UNIVERSITY OF CALIFORNIA,
IRVINE

An orthogonal DNA replication system for *in vivo* continuous directed evolution

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Engineering

by

Arjun Ravikumar

Dissertation Committee:
Professor Chang C. Liu, Chair
Professor John Chaput
Professor Arthur D. Lander

2018

Table of Contents

Page

List of figures iv

List of tables vi

Acknowledgements vii

Curriculum vitae ix

Abstract of the dissertation xi

Chapter 1: Introduction to *in vivo* continuous directed evolution 12

1.1 Traditional directed evolution
1.2 Challenges in achieving continuous directed evolution *in vivo*
1.3 Targeted mutagenesis in *E. coli* with error-prone DNA polymerase I
1.4 Yeast systems that do not use engineered DNA polymerases for mutagenesis
1.5 Somatic hypermutation as a means for targeted mutagenesis of GOIs
1.6 Phage assisted continuous evolution (PACE)
1.7 Orthogonal DNA replication (OrthoRep)
1.8 References

Chapter 2: Establishing an orthogonal DNA replication system (OrthoRep) 46

2.1 Traditional directed evolution
2.2 Increasing the p1 mutation rate
2.3 Verifying orthogonality of error-prone OrthoRep
2.4 Methods
2.5 References

Chapter 3: Engineering highly error-prone orthogonal DNAPs for OrthoRep 78

3.1 Screening single amino acid TP-DNAP1 mutants
3.2 Screening combinatorial libraries
3.3 Methods
3.4 References

Chapter 4: Crossing mutation-induced genomic error thresholds with OrthoRep 102

4.1 Crossing the haploid extinction threshold
4.2 Crossing intermediate genomic error thresholds
4.3 Methods

4.4 References

Chapter 5: High-throughput directed evolution of *Plasmodium falciparum* DHFR 110

5.1 Directed evolution of *Pf*DHFR drug resistance using OrthoRep
5.2 New adaptive pathways to resistance
5.3 Methods
5.4 References

Chapter 6: Discussion 131

6.1 OrthoRep as a new paradigm for directed evolution
6.2 Future OrthoRep technology development
6.3 Methods
6.4 References

Appendix A: Comparison of directed evolution systems 148

Appendix B: List of plasmids used in Chapter 2 149

Appendix C: List of primers used in Chapter 2 154

Appendix D: 99 TP-DNAP1 homologs generated via protein BLAST 157

Appendix E: All TP-DNAP1 variants characterized by fluctuation tests in Chapter 3 158

Appendix F: 210 TP-DNAP1 variants that replicate p1 at a higher copy number than WT TP-DNAP1 170

Appendix G: A list of all plasmids and yeast strains used in Chapter 3-6 175

List of figures

		Page
Figure 1.1	The traditional process of directed evolution	13
Figure 1.2	Targeted mutagenesis in *E. coli* with error-prone DNA polymerase I	20
Figure 1.3	Yeast systems that do not use engineered DNA polymerases for mutagenesis	25
Figure 1.4	Somatic hypermutation as a means for targeted mutagenesis	29
Figure 1.5	Phage assisted continuous evolution (PACE)	31
Figure 1.6	Orthogonal DNA replication (OrthoRep)	38
Figure 2.1	The p1/2 cytoplasmic plasmid system	47
Figure 2.2	DNA replication via protein-primed initiation as a mechanistic basis for orthogonal replication	48
Figure 2.3	Integration vectors used to introduce genes into p1	49
Figure 2.4	Gel analysis of cytoplasmic DNA plasmids from various strains	51
Figure 2.5	TP-DNAP1 variants screened for increased mutation rate	54
Figure 2.6	Replication of p1 by a recoded TP-DNAP1 expressed *in trans*	55
Figure 2.7	Expression of WT TP-DNAP1 *in trans*	56
Figure 3.1	Design of TP-DNAP1 mutants by homology analysis and by construction of a scanning saturation mutagenesis library	80
Figure 3.2	Functional purification of TP-DNAP1 sub-libraries	83
Figure 3.3	Engineering of highly error-prone orthogonal DNAPs for OrthoRep	86
Figure 4.1	OrthoRep crosses mutation-induced error thresholds of the yeast genome	103
Figure 5.1	High-throughput directed evolution of *Pf*DHFR resistance to pyrimethamine	112

Figure 5.2 Pilot studies of p1-*Pf*DHFR evolution 114

Figure 5.3 Five-mutation *Pf*DHFR fitness landscapes 117

Figure 5.4 Dynamics of *Pf*DHFR evolution in six representative populations 119

Figure 5.5 Alternative evolutionary trajectories directed by a synonymous initial mutation at S108. 121

Figure 6.1 Activity of TP-DNAP1 fusions to double-stranded DNA binding domains 137

Figure 6.2 Mn^{2+} increases WT p1 mutation rates 138

Figure 6.3 RNR1 mutants known to create imbalanced dNTP pools increase WT p1 mutation rates 139

Figure 6.4 Titratable control of p1 copy number 140

Figure 6.5 Characterization of promoter strengths of the p1/2 transcription system 141

List of tables

		Page
Table 2.1	Mutation rates measured in Chapter 2 using fluctuation analyses	52
Table 2.2	Copy numbers of p1-based plasmids	58
Table 2.3	Mutation rates of p1 replication by TP-DNAP1 (Y427A) in *format B* as measured using p1-Short-UL*(TGA) and p1-Short-U*L	59
Table 2.4	Substitution mutation preferences of p1 replication by TP-DNAP1 (Y427A) in *format B*	60
Table 3.1	Calibration curve of qPCR-determined p1 copy number to p1-encoded mKate2 fluorescence	82
Table 3.2	Mutation G410H broadly increases activity of TP-DNAP1 variants	85
Table 3.3	Substitution mutation preferences of highly error-prone TP-DNAP1s	87
Table 3.4	Yeast genomic substitution mutation rates in the presence of OrthoRep	88
Table 5.1	Stochastic extinction in revival experiments of *Pf*DHFR evolution	115
Table 6.1	Exonuclease mutants incorporated into highly error-prone TP-DNAP1s	136

Acknowledgements

I am indebted to many people that have taught and helped me over the years. This work would not have been possible without them.

I would like to thank my parents, Dr. Vasulinga Ravikumar and Sumathi Ravikumar. My father taught me to care about people and to work hard, in that order. And my mother, through her embrace of duty, taught me about responsibility. These are the principles that carried me through graduate school.

I would like to thank my brother, Arvind Ravikumar. Even though he is 10 years younger than me, often times I am the one learning from him. I am continually impressed by him.

I would like to thank my friends: Dr. David Li, Dr. Elena Liang, Erik Peterson, Kamran Ali, Dr. Phil Thomas, Pradeep Ramesh, and Dr. Siavash Ahrar. I was fortunate enough to find a kind and supportive group that were, frankly, a lot more mature than me. I grew tremendously from our time together.

I would like to thank Dr. Kim Makuch. Without a doubt, getting to know Kim was the best part of graduate school. She made me a better person; she made me better at research; and I owe her tremendously for both.

I would like to thank all of my lab mates. In particular, I would like to thank Garri Arzumanyan, who set up the entire *Pf*DHFR selection and carried the load with those experiments. This story would not have been complete without his efforts. I want to thank Muaeen Obadi, who shared the burden of the large DNA polymerase screen. It was mentally and physically exhausting, but Muaeen was always willing to push through. I would like to thank Ziwei Zhong and Alex Javanpour. We went through a couple of difficult years when experiments weren't working, but their persistence was essential for getting to the other side. I want to thank Adrian Arrieta who was critical in the early days, in terms of research and in terms of setting up the lab. I would like to thank Eamon Jahani and Ming Wang, who were fantastic undergraduates that I had the opportunity to work with. And I would like to thank Dr. Xiang Li. I have never seen someone learn as much, as quickly as he did in graduate school, and his efforts were a constant source of inspiration for me. His willingness to try and try again will set the standard in our lab for a long time.

I would like to thank my committee members: Dr. John Chaput, Dr. Arthur Lander, Dr. Nancy Da Silva, and Dr. Anthony Long. They have been very generous with their time, feedback, and encouragement.

I would like to thank my advisor, Dr. Chang Liu. For his time, for his care, and for instilling confidence in me, I will always be grateful.

Portions of this text have been committed to future publication by Wiley-VCH in a book titled, "Protein Engineering: Tools and Applications". Co-authors of this chapter include

Alon Wellner and Chang C. Liu. The publisher and co-authors have granted permission to reprint materials in this thesis.

Portions of this text are reprints of material as they appear in *Nature Chemical Biology* and *Cell*. Co-authors on these works include Adrian Arrieta, Garri A. Arzumanyan, Alex A. Javanpour, Chang C. Liu, and Muaeen K. A. Obadi.

Dr. Chang C. Liu directed and supervised the research which formed the basis for this thesis.

This research was funded by the NIH (1DP2GM119163-01), the Defense Advanced Research Projects Agency (HR0011-15-2-0031), the Sloan Research Fellowship, the Beckman Young Investigator Award, the Dupont Young Professor Award, startup funds from UC Irvine, and a US National Science Foundation Graduate Research Fellowship.

Curriculum vitae

Arjun Ravikumar

Education

***UC Irvine**, Irvine, California* **2012-2018**

PhD Biomedical Engineering
Advisor: Prof. Chang Liu
Research focus: An orthogonal DNA replication system for in vivo continuous directed evolution

***UC Berkeley**, Berkeley, California* **2011 – 2012**

PhD Bioengineering
Advisors: Prof. John Dueber; Prof. Adam Arkin and Dr. Chang Liu
Research focus: Xylose utilization pathway engineering; An orthogonal DNA replication system for in vivo continuous directed evolution

***California Institute of Technology**, Pasadena, California* **2007 - 2011**

B.S. Bioengineering
Advisors: Prof. Niles Pierce; Prof. Richard Murray
Research focus: Small, conditional RNAs for selective chemotherapies; Transcriptional oscillators in cell-free systems

Awards

- NSF Graduate Research Fellowship, 2011.
- Richter Scholar, Caltech, 2010.
- Summer Undergraduate Research Fellowship, Caltech, 2010.
- Summer Undergraduate Research Fellowship, Caltech, 2009.

Publications

- Wellner A, **Ravikumar A**, Liu CC. Continuous evolution of proteins *in vivo*. *Submitted* (2018).
- Zhong Z, **Ravikumar A**, Liu CC. Tunable expression systems for orthogonal DNA replication. *ACS Synthetic Biology*, accepted (2018).
- **Ravikumar A**, Arzumanyan G, Obadi MKA, Javanpour AA, Liu CC. Scalable, continuous evolution of genes at mutation rates above genomic error thresholds. *Cell*, in press (2018).
- Arzumanyan GA, Gabriel KN, **Ravikumar A**, Javanpour AA, Liu CC. Mutually orthogonal DNA replication systems *in vivo*. *ACS Synthetic Biology*, **7**, 1722-1729 (2018).

- **Ravikumar A**, Liu CC. Biocontainment through reengineered genetic codes. *ChemBioChem*, **16**, 1149-1151 (2015).
- **Ravikumar A**, Arrieta A, Liu CC. An orthogonal DNA replication system in yeast. *Nature Chemical Biology*, **10**, 175-177 (2014).

Selected Presentations

- Poster presentation at Synthetic Biology: Engineering, Evolution & Design (SEED), 2014.
- Conference talk at Engineering Biology Research Consortium (EBRC), 2015.
- Poster presentation at Synthetic Biology Gordon Research Conference, 2017.
- Poster presentation at Synthetic Biology: Engineering, Evolution & Design (SEED), 2018.

Teaching Experience

UC Berkeley
Graduate Student Instructor
- BioE 140L: Synthetic Biology Laboratory **Fall 2011, Fall 2012**
  Prof. John Dueber

UC Berkeley
Research Mentor
- Undergraduates: Jenna Yi **4/2012 – 12/2012**

UC Irvine
Research Mentor
- Undergraduates: Ming Wang, Eamon Jahani **4/2012 – 6/2018**
- Technician: Muaeen Obadi **9/2015 – 09/2016**

Abstract of the dissertation

An orthogonal DNA replication system for *in vivo* continuous directed evolution

By

Arjun Ravikumar

Doctor of Philosophy in Biomedical Engineering

University of California, Irvine, 2018

Professor Chang C. Liu, Chair

Directed evolution is a powerful approach for engineering biomolecules and understanding the basic principles of adaptation. However, experimental strategies for directed evolution are notoriously labor-intensive and low-throughput, limiting access to demanding functions, multiple functions in parallel, and the study of molecular evolution in replicate. Here, I describe OrthoRep, an orthogonal DNA polymerase-plasmid pair in yeast that stably mutates ~100,000-fold faster than the host genome *in vivo*, exceeding the error threshold of genomic replication that causes single-generation extinction. User-defined genes in OrthoRep continuously and rapidly evolve through serial passaging, a highly straightforward and scalable process. Using OrthoRep, I evolved malarial DHFR to strongly resist the drug pyrimethamine in 90 independent replicates. This large-scale experiment revealed a more complex fitness landscape than previously realized, including new resistant alleles, 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. OrthoRep enables a new paradigm of routine, high-throughput evolution of biomolecular and cellular function.

Chapter 1

Introduction to *in vivo* continuous directed evolution

1.1 Traditional directed evolution

Directed evolution is a powerful approach for engineering new biomolecular and cellular functions [1-3]. In contrast to rational design approaches, directed evolution exploits diversity and evolution to shape the behavior of biological matter by applying the Darwinian cycle of mutation, selection, and amplification to genes and genomes. By doing so, the field of directed evolution has generated important insights into the evolutionary process [4-6] as well as useful RNAs, proteins, and systems with wide-ranging applications across biotechnology and medicine [7-11].

To mimic the evolutionary process, classical directed evolution approaches carry out cycles of *ex vivo* diversification on genes of interest (GOIs), transformation of the resulting gene libraries into cells, and selection for the desired function (**Figure 1.1**). Each iteration of this cycle is defined as a round of evolution, and as selection stringency increases over rounds, either automatically through competition or manually through changing conditions (or both), this process can lead GOIs closer and closer to the desired function. This overall process makes practical sense for a number of reasons, especially for the goal of protein engineering (*i.e.* GOI encodes a protein). First, *ex vivo* diversification is appropriate, because test tube molecular biology techniques such as DNA shuffling, site-directed saturation mutagenesis, and error-prone PCR [2] are capable of generating exceptionally high and precise levels of sequence diversity for any GOI. Second, transforming diversified libraries of the GOI into cells is appropriate, because each GOI variant needs to be translated into a

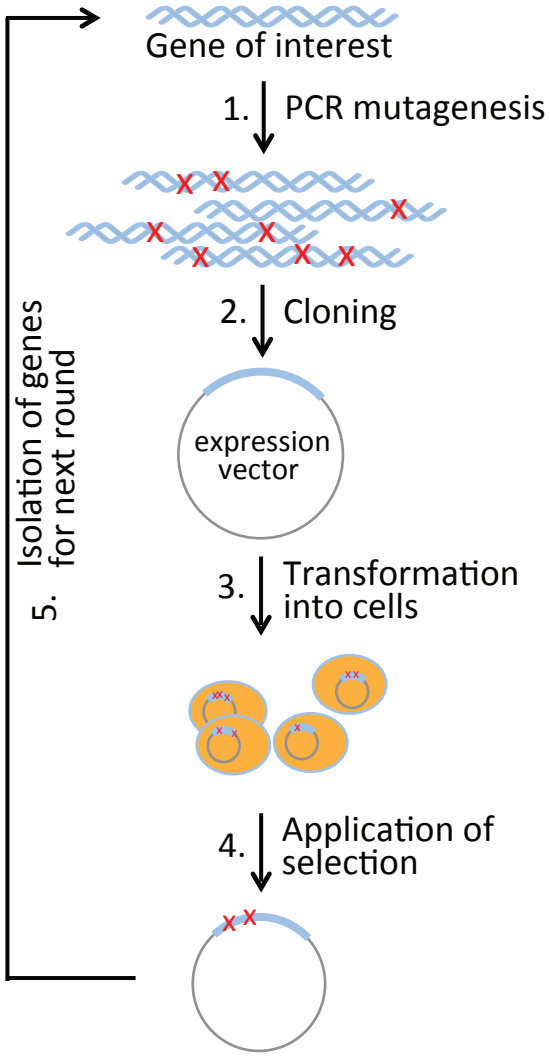


Figure 1.1 | The traditional process of directed evolution.
Genes of interest are mutated *in vitro*, cloned into an expression vector, transformed into cells, and subjected to selection conditions. Mutants surviving selection are the basis for the next round of mutation, cloning, transformation, and selection. Each round takes ~1 week, and transformation efficiencies of *Escherichia coli* or *Saccharomyces cerevisiae* limit standard library sizes to $\sim 10^9$ or $\sim 10^7$, respectively.

protein in order to express its function, and cells, especially model microbes, are naturally robust hosts for protein expression. Third, carrying out selection inside cells is appropriate, because 1) cells automatically maintain the genotype-phenotype connection between the GOI and expressed protein that is necessary for amplification of desired variants, 2) we often care about a GOIs function within the context of a cell, especially as metabolic engineering and cell-based therapy applications mature, and 3) the use of cell survival as the output for a desired protein function allows millions or billions of GOI variants to be simultaneously tested – it is easy to culture billions of cells under selection conditions – in contrast to *ex vivo* screens that are much lower throughput. Survival-based selections are not always immediately available, but with some ingenuity, one can usually find a way to link the desired function of a protein to cellular fitness.

While sensible, the practical requirement that diversification should occur *in vitro* but expression and selection should occur *in vivo* in this classical directed evolution pipeline creates significant suboptimalities. First, the number of steps that can be taken along an adaptive path becomes few, since each round of *in vitro* mutation, transformation, and *in vivo* selection takes several days or weeks to carry out. Second, limited DNA transformation efficiencies result in strong bottlenecking of diversity that can mitigate the probability of finding the most optimal solutions in sequence space. Third, the number of evolution experiments that can be run simultaneously is minimal, because *in vitro* mutagenesis, cloning, and transformation are experimentally onerous, demanding extensive researcher intervention [12]. These shortcomings keep two highly promising categories of experiments largely outside the grasp of classical methods: first is the directed evolution of genes toward highly novel functions that likely require long mutational paths to reach (*e.g.*

the optimization of multi-gene metabolic pathways or the *de novo* evolution of enzyme activity); and second is the large-scale replication of directed evolution experiments, needed in cases when many different functional variants of a gene are desired (*e.g.* the evolution of multiple synthetic receptors for a collection of ligands) or when statistical power is required in order to understand outcomes in experimental evolution (*e.g.* probing the scope of adaptive trajectories leading to resistance in a drug target).

The emerging field of *in vivo* continuous directed evolution seeks to overcome these shortcomings by performing both continuous diversification of the GOI and selection entirely within living cells [13]. In this way, GOIs can be rapidly and continuously evolved through basic serial passaging of cells under selective conditions. This removes the labor-intensive cycling between *in vitro* and *in vivo* steps and the DNA transformation bottlenecks associated with the classical pipeline, creating a new paradigm for directed evolution that is limited only by the generation time of the host cell and the number of cells that can be cultured. These limitations are usually negligible – in most host organisms for directed evolution such as *Escherichia coli* and *Saccharomyces cerevisiae*, generation time is fast (20-100 minutes) and the number of cells that can be cultured is massive (10^8-10^9 per mL) – so the potential power of continuous systems is enormous. Moreover, *in vivo* continuous directed evolution is amenable to high-throughput experiments, because serial passaging is straightforward and can be automated at scale or converted to continuous culture using bioreactors [14-16].

1.2 Challenges in achieving continuous directed evolution *in vivo*

Before discussing how *in vivo* continuous evolution can be realized, I shall first clarify why this is a challenging problem. The difficulty of achieving *in vivo* continuous evolution of

GOIs lies in the fundamental relationship between how fast one can mutate an information polymer and its length. Several theories predict that organisms face an "error threshold" at mutation rates on order 1/L (where L is the length of the genome), near which selection cannot maintain fitness, leading to gradual decline towards low fitness, or above which one is nearly guaranteed a lethal mutation every cycle of replication, leading to rapid extinction [17-20]. Because cellular genomes are large (*e.g.* $\sim 5 \times 10^6$ in *E. coli*, $\sim 1.2 \times 10^7$ in *S. cerevisiae*, and $\sim 3 \times 10^9$ in humans), this implies that evolution strongly favors low genomic mutation rates (*e.g.* $\sim 10^{-9}$ substitutions per base (s.p.b.) in *E. coli* and $\sim 10^{-10}$ s.p.b. in *S. cerevisiae* and human somatic cells). Experiments confirm this prediction. Drake observed empirically that mutation rates scale as 1/L across many organisms [17]; evolution experiments have shown that when mutator phenotypes do arise, they are accompanied by fitness costs and only transiently persist [19, 21-23]; and more direct tests in yeast find that there is indeed a mutation-induced extinction threshold at $\sim 1/L$, above which yeast cannot propagate [18]. Yet individual GOIs are small in comparison to genomes, so they are capable of tolerating much higher error rates. In fact, they *require* much higher per base error rates than genomes to generate the same amount of total mutational diversity, because they have fewer bases. Following the 1/L scaling, a typical 1kb GOI should be able to tolerate mutation rates on order $\sim 10^{-3}$ s.p.b.

