between these two replication systems, which we expect to be the case.

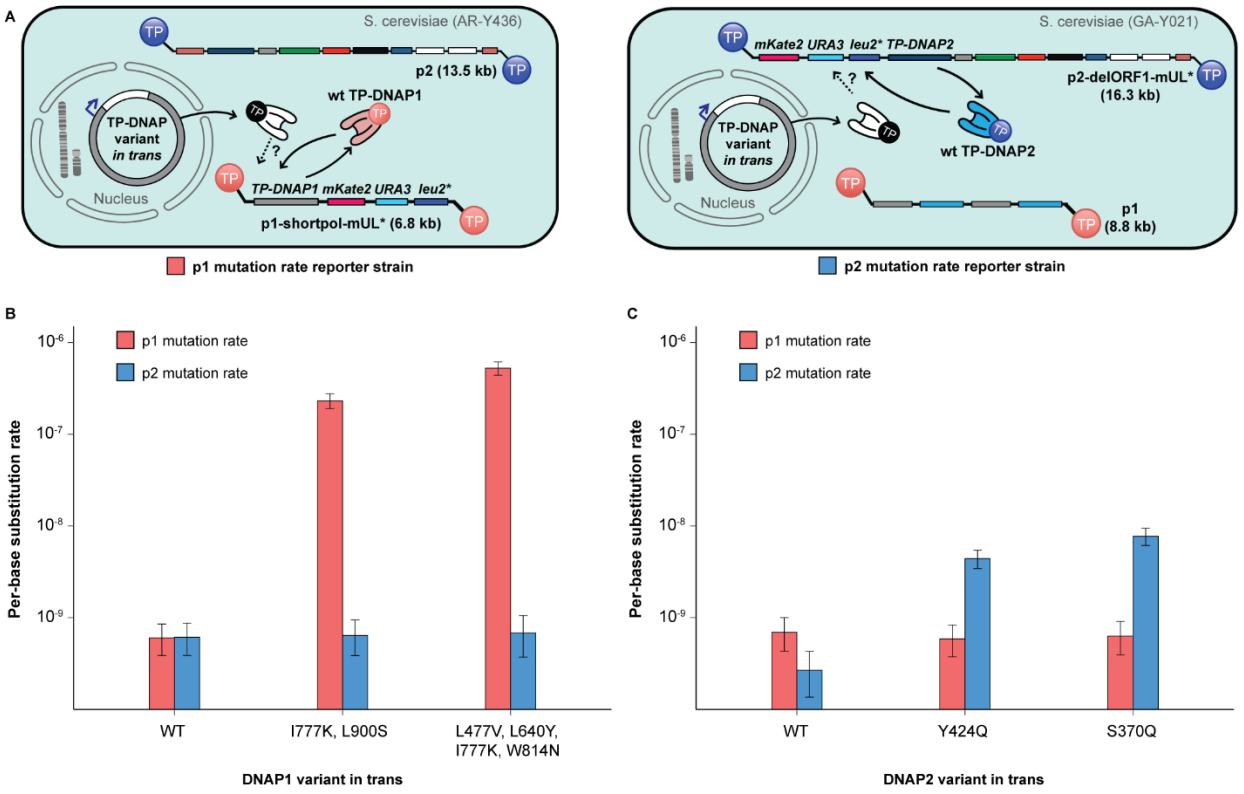


Figure 3. Mutual orthogonality between p1 and p2 replication. (A) Schematics detailing strains for measuring p1 mutation rate (left, red) and measuring p2 mutation rate (right, blue). Left: p1 mutation rate reporter strains derived from AR-Y304 encode *mKate2*, *URA3* and *leu2** on p1, enabling measurement of p1 copy number and mutation rate in presence of TP-DNAP variants *in trans*. Right: p2 mutation rate reporter strains derived from GA-Y021 encode *mKate2*, *URA3* and *leu2** on p2, enabling measurement of p2 copy number and mutation rate in presence of TP-DNAP variants *in trans*. **(B)** Mutagenic TP-DNAP1 variants increase p1 mutation rate (red) by 380- and 870-fold, without any increase in p2 mutation rate (blue). **(C)** Mutagenic TP-DNAP2 variants increase p2 mutation rate by 16- and 29-fold, without a similar increase in p1 mutation rate. Error bars correspond to 95% confidence intervals calculated according to the MSS method.[17-19] See Table S2 for exact mutation rate values.

Conclusion

In summary, p1 replication by TP-DNAP1 and p2 replication by TP-DNAP2 are both orthogonal to genomic replication and to each other, resulting in two mutually orthogonal DNA replication systems in the same cell. We envision that this pair of orthogonal replication systems will enable the *in vivo* evolution of multiple genes at different elevated mutation rates, molecular recording of biological signals in two distinct DNA channels, and the establishment of additional mutually orthogonal DNAP/plasmid pairs by engineering new TPs.

4.4 References

(1) Cadwell, R. C., and Joyce, G. F. (1992) Randomization of genes by PCR mutagenesis. *Genome Res. 2*, 28–33.

(2) Metzker, M. L. (2010) Sequencing technologies — the next generation. *Nature Reviews Genetics 11*, 31–46.

(3) Kranaster, R., and Marx, A. (2010) Engineered DNA Polymerases in Biotechnology. *ChemBioChem 11*, 2077–2084.

(4) Taylor, A. I., Pinheiro, V. B., Smola, M. J., Morgunov, A. S., Peak-Chew, S., Cozens, C., Weeks, K. M., Herdewijn, P., and Holliger, P. (2015) Catalysts from synthetic genetic polymers. *Nature 518*, 427–430.

(5) Pinheiro, V. B., Taylor, A. I., Cozens, C., Abramov, M., Renders, M., Zhang, S., Chaput, J. C., Wengel, J., Peak-Chew, S.-Y., McLaughlin, S. H., Herdewijn, P., and Holliger, P. (2012) Synthetic Genetic Polymers Capable of Heredity and Evolution. *Science 336*, 341–344.

(6) 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. *bioRxiv* 313338. DOI: 10.1101/313338

(7) Ravikumar, A., Arrieta, A., and Liu, C. C. (2014) An orthogonal DNA replication system in yeast. *Nature Chemical Biology 10*, 175–177.

(8) Perli, S. D., Cui, C. H., and Lu, T. K. (2016) Continuous genetic recording with self-targeting CRISPR-Cas in human cells. *Science 353*, DOI: 10.1126/science.aag0511.

(9) Malyshev, D. A., Dhami, K., Lavergne, T., Chen, T., Dai, N., Foster, J. M., Corrêa, I. R., and Romesberg, F. E. (2014) A Semi-Synthetic Organism with an Expanded Genetic Alphabet. *Nature 509*, 385–388.

(10) Ravikumar, A., and Liu, C. C. (2015) Biocontainment through Reengineered Genetic Codes. *ChemBioChem 16*, 1149–1151.

(11) Herr, A. J., Ogawa, M., Lawrence, N. A., Williams, L. N., Eggington, J. M., Singh, M., Smith, R. A., and Preston, B. D. (2011) Mutator Suppression and Escape from Replication Error–Induced Extinction in Yeast. *PLOS Genetics 7*, e1002282.

(12) Strak, M. J. R., Boyd, A., Mileham, A. J., and Ramonos, M. A. (1990) The plasmid-encoded killer system of Kluyveromyces lactis: A review. *Yeast 6*, 1–29.

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

(14) Rodríguez, I., Lázaro, J. M., Salas, M., and De Vega, M. (2004) phi29 DNA polymerase-terminal protein interaction. Involvement of residues specifically conserved among protein-primed DNA polymerases. *J. Mol. Biol. 337*, 829–841.