Therefore, the primary challenge in achieving rapid *in vivo* continuous evolution is how to develop molecular machinery or other strategies that target rapid mutagenesis to only GOIs, allowing the host genome to replicate at mutation rates below its low error thresholds but driving the GOI at the high mutation rate necessary for fast generation of sequence diversity. When considering the level of targeting in the ideal case, the

formidability of this challenge becomes quite apparent. Ideally, one should continuously mutate GOIs at rates close to their error threshold (~10^{-3}) to maximize diversification but leave the genomic error rate completely unchanged, as the genome's error rate is evolutionarily optimized for host fitness. In *E. coli*, *S. cerevisiae*, and human cells, this means that on-target versus off-target mutagenesis must differ by 10^6-fold, 10^7-fold, and 10^7-fold, respectively, which is much more than the 10- to 1000-fold targeting required in most synthetic biology problems involving molecular recognition. *How can we achieve such extreme precision in mutational targeting in the cell?*

There is yet another hard challenge in realizing *in vivo* continuous evolution, which has to do with the durability of mutagenesis. Ideally, one wants a high rate of mutagenesis on the GOI to persist indefinitely (or at least for as long as the experimenter cares), so that a protein can traverse long mutational pathways toward desired functions. Because one needs to achieve mutational targeting to the GOI, there is almost always a risk to durability: any mechanism for targeting the GOI over the rest of the genome will necessarily rely on some cis-elements in or surrounding the GOI to mediate the targeting. If these cis-elements become mutated, which is quite likely since they are in or near the GOI undergoing rapid mutation, then mutagenesis will slow or stop. Ideally, a continuous evolution system will limit the chance that a cis-element for mutational targeting gets degraded. In the case that it does, an ideal system will remove the GOI containing the mutated cis-element from the population so that it can't fix in the population (through gradual mutational accumulation or a selective sweep if mutagenesis comes with a fitness cost) and end the continuous evolution process prematurely. *How do we achieve architectures for durability?*

Other challenges for *in vivo* continuous evolution include generality across host organisms, the ability to mutate many genes simultaneously, and fine control over mutation rate and spectra; but the most defining ones are targeting and durability. In the remainder of this chapter, I describe several *in vivo* continuous directed evolution platforms within the framework of these challenges. In Section 1.6, I also discuss phage assisted continuous evolution (PACE), which has been remarkably successful for continuous biomolecular evolution. Although PACE is not an entirely *in vivo* system, it also achieves complete precision in mutational targeting and durability – in fact by not being entirely *in vivo,* as explained below. I conclude the chapter, in Section 1.7, by proposing orthogonal DNA replication (OrthoRep) as a system for *in vivo* continuous directed evolution that can uniquely achieve precision in mutational targeting and high durability for enforcing prolonged mutagenesis in GOIs. I do not discuss several powerful technologies for non-continuous *in vivo* diversification or streamlined diversification methods, such as MAGE [24], CREATE [25], DiVERGE [26], and CPR [27], but note that these are also promising approaches to protein evolution as they address some of the constraints of classical directed evolution methods. A summary of various characteristics of the systems discussed below is provided in Appendix A.

1.3 Targeted mutagenesis in *E. coli* with error-prone DNA polymerase I

The first system that was able to perform continuous targeted mutagenesis *in vivo* was published in 2000 by Fabret *et al.* [28]. It was designed based on developments in understanding the mechanism of ColE1 plasmid replication in *E. coli*. For plasmids that contain a ColE1 origin of replication, DNA polymerase I (Pol I) is responsible for elongating from the RNA primer that initiates replication at the origin. Pol I will extend for about 400-

2000 bp, after which DNA polymerase III (Pol III), responsible for bulk DNA replication in *E. coli*, replaces Pol I [29]. When using a genome-encoded proofreading-deficient Pol I, genes that were cloned near the ColE1 origin experienced a 6-20-fold higher degree of mutagenesis over genes at more remote areas in the plasmid, showing targeting. The system's components were further combined with mismatch repair mutants to raise the mutation rate on GOIs yet another 20- to 40-fold, although significant increases in genomic mutation rates of at least several hundred-fold were observed. As a proof of concept, the authors evolved dominant negative variants of LacI that would outcompete a genomically-encoded wild type (WT) LacI in binding its cognate operator, LacO. After 30 generations, LacI mutants that caused complete abolishment of WT LacI's binding to LacO were isolated. These variants were altered in their DNA binding domain but still formed tetramers with WT LacI, thereby abolishing LacI's repression at LacO.

Further improvement of the Pol I/ColE1 system was demonstrated in 2003 (**Figure 1.2A**) [30, 31]. Camps *et al.* modified the system to express the error-prone (ep) Pol I from a plasmid with a Pol I-independent origin of replication, so that it would not mutate its own gene. Then, they used a host *E. coli* strain (J2000) whose genomically-encoded WT Pol I was temperature sensitive (ts) [32]. At restrictive temperatures, the ts Pol I becomes inactive such that only the ep Pol I acts, preventing the high-fidelity ts Pol I from competing for replication at the ColE1 origin. Based on prior studies of Pol I from the same lab [33], Camps *et al.* engineered a Pol I variant that was exceptionally error-prone, leading to mutation rates as high as 8.1×10^{-4} s.p.b at the GOI when the ts Pol I was inactivated. Mutagenesis expanded to about 3 kb from the ColEI origin and was evenly distributed within this region, albeit with certain biases in mutational preference. As a proof of concept

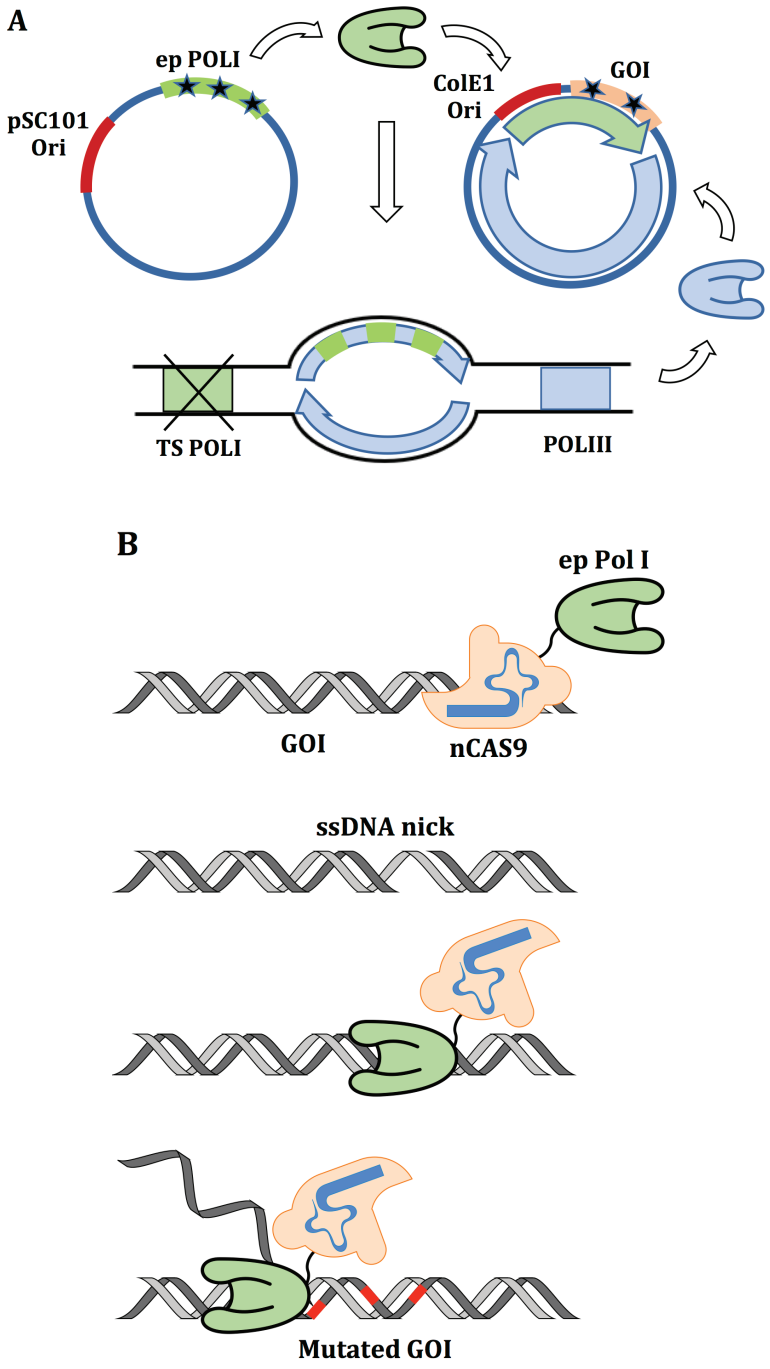


Figure 1.2 | Targeted mutagenesis in *E. coli* with error-prone DNA polymerase I. **(A)** An ep version of Pol I is expressed from a plasmid whose replication is driven by a non-ColE1 origin of replication (ori). The GOI is placed on the target plasmid near the ColE1 ori and thus targeted for mutagenesis. After 1-3 kb of ep replication, Pol III replaces Pol I to replicate the remainder of the plasmid with high fidelity. The genomic allele of *POL I* is temperature sensitive, such that the enhanced mutagenesis can be induced by growth at the restrictive temperature. **(B)** The EvolvR system is composed of a CRISPR-guided nickase that nicks the target GOI, fused to ep Pol I that performs nick translation.

experiment, Camps *et al.* demonstrated that their system could be used to evolve enzymes with diverged function by generating TEM-1 β-lactamase mutants that were able to hydrolyze a third-generation lactam antibiotic, aztreonam.

The ep Pol I/ColE1 system has subsequently been applied in a handful of additional directed evolution experiments. For example, Koch *et al.* used the system to prepare a library of terminal alkane hydroxylases with the aim of evolving variants that can oxidize butane [34]. Although they only used the system for the preparation of mutant libraries (*i.e.* as a mutator strain) and not for continuous evolution involving serial passaging under prolonged selection conditions, they demonstrated that one can create large libraries of GOI variants directly *in vivo*. In another application, an M13 phagemid with a ColE1 origin was made to encode LuxR and infect *E. coli* harboring the ep Pol I [35]. LuxR is a transcriptional activator and drove the transcription of an antibiotic resistance gene (β-lactamase) controlled by the *lux* promoter in the *E. coli*. Through several cycles of infecting fresh *E. coli*, antibiotic selection, lysis of *E. coli*, and phage isolation, LuxR evolved a 17-fold higher binding affinity to the *lux* promoter sequence.

While the ep Pol I/ColE1 system approaches *in vivo* continuous evolution, it is limited by off-target mutagenesis and low durability. Because Pol I is responsible for Okazaki fragment mending throughout the genome and also participates in DNA repair [36], expressing an ep Pol I causes substantial mutagenesis genome-wide. Targeting of mutations to the GOI does occur – owing to the ColEI origin, the limited role of Pol I in lagging strand replication, and special growth conditions optimized to time ep Pol I action with growth phases where genome replication activity is low – but is maximally only ~400-fold. Therefore, when highly ep Pol Is are used, it is possible that off-target mutagenesis

will lower the fitness of the cell, causing fixation of suppressor mutations that abrogate the activity of ep Pol I. Still, the Pol I/ColE1 system represents a landmark development that encouraged the field to pursue new strategies for realizing *in vivo* continuous evolution.

Perhaps the closest conceptual descendant of the ep Pol I/ColE1 system is a new *E. coli* continuous evolution system called EvolvR, which uses CRISPR-guided ep DNA polymerases (DNAPs) to continuously target mutations to GOIs (**Figure 1.2B**). Rather than rely on the natural targeting of Pol I to ColE1, Halperin *et al.* [37] fused ep Pol I variants (and other DNAPs) to a nickase Cas9 (nCas9) that would serve two purposes. First, nCas9 would bring the ep Pol I to any GOI encoded on a plasmid or the genome using a guide RNA (gRNA). Second, the nCas9 would nick the target strand, creating a free 3'-OH substrate from which the ep Pol I could extend. Once nCas9 releases the nicked product, it is believed that ep Pol I then latches on and carries out error-prone extension from the nick. This highly clever idea was demonstrated in *E. coli* with a number of ep Pol I variants spanning different mutation rates and activities, as well as with a moderately ep Phi29 DNAP with high processivity. Using the most mutagenic ep Pol I, Halperin *et al.* measured a mutation rate approaching 10^{-2} s.p.b. (a 7.7 million-fold elevation compared to WT cells) at the first nucleotide 3' of the nCas9-induced nick. While this extreme mutation rate quickly dropped when moving away from the nick, other Pol I and Phi29 DNAP variants with moderate error rates could achieve mutagenesis windows up to 350 bp. With these characteristics and with the potential to use multiple gRNAs to simultaneously target multiple parts of a gene, EvolvR could readily and efficiently generate sequence diversity on a GOI *in vivo* to support continuous evolution. Indeed, in a proof of principle experiment, Halperin *et al.*

used EvolvR to rapidly evolve spectinomycin resistance by targeting mutagenesis to the *rpsE* gene and found new resistance mutations that were previously unknown.

Future studies and improvements on EvolvR will clarify how well it drives *in vivo* continuous evolution for prolonged periods of time, needed to traverse long mutational pathways. Durability may be difficult in the current architecture, because the mutation rate is maximal at nucleotides within the target region of the gRNA, which if mutated, will reduce the ability of the system to continue inducing mutagenesis. Since the GOI can still be replicated (by high-fidelity host systems) in the absence of EvolvR function, this may result in the fixation of partially adapted GOI mutants that stop mutating, leading to premature cessation of evolution. In addition, EvolvR still has off-target elevations in mutation rate, presumably because ep Pol I or Phi29 can participate in genomic replication and/or because Cas9 has off-target binding. Strategies that use more processive ep DNA polymerases with no activity in normal genome replication and alternative CRISPR systems that nick outside the critical regions for gRNA targeting may overcome potential issues of targeting and durability. This system should also readily transfer to cell-types other than *E. coli*. Therefore, EvolvR is a highly promising new system for *in vivo* continuous evolution with enormous potential, especially for the multiplexed evolution of genes in their native genomic contexts rather than on a plasmid.

1.4 Yeast systems that do not use engineered DNA polymerases for mutagenesis

The first demonstration of continuous targeted mutagenesis *in vivo* in yeast was published in 2013 under the name TaGTEAM (**Figure 1.3A**), which stands for targeting glycosylases to embedded arrays for mutagenesis [38]. In TaGTEAM, mutagenesis at the GOI is initiated by recruiting a DNA glycosylase, which normally functions as the first step in the base

excision repair (BER) pathway responsible for removing chemically altered DNA bases [39]. The authors adopted the yeast 3-methyladenine glycosylase, Mag1p, and fused it to the tet repressor (tetR) that binds a 19-bp operator sequence, tetO. By introducing a non-recombinogenic tetO array (with each tetO site separated by 10–30 bp of random sequence), the tetR-Mag1p fusion could be targeted to GOIs in the chromosome or plasmid. It is presumed that tetR-Mag1p targeting generates a build-up of unprocessed abasic sites at target loci, leading to replication fork stalling and recruitment of ep translesion polymerases [40]. This faulty repair can lead to both point mutations and frameshifts. To test their system for its ability to generate mutagenesis at a GOI, Finney-Manchester *et al.* introduced a 240X tetO array upstream of a *URA3* auxotrophic marker in a region of chromosome 1 that does not contain nearby essential genes. The distance between the tetO array and the marker was titrated to assess the size of the area subjected to mutagenesis. The presence of tetR-Mag1p resulted in a >800-fold increase in mutation rate spanning a 10 kb region. However, the off-target mutation rate was also increased 40-fold in the absence of the array, indicating genome-wide mutagenesis by tetR-Mag1p. No direct applications of the system have been published to date, but this mutagenic strategy was important for opening new avenues of thought in the field.

ICE (*in vivo* continuous evolution) is another notable example of continuous evolution in yeast (**Figure 1.3B**), introduced in 2016 [41]. ICE adopts a strategy for DNA diversification that is based on the mutagenic properties of the Ty1 retrotransposon element. A GOI is cloned into the Ty1 cassette, which then gets transcribed into an RNA. Next, the RNA is reverse transcribed to form cDNA and reintegrated into the chromosome [42]. The mutagenic properties of the system stem from Ty1's self-encoded reverse

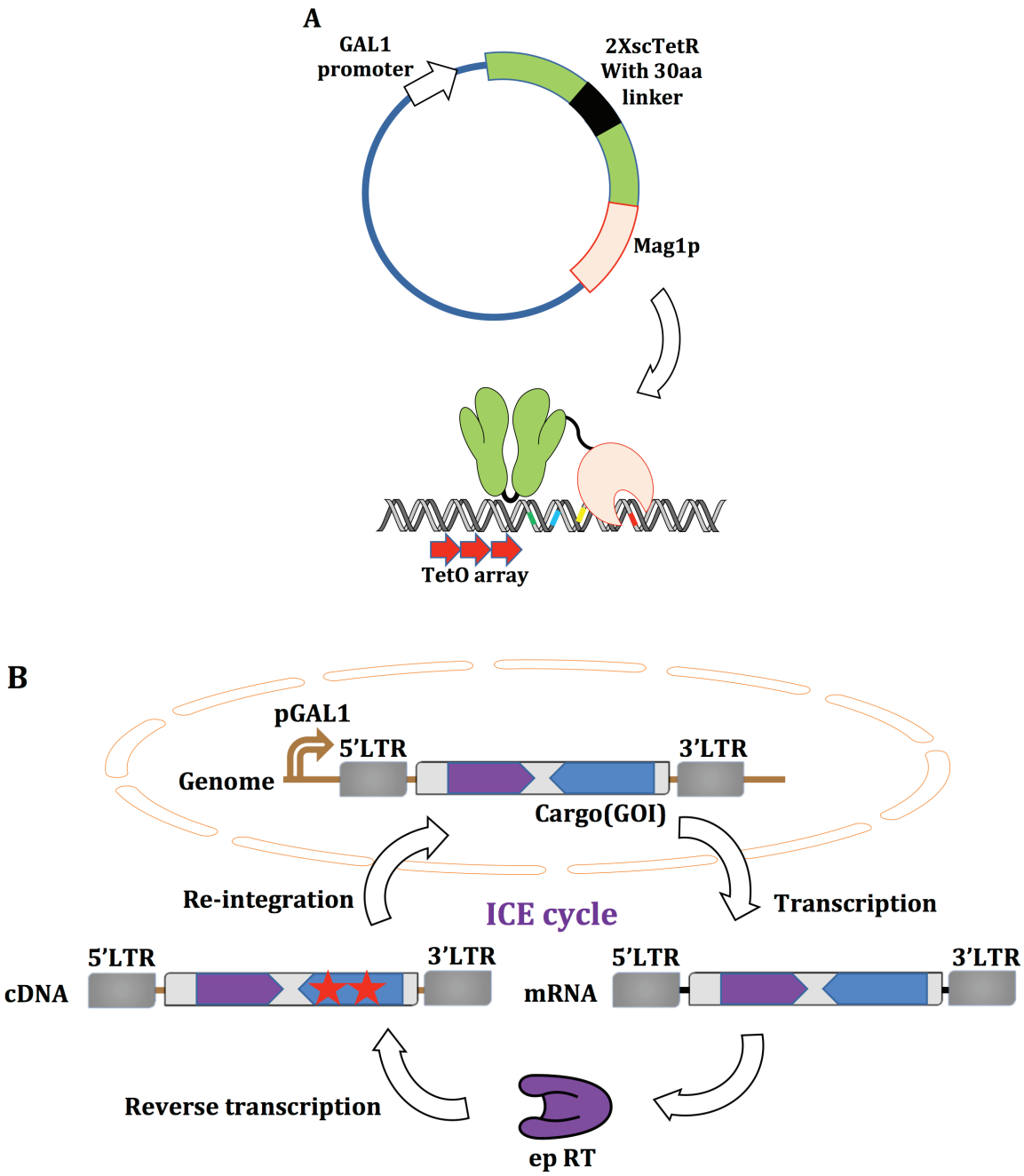


Figure 1.3 | Yeast systems that do not use engineered DNA polymerases for mutagenesis.
(A) TaGTEAM is achieved by fusing the yeast 3-methyladenine DNA glycosylase, Mag1p, to a tetR DNA-binding domain. Upon expression of the fusion from an inducible galactose promoter, the 20 kb region that is proximal to the tetO array experiences a high degree of mutagenesis. **(B)** In ICE, the GOI is cloned into an inducible Ty1 retrotransposon in the genome. The ICE cycle begins with inducible transcription of the retroelement followed by ep reverse transcription driven by Ty1's encoded rt. The cycle ends upon re-integration of the mutated cDNA into the genome.

transcriptase (rt), which introduces mutations at a rate of ~$2.5x10^{-5}$ to ~$1.5x10^{-4}$ per base per retrotransposition event [41, 43], thus allowing rapid mutagenesis of Ty1 and its embedded GOI. However, since mutagenesis depends on retrotransposition and the retrotransposition rate of Ty1 with a GOI inserted is low, the high mutation rate of Ty1's rt is only occasionally experienced on the GOI. Therefore, the authors carried out a series of experiments to increase the retrotransposition rate. By fine-tuning various parameters including the cargo's promoter strength, host genotype (*i.e.* deletions of certain host genes), cell density, temperature, initiator methionine tRNA expression (which acts to prime Ty1 replication), and inclusion of terminators, the authors were able to significantly increase retrotransposition rate. Altogether, the optimization process reached a mutation rate capable of generating up to $1.6x10^{7}$ distinct mutants of a GOI per round per liter cultured [41]. Crook *et al.* then used ICE in three independent experiments to test the system's ability to evolve genetic material. In the first demonstration, URA3 was evolved for increased resistance to 5-fluoroorotic acid (5-FOA); in the second example, the Spt15p global transcription regulator was evolved to confer a complex cellular phenotype of butanol resistance; and in the third example, a multigene pathway spanning 4.6 kb and containing two enzymes and a regulatory region was evolved for increased xylose catabolism. Additional experiments will clarify the extent to which ICE continuously mutates GOIs, since the ability for Ty1 elements to semi-randomly spread throughout the yeast genome [44, 45] could potentially complicate analysis, reduce mutational accumulation for the GOI, and diffuse the target of evolution. These issues could potentially be solved by somehow limiting Ty1 integration to a single location in the genome, turning the retrotransposon into a "retrocisposon," and then increasing the "retrocisposition" rate

to access high levels of diversification. In fact, the ability to achieve "retrocisposition" would also be important for reaching continuous evolution in other systems based on retroelement-mediated mutagenesis, such as a recently reported bacterial approach for *in vivo* genome editing and evolution [46]. Nevertheless, ICE is an important example of continuous evolution in yeast.

1.5 Somatic hypermutation as a means for targeted mutagenesis of GOIs

Some groups have harnessed one of nature's built-in mechanisms for generating targeted DNA diversity, somatic hypermutation (SHM). In SHM, B cells create point mutations in their immunoglobulins (Igs) to drive antibody affinity maturation [47]. The enzyme responsible for SHM is Activation Induced cytidine Deaminase (AID), which deaminates cytidine (C) to generate uridine (U). This triggers various mismatch repair mechanisms resulting in a mutation rate of $\sim 10^{-3}$ s.p.b. at Ig loci [48]. Several researchers have successfully hijacked this natural mechanism for diversifying and evolving non-antibody proteins. In 2001, Bachl *et al.* set the stage for SHM-based protein directed evolution [49]. They demonstrated a high rate of reversion of a premature stop codon in a GFP cloned into a hypermutator B cell line (18-81) that expresses endogenous AID. They concluded that elevated reversion rates depended on AID and were rate limited by transcriptional levels of the target gene, in agreement with previous findings on SHM mechanisms [50, 51]. In 2004, Wang *et al.* applied SHM to the directed evolution of an entire open reading frame [52] by integrating a single copy of red fluorescent protein (RFP) into Ramos cells, which express endogenous AID, using a lentivirus. Through iterative SHM and FACS (fluorescence-activated cell sorting), RFP mutants with enhanced photostability and far-red emissions were evolved. The study was conducted in the pre-CRISPR era, and thus the RFP GOI was

not targeted to an Ig locus but was rather integrated at various genomic locations within their cell population. However, the authors noted that the most evolved RFP variant, which they called mPlum, was located in the Ig heavy chain locus of chromosome 14, indicating that there is indeed a target locus where mutagenesis rates are highest, and that SHM is responsible for high levels of mutagenesis at the GOI. Yet it is expected that this targeting is incomplete, as mutation rates readily occur outside the Ig domain in cell lines that express endogenous AID [53].

Recently, a major development that avoids the use of hypermutator cell lines that express endogenous AID to mutate GOIs was independently published by two groups (**Figure 1.4**) [54, 55]. Hess *et al*. linked AID to a catalytically inactive dCas9 using MS2-modified sgRNAs, which achieved precise targeting of SHM to defined loci in HEK293 cells [54]. The system, which they called CRISPRx, allowed targeted mutagenesis of multiple genomic locations simultaneously. Their reported mutation rate was $\sim 5 \times 10^{-4}$ s.b.p., which is similar to that observed for somatic hypermutation [48]. In their first application, Hess *et al.* evolved GFP (excitation, 395 nm; emission, 509 nm) into EGFP (490/509 nm) by selecting for spectrum-shifted variants. Later, they mutated the target of the cancer therapeutic bortezomib, PSMB5, and identified known and novel mutations that confer bortezomib resistance. At the same time, Ma *et al*. developed a dCas9-AID fusion and targeted BCR-ABL for mutagenesis to efficiently identify known and new mutations conferring imatinib resistance mutations in chronic myeloid leukemia cells [55].

In both of these CRISPR-guided AID strategies, induction of mutagenesis at the GOI was followed directly by a single round of enrichment for the selected phenotype. Therefore, these studies do not directly demonstrate continuous evolution. However, multi-generation

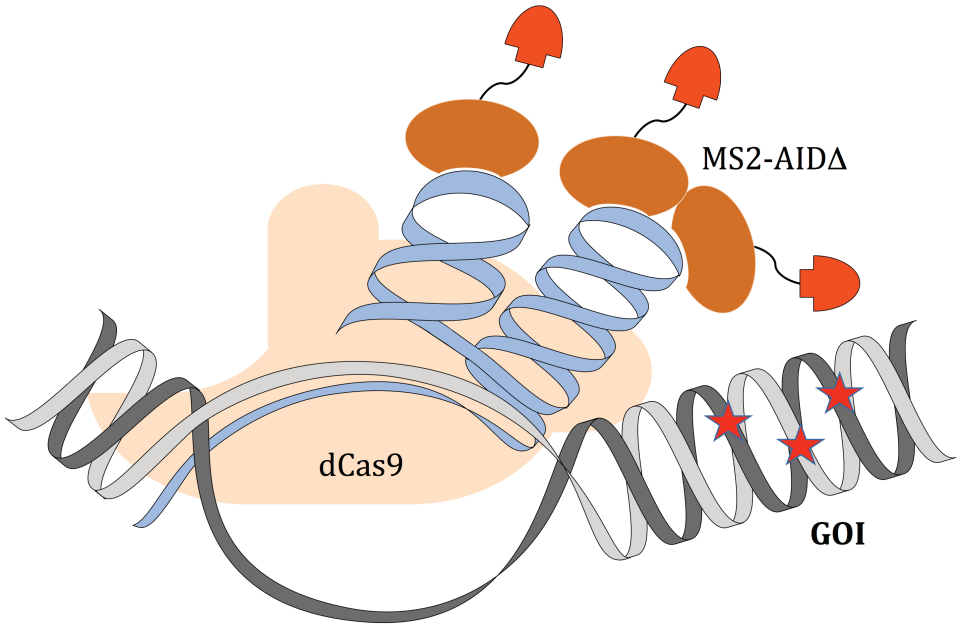


Figure 1.4 | Somatic hypermutation as a means for targeted mutagenesis of GOIs.
A region of ~100 bp is targeted for mutagenesis by a dCas9 complexed with sgRNA-MS2 to a hyperactive cytidine deaminase (AID).

continuous directed evolution could be carried out using cell lines stably transcribing sgRNAs that tile the GOI. Although Hess *et al.* observed some limited off-target mutagenesis, owing both to off-target activity of AID and off-target binding of sgRNAs [54], the durability of this system is predicted to be reasonably high, as the positions that are most prone to mutagenesis are outside of the spacer and protospacer adjacent motif (PAM) needed for sgRNA binding.

Another strategy for targeting AID to GOIs is based on fusing AID to T7 RNA polymerase (RNAP) [56]. The main advantage is that T7 RNAP induction could be precisely controlled in *E. coli*, and although not demonstrated by More *et al.*, could be largely transferred between various organisms. Mutations accumulate during induction of transcription, as the T7 RNAP carries AID over large stretches of DNA. Indeed, due to its high processivity, T7 RNAP can direct mutagenesis over several kb.

1.6 Phage Assisted Continuous Evolution (PACE)

The most successful method for continuous protein evolution thus far is the phage assisted continuous evolution (PACE) system developed in the lab of David Liu (**Figure 1.5**) [2, 12, 14, 15, 57-63]. PACE reimagines traditional 'rounds' of directed evolution as generations of the M13 bacteriophage life-cycle, thereby transforming a step-wise and labor-intensive procedure into a continuous biological process. In PACE, GOIs are encoded in the M13 genome, and the resulting phage continuously replicate in a vessel (termed 'lagoon') that experiences a constant influx of *E. coli* cells. To create a selection pressure for GOIs to evolve, the activity of interest is coupled to phage survival. This is achieved by deleting the essential gene III (gIII), encoding coat protein III (pIII), from the M13 genome. The host *E. coli* strain is engineered to encode gIII in a genetic circuit that makes pIII expression dose-

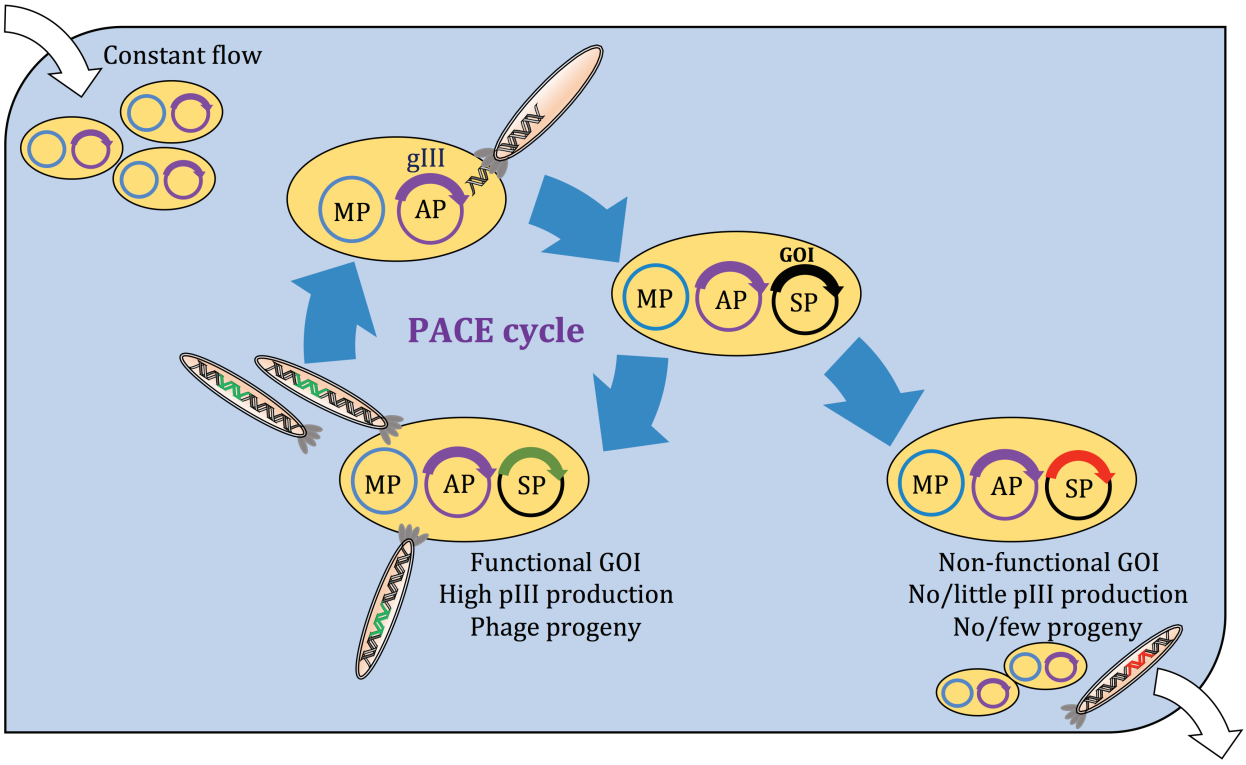


Figure 1.5 | Phage assisted continuous evolution (PACE).
Phage carrying the selection plasmid (SP) encoding the GOI propagate on *E. coli* cells which are constantly flowing into the 'lagoon' at a rate that does not permit their propagation but is longer than the phage life cycle, thus permitting phage replication. Upon infection, the SP (as well as the bacterial genome) experiences a high degree of mutagenesis due to the presence of a mutator plasmid (MP). In a PACE experiment, high GOI activity (green) is linked to drive strong gIII expression, resulting in progeny that can then infect incoming *E. coli*. No GOI activity (or a weak one, red) results in poor progeny production, becoming washed away from the lagoon at a larger rate (alongside bacterial cells). The system is designed to run for hundreds of generations without human intervention and result in the evolution of the GOI towards the desired activity.

dependent on the desired activity of the GOI (see below for examples); so only phage that successfully evolve the GOI can trigger pIII expression and continue propagating. Due to the rapid generation time of M13 (~10 minutes without selection), evolution in this manner can iterate hundreds of times in just a few days.

A key parameter in PACE is the *E. coli* flow-rate, which should exceed their doubling time but be slower than the phage life cycle, allowing only phage to replicate in the lagoon (on average). Consequently, only phage accumulate mutations whereas *E. coli* are physically prevented from doing so. High rates of mutation on the phage (and *E. coli*) genome is driven by a mutator plasmid (MP) that is carried by the *E. coli* cells and induced in the lagoon for error-prone M13 replication. The latest version of the MP is able to drive potent mutagenesis at $>10^{-3}$ s.p.b. by combining the effects of six different mutagenesis drivers [64].

Esvelt *et al.* first demonstrated proof of concept by evolving T7 RNA polymerase (RNAP) to initiate transcription from new promoter sequences [59]. pIII expression was bottlenecked at the level of transcription by encoding promoter sequences unrecognized by WT T7 RNAP (or any *E. coli* RNAPs), thus driving the selection to favor T7 RNAP variants that are able to efficiently recognize the new promoters. After 8 days and 200 'rounds' of PACE, new T7 RNAPs emerged that could transcribe from the distant T3 RNAP promoter as efficiently as WT T7 RNAP does from its cognate promoter [59]. Similarly, T7 RNAP variants that efficiently initiate transcription with ATP or CTP, instead of GTP were evolved. Since that landmark study, the ability to couple T7 RNAP activity to PACE has been exploited in a number of ways, ranging from basic adaptation studies to selections for split T7 RNAP [14, 15, 61-64].

In principle, PACE is applicable for the evolution of any biomolecular function that can be linked to pIII expression; and in just a few years since its inception, this has been realized in a wide range of applications beyond RNAP evolution. A notable example is the evolution of new DNA binding domains. Hubbard *et al.* employed the classic one-hybrid selection with PACE to evolve TALENs (transcription activator-like effector nucleases) with broadly improved DNA cleavage specificity [60]. Although TALENs are highly promising for gene editing, their major limitation is that they require the 5′ nucleotide of the target sequence to be T [65]. New TALEs (TALENs without the fused nuclease) were evolved with PACE by fusing the DNA binding domain of the canonical CBX8-targeting TALE to the ω subunit of *E. coli* RNAP. The PACE system was designed to include the TALE target sequence upstream of gIII. TALEs that successfully bind the target DNA recruit holoenzyme RNAP around the ω subunit, resulting in subsequent pIII expression. With this TALE selection, the identity of the target sequence can be custom-tailored, in this case, to encode noncanonical 5' nucleotides. After using an additional negative selection (see below) that inhibited variants with promiscuous substrate specificity, Hubbard *et al.* were able to evolve TALE variants that displayed 2- to 4-fold increases in specificity for 5' A, 5' C, or 5' G vs 5' T, relative to WT TALE.

The one-hybrid PACE format was also used for overcoming one of Cas9's main limitations, restricted PAM (protospacer adjacent motif) compatibility. This time, Hu *et al.* fused a catalytically dead variant of *Streptococcus pyogenes* Cas9 (dCas9) to the ω subunit of *E. coli* RNAP [66]. Then, the authors cleverly fed the lagoon with a mixture of host *E. coli* cells bearing a library of target sequences that covers all 64 possible PAM sequences, to select for broadened PAM compatibility. After PACE, several variants were isolated that

could efficiently recognize NG, GAA and GAT as PAMs. Upon restoration of nuclease catalytic activity to these evolved dCas9 variants, the authors remarkably found that one of them, xCas9, exhibited greater DNA specificity than WT Cas9, even with its newly-gained broad PAM compatibility. This result challenges the widely-held assumption that there must be a trade-off between editing specificity and PAM compatibility, and suggests that Cas9 can be improved through laboratory evolution to meet the most demanding challenges of CRISPR-Cas9 applications.

Another important form of PACE is its use with two-hybrid selection for the evolution of high-affinity protein-binders [57]. In the bacterial two-hybrid system, the ω subunit of *E. coli* RNAP is fused to a protein of interest, which is recruited to DNA through its interaction with a target protein. This target protein is fused to a DNA binding domain that localizes the complex at its cognate sequence encoded upstream of a reporter gene. If the protein of interest binds the target protein, then the RNAP holoenzyme can reconstitute around the ω subunit and drive expression of the downstream reporter. Badran *et al.* adapted this system for PACE using gIII as the reporter. After extensive optimization, Badran *et al.* were able to use this PACE format to evolve the insecticidal protein, *Bacillus thuringiensis* δ-endotoxin (Bt toxin) Cry1Ac, to bind and inhibit a new receptor in the gut of the insect pest *Trichoplusia ni* (TnCAD) [57]. Although WT Cry1Ac did not detectably bind TnCad, the evolved variants were able to bind with nM affinity. Significantly, this strategy could overcome widespread Bt toxin resistance, which primarily occurs through mutational changes that inhibit binding to the native receptor of WT Cry1Ac. Badran *et al.* demonstrated this by showing that evolved Cry1Ac is highly potent at killing *T. ni* that are resistant to WT Cry1Ac. An exciting possibility for the future would be to evolve TnCAD to

resist the new Cry1Ac variant, and then iterate this cycle in a study of molecular co-evolution.

Additional positive selections developed for PACE have enabled evolution of proteases that are drug resistant [58] or have altered substrate specificities [67], aminoacyl-tRNA synthetases (aaRSs) that can accept noncanonical amino acids [68], and protein variants with improved soluble expression [69]. Negative selections are also compatible with PACE, and are useful in cases where it is desirable to evolve high specificity towards the target substrate and restrict promiscuity towards others (especially the native substrate). This can be achieved by introducing a dominant negative allele of pIII, pIII-neg, that inhibits phage propagation [14]. The expression of pIII-neg can then be linked to the unwanted activity (*e.g.* recognition of the T7 promoter by T7 RNAP) for negative selection. (This strategy was successfully employed during TALEN and aaRS evolution.) Selection stringency and mutation rate are also important determinants of PACE outcomes and can be titrated [14, 61]. Lastly, we note that the Isalan lab developed a system related to PACE that accommodates the evolution of multiple genes, starting from combinatorial libraries. With this system, they were able to evolve a panel of orthogonal dual promoter-transcription factor pairs that were used to make multi-input logic gates [70, 71].

Clearly, PACE is a powerful method for continuous protein evolution, but it is not an entirely *in vivo* system. Rather, M13 serves as a biological carrier of the GOI from one *E. coli* host cell to the next, with a given cell serving as a host of error-prone replication just once (on average). This ingenious design circumvents the challenges of *in vivo* mutational targeting. Since mutagenesis is induced in the lagoon, where *E. coli* briefly reside without doubling, mutation rates can be elevated entirely through untargeted mechanisms (and

temporarily induced to be as high as desired), without consideration for replication of the *E. coli* genome. Even if *E. coli* cells stochastically replicate in the lagoon and become a source of cheater mutations (*e.g.* constitutive gIII expression), the flow-rate ensures that any progeny are quickly diluted out. What's left in the lagoon is a population of M13 that selectively undergoes error-prone replication. In effect, targeting of mutations to the phage genome containing the GOI is complete, as the host *E. coli* is constantly replaced.

PACE also achieves durable mutagenesis by enforcing continuity. Replication of GOIs is intrinsically coupled to mutagenesis, through error-prone replication of the M13 genome. Any phage that escapes mutagenesis through a mutation in the phage genome's origin of replication, for example, must do so at the expense of being replicated. Only variants that continue to accumulate mutations can survive and propagate. And since *E. coli* cells do not persist long enough in the lagoon to evolve, the mutation rate experience by phage remains unchanged. The durability of PACE is best evidenced by the long mutational trajectories traversed during evolution experiments, which have yielded protein variants with up to 16 mutations [57].

Because PACE is not entirely *in vivo*, it suffers two major limitations. First, it requires continuous propagation of phage in a population of freshly diluted *E. coli* cells, which has only been achieved thus far with a chemostat or turbidostat setup. This greatly limits the throughput of PACE experiments, typically to fewer than ten replicates, and impedes its widespread use in the community. Second, PACE is restricted to selections that are linked to phage propagation. This precludes selections for *in vivo* phenotypes like tolerance or metabolism, as well as cell-based selections like FACS or droplet sorting. These limitations

notwithstanding, PACE is currently the gold standard among continuous evolution systems and serves as model for the field.

1.7 Orthogonal DNA replication (OrthoRep)

In the following chapters, I describe a new system for *in vivo* continuous directed evolution, termed OrthoRep, based on orthogonal DNA replication (**Figure 1.6**) [72, 73]. Fundamentally, OrthoRep can be described as a cell harboring a synthetic DNA replication system that propagates without affecting endogenous replication of the host genome. Specifically, this can be implemented in the form of an orthogonal DNAP/plasmid pair, where orthogonality means that the DNAP is dedicated to the cognate plasmid and does not participate in genomic replication (unlike Pol I in the Pol I/ColE1 systems). This allows us, broadly, to engineer DNA replication *in vivo* for user-defined purposes without harming the host. For the purpose of *in vivo* continuous directed evolution, the orthogonal DNAP can be made as error-prone as desired, since the genome is completely spared from mutation. Then, GOIs can simply be encoded on the orthogonal plasmid and rapidly and continuously mutated by the orthogonal ep DNAP during evolution.

OrthoRep is uniquely positioned to address the long-term technological challenges associated with *in vivo* continuous directed evolution. Although many of the approaches described above are capable of elevating mutagenesis of GOIs, they often introduce mutations into genomic genes as well; but the level of targeting in OrthoRep, in theory, should be complete. As detailed in Chapter 3, the OrthoRep system I have built achieves *at least* ~100,000-fold targeting, which is currently unmatched. Furthermore, continuous mutagenesis in OrthoRep should be extraordinarily durable. Because replication of GOIs occurs exclusively through the action of the ep DNAP, inheritance of GOIs is intrinsically

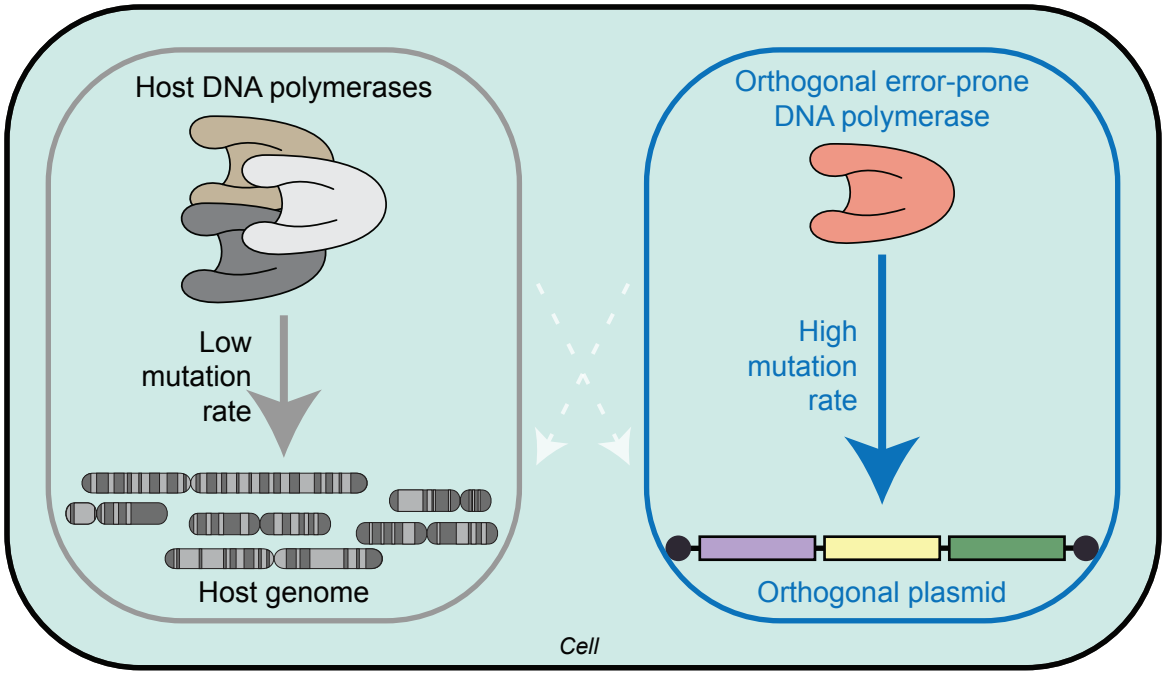


Figure 1.6 | Orthogonal DNA replication (OrthoRep).
In OrthoRep, GOIs encoded on the orthogonal plasmid are replicated by the orthogonal ep DNAP. The genome is fully spared from mutation by the orthogonal ep DNAP.

coupled to mutagenesis. Put in more specific terms, cells may acquire a disabling mutation in the plasmid's origin of replication that ceases its mutagenesis, but these mutant plasmids will no longer get replicated, and are immediately removed from the population. In addition, as explained in Chapter 1, the origin of replication I employed in OrthoRep is mostly proteinaceous and cannot be mutated by the orthogonal DNAP, so the chance of a disabling mutation at the origin is low to begin with. In short, continuity of mutagenesis is enforced in the orthogonal DNA replication architecture. This durability will become increasingly important as the field makes headway towards more and more difficult protein functions that require many mutations to access. Such problems include the *de novo* evolution of enzymes, evolution of protein-protein interactions, and evolution of high-affinity therapeutic antibodies against difficult targets, all of which are at the forefront of the protein engineering field.