(15) Mysiak, M. E., Holthuizen, P. E., and van der Vliet, P. C. (2004) The adenovirus priming protein pTP contributes to the kinetics of initiation of DNA replication. *Nucleic Acids Res 32*, 3913–3920.

(16) Schaffrath, R., Stark, M. J. R., Gunge, N., and Meinhardt, F. (1992) *Kluyveromyces lactis* killer system: *ORF1* of pGKL2 has no function in immunity expression and is dispensable for killer plasmid replication and maintenance. *Curr Genet 21*, 357–363.

(17) Foster, P. L. (2006) Methods for Determining Spontaneous Mutation Rates, in *Methods in Enzymology*, pp 195–213. Academic Press.

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

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

(20) Lang, G. I., and Murray, A. W. (2008) Estimating the Per-Base-Pair Mutation Rate in the Yeast Saccharomyces cerevisiae. *Genetics 178*, 67–82.

(21) Zheng, Q. (2016) Comparing mutation rates under the Luria-Delbrück protocol. *Genetica 144*, 351–359.

(22) Gibson, D. G. (2011) Enzymatic assembly of overlapping DNA fragments. *Meth. Enzymol. 498*, 349–361.

(23) Gietz, R. D., and Schiestl, R. H. (2007) High-efficiency yeast transformation using the LiAc/SS carrier DNA/PEG method. *Nat Protoc 2*, 31–34.

(24) Amberg, D. C. (2005) Methods in yeast genetics: a Cold Spring Harbor Laboratory course manual / 2005 ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y.

(25) 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*.

(26) Altschul, S. F., Madden, T. L., Schäffer, A. A., Zhang, J., Zhang, Z., Miller, W., and Lipman, D. J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res. 25*, 3389–3402.

Author Contributions:

G.A.A., A.R., and C.C.L. conceived of the project and experimental strategy. G.A.A. designed and performed all experiments. G.A.A. and K.N.G. designed and performed TP-DNAP2 library screening experiments. A.A.J. developed the Cas9 method for curing linear plasmids. All authors analyzed data and results. G.A.A. and C.C.L. wrote the manuscript. C.C.L. supervised the research.

G.A.A. – Garri A Arzumanyan

A.R. – Arjun Ravikumar

K.N.G – Kristin N Gabriel

A.A.J. – Alex A. Javanpour

C.C.L – Chang C. Liu

CHAPTER 5
Conclusions and future directions

5.1 OrthoRep for studying drug resistance

In Chapter 2, we described a large scale evolution experiment of *P. falciparum* DHFR using OrthoRep, which led to insights into the adaptive trajectories and fitness landscape of pyrimethamine resistance. This experiment produced a rich data set despite relying on a relatively simple selection regime where pyrimethamine concentration is increased quickly to maintain strong selection. However, more nuanced selection conditions could be used to explore the effects of selection strength and population size on protein evolution. For example, a less steep increase in pyrimethamine concentration may shift preference away from the dominant S108N-based path towards rare adaptive trajectories that are outcompeted in regimes of strong selection. A similar shift in path preference may also be achievable by decreasing population size (< 1 uL cultures), where evolution becomes stochastic and clonal interference is minimal. This less competitive environment affords an opportunity for weak genotypes to sweep and lead to previously unobserved mutational trajectories.

The OrthoRep system for evolving PfDHFR also presents opportunities for studying the utility of multi-drug combinations in preventing drug resistance. After pyrimethamine resistance became widespread in wild *P. falciparum* populations, experimental drugs were designed to target the pyrimethamine-resistant active sites of PfDHFR mutants. One such experimental drug, WR99210, was designed specifically to bind with higher affinity to mutants with the common S108N mutation. Applying alternating or combined selection using pyrimethamine and WR99210 could potentially trap the enzyme into genotypes that cannot evolve high levels of resistance, with the

premise that resistance to one drug would result in sensitivity to the other. could be used to study the efficacy of multi-drug regimes in preventing high levels of resistance.

Due to its pivotal role in DNA precursor synthesis, DHFR has been an attractive drug target in pathogens other than *P. falciparum*. However, widespread use of DHFR inhibitors has resulted in drug resistance through mutations in DHFR orthologs other than PfDHFR. For example, pyrimethamine resistance has not only evolved in malarial parasites like *P. falciparum* and *P. vivax*, but also in *T. gondii*, a parasite that causes toxoplasmosis in immunocompromised patients. The prolonged nature of prophylactic administration of pyrimethamine to HIV patients meant that arisal of resistance was simply a matter of time. The ease with which our OrthoRep model system can be expanded to DHFR orthologs provides a great opportunity for studies into wiser design of drug schedules and combinations that could prevent such cases of resistance. Other cases of DHFR resistance to explore include trimpethoprim resistance in *S. aureus* and methotrexate resistance acquired by tumors through DHFR mutations.

5.2 High-throughput evolution of intracellular protein-protein interactions

The continuous evolution system for intracellular protein-protein interactions described in Chapter 3 lays the groundwork for high-throughput discovery of binding proteins. However, key challenges prevent more sophisticated applications. For one, the design, build, test cycle is hindered by poor growth of YH5 in synthetic media (5-7 days on plates, 3-4 days in liquid culture). This poor viability is likely due to inefficient import of dTMP by yeast grown in synthetic media, as suggested by significantly faster growth in YPD media supplemented with equal amounts of dTMP (2 days). We found that growth rate can be improved slightly by maintaining SC media in the 5.5-

6.0 pH range. However, more impactful measures will most likely rely on supplementing folate pathway metabolites are more readily imported than dTMP, such as folic or folinic acid. Previous data also suggests that the rate of dTMP degradation by intracellular phosphatases is affected by phosphate concentration in media, which can be tuned to improve dTMP utilization. Improvement of viability and growth rate will likely also increase transformation efficiency of DNA into YH5, opening avenues into applications requiring libraries. Aside from this technical issue, the current version of the system is missing a negative selection capability. With this feature, binding proteins can be evolved to distinguish between two target proteins whose differences are minute (i.e. several residues residues, post-translational modifications, etc.). A simple negative selection can be implemented by encoding the off-target protein without fusion to the mDHFRII fragment, so that non-specific binders, or "preys", are sequestered do not result in reconstituted DHFR activity.

After solving these challenges, this system can be applied to the isolation and continuous affinity maturation of novel binding proteins from a library. A diverse library encoded as "prey" on OrthoRep would provide many starting points for evolution, with the premise that any member with affinity for the target of interest can be affinity matured simply by continuous passaging of yeast cultures. Though the initial library integration step onto p1 is challenging due to low transformation efficiency, it need only be performed once, after which it can be reused to isolate binders for many different targets by transforming them as "baits" on a nuclear plasmid. This approach can be easily scaled to discover novel binders towards many targets in parallel. In preliminary attempts to demonstrate this capability, we encoded a nanobody library on OrthoRep and attempted selection of binding variants against GFP. To achieve 5-fold coverage of a ~200,000 member nanobody library, yeast transformation was scaled up 600x and further improved by using YPDA for outgrowth and YPD plus for recovery. Unfortunately, we encountered issues with

transforming bait plasmids into this prepared library (no transformants yielded), despite successes with earlier control transformations of bait plasmids to cover smaller p1 libraries. This remains a potentially lucrative application but clearly challenges with YH5 viability and transformation efficiency need to be overcome first.

Our scalable continuous evolution platform may also be well-suited to affinity mature *de novo* designed protein-protein interactions. Due to the hit and miss nature of *de novo* designs, the capability to evolve many designs in parallel will increase chances of achieving the desired binding interface. Such experiments can quickly become burdensome or even impossible to scale with traditional techniques like phage and yeast display techniques. In a recent paper, an existing rapamycin-dependent PPI was altered through computational design to detect novel small molecules, effectively turning the PPI into a biosensor for small-molecules. In combination with our OrthoRep selection, such a system can become a platform for evolving biosensors for many small-molecules in parallel with low effort.

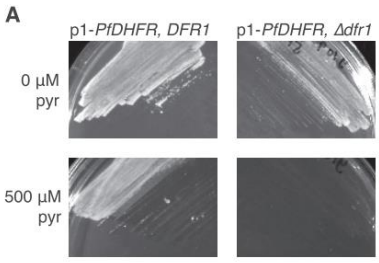
5.3 Future applications of mutually orthogonal DNA replication systems

In the future, the mutually orthogonal DNA replication systems of the linear plasmids that comprise OrthoRep may enable new applications in DNA recording. Exposure to environmental events like presence of toxins can be tied to expression of mutagenic TP-DNAP1 and TP-DNAP2 polymerases, which would create a permanent record of the exposure events through physical mutations in the p1 and p2 linear plasmids. The 10-fold higher mutation rate of p2 replication achieved in Chapter 4 is not sufficient to achieve efficient recording, however, and leaves room for future work to screen more mutagenic TP-DNAP2 variants. Ideally, the essential

genes of p2 must also be refactored and expressed genomically to prevent instability of the linear plasmid system due to high mutagenesis. Furthermore, the discovery of mutual orthogonality between p1 and p2 replication may spark future studies into the mechanism of protein-primed replication and the role of terminal proteins in determining specificity of replication.

Appendix A

Supplementary information for Chapter 2



B

Plasmid (DNAP)	Mutation rate (s.p.b.)	Average plasmid copy number	Number of cells plated on 500 μM pyr	Number of resistant colonies
CEN6/ARS4 (genomic DNAPs)	1.86×10^{-10}	1-5	$\sim 1 \times 10^{9}$	<50
p1 (N423E)	1.81×10^{-7}	15.05	8×10^{7}	~150
p1 (I777K, L900S)	2.18×10^{-6}	16.00	4×10^{6}	>500

C

		PfDHFR mutations	
TP-DNAP1	Mutation rate (s.p.b.)	1:25 passaging	1:250 passaging
WT	2.23×10^{-9}	S108N	(extinction)
N423E	1.81×10^{-7}	S108N	S108N
I777K, L900S	2.18×10^{-6}	prom -87 T->C, D54N, Y57H, C59R	D54N, Y57H, I163V, V168I, G264D

Figure S4. Pilot Studies of p1-*Pf*DHFR Evolution, Related to Figures 3, 4, and 5

(A) Yeast strains dependent on p1-encoded *PfDHFR* acquired sensitivity to pyrimethamine (pyr). *Pf*DHFR was expressed from p1 in strains that retain or lack *DFR1*, which encodes yeast's endogenous DHFR. Strains were grown in selective SC media and plated on solid media with or without 500 μM pyrimethamine. Plates were incubated at 30°C for 5 days prior to imaging.

(B) Pyrimethamine resistant clones arise in small culture volumes. A yeast strain that encodes *PfDHFR* on a nuclear plasmid and two OrthoRep strains that encode *PfDHFR* on p1 at for rapid mutation were grown to saturation in selective SC media and plated on solid media supplemented with pyrimethamine. After 5-6 days of growth at 30°C, resistant colonies were counted. p1-encoded *PfDHFR*s carried resistance mutations in all 30 resistant clones sequenced.

(C) OrthoRep strains evolved pyrimethamine resistance in batch culture by rapidly mutating p1-encoded *PfDHFR*. OrthoRep strains with varying p1 mutation rates were serially passaged in 25 mL cultures, at 1:25 or 1:250 dilutions, in selective SC media initially supplemented with 500 μM pyrimethamine. OD600 was monitored daily and saturated cultures were passaged into gradually increasing drug concentrations as cultures adapted. After strains evolved resistance to 2 mM pyrimethamine, bulk populations of p1 plasmids were extracted and subject to Sanger sequencing. The OrthoRep strain containing WT TP-DNAP1 stopped growing in the 500 μM pyrimethamine condition when passaged at 1:250 dilutions.

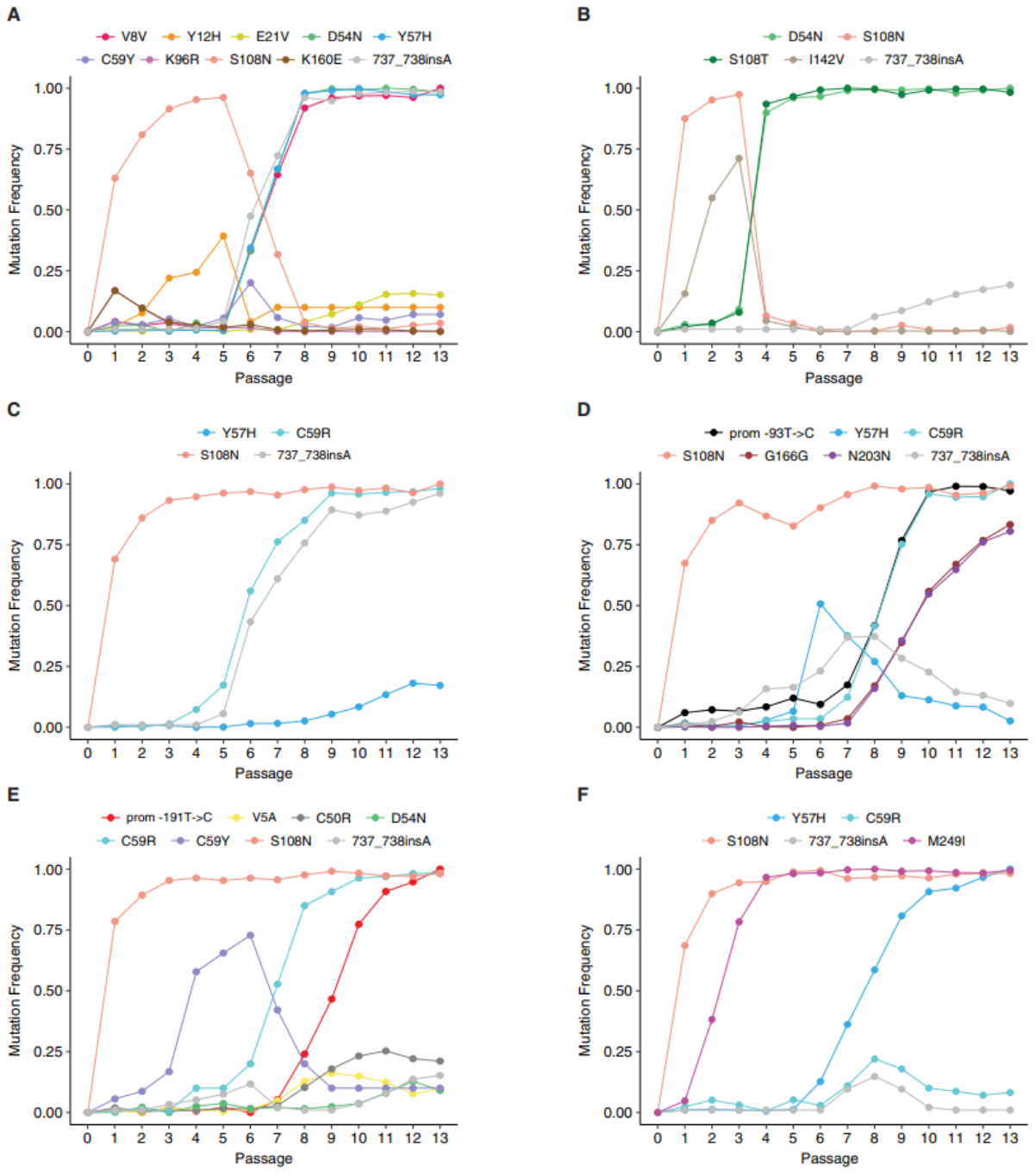


Figure S5. Dynamics of PfDHFR Evolution in Six Representative Populations, Related to Figures 3, 4, and 5