1.8 References

[1] Turner N.J. (2009) Directed evolution drives the next generation of biocatalysts. *Nature Chemical Biology*, **5** (8), 567-573.

[2] Packer M.S., Liu D.R. (2015) Methods for the directed evolution of proteins. *Nature Reviews Genetics*, **16** (7), nrg3927.

[3] Cobb R.E., Chao R., Zhao H. (2013) Directed evolution: Past, present, and future. *AICHE Journal*, **59** (5), 1432-1440.

[4] Bloom J.D., Arnold F.H. (2009) In the light of directed evolution: Pathways of adaptive protein evolution. *Proceedings of the National Academy of Sciences*, **106** (Supplement 1), 9995-10000.

[5] Wellner A., Gurevich M.R., Tawfik D.S. (2013) Mechanisms of Protein Sequence Divergence and Incompatibility. *PLoS Genetics*, **9** (7), e1003665.

[6] Soskine M., Tawfik D.S. (2010) Mutational effects and the evolution of new protein functions. *Nature Reviews Genetics*, **11** (8), 572-582.

[7] Jespers L.S., Roberts A., Mahler S.M., Winter G., Hoogenboom H.R. (1994) Guiding the Selection of Human Antibodies from Phage Display Repertoires to a Single Epitope of an Antigen. *Nature Biotechnology*, **12** (9), 899-903.

[8] Coelho P.S., Brustad E.M., Kannan A., Arnold F.H. (2013) Olefin Cyclopropanation via Carbene Transfer Catalyzed by Engineered Cytochrome P450 Enzymes. *Science*, **339** (6117), 307-310.

[9] McIsaac R.S., Engqvist M.K.M., Wannier T., Rosenthal A.Z., Herwig L., Flytzanis N.C., et al. (2014) Directed evolution of a far-red fluorescent rhodopsin. *Proceedings of the National Academy of Sciences*, **111** (36), 13034-13039.

[10] Goldsmith M., Eckstein S., Ashani Y., Greisen P., Leader H., Sussman J.L., et al. (2016) Catalytic efficiencies of directly evolved phosphotriesterase variants with structurally different organophosphorus compounds in vitro. *Archives of Toxicology*, **90** (11), 2711-2724.

[11] Joyce G.F. (2004) Directed evolution of nucleic acid enzymes. *Annual Review of Biochemistry*, **73** (1), 791-836.

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

[13] d'Oelsnitz S., Ellington A. (2018) Continuous directed evolution for strain and protein engineering. *Current Opinion in Biotechnology*, **53**, 158-163.

[14] Carlson J.C., Badran A.H., Guggiana-Nilo D.A., Liu D.R. (2014) Negative selection and stringency modulation in phage-assisted continuous evolution. *Nature Chemical Biology*, **10** (3), 216-222.

[15] Dickinson B.C., Leconte A.M., Allen B., Esvelt K.M., Liu D.R. (2013) Experimental interrogation of the path dependence and stochasticity of protein evolution using phage-assisted continuous evolution. *Proceedings of the National Academy of Sciences*, **110** (22), 9007-9012.

[16] Wong B.G., Mancuso C.P., Kiriakov S., Bashor C.J., Khalil A.S. (2018) Precise, automated control of conditions for high-throughput growth of yeast and bacteria with eVOLVER. *Nature Biotechnology*, **36** (7), 614.

[17] Drake J.W. (1991) A constant rate of spontaneous mutation in DNA-based microbes. *Proceedings of the National Academy of Sciences*, **88** (16), 7160-7164.

[18] Herr A.J., Ogawa M., Lawrence N.A., Williams L.N., Eggington J.M., Singh M., et al. (2011) Mutator Suppression and Escape from Replication Error–Induced Extinction in Yeast. *PLoS Genetics*, **7** (10), e1002282.

[19] Wilke C.O., Wang J.L., Ofria C., Lenski R.E., Adami C. (2001) Evolution of digital organisms at high mutation rates leads to survival of the flattest. *Nature*, **412** (6844), 331.

[20] Nowak M., Schuster P. (1989) Error thresholds of replication in finite populations mutation frequencies and the onset of muller's ratchet. *Journal of Theoretical Biology*, **137** (4), 375-395.

[21] (2006) Quasispecies: Concept and Implications for Virology. **299**.

[22] Giraud A., Matic I., Tenaillon O., Clara A., Radman M., Fons M., et al. (2001) Costs and Benefits of High Mutation Rates: Adaptive Evolution of Bacteria in the Mouse Gut. *Science*, **291** (5513), 2606-2608.

[23] Notley-McRobb L., Seeto S., Ferenci T. (2002) Enrichment and elimination of mutY mutators in Escherichia coli populations. *Genetics*, **162** (3), 1055-1062.

[24] Wang H.H., Isaacs F.J., Carr P.A., Sun Z.Z., Xu G., Forest C.R., et al. (2009) Programming cells by multiplex genome engineering and accelerated evolution. *Nature*, **460** (7257), 894.

[25] Liang L., Liu R., Garst A.D., Lee T., Nogué V.S.i., Beckham G.T., et al. (2017) CRISPR EnAbled Trackable genome Engineering for isopropanol production in Escherichia coli. *Metabolic Engineering*, **41**, 1-10.

[26] Nyerges Á., Csörgő B., Draskovits G., Kintses B., Szili P., Ferenc G., et al. (2018) Directed evolution of multiple genomic loci allows the prediction of antibiotic resistance. *Proceedings of the National Academy of Sciences*, **115** (25), 201801646.

[27] Abil Z., Ellefson J.W., Gollihar J.D., Watkins E., Ellington A.D. (2017) Compartmentalized partnered replication for the directed evolution of genetic parts and circuits. *Nature Protocols*, **12** (12), nprot.2017.2119.

[28] Fabret C., Poncet S., Danielsen S., Borchert T.V., Ehrlich S.D., Jannière L. (2000) Efficient gene targeted random mutagenesis in genetically stable Escherichia coli strains. *Nucleic Acids Research*, **28** (21), e95-e95.

[29] Troll C., Yoder J., Alexander D., Hernández J., Loh Y., Camps M. (2014) The mutagenic footprint of low-fidelity Pol I ColE1 plasmid replication in E. coli reveals an extensive interplay between Pol I and Pol III. *Current Genetics*, **60** (3), 123-134.

[30] Alexander D.L., Lilly J., Hernandez J., Romsdahl J., Troll C.J., Camps M. (2014) Methods in Molecular Biology. **1179**, 31-44.

[31] Camps M., Naukkarinen J., Johnson B.P., Loeb L.A. (2003) Targeted gene evolution in Escherichia coli using a highly error-prone DNA polymerase I. *Proceedings of the National Academy of Sciences*, **100** (17), 9727-9732.

[32] Alexander D.L., Lilly J., Hernandez J., Romsdahl J., Troll C.J., Camps M. (2014) Methods in Molecular Biology. *Methods in molecular biology (Clifton, NJ)*, **1179**, 31-44.

[33] Shinkai A., Loeb L.A. (2001) In vivo mutagenesis by Escherichia coli DNA polymerase I. Ile(709) in motif A functions in base selection. *Journal of Biological Chemistry*, **276** (50), 46759-46764.

[34] Koch D.J., Chen M.M., Beilen J.B.v., Arnold F.H. (2009) In Vivo Evolution of Butane Oxidation by Terminal Alkane Hydroxylases AlkB and CYP153A6. *Applied and Environmental Microbiology*, **75** (2), 337-344.

[35] Na D., Lee S., Yi G.-S., Lee D. (2011) Synthetic inter-species cooperation of host and virus for targeted genetic evolution. *Journal of Biotechnology*, **153** (1-2), 35-41.

[36] Allen J.M., Simcha D.M., Ericson N.G., Alexander D.L., Marquette J.T., Biber B.P.V., et al. (2011) Roles of DNA polymerase I in leading and lagging-strand replication defined by a high-resolution mutation footprint of ColE1 plasmid replication. *Nucleic Acids Research*, **39** (16), 7020-7033.

[37] Halperin S.O., Tou C.J., Wong E.B., Modavi C., Schaffer D.V., Dueber J.E. (2018) CRISPR-guided DNA polymerases enable diversification of all nucleotides in a tunable window. *Nature*, **560** (7717), 248-252.

[38] Finney-Manchester S.P., Maheshri N. (2013) Harnessing mutagenic homologous recombination for targeted mutagenesis in vivo by TaGTEAM. *Nucleic Acids Research*, **41** (9), e99-e99.

[39] Nilsen H., Krokan H.E. (2001) Base excision repair in a network of defence and tolerance. *Carcinogenesis*, **22** (7), 987-998.

[40] Boiteux S., Guillet M. (2004) Abasic sites in DNA: repair and biological consequences in Saccharomyces cerevisiae. *DNA Repair*, **3** (1), 1-12.

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

[42] Wilhelm F.X., Wilhelm M., Gabriel A. (2005) Reverse transcriptase and integrase of the Saccharomyces cerevisiae Ty1 element. *Cytogenetic and Genome Research*, **110** (1-4), 269-287.

[43] Boutabout M., Wilhelm M., Wilhelm F.-X. (2001) DNA synthesis fidelity by the reverse transcriptase of the yeast retrotransposon Ty1. *Nucleic Acids Research*, **29** (11), 2217-2222.

[44] Carr M., Bensasson D., Bergman C.M. (2012) Evolutionary Genomics of Transposable Elements in Saccharomyces cerevisiae. *PloS One*, **7** (11), e50978.

[45] Devine S.E., Boeke J.D. (1996) Integration of the yeast retrotransposon Ty1 is targeted to regions upstream of genes transcribed by RNA polymerase III. *Genes & Development*, **10** (5), 620-633.

[46] Simon A.J., Morrow B.R., Ellington A.D. (2018) Retroelement-based genome editing and evolution. *ACS Synth Biol*.

[47] Wu X., Feng J., Komori A., Kim E.C., Zan H., Casali P. (2003) Immunoglobulin Somatic Hypermutation: Double-Strand DNA Breaks, AID and Error-Prone DNA Repair. *Journal of Clinical Immunology*, **23** (4), 235-246.

[48] Rajewsky K., Forster I., Cumano A. (1987) Evolutionary and somatic selection of the antibody repertoire in the mouse. *Science*, **238** (4830), 1088-1094.

[49] Bachl J., Carlson C., Gray-Schopfer V., Dessing M., Olsson C. (2001) Increased Transcription Levels Induce Higher Mutation Rates in a Hypermutating Cell Line. *The Journal of Immunology*, **166** (8), 5051-5057.

[50] Chaudhuri J., Tian M., Khuong C., Chua K., Pinaud E., Alt F.W. (2003) Transcription-targeted DNA deamination by the AID antibody diversification enzyme. *Nature*, **422** (6933), 726-730.

[51] Yu K., Huang F.-T., Lieber M.R. (2004) DNA Substrate Length and Surrounding Sequence Affect the Activation-induced Deaminase Activity at Cytidine. *Journal of Biological Chemistry*, **279** (8), 6496-6500.

[52] Wang L., Jackson W.C., Steinbach P.A., Tsien R.Y. (2004) Evolution of new nonantibody proteins via iterative somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America*, **101** (48), 16745-16749.

[53] Wang C.L., Harper R.A., Wabl M. (2004) Genome-wide somatic hypermutation. *Proceedings of the National Academy of Sciences of the United States of America*, **101** (19), 7352-7356.

[54] Hess G.T., Frésard L., Han K., Lee C.H., Li A., Cimprich K.A., et al. (2016) Directed evolution using dCas9-targeted somatic hypermutation in mammalian cells. *Nature Methods*, **13** (12), 1036-1042.

[55] Ma Y., Zhang J., Yin W., Zhang Z., Song Y., Chang X. (2016) Targeted AID-mediated mutagenesis (TAM) enables efficient genomic diversification in mammalian cells. *Nature Methods*, **13** (12), 1029-1035.

[56] Moore C.L., Papa L.J., 3rd, Shoulders M.D. (2018) A Processive Protein Chimera Introduces Mutations across Defined DNA Regions In Vivo. *Journal of the American Chemical Society*.

[57] Badran A.H., Guzov V.M., Huai Q., Kemp M.M., Vishwanath P., Kain W., et al. (2016) Continuous evolution of Bacillus thuringiensis toxins overcomes insect resistance. *Nature*, **533** (7601), 58.

[58] Dickinson B.C., Packer M.S., Badran A.H., Liu D.R. (2014) A system for the continuous directed evolution of proteases rapidly reveals drug-resistance mutations. *Nature Communications*, **5** (1), 5352.

[59] Esvelt K.M., Carlson J.C., Liu D.R. (2011) A system for the continuous directed evolution of biomolecules. *Nature*, **472** (7344), 499.

[60] Hubbard B.P., Badran A.H., Zuris J.A., Guilinger J.P., Davis K.M., Chen L., et al. (2015) Continuous directed evolution of DNA-binding proteins to improve TALEN specificity. *Nature Methods*, **12** (10), 939-942.

[61] Leconte A.M., Dickinson B.C., Yang D.D., Chen I.A., Allen B., Liu D.R. (2013) A Population-Based Experimental Model for Protein Evolution: Effects of Mutation Rate and Selection Stringency on Evolutionary Outcomes. *Biochemistry*, **52** (8), 1490-1499.

[62] Pu J., Zinkus-Boltz J., Dickinson B.C. (2017) Evolution of a split RNA polymerase as a versatile biosensor platform. *Nature Chemical Biology*, **13** (4), 432-438.

[63] Pu J., Kentala K., Dickinson B.C. (2017) Multidimensional control of Cas9 by evolved RNA polymerase-based biosensors. *ACS Chemical Biology*, **13** (2), 431-437.

[64] Badran A.H., Liu D.R. (2015) Development of potent in vivo mutagenesis plasmids with broad mutational spectra. *Nature Communications*, **6** (1), 8425.

[65] Deng D., Yan C., Pan X., Mahfouz M., Wang J., Zhu J.-K., et al. (2012) Structural Basis for Sequence-Specific Recognition of DNA by TAL Effectors. *Science*, **335** (6069), 720-723.

[66] Hu J.H., Miller S.M., Geurts M.H., Tang W., Chen L., Sun N., et al. (2018) Evolved Cas9 variants with broad PAM compatibility and high DNA specificity. *Nature*, **556** (7699), 57.

[67] Packer M.S., Rees H.A., Liu D.R. (2017) Phage-assisted continuous evolution of proteases with altered substrate specificity. *Nature Communications*, **8** (1), 956.

[68] Bryson D.I., Fan C., Guo L.-T., Miller C., Söll D., Liu D.R. (2017) Continuous directed evolution of aminoacyl-tRNA synthetases. *Nature Chemical Biology*, **13** (12), 1253.

[69] Wang T., Badran A.H., Huang T.P., Liu D.R. (2018) Continuous directed evolution of proteins with improved soluble expression. *Nature Chemical Biology*.

[70] Brödel A.K., Isalan M., Jaramillo A. (2018) Engineering of biomolecules by bacteriophage directed evolution. *Current Opinion in Biotechnology*, **51**, 32-38.

[71] Brödel A.K., Jaramillo A., Isalan M. (2016) Engineering orthogonal dual transcription factors for multi-input synthetic promoters. *Nature Communications*, **7**, ncomms13858.

[72] Ravikumar A., Arrieta A., Liu C.C. (2014) An orthogonal DNA replication system in yeast. *Nature Chemical Biology*, **10** (3), 175-177.

[73] Ravikumar A., Arzumanyan G.A., Obadi M.K.A., Javanpour A.A., Liu C.C. (2018) Scalable continuous evolution of genes at mutation rates above genomic error thresholds. *Cell*, in press.

Chapter 2

Establishing an orthogonal DNA replication system (OrthoRep)

2.1 The p1/2 plasmid system

To establish OrthoRep, we chose to exploit the pGKL1/2 (or p1/2) cytoplasmic plasmid system of *Kluveromyces lactis* (**Figure 2.1**). p1 (~8.9 kb) and p2 (~13.5 kb) are both linear, high-copy, double-stranded DNA plasmids that replicate autonomously, via protein-primed replication, in the cytoplasm of certain *K. lactis* and *S. cerevisiae* strains [1, 2]. The 5' ends of p1 and p2 are covalently linked to terminal proteins (TPs), which serve as replication origins in TP-primed DNA amplification (**Figure 2.2**) [3-5]. All components required for replication and transcription are encoded on p2, except for a second DNA polymerase, TP-DNAP1, encoded on p1 (**Figure 2.1**). TP-DNAP1 replicates p1. Due to the mechanistic nuances of TP-primed DNA replication (**Figure 2.2**), and the physical separation between cytoplasmic p1/2 DNA and nuclear genomic DNA, we expected no productive interactions between TP-DNAPs and nuclear DNA in yeast. Therefore, we chose to repurpose the p1-TP-DNAP1 plasmid-polymerase pair for OrthoRep.

OrthoRep must be able to encode and express user-selected genes of interest. Prior work has demonstrated integration of genes into p1 via homologous recombination [6]. To manipulate p1, we designed various integration cassettes containing cloning sites flanked by regions of homology to p1 (**Figure 2.3**). Using these cassettes, DNA encoding several markers (*e.g. LEU2*, *URA3*, and the fluorescent protein *mKate*) were integrated into p1 in *S. cerevisiae*. The resulting cells displayed the expected phenotypes associated with expression of the integrated genes (*e.g.* survival under selection conditions or red fluorescence), and gel analysis of extracted DNA showed p1-derived cytoplasmic plasmids

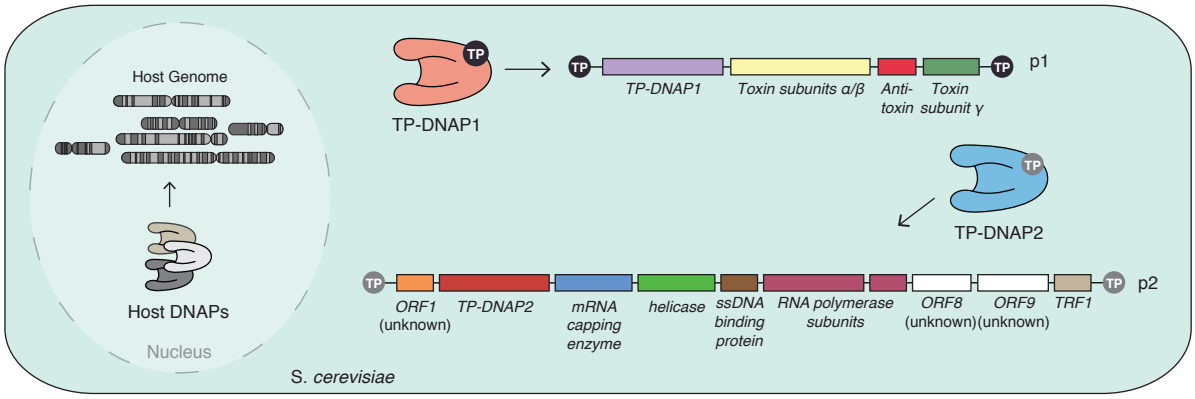


Figure 2.1 | The p1/2 cytoplasmic plasmid system.
The specific basis for OrthoRep is the p1/2 (also known as the pGKL1/2) plasmid system. p1 and p2 are linear, high-copy, double-stranded DNA plasmids that propagate autonomously as selfish elements in the cytoplasm of certain strains of *K. lactis* and *S. cerevisiae*. Each plasmid encodes its own, dedicated DNAP. In addition, p1 encodes a toxin/antitoxin system and p2 encodes replication and transcription factors for both plasmids. ORFs with unknown function are indicated.

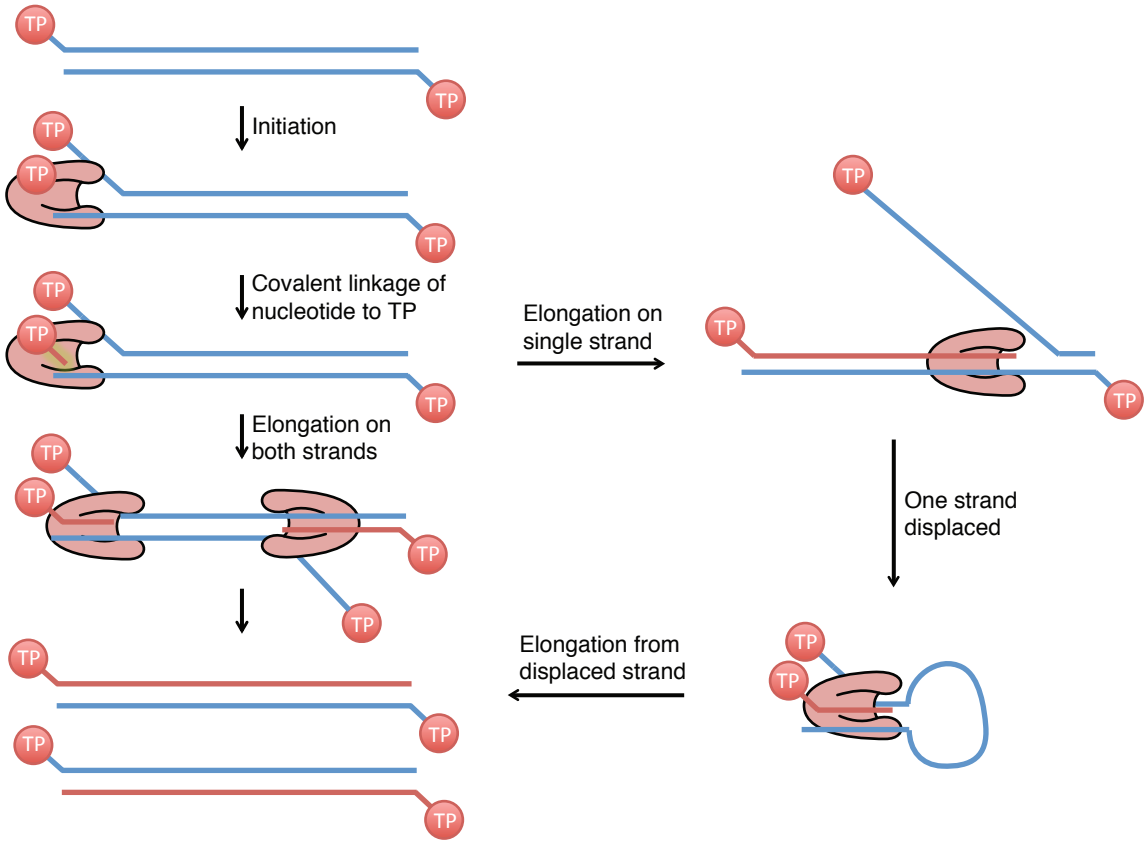


Figure 2.2 | DNA replication via protein-primed initiation as a mechanistic basis for orthogonal replication.
Initiation involves the dimerization of TP-DNAP with a covalently linked TP on one end of the DNA, with one molecule of TP being used for every strand of DNA replicated. The dimer complex triggers the start of DNA replication. The first step is the covalent addition of a nucleotide onto the TP. Then the DNA polymerase continues replication from this first nucleotide in a highly processive manner (limiting the existence of replication intermediates). This occurs asynchronously from both ends, and the result is a duplicated linear plasmid with two covalently attached TPs that act to prime the initiation of the next round of replication [3]. (A "panhandle" mechanism is also possible.) The natural DNAPs of yeast cannot initiate replication from TP-flanked genomes due to their unusual protein origin of replication. The canonical TP-primed DNA replication system of Phi29 [4], and some members of the Phi29 family of DNA polymerases, can only initiate replication from the TP (and not from oligonucleotide primers) [5]. Unlike Phi29, in the p1 replication system, the TP and the DNAP are encoded as a fusion, in one open reading frame. During initiation, the TP domain is cleaved from the DNAP domain (unpublished work), suggesting that the free DNAP is not catalytic in replication initiation.