(A–F) Mutation frequencies tracked across all 13 passages from lines 52, 64, 74, 62, 70, and 73 shown in Figure 3, respectively. Populations from each passage were revived from glycerol stocks in the same media condition that they were initially grown in. Mutation frequencies were calculated from Sanger sequencing of revived populations. See STAR Methods for SNP analysis details.

Appendix B

Supplementary information for Chapter 3

Supplementary Table 1. List of plasmids used in this study. Plasmids used and cloned in this work are listed here.

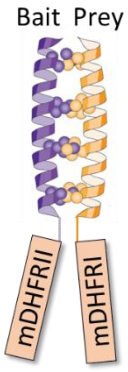
#	Name	Origin of replication (yeast, bacterial)	Selection marker (yeast, bacterial)	Notes
1	pGA-YTK001	CEN6/ARS4, ColE1	URA3/Kan	Expresses mDHFR under pREV1 promoter
2	pGA-YTK002	CEN6/ARS4, ColE1	URA3/Kan	Expresses mDHFR under pRNR2 promoter
3	pGA-YTK003	CEN6/ARS4, ColE1	URA3/Kan	Expresses mDHFR under pRpl18b promoter
4	pGA-YTK004	CEN6/ARS4, ColE1	URA3/Kan	Expresses mDHFR under pTDH3 promoter
5	pGA-YTK005	CEN6/ARS4, ColE1	LEU2/Kan	Expresses mDHFRI under pRNR2
6	pGA-YTK006	CEN6/ARS4, ColE1	URA3/Kan	Expresses mDHFRII under pRNR2
7	pGA-YTK013	CEN6/ARS4, ColE1	URA3/Kan	Expresses EE34 leucine zipper fused to mDHFRII
8	pGA-YTK014	CEN6/ARS4, ColE1	LEU2/Kan	Expresses EE34 leucine zipper fused to mDHFRI
9	pGA-YTK015	CEN6/ARS4, ColE1	URA3/Kan	Expresses RR34 leucine zipper fused to mDHFRII
10	pGA-YTK016	CEN6/ARS4, ColE1	LEU2/Kan	Expresses RR34 leucine zipper fused to mDHFRI
11	pGA-YTK017	CEN6/ARS4, ColE1	URA3/Kan	Expresses EE12RR345L->A leucine zipper fused to mDHFRII
12	pGA-YTK018	CEN6/ARS4, ColE1	LEU2/Kan	Expresses RR12EE345L leucine zipper fused to mDHFRI
13	pGA-YTK021	CEN6/ARS4, ColE1	URA3/Kan	Expresses EE12RR345L->A leucien zipper fused to mDHFRII (I114A variant)
14	pGA-YTK025b	CEN6/ARS4, ColE1	URA3/Kan	Expresses mDHFRII(I114A) under pRpl18b promoter
15	pGA-YTK026	CEN6/ARS4, ColE1	URA3/Kan	Expresses EE34 leucine zipper fused to mDHFRII(I114A), under pRNR2 promoter
16	pGA-YTK027	CEN6/ARS4, ColE1	URA3/Kan	Expresses RR34 leucine zipper fused to mDHFRII(I114A), under pRNR2 promoter

17	pGA-YTK026b	CEN6/ARS4, ColE1	URA3/Kan	Expresses EE34 leucine zipper fused to mDHFRII(I114A), under pRpl18b promoter
18	pGA-YTK027b	CEN6/ARS4, ColE1	URA3/Kan	Expresses RR34 leucine zipper fused to mDHFRII(I114A), under pRpl18b promoter
19	pGA-YTK028	CEN6/ARS4, ColE1	URA3/Kan	expresses mDHFRII(I114A) under pRNR2 promoter
20	pGA187-193	CEN6/ARS4, ColE1	LEU2/Amp	p1 integration cassettes, integrating various leucine zippers fused to mDHFRI
21	pGA194	CEN6/ARS4, ColE1	URA3/Kan	Bait plasmid with EE12RR345L leucine zipper, and p1 mutagenic polymerase "633"
22	pGA195	CEN6/ARS4, ColE1	URA3/Kan	Bait plasmid with EE12RR345L leucine zipper, and p1 mutagenic polymerase "633"
23	pGA196	CEN6/ARS4, ColE1	URA3/Kan	Bait plasmid with EE12RR345L leucine zipper, and p1 mutagenic polymerase "633"
24	pGA199	CEN6/ARS4, ColE1	LEU2/Amp	p1 integration cassette, integrating RR12EE345L leucine zipper fused to mDHFRI, with CFN scar as a control
25	pGA200	CEN6/ARS4, ColE1	LEU2/Amp	p1 integration cassette, integrating RR34 leucine zipper fused to mDHFRI, with CFN scar as a control
26	pGA201	CEN6/ARS4, ColE1	LEU2/Amp	p1 integration cassette integrating RR12EE345L leucine zipper fused to CfaN split intein
27	pGA202	CEN6/ARS4, ColE1	LEU2/Amp	p1 integration cassette integrating RR34 leucine zipper fused to CfaN split intein
28	pGA204	CEN6/ARS4, ColE1	LEU2/Kan	prey plasmid, expressing 3G61 anti-GFP DARPin
29	pGA205	CEN6/ARS4, ColE1	LEU2/Kan	prey plasmid expressing YKKD anti-GFP DARPin
30	pGA206	CEN6/ARS4, ColE1	URA3/Kan	bait plasmid expressing GFP-mDHFRII fusion
31	pGA207	CEN6/ARS4, ColE1	URA3/Kan	bait plasmid expressing mDHFRII-GFP fusion
32	pGA208-212	CEN6/ARS4, ColE1	varies	Various plasmids expressing ECC as prey or bait, used to test ECC homodimerization in yeast cytoplasm
33	pGA254	CEN6/ARS4, ColE1	LEU2/Kan	Prey plasmid expressing EE34(K25Q, K36Q) evolved variant
34	pGA255	CEN6/ARS4, ColE1	LEU2/Kan	Prey plasmid expressing RR34(V34I) evolved variant
35	pGA256	CEN6/ARS4, ColE1	LEU2/Kan	Prey plasmid expressing (-3e-g') variant
36	pGA257	CEN6/ARS4, ColE1	LEU2/Kan	Prey plasmid expressing (-3e-g')-E24T variant

Supplementary Table 2. List of yeast strains used in this study. All yeast strains used and created in this work are listed here.