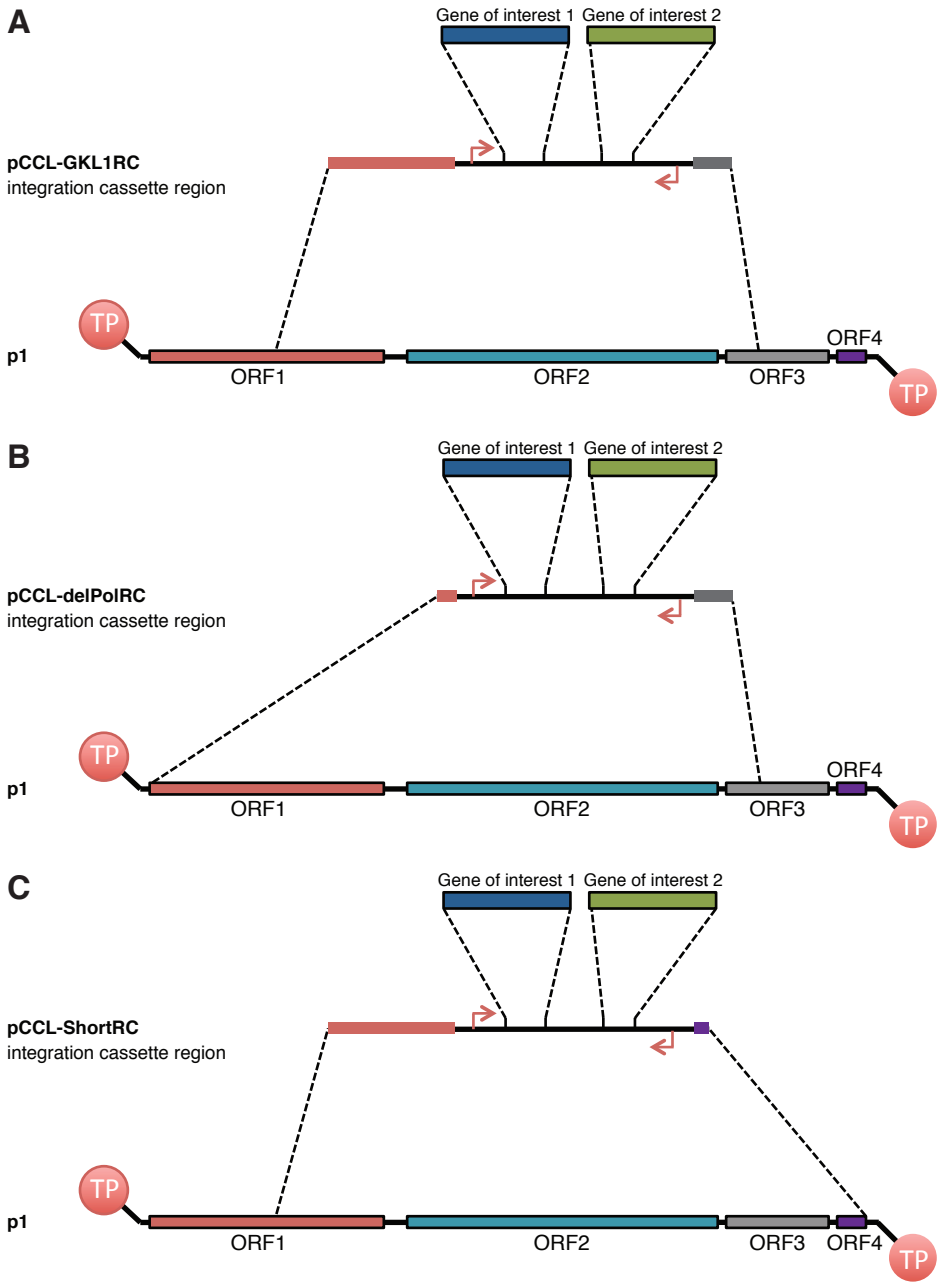


Figure 2.3 | Integration vectors used to introduce genes into p1.
(A) Genes of interest (GOIs), cloned into the pCCL-GKL1RC integration vector, can be inserted into p1, yielding a p1-based variant plasmid that has ORF2 replaced. **(B)** GOIs, cloned into the pCCL-delPolRC integration vector, can be inserted into p1, yielding a p1-based variant plasmid that has ORF2 and the majority of ORF1 (TP-DNAP1) replaced. **(C)** GOIs, cloned into the pCCL-ShortRC integration vector, can be inserted into p1, yielding a p1-based variant plasmid that has ORF2, ORF3, and the majority of ORF4 replaced. In **(A)**, **(B)**, and **(C)** GOI 1 is driven by a promoter derived from the upstream sequence of p1's ORF2; GOI 2 is driven by a promoter derived from the upstream sequence of p2's ORF5. Linear integration cassette regions are generated from plasmids by PCR amplification or digestion by ScaI, as integration regions are flanked by ScaI sites.

of the expected sizes (**Figure 2.4**). These plasmids contained covalently linked TPs, as evidenced by their inability to separate properly without proteinase K treatment (**Figure 2.4**). After serial passaging under selection, DNA gel analysis showed displacement of all copies of p1 by the p1-derived cytoplasmic plasmids (**Figure 2.4**), stably maintained thereafter.

The error rate of p1 replication by TP-DNAP1 was determined via fluctuation tests. We designed and integrated a cassette containing *URA3* and *leu2* (*Q180**) into p1. *leu2* (*Q180**) contains a nonsense mutation at a permissive site in *LEU2*, the reversion of which can be observed on media lacking leucine. We note that only substitution mutations can successfully reinstate *LEU2*, making this assay insensitive to in/dels, and that only substitution mutation rates are analyzed throughout this study, as they are most relevant to gain-of-function changes for directed evolution applications. The resulting strain was used to measure error rate via fluctuation analyses [7]: scores of parallel cultures were expanded in liquid media containing leucine and subsequently plated on solid media lacking leucine. The number distribution of mutants across replicates was used to determine mutation rate by the Ma-Sandri-Sarkar (MSS) method [8, 9]. The mutation rate of genes encoded on p1 was 1.39×10^{-9} s.p.b. (**Table 2.1**). This is higher than yeast genomic substitution mutation rates [10], but still low.

2.2 Increasing the p1 mutation rate

To increase the mutation rate of p1 replication, we generated 32 variants of TP-DNAP1, each containing a single amino acid change chosen based on fidelity studies of homologous B family DNAPs (**Figure 2.5**) [11, 12]. The corresponding integration cassettes containing variant *TP-DNAP1*s alongside *URA3* and *leu2* (*Q180**) were used to replace the WT *TP-*

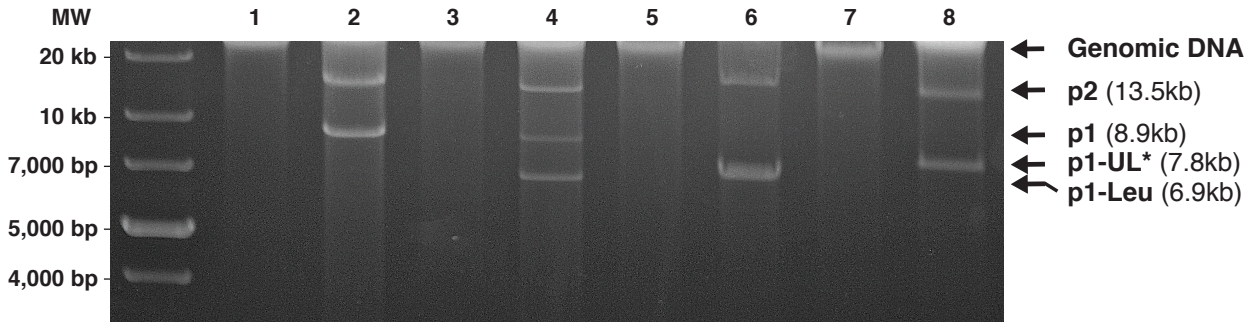


Figure 2.4 | Gel analysis of cytoplasmic DNA plasmids from various strains.
Odd lanes have no proteinase K added. Even lanes have proteinase K added. Lanes 1 and 2: p1 and p2 cytoplasmic plasmids; lanes 3 and 4: cytoplasmic plasmids from a yeast strain containing *LEU2* integrated into p1 (p1-Leu) after a single 1:10,000 passage; lanes 5 and 6: cytoplasmic plasmids from a yeast strain containing p1-Leu after three 1:10,000 passages; lanes 7 and 8: cytoplasmic plasmids from a yeast strain containing *URA3* and *leu2 (Q180*)* integrated into p1 (p1-UL*) after three 1:10,000 passages. The gel is 1% agarose and stained with EtBr.

Entry	Description	Copy #	Phenotypic mutation rate ($\times 10^{-7}$)	Per-base substitution rate ($\times 10^{-10}$)	# of replicates
1	*p1-UL** (day 8, TDF=10^{16})	124	4.00	13.9 (10.9-17.1)	92
2	*format A* (day 6, TDF=10^{12})	12	3.42	122 (91-158)	74
3	*format A* (day 8, TDF=10^{16})	12	2.50	89.5 (69.9-111)	95
4	*format A* (day 18, TDF=10^{36})	12	3.80	136 (107-167)	96
5	*format B* (day 6, TDF=10^{12})	52	42.4	350 (285-421)	68
6	*format B* (day 8, TDF=10^{16})	52	38.2	316 (258-378)	60
7	*format B* (day 10, TDF=10^{20})	52	48.4	399 (351-447)	192
8	*format B* (day 18, TDF=10^{36})	52	44.9	371 (313-433)	94
9	Genomic F102-2 (no TP-DNAP1 *in trans*)	1	0.190	1.29 (0.98-1.61)	183
10	Genomic AH22 (no TP-DNAP1 *in trans*)	1	0.234	1.59 (1.07-2.18)	92
11	Genomic F102-2 (WT TP-DNAP1 *in trans*)	1	0.152	1.03 (0.77-1.32)	171
12	Genomic F102-2 (TP-DNAP1(Y427A) *in trans*)	1	0.115	0.78 (0.57-1.01)	177

Table 2.1 | Mutation rates measured in Chapter 2 using fluctuation analyses.
Entries 1-8 are substitution mutation rates measured using *leu2* (*Q180**) encoded on the p1 orthogonal plasmid. Entries 9-12 are substitution mutation rates measured using the *URA3* locus of the yeast genome. TDF refers to the total dilution factor cells were passaged through before measuring mutation rate. Ranges in parentheses correspond to 95% confidence intervals, determined according to the MSS method. For entries 1-8, copy # was determined by qPCR on total DNA extracted from yeast cells containing p1-UL*, *format A* plasmids, or *format B* plasmids. For entries 9-12, copy # was set to 1, as this is the genomic copy # of haploid yeast. Copy # for *format A* is likely an overestimate relative to other copy numbers, as the lower processivity of the TP-DNAP1 (Y427A) mutant likely leaves a large fraction of p1 replicons unfinished. Yet qPCR still may detect this unfinished fraction as full copies since the qPCR primers only bind within a limited region. The per-base substitution rate in *format A* is therefore likely underestimated.

DNAP1 gene encoded on p1, yielding 32 p1-MutX-UL* cytoplasmic plasmids (X denotes the specific mutation) in their corresponding strains. Small-scale fluctuation analyses using *leu2* (*Q180**) reversion followed by larger-scale fluctuation analyses on promising mutants revealed p1-MutY427A-UL*, containing a Y427A mutation in TP-DNAP1, to have the greatest increase in mutation rate (**Figure 2.5**). However, this increase was modest. Upon further inspection, we observed that p1-MutX-UL* cytoplasmic plasmids frequently recombined with copies of p1 still inside cells during serial passaging, resulting in the generation of p1-UL* containing both WT *TP-DNAP1* and the selectable *URA3* marker. After such events, recombinants consistently displaced precursor plasmids – presumably, the WT TP-DNAP1 was more processive than our tested mutants – thereby progressively lowering mutation rate throughout growth and complicating our measurements.

To overcome this issue, we expressed DNAPs *in trans* by encoding them on a standard yeast nuclear plasmid containing a *CEN6/ARS4* origin of replication (**Figure 2.6**). Functional complementation was assayed by replacing the WT *TP-DNAP1* gene encoded on p1 with the selectable *URA3* gene, yielding p1-delPol-UL*. Although initial attempts to functionally express the AT-rich *TP-DNAP1* gene *in trans* failed (**Figure 2.7A**), a synthetic *TP-DNAP1*, recoded for efficient yeast nuclear transcription, was fully capable of replicating p1-derived cytoplasmic plasmids. This was evidenced by the simultaneous loss of p1 (and the associated p1-encoded TP-DNAP1) and maintenance of p1-delPol-UL* after passaging in selective media (**Figure 2.7B, C**). In contrast, cells that did not express TP-DNAP1 were unable to maintain p1-delPol-UL* after passaging (**Figure 2.7A**).

We next mutated our recoded nuclear *TP-DNAP1* to create pCCL-ReDNAP(Y427A), which expresses the error-prone TP-DNAP1 (Y427A) mutant *in trans*. In conjunction, we

	TP-DNAP1	Relative Mutation Rate
	WT	+
ExoI	S359A	-
	S359I	+++
	S359T	+
ExoII (Candidate 1)	N388A	-
	N388D	+
	F392A	-
	F392S	-
	F392Y	-
ExoII (Candidate 2)	N423A	-
	N423D	++
	Y427A	+++
	Y427S	++
	Y427F	+
Motif A	Y646X (where X is all 19 amino acids)	-

Figure 2.5 | TP-DNAP1 variants screened for increased mutation rate.
Alignments of conserved ExoII motifs from family B DNAPs to TP-DNAP1 gave competing candidates, so mutants from both were generated and tested. Numbering does not correspond to the annotation of pGKL1 ORF1 in GenBank (X01095.1), as our analysis of conserved upstream sequences of all ORFs of pGKL1 and pGKL2 as well as Kozak sequences suggests it is more likely that the start position is actually position 9 of ORF1. Therefore, we treat position 9 of ORF1 as the actual start codon for all experiments with TP-DNAP1, except as noted. For each variant, at least 8 replicates were used in preliminary fluctuation analyses to determine relative mutation rates. (+), (++), and (+++) correspond to observed phenotypic *leu2 (Q180*)* reversion frequencies similar to, slightly greater than (~2-fold increase), and greater than (~3-5-fold increase) that achieved with WT TP-DNAP1. (-) refers to phenotypic *leu2 (Q180*)* reversion frequencies lower than that achieved with WT TP-DNAP1. We believe this is due to a drop in copy number of p1-MutX-UL* when certain TP-DNAP1 variants with substantially lower activity are expressed as lower activity of replication likely drops copy number, thereby lowering the number of opportunities for reversion of *leu2 (Q180*)*.

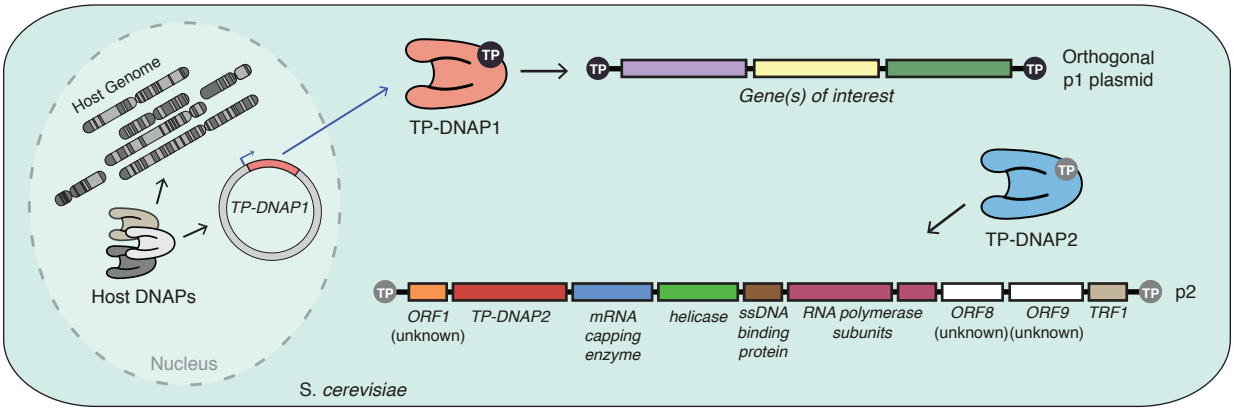


Figure 2.6 | Replication of p1 by a recoded TP-DNAP1 expressed *in trans*.
Schematic of OrthoRep with a recoded TP-DNAP1 encoded and expressed from a nuclear *CEN6/ARS4* plasmid.

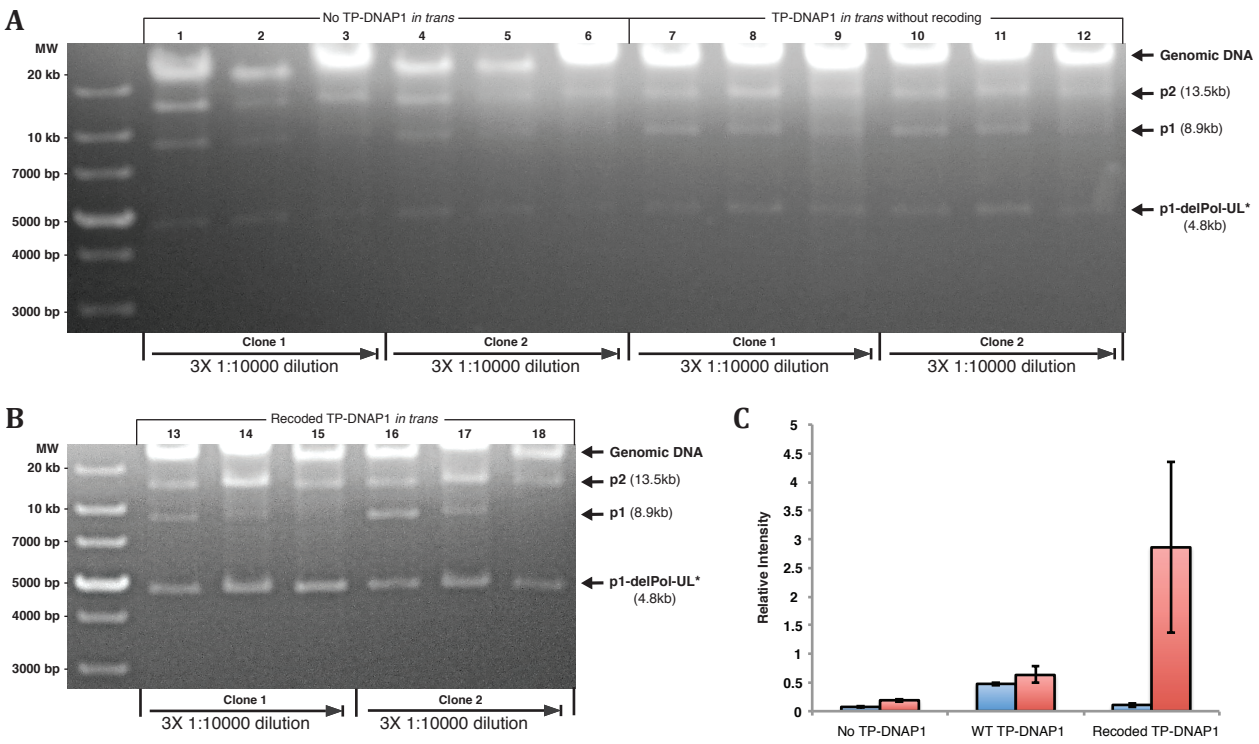


Figure 2.7 | Expression of WT TP-DNAP1 *in trans*.
(A) and **(B)** DNA gel analysis (1% agarose, EtBr stained) corresponds to cytoplasmic plasmids extracted from replicate clones containing p1-delPol-UL* (a p1-based plasmid generated by replacing the p1-encoded TP-DNAP1 with *URA3* and *leu2 (Q180*)*) and a *CEN6/ARS4* nuclear plasmid expressing no *TP-DNAP1*, a *TP-DNAP1* gene that was not recoded, or a recoded *TP-DNAP1* gene. Lanes 1, 4, 7, 10, 13, and 16 correspond to DNA content of cells extracted directly after deletion of *TP-DNAP1* from p1 via integration of the *TP-DNAP1* deletion cassette to generate p1-delPol-UL*. Lanes 2-3, 5-6, 8-9, 11-12, 14-15, and 17-18 correspond to DNA content of cells passaged in SC without uracil and without histidine. Only when the recoded TP-DNAP1 is expressed *in trans* does p1-delPol-UL* copy number remain high. In other cases, a dramatic drop in both p1 and p1-delPol-UL* is observed, suggesting that insufficient functional TP-DNAP1 is expressed *in trans*. **(C)** Image analysis of gel shown in **(B)**, depicting mean intensity of p1 (blue bars) and p1-delPol-UL* (red bars) bands relative to the corresponding intensity of p2 for the third passage of each sample. Error bars span the range of intensities for the two clones of each sample. Except in the case of recoded TP-DNAP1, p1 intensity is similar to p1-delPol-UL* intensity because p1 encodes a TP-DNAP1 necessary for the maintenance of p1-delPol-UL* in the absence of a TP-DNAP1 properly expressed *in trans*. Also, except in the case of recoded TP-DNAP1, the overall intensities of the p1 and p1-delPol-UL* bands drop substantially after passaging. This is because as selection in SC without uracil and without histidine increases the copy number of p1-delPol-UL*, it also decreases the amount of functional TP-DNAP1 expressed from p1 (since they share the same replication system). This begins a positive feedback loop that drops the copy number of all p1-based plasmids. This is not the case with the recoded TP-DNAP1, because it is functionally expressed *in trans* at a constant level.

integrated *URA3* and *leu2* (*Q180**) into p1, replacing the p1-encoded *TP-DNAP1* to create p1-delPol-UL*, or retaining the p1-encoded *TP-DNAP1* to create p1-Short-UL*. Fluctuation analyses based on *leu2* (*Q180**) reversion were used to determine the substitution mutation rates of genes encoded on p1 in both formats: pCCL-ReDNAP(Y427A) + p1-delPol-UL* (*format A*) and pCCL-ReDNAP(Y427A) + p1-Short-UL* (*format B*). *Format A* resulted in a low-copy error-prone p1 replication system, while *format B* resulted in a high-copy error-prone p1 replication system (**Tables 2.1, 2.2**). We attribute the lowering of copy number in *format A* to the fact that TP-DNAP1 (Y427A) is likely less active than WT TP-DNAP1. In both formats, the per-base mutation rate was higher, by a factor of ~9 or ~25, respectively, than the mutation rate of WT TP-DNAP1, which itself is several times higher than the yeast genomic mutation rate. (Exact substitution preferences of TP-DNAP1 (Y427A) are shown in **Tables 2.3, 2.4**.) Thus, we created an error-prone OrthoRep whose error rate can be modulated through engineering the TP-DNAP1.