Strain	Genotype	Parent strain	Source	Notes
YH1 (GA-Y292)	*MATa* HIS3 *ura3-52 leu2-3,112 trp1 tup*	UTL-7A	Tun Huang, Carol Sibley, Little and Haynes (1979)	Yeast strain with dTMP uptaking phenotype (*tup*). Generated by inhibiting DHFR activity on plates using methotrexate and sulfanilamide, and supplementing dTMP to rescue growth
YH5	*MATa* HIS3 *ura3-52 leu2-3,112 trp1 tup dfr1::URA3*	YH1	This work	
GA-Y294	*MATa his3 ura3-52 leu2 trp1 tup dfr1::URA3*	YH5	This work	Full deletions of HIS3 and LEU2 and TRP1 made in YH5 background, using CRISPR Cas9
GA-Y090	F102 HIS3 p1-shortpol-TRP1-mTurq2	F102	Alex Javanpour	F102 strain harboring p1-shortpol-TRP1-mTurq2, which was protoplast fused into YH5 to create a landing pad
GA-Y299	*MATa his3 ura3-52 leu2 trp1 tup dfr1::URA3* + p1-shortpol-TRP1-mTurq2	GA-Y294 and GA-Y090	This work	Created by protoplast fusing GA-Y294 with GA-Y090. This was done to transfer the OrthoRep plasmids into the dTMP uptaking strain, enabling split DHFR experiments with OrthoRep
GA-Y321 (also GA-Y399 and 400)	*MATa his3 ura3-52 leu2 trp1 tup dfr1::URA3* HO:Rpl18b-CfaC-mDHFRI-KanMX + p1-shortpol-TRP1-mTurq2	GA-Y299	This work	Split DHFR selection strain with split intein version of mDHFRI in the genome. A cassette encoding pRpl18b-CfaC-mDHFRI-KanMX was integrated into the HO locus of GA-Y299.

a)



b)

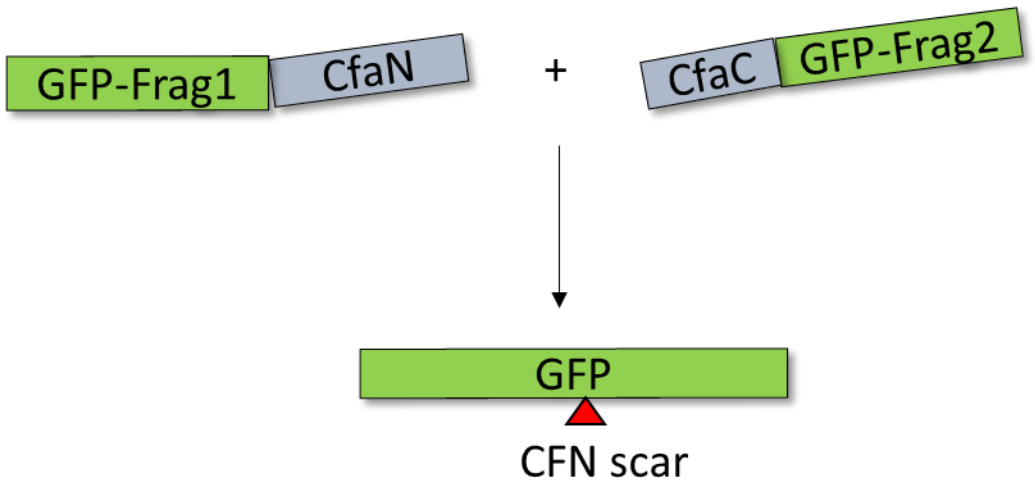
Type	X	Y	+dTMP	-dTMP
Base strain	No mDHFR or fragments			
+ control	Full WT mDHFR			
- control	-	-		
- control	-	-		
- control	- (I114A)	-		
- control	- (I114A)	-		
- control	Zip1 (I114A)	-		
- control	-	Zip2		
Exp. 1	Zip1 (I114A)	Zip2		
Exp. 2	Zip1 (I114A)	Zip2		

Expression level:

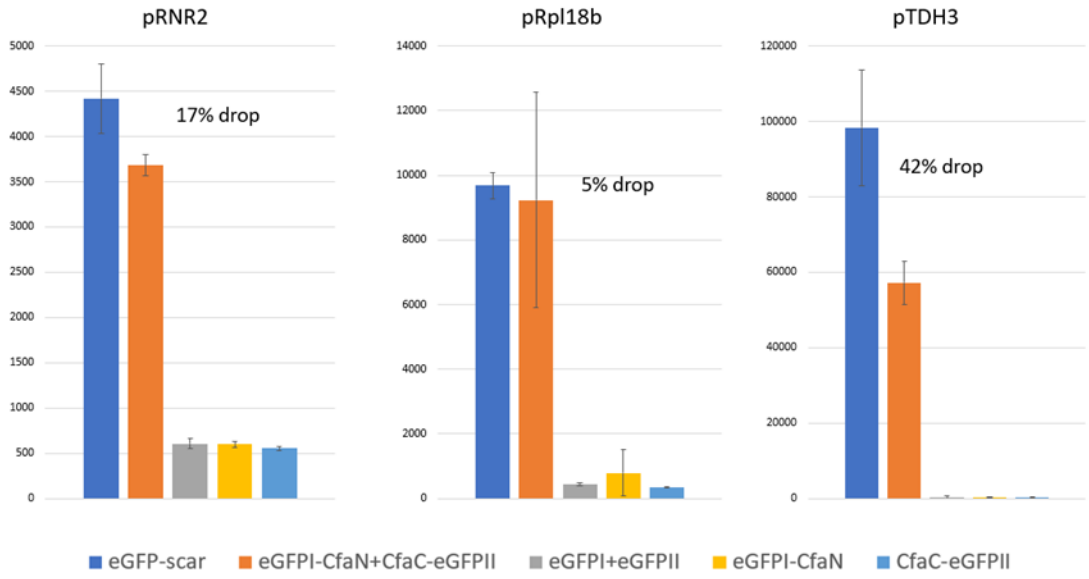
low (pRNR2)
medium (pRpl18b)

Supplemental Figure 1. Spontaneous reconstitution of split DHFR fragments is minimized by introduction of the I114A mutation into mDHFRII. a) Schematic of interaction between bait and prey proteins fused to split DHFR fragments. b) Expression of full murine DHFR restores GA-Y294's ability to grow without exogenous dTMP growth in GA-Y294. Expressions of mDHFRI and mDHFRII fragments without bait or prey results in some restored growth, suggesting spontaneous reconstitution of DHFR fragments, independent of bait or prey. This spontaneous reconstitution is minimized by introducing the I114A mutation into mDHFRII, which breaks contacts with I51 and L93 residues in mDHFRI. Fusion of strongly interacting leucine zippers (Zip1 and Zip2, K_d ~ 1 nM) results in full growth complementation in presence of the I114A mutation, suggesting that it has no deleterious effect on DHFR activity.

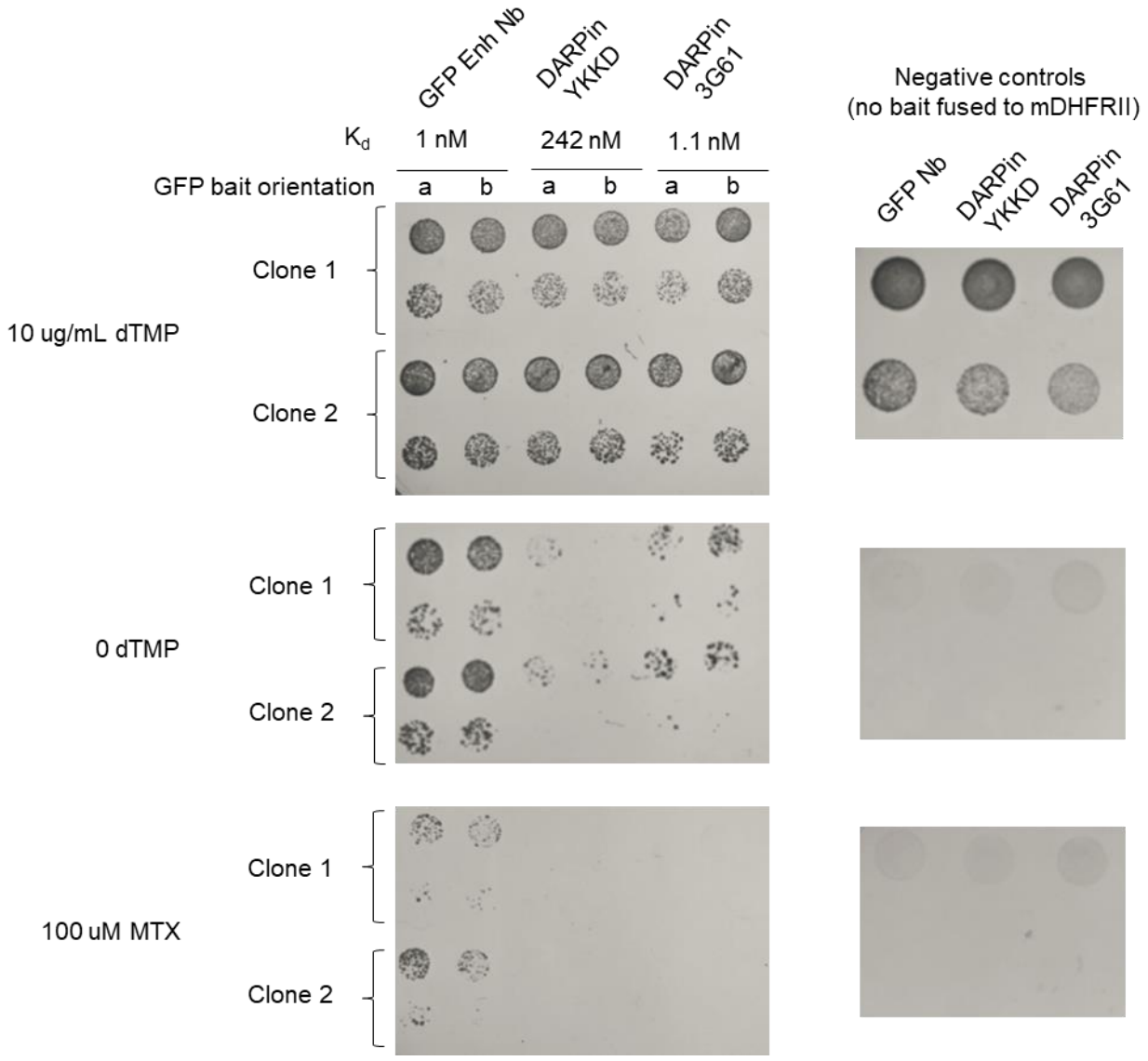
a)



b)



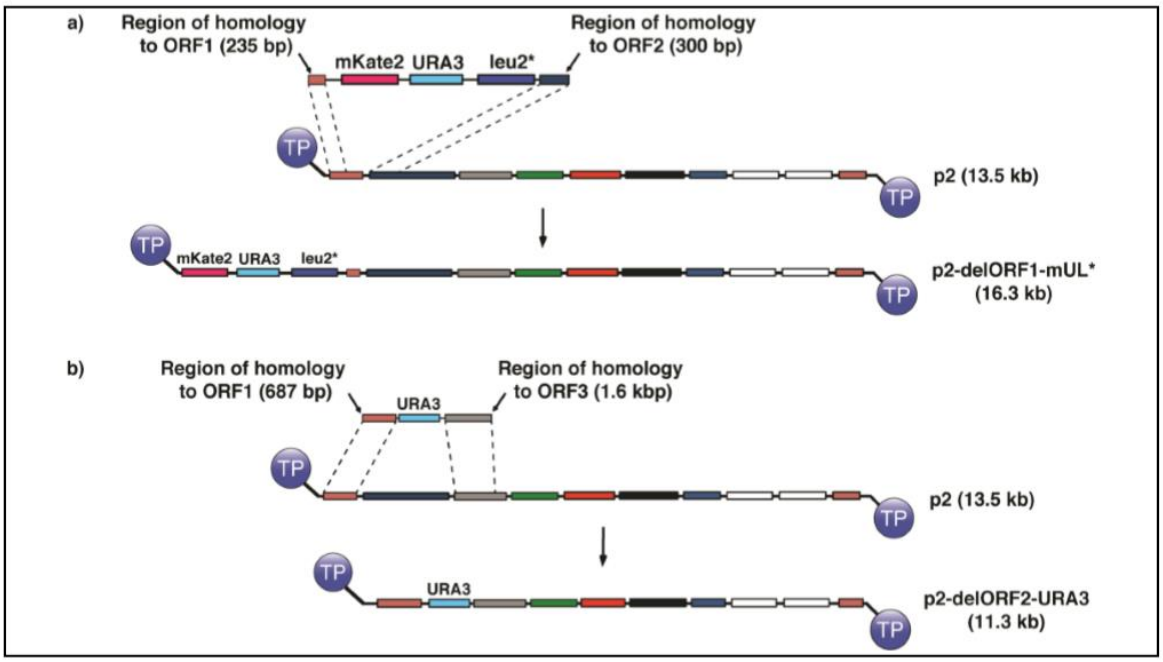
Supplemental Figure 2. Cfa split intein effectively performs post-translational fusion of split GFP in the yeast cytoplasm. a) GFP was split at residue 157, resulting in GFP-Frag1 (beta strands 1-7) and GFP-Frag2 (beta strands 8-11). The fragments were expressed as fusions with the CfaN and CfaC fragments of the Cfa split intein at low (pRNR2), medium (pRpl18b) and high (pTDH3) expression levels. Fluorescence was compared to a genetically encoded eGFP-scar protein, that simulates the final product of a successful split intein reaction, resulting in a 3-amino acid (CFN) scar. Use of the split-intein resulted in relatively small drops in total GFP produced (17% at low expression and 5% at medium expression), suggesting it may be an effective way to separately encode prey and mDHFRI to prevent cheater mutants.