OrthoRep should stably generate mutations in genes of interest. We passaged cells containing the error-prone OrthoRep (*format A* and *format B*) for 18 days, representing a 10^{36}-fold total dilution factor. On days 6, 8, and 18, we conducted fluctuation analyses to track error rates. In both formats, the mutation rate remained steady (**Table 2.1**).

2.3 Verifying orthogonality of error-prone OrthoRep

OrthoRep must spare the host genome from rapid mutation. Since our engineered TP-DNAP1 (Y427A) continuously mutated p1 at an error rate much higher than the known genomic mutation rate of yeast [10], we had the necessary ingredients to test orthogonality. We transformed vectors pCCL-ReDNAP or pCCL-ReDNAP(Y427A), expressing WT TP-DNAP1 or the error-prone mutant TP-DNAP1 (Y427A), respectively, into

	Copy Number of *LEU3*	Copy Number of *leu2 (Q180*)*	Normalized Copy Number of *leu2 (Q180*)*
p1-UL + WT TP-DNAP1 in trans*	1121.83 ± 39.54	138996.36 ± 1815.85	123.9 ± 6.11
format A	1598.41 ± 139.96	18752.81 ± 395.87	11.73 ± 0.38
format B	1476.14 ± 53.37	76411.52 ± 12564.68	51.76 ± 9.83

Table 2.2 | Copy numbers of p1-based plasmids.
Copy number was determined by qPCR from total DNA isolated from various samples (*n* = 3, error-bars are ± standard deviation). Primers specific to the *LEU3* gene, which is present only in genomic DNA, were used to determine the amount of genomic DNA in each sample. Primers specific to the *leu2 (Q180*)* gene encoded on p1-based plasmids were used to determine the amount of p1-based cytoplasmic plasmids. Assuming a genomic copy number of 1 copy per cell, the ratio of *leu2 (Q180*)* to *LEU3* DNA amounts gives the copy number of p1-based plasmids for each condition.

Description	# of replicates	Copy #	Phenotypic mutation rate (x 10^{-7})	Per-base substitution rate (x 10^{-10})
p1-Short-UL(TGA)*	161	52	51.3	371 (327-415)
*p1-Short-U*L*	83	52	3.09	180 (115-255)

Table 2.3 | Mutation rates of p1 replication by TP-DNAP1 (Y427A) in *format B* as measured using p1-Short-UL*(TGA) and p1-Short-U*L.
p1-Short-U*L measures only the mutation rate for the G:C→A:T transition (see **2.4 Methods**), which is lower than the mutation rate for other mutation types (see **Table 2.4**). Ranges in parentheses correspond to 95% confidence intervals, as determined using the MSS method [8, 9].

Mutation	*N*	Normalized *N*	Normalized % total
A:T→G:C (transition)	8	4	6.6%
A:T→T:A (transversion)	25	12.5	20.5%
A:T→C:G (transversion)	21	10.5	17.2%
G:C→A:T (transition)	5	5	8.2%
G:C→T:A (transversion)	28	28	45.9%
G:C→C:G (transversion)	1	1	1.6%

Table 2.4 | Substitution mutation preferences of p1 replication by TP-DNAP1 (Y427A) in *format B*.
N refers to the raw number of sequences where the corresponding mutation was detected. Normalized *N* refers to *N* divided by the number of possible locations where the mutation could be observed. All mutations except for G:C→A:T were detected from a single fluctuation analysis experiment using p1-Short-UL*(TGA). G:C→A:T data, determined using p1-Short-U*L, were manually added into this table in proportion to the individual error rates calculated for each type of mutation (**Table 2.3**). We note that this assay is only capable of measuring substitution mutations, and therefore does not include the preferences for frameshifts or indels.

yeast cells and measured the corresponding genomic mutation rates. We also measured the genomic mutation rate in the presence and absence of the entire p1/2 replication system by comparing the p1/2-containing *S. cerevisiae* strain, F102-2, to its parent strain, AH22 [1], lacking cytoplasmic plasmids. Mutation rates were characterized via fluctuation analyses on the genomic *URA3* locus, using 5-fluoroorotic acid (5-FOA) selection, because the spectrum of *URA3* disabling mutations is substitution-rich, thereby allowing us to robustly assess whether the high substitution rate of TP-DNAP1 (Y427A) contributes to genomic mutation. The number distribution of mutants across replicates was used to calculate the phenotypic mutation rate, which was converted to a per-base mutation rate following previously described parameters [10]. We observed no increase in the genomic mutation rate in the presence of WT TP-DNAP1 or mutagenic TP-DNAP1 (Y427A) nor did we observe an increase in the genomic mutation rate due to the presence of the p1/2 system itself (**Table 2.1**). This means that error-prone TP-DNAP1 (Y427A) achieves complete mutational targeting to p1.

We conclude here that OrthoRep with the p1-TP-DNAP1 (Y427A) plasmid-DNAP pair is durably error-prone and orthogonal. With TP-DNAP1 (Y427A), the per-base substitution mutation rate of genes on p1 (in *format B*) is ~400-fold greater than the per-base substitution mutation rate of the host genome. This demonstrates that p1-TP-DNAP1 can serve as the basis for OrthoRep, but much higher mutation rates are required for rapid and high-throughput directed evolution experiments.

2.4 Methods

Yeast strains. *Saccharomyces cerevisiae* strain F102-2 (**a** *leu2-3, 2-112 his4-519 can1* ρ^0) was obtained from ATCC (Catalog #200585). F102-2u was generated by disabling the *URA3*

gene of F102-2. This was done by PCR amplification of *ura3* from plasmid 36 using primers 51 and 52, transforming the linear fragment into F102-2, and selecting on solid SC media containing 5-fluoroorotic acid (5-FOA) (1g/L). Eight colonies were replica plated on solid SC media without uracil and solid SC media with 5-FOA for verification of uracil auxotrophs. The *URA3* locus of these 8 strains were then amplified using primers 51 and 52, and sequenced. Although none had the desired mutation, one strain had a deletion of bases 374-614 in the *URA3* gene and was designated F102-2u. F102-2u was transformed with plasmid 35 and selected on solid SC media lacking uracil, to verify that expression of *URA3* can complement uracil auxotrophy.

DNA cloning. A list of plasmids and primers appear as **Appendix B** and **Appendix C**, respectively. *E. coli* strains TG1 (Lucigen) and SS320 (Lucigen) were used for all DNA cloning steps. All primers were purchased from IDT. All restriction enzymes, DNA polymerases, ligases, and PCR reagents were obtained from NEB except where otherwise noted.

To clone plasmid 1, a synthetic DNA sequence containing, in order, a ScaI restriction site, a homology region of bases 2328-3200 of p1, a putative promoter sequence of bases 3201-3229 of p1, a multiple cloning site containing NsiI, KpnI, BamHI, SacI and NdeI restriction sites, a putative promoter sequence of the reverse complement of bases 7923-7481 of p2, a homology region of bases 6571-7309 of p1, and a ScaI restriction site was custom synthesized by GenScript. This synthetic construct was amplified using primers 1 and 2, plasmid 2 was amplified using primers 3 and 4, and PCR products were digested with enzymes XhoI and XbaI and ligated to yield plasmid 1.

Plasmid 1 served as the parent vector for plasmids 3-9, 13, and 39. To obtain plasmid 3, the *LEU2* gene was amplified from plasmid 75 using primers 5 and 6. Subsequent digestion with NsiI and SacI and ligation with a similarly digested plasmid 1 yielded plasmid 3. To obtain plasmids 4 and 5, the *URA3* gene was amplified from plasmid 35 using primers 7 and 8, digested with NdeI and SacI, and ligated with similarly digested plasmids 1 and 3. To obtain plasmids 6, 7, and 8, plasmids 5, 6, and 7 were amplified using 5'-phosphorylated primers 9 and 10, 11 and 12, and 13 and 14. The PCR products were blunt ligated to yield desired products. To obtain plasmid 9, a fragment containing *leu2 (Q180*)* and *URA3* was digested from plasmid 8 with NsiI and NdeI. Subsequent ligation into a similarly digested plasmid 1 yielded plasmid 9. To obtain plasmid 13, a fragment containing the mKate gene was amplified from plasmid 10 using primers 15 and 16. Subsequent digestion with SacI and NdeI and ligation with a similarly digested plasmid 3 yielded plasmid 13. To obtain plasmid 39, a fragment containing the *KanMX* gene was amplified from plasmid 74 using primers 49 and 50. Subsequent digestion with NdeI and SacI and ligation into a similarly digested plasmid 1 yielded plasmid 39. Plasmids 3, 4, 9, 13, and 39 can be linearized by digestion with ScaI, and integrated into p1 to express *LEU2*, *URA3*, *URA3* and *leu2 (Q180*)*, *LEU2* and *mKate*, and *KanMX*, respectively.

Plasmid 9 served as the parent vector for plasmid 14. To obtain plasmid 14, a p1 homology region was amplified from DNA extracted from F102-2 using primers 101 and 102, and a fragment was amplified from plasmid 9 using primers 103 and 104. Both PCR products were digested with XhoI and NsiI and ligated to yield plasmid 14.

Plasmid 14 served as the parent vector for plasmids 15-27. To obtain plasmids 15-27, plasmid 14 was amplified using 5'-phosphorylated primers 17 and 20, 18 and 20, 19

and 20, 21 and 23, 22 and 23, 24 and 27, 25 and 27, 26 and 27, 28 and 30, 29 and 30, 31 and 34, 32 and 34, and 33 and 34, respectively. The PCR products were blunt ligated to create plasmids 15-27. Therefore, plasmids 15-27 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-MutX-UL* cytoplasmic plasmids.

Plasmid 28 is a standard yeast expression vector, containing the *CEN6/ARS4* origin of replication, and a multiple cloning site flanked by the TDH3 promoter and ADH1 terminator. Plasmid 28 served as the parent vector for plasmids 29-34. To obtain plasmid 29, a fragment containing the *LEU2* gene was amplified from plasmid 3 using primers 35 and 36. Subsequent digestion with BglII and XhoI and ligation into similarly digested plasmid 28 yielded plasmid 29. To create plasmids 30-34, a region of plasmid 29 was amplified using 5'-phosphorylated primers 37 and 38, 37 and 39, 37 and 40, 37 and 41, and 37 and 42, respectively. The PCR products were blunt ligated to create plasmids 30-34. Therefore, plasmids 29-34 encode *LEU2*, and the Q180E, Q180K, Q180S, Q180L, and Q180Y mutants of *LEU2*.

Plasmid 35 encoding the *URA3* gene served as the parent vector for plasmid 36. To create plasmid 36, plasmid 35 was amplified using 5'-phosphorylated primers 43 and 44, and the PCR product was blunt ligated. Therefore, plasmid 36 encodes *ura3* containing a nonsense mutation in the *URA3* gene.

Plasmid 9 served as the parent vector for plasmid 37. To clone plasmid 37, a region of plasmid 9 was amplified using 5'-phosphorylated primers 45 and 46. The PCR product was blunt ligated to create plasmid 37. Therefore, plasmid 37 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-delPol-UL*.

Plasmid 9 served as the parent vector for plasmid 38. To clone plasmid 38, plasmid 9 was amplified using primers 47 and 48. The amplified products contained overlapping regions, which were resolved when transformed directly into *E. coli* strain SS320, yielded plasmid 38. Therefore, plasmid 38 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-Short-UL*.

Plasmid 13 served as the parent vector for plasmids 40-52. To clone plasmid 40, plasmid 13 was amplified using 5'-phosphorylated primers 53 and 54. The PCR product was blunt ligated to create plasmid 40. To clone plasmid 41, plasmid 40 was amplified using 5'-phosphorylated primers 55 and 56. The PCR product was blunt ligated to create plasmid 41. Primers 57 and 58, 59 and 60, 61 and 62, 63 and 64, 65 and 66, 67 and 68, 69 and 70, 71 and 72, 73 and 74, 75 and 76, and 77 and 78 were 5'-phosphorylated by treatment with T4 Polynucleotide Kinase (NEB), and used to amplify plasmid 41. The PCR products were blunt ligated to create plasmids 42-52, respectively. Therefore, plasmids 41-52 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-Leu-mKate variants, with no upstream sequence (US) (plasmid 41) and 100bp USs (plasmids 42-52) of each open reading frame (ORF) of the p1/2 system cloned directly upstream of the coding region for *mKate*.

Plasmids 41-52 served as the parent vectors for plasmids 53-64, respectively. To clone plasmids 53-64, primers 79 and 80 were annealed and amplified. Subsequent digestion of the PCR product with MluI and XbaI and ligation into similarly digested plasmids 41-52 yielded plasmids 53-64. Therefore, plasmids 53-64 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-Short-Leu-mKate variants, with no

US (plasmid 53) and 100bp USs (plasmids 54-64) of each ORF of the p1/2 system cloned directly upstream of the coding region for *mKate*.

Plasmid 29 served as the parent vector for plasmids 65-67. To obtain plasmid 65, plasmid 29 was amplified using 5'-phosphorylated primers 11 and 12. The PCR product was blunt ligated to create plasmid 65. To obtain plasmid 66, the *HIS4* gene was amplified using primers 81 and 82 from F102-2 genomic DNA purified using the YeaStar Genomic DNA Kit (ZymoResearch), and a yeast expression vector plasmid backbone was amplified using primers 83 and 84 from plasmid 65. The PCR fragments were assembled using the Gibson method [13], yielding plasmid 66. To obtain plasmid 67, plasmid 66 was amplified using 5'-phosphorylated primers 85 and 86. The PCR product was blunt ligated to create plasmid 67. Therefore, yeast nuclear promoters and genes can be cloned into the SacI and SpeI sites of plasmid 67 for expression on a *CEN6/ARS4* plasmid with the *HIS4* marker.

Plasmid 68 is a yeast expression vector encoding the *CEN6/ARS4* origin of replication, and a multiple cloning site flanked by the REV1 promoter and ADH1 terminator. Plasmid 68 served as the parent vector for plasmid 69. To obtain plasmid 69, the *TP-DNAP1* gene was amplified from F102-2 cytoplasmic plasmids using primers 87 and 88. Subsequent digestion with BglII and XhoI and ligation into a similarly digested plasmid 68 yielded plasmid 69.

Plasmid 67 served as the parent vector for plasmids 70-72. To obtain plasmid 70, a fragment containing the REV1 promoter and the *TP-DNAP1* gene was amplified using primers 89 and 90. The PCR product with digested with SacI and SpeI and ligated into a similarly digested plasmid 67 to yield plasmid 70. To obtain plasmid 71, plasmid 70 was amplified using 5'-phosphorylated primers 91 and 92. The PCR product was blunt ligated

to create plasmid 71. To obtain plasmid 72, three fragments spanning a recoded *TP-DNAP1* (synthesized by Life Technologies) were assembled using the Gibson method [13], and the resulting fragment was amplified using primers 93 and 94. Subsequent digestion with BglII and XhoI and ligation into a similarly digested plasmid 71 yielded plasmid 72.Therefore, plasmids 71 and 72 encode the WT and recoded *TP-DNAP1*, respectively, driven by the REV1 promoter, the *HIS4* marker, and a *CEN6/ARS4* nuclear replication origin.

Plasmid 72 served as the parent vector for plasmid 73. To obtain plasmid 73, plasmid 72 was amplified using primers 95 and 96. The amplified product contained overlapping regions, which were resolved when transformed directly into *E. coli* strain SS320, yielding plasmid 73. Therefore, plasmid 73 encodes the recoded *TP-DNAP1* (Y427A) driven by the REV1 promoter, the *HIS4* marker, and a *CEN6/ARS4* nuclear replication origin.

To clone plasmid 76, plasmid 38 was amplified using primers 106 and 107. The product contained overlapping regions, which were resolved when transformed directly into *E. coli* strain SS320, yielding plasmid 76. Therefore, plasmid 76 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-Short-UL*(TGA).

Plasmids 37 served as the parent vector for plasmid 77. To obtain plasmid 77, a fragment containing a portion of the *LEU2* gene was digested from plasmid 5 with KpnI and SacI. Subsequent ligation with a similarly digested plasmid 37 yielded plasmid 77. Therefore, plasmid 77 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-delPol-UL.

Plasmid 77 served as the parent vector for plasmid 78. To obtain plasmid 78, plasmid 77 was amplified using primers 108 and 109. The amplified product contained

overlapping regions, which were resolved when transformed directly into *E. coli* strain SS320, yielding plasmid 78. Therefore, plasmid 78 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-delPol-U*L.

Plasmid 37 served as the parent vector for plasmid 79. To obtain plasmid 79, the *URA3*(K93R) and *LEU2* genes were digested from plasmid 78 using NsiI and NdeI. Subsequent ligation with a similarly digested plasmid 37 yielded plasmid 79. Therefore, plasmid 79 can be linearized by digestion with ScaI, and integrated into p1 to yield p1-Short-U*L.

Yeast transformation and integration. All transformations were performed using the Yeastmaker Yeast Transformation System 2 kit (Clontech). For integrations into p1, 10µg of plasmid was linearized by digestion with ScaI, which generated blunt ends containing homologous regions as per the design of our integration cassettes. DNA fragments were transformed directly into F102-2 or F102-2u cells. Co-transformations were performed with 500ng of supercoiled plasmid DNA and 10µg of ScaI-linearized DNA, for concomitant integration into p1 and expression of TP-DNAP1s *in trans*. Transformants were selected on the appropriate selective solid SC media. The appearance of colonies following integration typically took 4-6 days of growth at 30°C.

Yeast passaging. Following the appearance of colonies from integration, 3-6 colonies were picked from plates directly into the appropriate selective liquid SC media. Saturation of liquid cultures following inoculation typically took 2 days of growth at 30°C with orbital shaking (200rpm). Upon saturation, cultures were passaged with 1:10,000 dilutions into the same media. The remaining saturated culture volume was subjected to cytoplasmic

plasmid extraction (see Cytoplasmic plasmid extraction section), and agarose gel electrophoresis of extracted plasmids was performed to track displacement of p1 by p1-derived integrants. This was repeated until pure populations of cells with only p1-derived integrants were obtained, typically requiring three passages at 1:10,000 dilutions.

Cytoplasmic plasmid extraction. p1, p2, and p1-derived plasmids were extracted following the yeast DNA miniprep procedure as summarized in *Methods in Yeast Genetics* [14]. We note the following modifications: (1) cells were washed in 0.9% NaCl prior to treatment with Zymolyase (US Biological); (2) 200μg/mL proteinase K (Sigma) was supplemented during SDS treatment for degradation of TP, except where noted; (3) Rotation at ~10rpm was used during Zymolyase and proteinase K treatments. Smaller-scale preparations, sufficient for culture volumes between 1-10mL, were performed by scaling down the large-scale protocol by 32-fold. Extracted DNA was still enough to be observed by agarose gel electrophoresis.

DNA gel analysis. DNA gels were analyzed using the ImageJ software. Band intensities were quantified using the mean gray value measurement tool. The surrounding background for each band was analyzed individually to account for differences within the image, and was defined as the area within one band-width's distance. The intensity of a band was then calculated as the magnitude of the difference between the mean gray value of the band and the mean gray value of its surrounding background.

Fluctuation analysis using p1-encoded *leu2 (Q180*)*. Fluctuation analyses on p1-encoded *leu2 (Q180*)* were performed to determine per-base substitution rates of p1 replication. *leu2 (Q180*)* contains a C→T mutation at base 538 in *LEU2* that results in an

ochre nonsense mutation at a site permissive to all single point mutants that generate missense mutations. This was verified by cloning each single point missense mutant – Q180Y, Q180K, Q180E, Q180L, and Q180S (plasmids 30-34) – transforming mutants into F102-2u, and verifying complementation of leucine auxotrophy by growth on solid SC media lacking leucine.

To perform fluctuation analyses, F102-2u strains encoding p1-UL*, p1-MutX-UL*, p1-Short-UL*, and p1-delPol-UL* were first passaged until complete displacement of p1 (see Yeast passaging section). Once this occurred, a small culture was grown to saturation in SC lacking uracil (and also lacking histidine for strains with *CEN6/ARS4* plasmids), diluted 1:10,000 into the same media and split into replicate cultures in 96-well trays. All strains were grown using 100μL volumes, except for the *format B* strain, which was grown in 15μL volumes due to its high mutation frequency. Cultures were grown to saturation for 2-2.5 days. Cultures were washed with 500μL 0.9% NaCl and resuspended in 35μL 0.9% NaCl. Four cultures from each strain were pooled, diluted, and titered on solid YPD media to obtain the total number of cells. 20μL of all remaining cultures were spot plated onto solid SC media lacking leucine, and spots were allowed to dry before incubation. Plates were incubated at 30°C. Colonies were counted from YPD titer plates after 2 days, and from spot plates after 5 days. The choice of 5 days was based on the fact that it takes ~5 days after transformation for cells containing *LEU2* integrated into p1 to clearly appear as colonies on a plate, a process that approximates the appearance of *LEU2* colonies through spontaneous reversion of *leu2 (Q180*)*.

Fluctuation data were analyzed by the Ma-Sandri-Sarkar (MSS) maximum-likelihood estimator method[9], whereby the best *m*, the mean number of mutations resulting in the

measured number distribution of phenotypic mutants, is found. This was implemented using the FALCOR tool [8] to calculate *m* values and 95% confidence intervals, and the partial plating correction frequency as described by Foster [7] was applied. The phenotypic mutation rate was calculated by dividing *m* by the average number of cells per culture, as determined by the cell titers. The per-base substitution rate was then calculated by dividing the phenotypic mutation rate by the measured copy number of p1-encoded *leu2 (Q180*)* (see Determination of copy number via quantitative PCR section) and by the number of ways *leu2 (Q180*)* can revert to *LEU2* (2.33 for the ochre codon). 95% confidence intervals were similarly scaled by these factors.

To verify that growth of revertant colonies was due to reversion of the p1-encoded *leu2 (Q180*)*, 11 revertant colonies from independent spot platings were analyzed by sequencing. To do this, colonies were expanded in SC lacking leucine, passaged 8 times at 1:1,000, and subjected to small-scale cytoplasmic plasmid extraction. *leu2 (Q180*)* was amplified using primers 5 and 105, and sequenced. All 11 clones showed reversions of the ochre mutation to a functional codon.

Fluctuation analysis using genomic *URA3*. Fluctuation analyses on the genomic *URA3* gene were performed to determine genomic per-base substitution rates, as previously described [10]. AH22 or F102-2 strains encoding appropriate *CEN6/ARS4* plasmids were grown to saturation in SC lacking uracil and histidine, diluted 1:5,000 into SC lacking histidine, and split into replicate 200μL cultures in 96-well trays. Cultures were grown for 2-2.5 days until saturation. Cultures were washed with 400μL 0.9% NaCl and resuspended in 210μL 0.9% NaCl. Four cultures from each strain were pooled, diluted, and titered on solid YPD media. 200μL of all remaining cultures were spot plated onto solid SC media

supplemented with 5-FOA (1g/L), and spots were allowed to dry before incubation. Plates were incubated at 30°C. Colonies were counted from titer plates after 2 days, and from spot plates after 5 days.

Phenotypic mutation rates were calculated as described above (see Fluctuation analysis using p1-encoded *leu2 (Q180*)* section). Per-base substitution rates were then calculated from the mutation frequency using parameters previously described[11]. Specifically, phenotypic mutation rate was multiplied by the fraction of 5-FOA resistant clones that contain mutations in the *URA3* gene and divided by the target size for loss of function of *URA3 via* base pair substitution. 95% confidence intervals were similarly scaled by these factors.

Determination of copy number *via* quantitative PCR. The copy numbers of p1-encoded *leu2 (Q180*)* were determined by quantitative PCR. Whole-cell DNA extracts were prepared by the large-scale cytoplasmic plasmid extraction protocol described above. During extraction, ethanol- and isopropanol-precipitated pellets were thoroughly resuspended to ensure complete extraction. All extracts were diluted 20-fold for quantitative PCR reactions.

Primers 97 and 98 were designed to bind in the *leu2 (Q180*)* gene of p1-UL* and p1-delPol-UL*. The genomic sequence of *S. cerevisiae* was obtained from the *Saccharomyces* Genome Database (http://www.yeastgenome.org), and primers 99 and 100 were designed to bind in the genomic *LEU3* gene as a reference. Both primer pairs were designed with melting temperatures of ~60°C, and for amplification of ~150bp long amplicons.

Quantitative PCR reactions were performed in 20µL mixtures containing 1µL template DNA, 2µL forward primer (5µM), 2µL reverse primer (5µM), 0.1µL HotStar Taq

DNA Polymerase (Qiagen), 10µl SyBR Green master mix, and 4.9µL ddH_2O. SyBR Green master mix contained 100mM Tris (pH 8.3), 6mM $MgCl_2$, 1mg/mL BSA, 400µM dNTP mix, and 0.66X SyBR Green (Invitrogen).

Standard curves for both primer pairs were prepared using 6 serial 5-fold dilutions of DNA extracted from F102-2u, and a non-template control with only ddH_2O. All samples were measured in triplicate with both primer pairs. Quantitative PCR was performed using an iCycler iQ5 Multicolor Real-Time PCR Detection System (Bio-Rad). The PCR cycling conditions used were:

95°C for 8m30s

95°C for 15s, 60°C for 1m – 40X

95°C for 1m

55°C for 1m

60°C for 10s – 71X

4°C - Hold

Cycle threshold (C_t) values were determined for a fluorescent signal of 700au, for both primer pairs. C_t values from both standard curves were fit to a semi-log regression line plot of C_t *versus log*([DNA]). From this, a slope and y-intercept were calculated, and relative copy numbers of samples were determined by the equation, copy number = $10^{((\text{sample Ct} - \text{y-intercept})/\text{slope})}$. Triplicate samples were averaged, and p1-derived plasmid copy numbers were normalized to those of the genome by dividing the average *leu2 (Q180*)* copy number by the average *LEU3* copy number. Variance was calculated using equation 5.2 of Frishman [15].

Promoter characterization. Cassettes corresponding to plasmids 53-64 were linearized and transformed into F102-2. Plasmids 8-10, representing mKate driven by different nuclear promoters, were also transformed into F102-2 for comparison. Six colonies were selected from each transformation and passaged four times at 1:10,000 in SC without leucine to completely displace plasmid p1. 200µl saturated cultures were transferred into a 96-well polystyrene plate (Sigma-Aldrich) and mixed thoroughly. mKate fluorescence (ex/em = 565nm/630nm) was immediately measured using a microplate reader (Fluoroskan Ascent FL). Fluorescence was normalized to OD_{600}, measured using a spectrophotometer (Bio-Rad SmartSpec 3000). The average and standard deviation of normalized fluorescence was calculated for each set of six replicates.

Characterizing the mutational preferences of TP-DNAP1 (Y427A) *in vivo*. Two additional fluctuation analyses, coupled with sequencing, were used to determine the mutational preferences of TP-DNAP1 (Y427A) in *format B*. In one experiment, p1-Short-UL*(TGA), generated by inserting the integration cassette from plasmid 76 into p1, was used for fluctuation analysis. p1-Short-UL*(TGA) encodes *leu2 (Q180*(TGA))*, which contains a C→T mutation at base 538 and an A→G mutation at base 539, resulting in an opal nonsense mutation in *LEU2*. Fluctuation analysis can be used to determine the frequency of reversion from *leu2 (Q180*(TGA))* to *LEU2*. Single point mutations that restore *LEU2* from *leu2 (Q180*(TGA))* include T:A→A:T, T:A→C:G, T:A→G:C, G:C→T:A, and G:C→C:G. This leaves the G:C→A:T mutation unrepresented. To measure the rate of this final mutation, we designed a second fluctuation analysis experiment where we used p1-Short-U*L, encoding a *ura3** gene containing an A→G mutation at base 278. p1-Short-U*L was generated by inserting the integration cassette from plasmid 79 into p1. For function,

the codon in *URA3* corresponding to our A→G mutation must encode the amino acid lysine, based on previous studies [10, 16, 17]. Given this, the only way to restore *URA3* from our *ura3** is *via* the G:C→A:T mutation.

To perform fluctuation analyses, F102-2u strains containing plasmid 73 along with p1-Short-UL*(TGA) or p1-Short-U*L were first passaged until complete displacement of p1 occurred (see Yeast passaging section). Replicate cultures (15mL) were then inoculated at 1:10,000 and grown to saturation for 2-2.5 days in appropriate media (SC without uracil for p1-Short-UL*(TGA) and SC without leucine for p1-Short-U*L). Cultures were washed, spotted, titered, and counted as described (see Fluctuation analysis using p1-encoded *leu2 (Q180*)* section), except that cells containing p1-Short-U*L were plated on solid media lacking uracil.

Fluctuation analyses data were analyzed as described (see Fluctuation analysis using p1-encoded *leu2** section). The per-base substitution rates were calculated by dividing the phenotypic mutation rates by the measured copy number of p1-encoded *leu2 (Q180*)* (see Determination of copy number via quantitative PCR section) and by the number of ways *leu2 (Q180*)* can revert to *LEU2* (2.66 for the opal codon) or by the number of ways *ura3** can revert to *URA3* (0.33 for the A→G mutation at base 278). 95% confidence intervals were similarly scaled by these factors. In the case of p1-Short-U*L, the per-base substitution rate calculated (**Table 2.3**) is equal to the mutation rate of G:C→A:T. Sequencing of *URA3* from 6 revertent colonies, each from one replicate in the fluctuation analysis with p1-Short-U*L, showed the G→A mutation to be present in all cases. In the case of p1-Short-UL*(TGA), the per-base substitution rate calculated (**Table 2.3**) is more specifically the mean mutation rate of T:A→A:T, T:A→C:G, T:A→G:C, G:C→T:A, and

G:C→C:G, each weighted in proportion to the number of possible sites at which each mutation can occur and be detected. Sequencing of *LEU2* from 83 revertent colonies, each from one replicate in the fluctuation analysis with p1-Short-UL*(TGA), was used to determine the relative preferences of the T:A→A:T, T:A→C:G, T:A→G:C, G:C→T:A, and G:C→C:G mutations. To properly include the G:C→A:T mutation into preferences, the mutation rate measured from p1-Short-U*L was proportionately added. Full substitution mutation preferences are compiled as **Table 2.4**.

2.5 References

[1] 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. Journal of Bacteriology *147*, 155-160.

[2] Klassen, R., and Meinhardt, F. (2007). Linear Protein-Primed Replicating Plasmids in Eukaryotic Microbes. In Microbial Linear Plasmids, F. Meinhardt, R. Klassen, eds. (Berlin, Heidelberg: Springer-Verlag), pp. 187-226.

[3] Blanco, L., Bernad, A., Lazar, J.M., Martin, G., Garmendia, C., and Salas, M. (1989). Highly efficient DNA synthesis by the phage phi 29 DNA polymerase. J. Biol. Chem*. 264,* 8935-8940.

[4] Meijer, M.J., Horcajadas, J.A. and Salas, M. (2001). Phi29 family of phages. Microbiol. Mol. Biol. Rev*. 65,* 261-287.

[5] Longás, E., de Vega, M., Lázaro, J.M. and Salas, M. (2006). Functional characterization of highly processive protein-primed DNA polymerases from phages Nf and GA-1, endowed with a potent strand displacement activity. Nucl. Acids Res*. 34,* 6051-6063.

[6] Kämper, J., Esser, K., Gunge, N., and Meinhardt, F. (1991). Heterologous gene expression on the linear DNA killer plasmid from Kluyveromyces lactis. Current Genetics *19*, 109-118.

[7] Foster, P.L. (2006). Methods for determining spontaneous mutation rates. Methods in Enzymology *409*, 195–213.

[8] 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-Delbruck fluctuation analysis. Bioinformatics *25*, 1564-1565.

[9] Ma, W.T., Sandri, G. vH., and Sarkar, S. (1992). Analysis of the Luria-Delbrück distribution using discrete convolution powers. Journal of Applied Probability *29*, 255-267.

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

[11] Jung, G.H., Leavitt, M.C., and Ito, J. (1987). Yeast killer plasmid pGKL1 encodes a DNA polymerase belonging to the family B DNA polymerases. Nucleic Acids Research *15*, 9088.

[12] Uil, T.G., Vellinga, J., de Vrij, J., van den Hengel, S.K., Rabeling, M.J., Cramer, S.J., Eekels, J.J., Ariyurek, Y., van Galen, M., and Hoeben, R.C. (2011). Directed adenovirus evolution using engineered mutator viral polymerases. Nucleic Acids Research *39*, e30.

[13] Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison C.A. 3rd, Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6,* 343-345.

[14] Burker, D.J., Amberg, D.C. and Strathern, J.N. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*, 2005 edition, CSHL Press, New York.

[15] Frishman, F. (1975). *A Modern Course on Statistical Distributions in Scientific Work 17,* 401-406.

[16] Smiley, J.A and Jones, M.E. (1992). A unique catalytic and inhibitor-binding role for Lys93 of yeast orotidylate decarboxylase. Biochemistry *31,* 12162-12168.

[17] Miller, B.G. and Wolfenden, R. (2002). Catalytic proficiency: the unusual case of OMP decarboxylase. Annu. Rev. Biochem*. 71,* 847-885.

[18] Tan, S. (2001). A modular polycistronic expression system for overexpression protein complexes in *Escherichia coli*. Protein Expr. Purif. *21*, 224-234.

Chapter 3

Engineering highly error-prone orthogonal DNAPs for OrthoRep

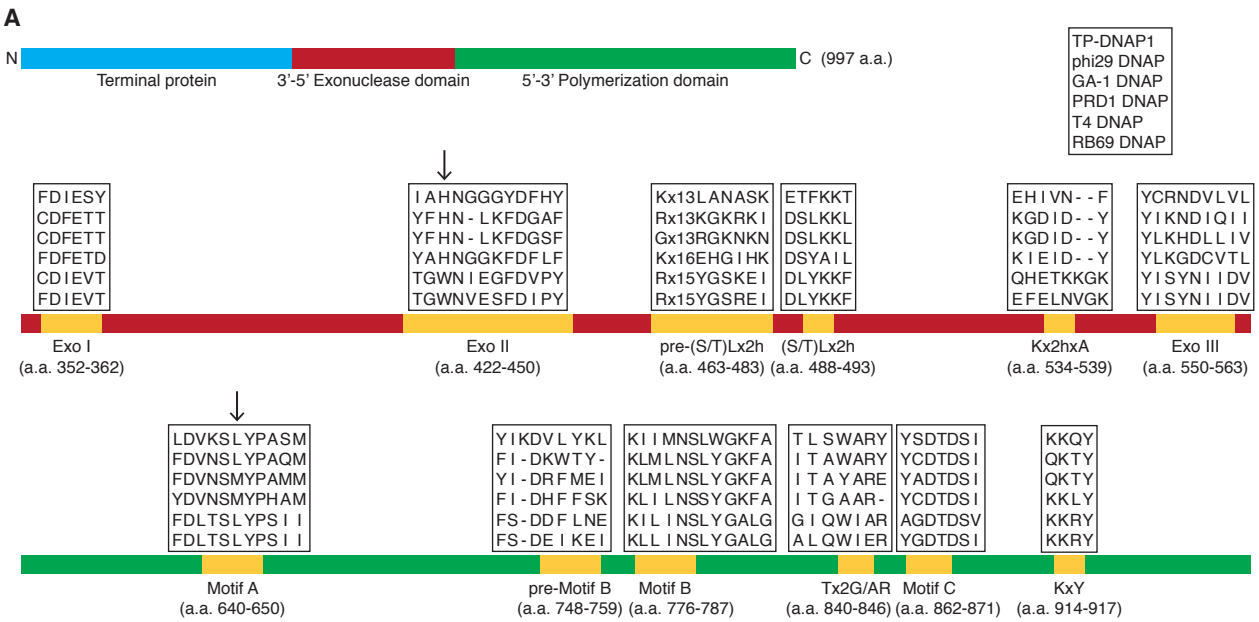
3.1 Screening single amino acid TP-DNAP1 mutants

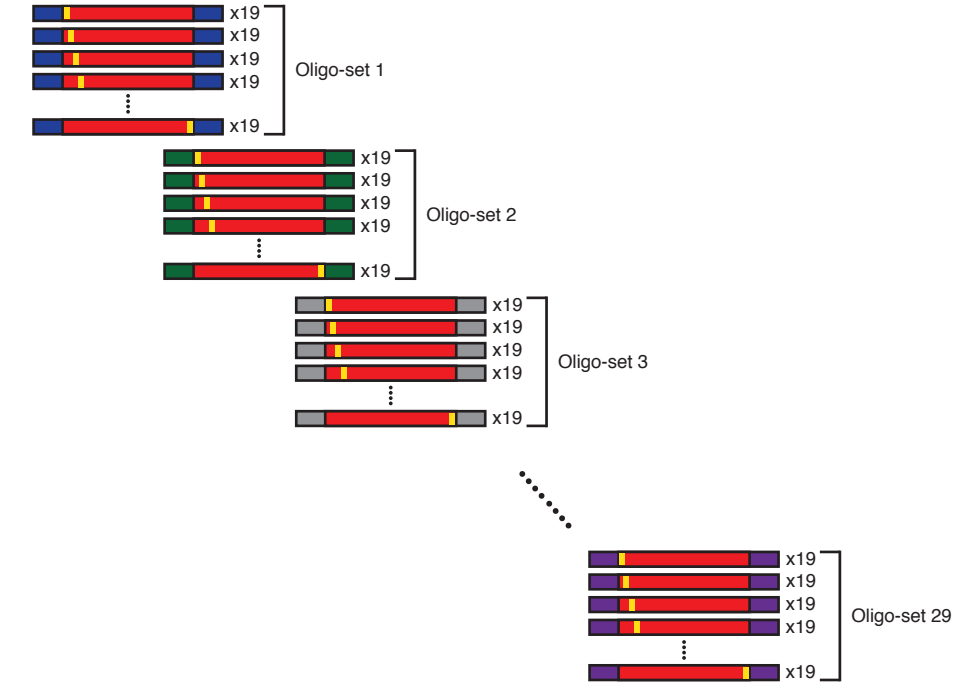
In order to enable accelerated directed evolution experiments in small culture volumes amenable to high-throughput serial passaging, we drastically increased the mutation rate of p1 replication by engineering highly error-prone variants of TP-DNAP1. First, we sought to find a collection of single amino acid mutant TP-DNAP1s with elevated mutation rates. We anticipated that single amino acid changes would yield modest mutators that could then serve as a basis set for building TP-DNAP1s with multiple mutations that act together to reduce fidelity. Indeed, in a study of Escherichia coli Pol I, a combination of three moderately fidelity-reducing mutations in the exonuclease domain, active site, and O-helix resulted in an 80,000-fold increase in Pol I's error rate [1]. Initial attempts to populate the TP-DNAP1 basis set, based on homology analysis to related family B DNAPs, (**Figure 3.1A**, **Appendix D**) mostly yielded low activity variants (**Appendix E**), due to the idiosyncrasies of protein-priming DNAPs. Mutators identified from this effort (referred to below as Rd1 mutants) seeded our basis set, but we pursued a comprehensive approach to find high-activity variants suitable for combination.

In the largest fidelity screen of a DNAP to date, ~14,000 clones from a scanning saturation mutagenesis library of TP-DNAP1 were assayed for p1 replication activity and mutation rate, from which a set of active mutators was found. To construct the TP-DNAP1 library, we obtained an oligonucleotide pool designed to evenly sample every single-substitution amino acid variant of TP-DNAP1 without degeneracy. Oligos were designed as

29 sets, each diversifying a 20-50 amino acid region flanked by ~25 bp constant regions (**Figure 3.1B-D**). We amplified each oligo set and assembled them into full-length TP-DNAP1 variants, yielding 29 sub-libraries. Plasmid sub-libraries were screened in the OrthoRep strain, OR-Y24. OR-Y24 contains recombinant p1 that lacks WT *TP-DNAP1*, and instead, encodes a standardized fluorescence reporter of p1 copy number (**Table 3.1**), and a disabled version of the *LEU2* selection marker (*leu2* (*Q180**)), which serves as a reporter of substitution mutation rate in fluctuation tests that measure reversion to functional *LEU2* (see **3.3 Methods** for details). We first implemented a weak selection for TP-DNAP1 activity in OR-Y24 to eliminate frame-shifted TP-DNAP1 variants, which were common due to errors in oligonucleotide synthesis (**Figure 3.2**). Purified yeast sub-libraries were plated on solid media and arrayed at 1-fold coverage, totaling 13,625 clones. (Sub-libraries 1-10, corresponding to the putative N-terminal TP of TP-DNAP1, which should not influence fidelity, were omitted.) All clones were expanded in quadruplicate and subject to p1 copy number and mutation rate measurements, using OR-Y24's p1-encoded reporters. We carried forward 95 promising candidates and measured their mutation rates more accurately through large-scale fluctuation tests. From this, we identified 41 unique variants (Rd2 mutants) with error rates as high as ~$2x10^{-7}$ s.p.b. (**Appendix E**). Unlike Rd1 mutants, Rd2 mutants retained high activity, and on average replicated p1 at only a 2-fold lower copy number than did WT TP-DNAP1. Only 9 of the Rd2 hits contained mutations at positions considered in the homology-based library design that generated Rd1 hits, indicating that fidelity determinants of TP-DNAP1 can lie outside of the most-conserved regions of DNAPs. Incidentally, we also discovered 210 TP-DNAP1 variants that replicated p1 at a higher copy number than did WT TP-DNAP1 (**Appendix F**), and added the mutation

Figure 3.1 | Design of TP-DNAP1 mutants by homology analysis and by construction of a scanning saturation mutagenesis library.
(A) The architecture of TP-DNAP1 consists of a fusion between the terminal protein, a 3'-5' proofreading exonuclease domain, and a DNA polymerization domain. Motifs responsible for fidelity in the exonuclease and proofreading domains are highlighted. A multiple sequence alignment between TP-DNAP1 and five closely related family B DNAPs is shown. In the larger homology study described in the main text, multiple sequence alignment with 99 closely related DNAPs (**Appendix D**) was used to identify positions that exhibit amino acid (a.a.) variation and are flanked by conserved residues (see **3.3 Methods** for details). Two candidate positions identified from this study are denoted with arrows. Amino acid variations found at these positions were transplanted into the corresponding location in TP-DNAP1. A total of 87 such TP-DNAP1 mutants were generated and screened in OR-Y24. Twenty-four of the TP-DNAP1 variants displayed elevated mutation rates, but almost 60% of these suffered from low activity, judging by the copy number of p1 (**Appendix E**). **(B)** A pool of ~19,000 oligonucleotides ranging in length from 130-200 nt were designed as 30 sets, each encoding a 20-50 amino acid variable region flanked by ~25 bp constant regions. Variable regions mutate each w.t. codon to 19 codons, representing all single amino acid substitutions. Each oligo set was PCR amplified and assembled with corresponding TP-DNAP1 plasmid backbones, yielding 30 full-length TP-DNAP1 plasmid sub-libraries. **(C)** The genetic code used for mutagenesis, chosen to maximize the codon adaptation index in *S. cerevisiae*. **(D)** Lengths of variable regions from each oligo set. Oligo sets 3 and 4, which were synthesized separately from the rest, have overlapping variable regions. Oligo sets 1-10 correspond to the putative TP portion of TP-DNAP1.