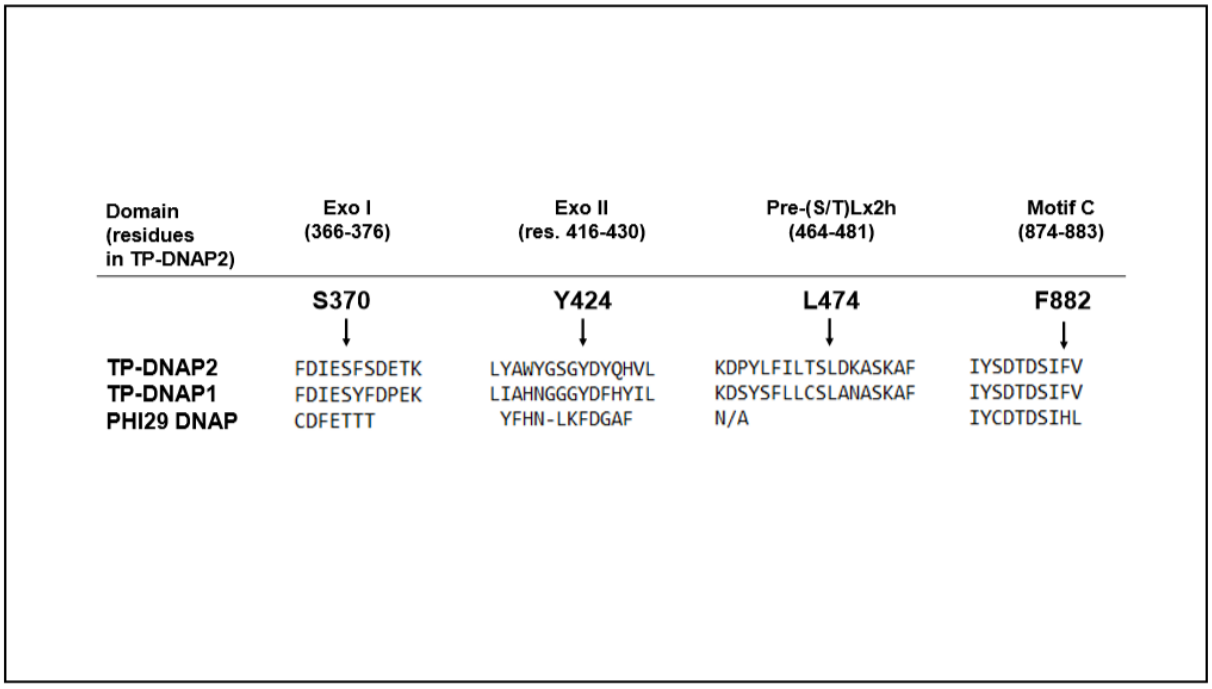
Supplemental Figure 3. Nanobody and DARPin based GFP binders are functional with the split DHFR system. Three GFP binders (GFP Enhancer nanobody, DARPin YKKD, and DARPin 3G61) were encoded as prey fusions to mDHFRI, and tested for binding against GFP as bait in two orientations (a: eGFP-mDHFRII(I114A), and b: mDHFRII(I114A)-eGFP). Binding was detected in all variants in both bait orientations resulted, and growth on media without dTMP was proportional to the binding affinity. Negative controls expressing the GFP binders fused to mDHFRI, along with bait-less mDHFRII(I114A) resulted in no growth complementation, suggesting that the binders have no off-target affinity for mDHFRII. Note: bait and prey were encoded on nuclear plasmids.

Appendix C

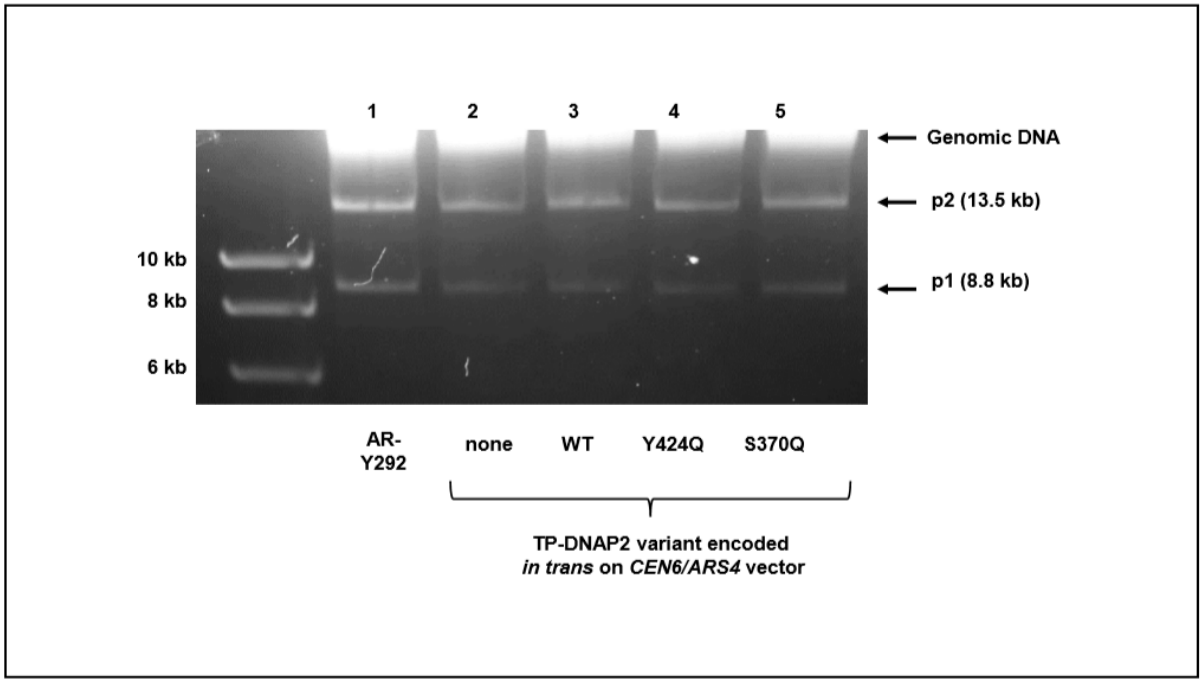
Supplementary information for Chapter 4



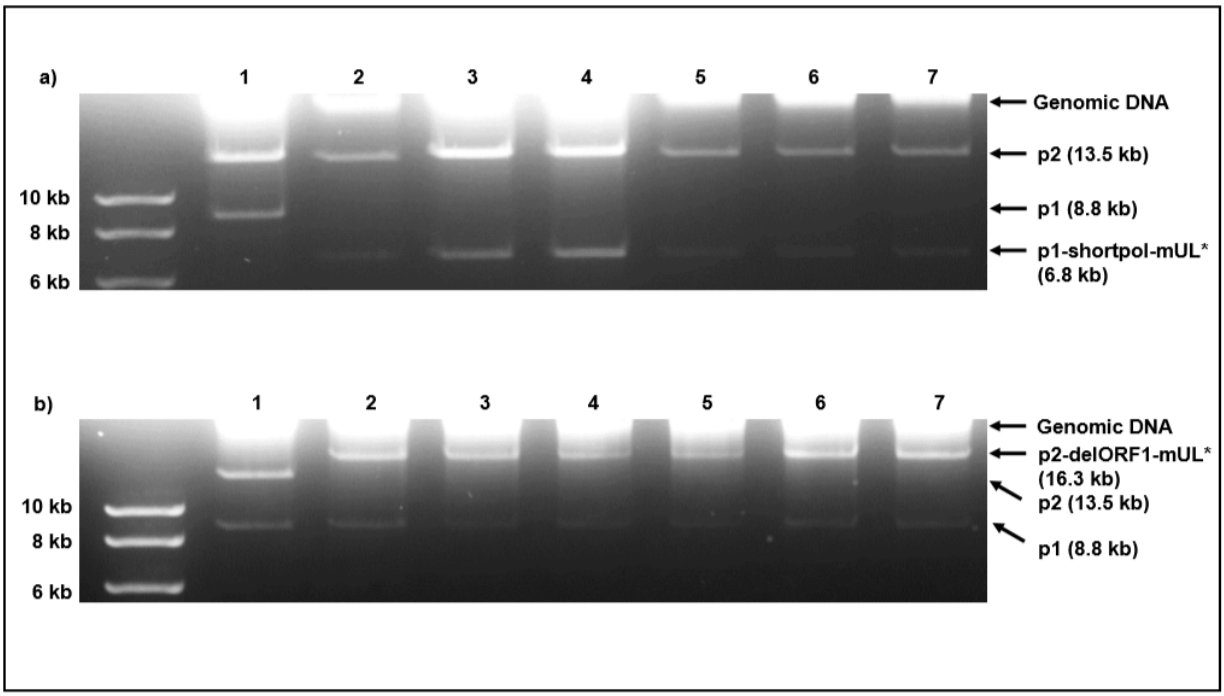
Supplementary Figure 1. Homologous recombination cassettes for p2 integration. Top: homologous recombination cassette encoding *mKate2, URA3,* and *leu2*,* and appropriate flanking regions for integration over ORF1 of p2. This linearized cassette is generated by digestion of pGA1 with ScaI. Bottom: homologous recombination cassette encoding *URA3* with flanking regions for integration over ORF2 of p2, resulting in full deletion of *TP-DNAP2.* This linearized cassette is generated by digestion of pGA59 with ScaI.



Supplementary Figure 2. Alignment between TP-DNAP2, TP-DNAP1 and Phi29 DNAP at chosen library sites. Selected regions from the alignment of TP-DNAP2, TP-DNAP1, and Phi29 DNA polymerase are shown, which highlights the homology between conserved domains at the chosen library sites.[26] While these conserved domains from TP-DNAP1 and TP-DNAP2 show high homology, the homology to Phi29 is limited. Due to this reason, library sites were chosen based on previously discovered residues that govern TP-DNAP1 fidelity, rather than introducing mutations from the well-studied Phi29 DNAP which increase mutation rate.[6,7] N/A indicates poor alignment.

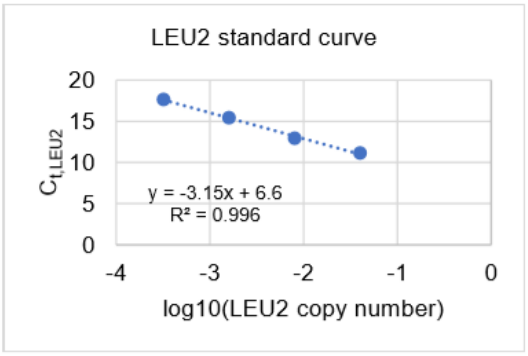


Supplementary Figure 3. Cytoplasmic plasmids from strains used to test orthogonality between TP-DNAP2 variants and genomic replication. Agarose gel (0.9%) analysis of DNA extracted after measuring genomic mutation rate with *URA3* fluctuation tests, in presence of mutagenic TP-DNAP2 variants. Lane 1: DNA extracted from parent strain AR-Y292, containing wild type p1 and p2. Lanes 2-5: DNA extracted from strains corresponding to Entries 1-4 in Table 2. Replication of p1 and p2 is sustained in all strains throughout the fluctuation test procedure, ensuring that TP-DNAP2 variants had the opportunity to replicate p2 and lose their terminal proteins, which may play a role in determining specificity for p2 versus the host genome.

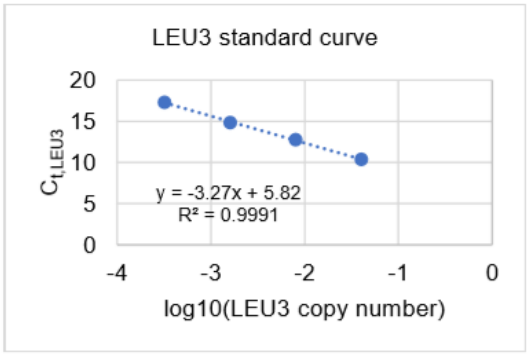


Supplementary Figure 4. Cytoplasmic plasmids from strains used to test mutual orthogonality. a) Gel analysis (0.9% agarose) of DNA extracted from p1 mutation rate reporter strains used to measure p1 mutation rate in presence of 6 TP-DNAP variants. Lane 1: parental strain AR-Y292. Lanes 2-7: In order, TP-DNAP1^{WT}, TP-DNAP1$^{I777K, L900S}$, TP-DNAP1$^{L477V, L640Y, I777K, W814N}$, TP-DNAP2^{WT}, TP-DNAP2^{Y424Q}, TP-DNAP2^{S370Q}. Wild type p2 is present in all strains, along with the recombinant p1-shortpol-mUL* which encodes *mKate2, URA3,* and *leu2*.* b) Gel analysis (0.9% agarose) of DNA extracted from strains used to measure p2 mutation rate in presence of 6 TP-DNAP variants. These strains correspond to p2 mutation rate reporter strains described in Fig. 3a. Wild type p1 is present in all strains, along with the recombinant p2-delORF1-mUL* which encodes *mKate2, URA3,* and *leu2*.* Lanes 2-7: In order, TP-DNAP1^{WT}, TP-DNAP1$^{I777K, L900S}$, TP-DNAP1$^{L477V, L640Y, I777K, W814N}$, TP-DNAP2^{WT}, TP-DNAP2^{Y424Q}, TP-DNAP2^{S370Q}.

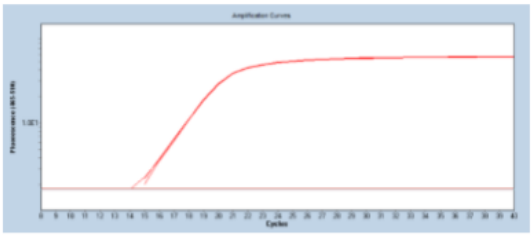
A



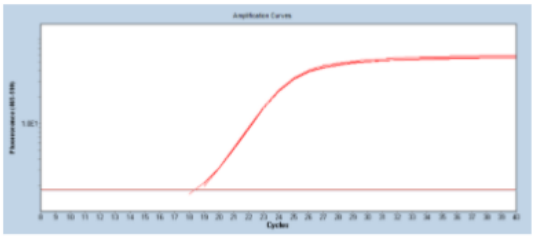
B



C



D



E

	LEU2 C_t	LEU3 C_t
Replicate 1	13.74	17.96
Replicate 2	13.45	18.19
Replicate 3	13.14	17.85
Mean	13.44	18.00

LEU2 copy number calculation:	**LEU3 copy number calculation:**
y = -3.15x + 6.6; where y = $C_{t,LEU2}$; x = log10(LEU2 copy number, 4000x diluted) 13.44 = -3.15x + 6.6 x = -2.17 LEU2 copy number (4000x diluted) = 10^x = $10^{-2.17}$ = 0.0068 LEU2 copy number (1x) = 4000*0.0068 – 1 genomic copy of LEU2 = **26.2**	y = -3.27x + 5.82; where y = $C_{t,LEU3}$; x = log10(LEU3 copy number, 4000x diluted) 18.00 = -3.27x + 5.82 x = -3.72 LEU3 copy number (4000x diluted) = 10^x = $10^{-3.72}$ = 0.000191 LEU3 copy number (1x) = 4000*0.000195 = **0.76**
Normalized LEU2 copy number = LEU2 copy number/LEU3 copy number = 26.2/0.76 = **34.5 copies**	

Supplementary Figure 5. qPCR data analysis. (A) and (B): Standard curves prepared by performing triplicate qPCR reactions with primers specific to LEU2 (A) or LEU3 (B), where the template DNA is a dilution series of DNA extracted from AR-Y003. AR-Y003 contains a single genomic copy of LEU2 and LEU3, allowing a standard curve of C_t vs. copy number to be constructed. (C) and (D): Triplicate qPCR reactions performed on a sample strain, with primers specific to LEU2 (C) or LEU3 (D), with C_t threshold shown at Fluorescence (465-510) = 1.75. (E) Each sample's mean $C_{t,LEU2}$ and $C_{t,LEU3}$ value is first converted to copy numbers, respectively, using the linear fits from standard curves shown in C and D. LEU2 copy number is then normalized by LEU3 copy number to correct for potential differences in DNA extraction efficiency between samples.