C

a.a.	Codon
A	GCT
R	AGA
N	AAC
D	GAC
C	TGT
Q	CAA
E	GAA
G	GGT
H	CAC
I	ATC
L	TTG
K	AAG
M	ATG
F	TTC
P	CCA
S	TCT
T	ACC
W	TGG
Y	TAC
V	GTT

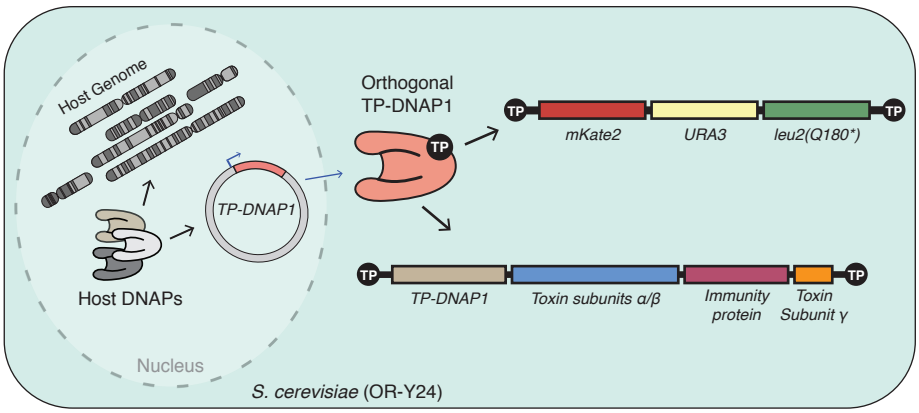
D

Oligo-set	Variable region (# a.a.)
1	30
2	30
3	50
4	50
5	30
6	30
7	30
8	30
9	30
10	30
11	35
12	30
13	30
14	20
15	30
16	50
17	30
18	30
19	30
20	45
21	25
22	50
23	30
24	30
25	45
26	20
27	50
28	50
29	50

TP-DNAP1	mKate2 fluorescence (a.u.)	Average *leu2 (Q180*)* copy number per cell	Ratio
WT	1455.5	68.9	0.047
	1650.14	88.1	0.053
I777K	1192.83	54.1	0.045
	1064.79	45.8	0.043
I777K, L900S	601.91	22.5	0.037
	617.55	38.6	0.063
L474W, L640Y,	298.97	8.63	0.029
I777K	180.05	11.1	0.062
L477V, L640Y,	288.24	16.9	0.058
I777K, W814N	233.35	15.6	0.067

Table 3.1 | Calibration curve of qPCR-determined p1 copy number to p1-encoded mKate2 fluorescence.
Five TP-DNAP1 variants were chosen to represent the range of p1 copy numbers observed across all experiments. OR-Y24 strains containing these TP-DNAP1s were grown in biological duplicates. Each biological duplicate was expanded in technical triplicates for mKate2 fluorescence measurements. In parallel, biological duplicates were expanded in single large volume cultures and subject to DNA extraction and qPCR measurements (see **3.3 Methods**). Data shown are mean fluorescence in arbitrary units (a.u.) and qPCR-determined p1 copy numbers. The ratio of copy number to mKate2 fluorescence is shown for each sample. The linear regression has low background and a strong fit ($y = 0.048x + 0.206$, $r^2 = 0.954$).

A



B

	% full-length clones (n)	
Selection Stringency	Sub-lib. 14	Sub-lib. 21
Initial outgrowth (SC-H)	70% (20)	35% (20)
Initial outgrowth (SC-UH)	94% (18)	76% (21)
Passage 1 (SC-UH)	n.d.	87% (16)
Passage 3 (SC-UH)	100% (5)	100% (17)
Passage 4 (SC-UH)	100% (6)	100% (7)

Figure 3.2 | Functional purification of TP-DNAP1 sub-libraries.
(A) A conceptual illustration of OR-Y24, which serves as parent strain for screening TP-DNAP1 variants. OR-Y24 contains a mixture of WT p1 and recombinant p1 encoding *mKate2*, *URA3*, and *leu2(Q180*)*. *mKate2* serves as a reporter of p1 copy number (**Table 3.1**) and *leu2(Q180*)* serves as a reporter of p1 substitution mutation rates. TP-DNAP1 mutants are encoded on a *CEN6/ARS4* plasmid and transformed into OR-Y24. If the resulting strain is subject to selection for URA3 in media lacking uracil, recombinant p1 increases in copy number and WT p1 decreases in copy number, because they compete for the same source of TP-DNAP1 (**Figure 2.7**). After several passages in media lacking uracil, WT p1 is fully cured, and then the strain is used for screening. If, however, an inactive TP-DNAP1 mutant is transformed into OR-Y24, then the resulting strain is driven to extinction under selection for URA3. This is because selection for recombinant p1 causes WT p1 to decrease in copy number, which decreases the amount of functional TP-DNAP1 available. This initiates a feedback loop that drops the copy number of both plasmids, leading to extinction (**Figure 2.7**). (Functional TP-DNAP1 variants expressed in trans from the nuclear *CEN6/ARS4* plasmid can replicate recombinant p1 at a constant level and thereby rescue growth.) **(B)** Due to a high fraction of non-functional frame-shifted variants present in the scanning saturation mutagenesis library of TP-DNAP1, OR-Y24 was used to functionally purify sub-libraries prior to screening. Two representative TP-DNAP1 sub-libraries, 14 and 21, were subject to purification. OR-Y24 cells transformed with sub-libraries 14 or 21 were split and inoculated into outgrowth conditions that were selective only for plasmid uptake (SC-H) or for plasmid uptake and URA3 expression (SC-UH). Selection in SC-UH was maintained for four 1:100 serial passages. After each passage, TP-DNAP1 plasmids were isolated from individual clones and subject to Sanger sequencing. Data shown are the percentage of (n) clones encoding full-length TP-DNAP1 variants.

from one of these variants to several low-activity mutator TP-DNAP1s to confirm the generality of the activity-boosting phenotype (**Table 3.2**). (These variants were not included in subsequent experiments here, but should prove useful in future TP-DNAP1 engineering efforts.) Rd2 hits were combined with Rd1 hits to form a 65-member basis set (**Figure 3.3A**).

3.2 Screening combinatorial libraries

Using this basis set, we designed, cloned, and screened combinatorial libraries in order to find highly error-prone TP-DNAP1s. To limit combinatorial diversity in our designs, we grouped basis set mutations according to their proximity to DNAP motifs known to affect fidelity (i.e., the A and C motifs in the palm domain, the B motif in the fingers domain, and the Exo I, II, and III motifs in the exonuclease domain) [2] and cloned only inter- motif combinations. We expected that synergy between inter-motif mutations from different domains (e.g., motif A mutations with motif B mutations) would yield super-additive or super-multiplicative reductions in fidelity, as observed with RB69 DNAP and E. coli Pol I, respectively [1, 3]. We screened a library of motif B mutants crossed with motif A and C mutants and found 46 mutators (**Appendix E**). The most error-prone of these 46 include three TP-DNAP1 variants (Rd3 mutants) with mutation rates of $\sim 1 \times 10^{-6}$ s.p.b., representing a ~400-fold increase over the WT TP-DNAP1 mutation rate and a ~10,000-fold increase over the yeast genomic mutation rate. We then crossed these Rd3 mutants with all of the exonuclease domain mutants from our basis set. After screening the resulting library, we obtained four hits (Rd4 mutants), including two highly error-prone variants, TP-DNAP1-4-1 (V574F, I777K, L900S) and TP-DNAP1-4-2 (L477V, L640Y, I777K, W814N), that replicate p1 at mutation rates of $\sim 7 \times 10^{-6}$ s.p.b. and $\sim 1 \times 10^{-5}$ s.p.b., respectively, and that both sustain

Fold-change of p1 copy number over WT		
	G410 (WT)	G410H
WT	1.0 ± 0.27	1.53 ± 0.33
N423R	0.59 ± 0.11	1.03 ± 0.21
N423Q	0.44 ± 0.22	0.93 ± 0.18
L640A	0.069 ± 0.015	1.19 ± 0.22
N423R, L640A	0.092 ± 0.056	0.24 ± 0.082

Table 3.2 | Mutation G410H broadly increases activity of TP-DNAP1 variants.
Mutation G410H was added to several low activity TP-DNAP1s. Mutants were transformed into OR-Y24 and subject to p1 copy number measurements. Data shown are mean fold-change ± standard deviation for biological triplicates.

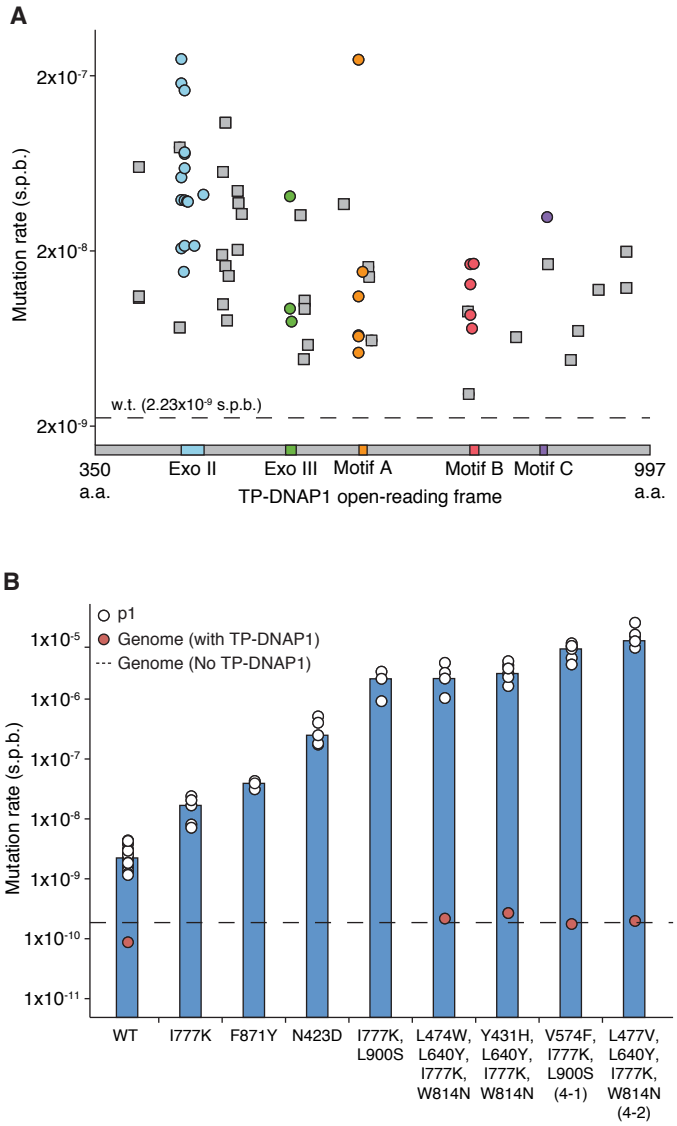


Figure 3.3 | Engineering of highly error-prone orthogonal DNAPs for OrthoRep. **(A)** Mutation rates of 65 basis set TP-DNAP1 variants found from a homology study and a TP-DNAP1 library screen. Variants are ordered by amino acid position in the TP-DNAP1 open-reading frame. Residues 1-350, corresponding to the putative TP domain, are not shown. Color-coding indicates regions known to determine fidelity. **(B)** Mutation rates of a representative panel of TP-DNAP1s and genomic substitution rates in the presence of highly error-prone variants. TP-DNAP1 substitution rates shown in **(A)** and **(B)** were measured with fluctuation tests using p1-encoded *leu2 (Q180*)*. Open circles in **(C)** represent measurements from independent fluctuation tests, and bars denote median measurements. Genomic substitution rates shown in **(C)** were determined for strains harboring p1 and each TP-DNAP1 variant as well as for the OrthoRep parent strain, AH22, which lacks p1 and TP-DNAP1. Genomic substitution rates were measured at the *URA3* locus in large-scale fluctuation tests and are shown as individual measurements. See **Appendix E** and **Table 3.3** for all mutation rate values, confidence intervals, and other information.

Substitution	L474W, L640Y, I777K, W814N	V574F, I777K, L900S	L477V, L640Y, I777K, W814N
A:T→T:A (transversion)	<1.2%	<1%	<1.1%
A:T→C:G (transversion)	<1.2%	<1%	<1.1%
A:T→G:C (transition)	65.2%	69%	73.1%
G:C→T:A (transversion)	<1.2%	1%	<1.1%
G:C→C:G (transversion)	1.2%	<1%	2.2%
G:C→A:T (transition)	29.9%	26.8%	21.3%

Table 3.3 | Substitution mutation preferences of highly error-prone TP-DNAP1s
All substitution preferences except for the G:C→A:T transition were measured from reversion of *leu2 (538C>T, 540A>G)*. The substitution rate of G:C→A:T was determined from fluctuation tests of *ura3 (278A>G)*. The per-base-normalized G:C→A:T substitution rates of TP-DNAP1 (L474W, L640Y, I777K, W814N), TP-DNAP1 (V574F, I777K, L900S), and TP-DNAP1 (L477V, L640Y, I777K, W814N) are 6.41×10^{-6} s.p.b, 1.73×10^{-5} s.p.b, and 1.77×10^{-5} s.p.b, respectively. These rates were incorporated in proportion to the individual error rates calculated for each of the other five substitutions (see **3.3 Methods** for details). Data shown are normalized percentages of each substitution mutation.

TP-DNAP1	Mutation rate (s.p.b.)	Lower 95% C.I. (s.p.b.)	Upper 95% C.I. (s.p.b.)	Number of replicates
None (AH22)	1.86×10^{-10}	1.34×10^{-10}	2.50×10^{-10}	187
WT	8.73×10^{-11}	3.47×10^{-11}	1.77×10^{-10}	90
L474W, L640Y, I777K, W814N	2.17×10^{-10}	7.80×10^{-11}	4.67×10^{-10}	90
Y431H, L640Y, I777K, W814N	2.68×10^{-10}	1.53×10^{-11}	1.18×10^{-9}	90
V574F, I777K, L900S	1.76×10^{-10}	4.37×10^{-11}	4.55×10^{-10}	90
L477V, L640Y, I777K, W814N	1.98×10^{-10}	1.40×10^{-10}	2.71×10^{-10}	186

Table 3.4 | Yeast genomic substitution mutation rates in the presence of OrthoRep. Per-base substitution rates were measured at the *URA3* locus in the presence of TP-DNAP1 variants. Genomic fluctuation tests were performed in the presence of p1 replication by TP-DNAP1 variants by selecting for a p1-encoded marker. AH22 is the parent OrthoRep strain and lacks p1 and TP-DNAP1. Data shown are individual mutation rate measurements, each with a corresponding 95% confidence interval (C.I.) and a count of the number of replicates performed for the fluctuation test.

a p1 copy number of ~5 (**Figure 3.3B**, **Appendix E**). Substitution mutation preferences of these highly error-prone variants are shown in **Table 3.3**. Additional rounds of library design and screening should reach even higher error rates, but these two Rd4 mutants are already exceptionally error-prone, so we ended our polymerase engineering effort at this point. As a practical guide, for facile generation of DNA libraries in vivo with TP-DNAP1-4-2, a 1-μL saturated yeast culture is theoretically sufficient for 1-fold coverage of all single mutants of a 1-kb gene and a 200-mL culture is sufficient for all double mutants. With mutational accumulation, highly diverse libraries can be generated in even smaller volumes: 1-fold coverage of all double mutants of a 1-kb gene can be achieved in a 650-μL culture with just 50 generations of propagation.

We found that the high p1 mutation rates driven by error-prone TP-DNAP1s remained completely stable for the longest duration tested (90 generations; **Appendix E**), and genomic mutation rates remain unchanged in the presence of p1 replication by the most error-prone TP-DNAP1, TP-DNAP1-4-2 (**Figure 3.3B**, **Table 3.4**). Therefore, OrthoRep can durably sustain in vivo mutagenesis with complete orthogonality (i.e., at least ~100,000-fold mutational targeting) to enable continuous evolution experiments.

3.3 Methods

Yeast strains. All *Saccharomyces cerevisiae* parent strains used in this chapter are listed in **Appendix G**. Parent strains do not include OR-Y24-derived strains containing TP-DNAP1 variants. Strains AH22 and F102-2 are described previously [4]. All genetic modifications that were made during strain construction were verified by sequencing and phenotyping.

DNA cloning. Plasmids used in this Chapter are listed in **Appendix G**. *E. coli* strain TG1 (Lucigen) was used for all of the DNA cloning steps. All primers used in this Chapter 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 [5].

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 3.1B**). The variable region consisted of a replacement of each amino acid in the WT 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 3.1C**). 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 [5]. 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 [5]. 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 [5]. 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 transformations were performed using the Yeastmaker Yeast Transformation System 2 kit (Clontech). For integrations into p1, 10μg of plasmid was linearized by digestion with ScaI, which generated blunt ends containing homologous regions as per the design of our integration cassettes. DNA fragments were transformed directly into p1-containing cells. Transformants were selected on the appropriate selective

solid SC media. The appearance of colonies following integration typically took 4-6 days of growth at 30°C. Genomic modifications were made using a CRISPR-Cas9 system for *S. cerevisiae* [6].

The following protocol modifications were made 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 as summarized in *Methods in Yeast Genetics* [7]. We note the following modifications: (1) cells were washed in 0.9% NaCl prior to treatment with Zymolyase (US Biological); (2) 200µg/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 treatments. Small-scale preparations, sufficient for culture volumes between 1-10mL, were performed by scaling down the large-scale protocol by 32-fold. Extracted DNA was still enough to be observed by agarose gel electrophoresis.

TP-DNAP1 homology analysis. A list of 99 homologs to TP-DNAP1 (EMBL accession number: CAA25568.1) was generated via protein BLAST [8] with default settings (**Appendix D**). A multiple sequence alignment of TP-DNAP1 to these homologs was performed using Clustal Omega [9] and the resulting alignment was analyzed using Jalview [10].

Amino acid mutations were selected based on three criteria. First, candidate positions should be flanked on both sides by residues with sequence alignment to >75% of homologs. Second, the TP-DNAP1 amino acid at a candidate position should be represented across >25% of homologs. Third, amino acids not present in TP-DNAP1 at a candidate position should be conserved across >25% of homologs. If these criteria were met, then amino acids identified from the third criterion were introduced at the candidate position in TP-DNAP1.

Small-scale p1 fluctuation tests. All three TP-DNAP1 libraries were screened through small-scale p1 fluctuation tests in the metastable OrthoRep strain, OR-Y24. OR-Y24 contains WT p1 and recombinant p1 that lacks WT *TP-DNAP1*, and instead, encodes a standardized fluorescence reporter of p1 copy number (**Table 3.1**), and a disabled version of the *LEU2* selection marker (*leu2* (*Q180**)). As described in Chapter 2, *leu2 (Q180*)* contains a C→T mutation at base 538 in *LEU2* at a site permissive to all single point mutants that generate missense mutations. Reversion to functional *LEU2* can be detected on medium lacking leucine.

Generally, to screen TP-DNAP1 mutants, OR-Y24 strains were transformed with TP-DNAP1 plasmids and the resulting yeast strains were passaged 3-4 times at 1:100 dilutions in SC-UH to fully cure WT p1 (**Figure 3.2**). Cured strains were diluted 1:10,000 into selective SC media for fluctuation tests. Selective SC media used for p1 fluctuation tests lacked uracil, histidine and tryptophan, and was adjusted to pH 5.8 with NaOH (SC-UHW, pH 5.8). Absence of tryptophan and pH adjustment inhibited growth on reversion medium resulting from nonsense suppression of *leu2 (Q180*)*. Dilutions were split into three 100 µL cultures and one 200 µL culture in 96-well trays, and cultures were grown to saturation for

2-2.5 days. Saturated 200 µL cultures were subject to a copy number measurement, as described below. The remaining three replicates were washed and resuspended in 35 µL 0.9% NaCl. 10 µL was spot plated onto solid SC medium selective for *LEU2* revertants. Solid SC medium used for p1 fluctuation tests lacked uracil, histidine, tryptophan and leucine and was adjusted to pH 5.8 with NaOH (SC-UHLW, pH 5.8). Plates were incubated at 30 °C for 5-6 days, and afterwards, colony-count was determined for each spot. Colony-count data were analyzed by the MSS maximum-likelihood estimator method [11]. Measuring cell titers was infeasible due to the large number of strains, so the average number of cells per culture was assumed to remain constant. Relative phenotypic mutation rates were calculated by normalization to p1 copy number.

Prior to screening the TP-DNAP1 scanning saturation mutagenesis library, a functional purification was imposed in OR-Y24 to eliminate frame-shifted TP-DNAP1 variants, which were common due to errors in oligonucleotide synthesis. The pilot study shown in **Figure 3.2** confirmed that sub-libraries transformed into OR-Y24 are enriched for full-length TP-DNAP1 variants after two passages in SC media lacking uracil and histidine (SC-UH). To purify the entire scanning saturation mutagenesis library, the remaining 27 TP-DNAP1 plasmid sub-libraries were individually transformed into OR-Y24, and the resulting yeast sub-libraries were passaged twice at 1:100 dilutions in SC-UH. Passaged yeast sub-libraries were individually plated on solid SC medium lacking histidine. For each sub-library, 24 colonies were propagated in small cultures of SC-UH, in order to verify that >90% of clones robustly grow under selection for p1 replication. Afterwards, purified yeast sub-libraries were plated on solid media and colonies from each were individually inoculated into small cultures of SC-UH at ~1-fold coverage of theoretical sub-

library diversity. This resulted in a total of 13,625 clones. (This does not include sub-libraries 1-10, which correspond to the putative N-terminal TP of TP-DNAP1, which should not influence fidelity. These sub-libraries were cloned and purified, but omitted from the screen.) The arrayed clones were then cured of WT p1 and subject to small-scale p1 fluctuation tests, as described above. Then, 376 clones with the highest relative phenotypic mutation rates were subject to an additional small-scale p1 fluctuation test with six replicates. TP-DNAP1 expression vectors were isolated from 95 yeast clones with the highest relative phenotypic mutation rates and subject to Sanger sequencing. These TP-DNAP1s were characterized with large-scale p1 fluctuation tests, as described below.

Shuffling libraries were screened in a similar manner. From the first combinatorial library, 1520 yeast clones were subject to small-scale p1 fluctuation tests. 188 clones were subject to additional small-scale p1 fluctuation tests and isolated for Sanger sequencing. 46 unique variants with the highest relative phenotypic mutation rates were characterized with large-scale p1 fluctuation tests, as described below. From the second combinatorial library, 744 yeast clones were screened and 58 clones were characterized with additional small-scale p1 fluctuation tests. Four clones were extracted, subject to Sanger sequencing and characterized with large-scale p1 fluctuation tests, as described below.

Large-scale p1 fluctuation tests. Large-scale fluctuation tests of p1-encoded *leu2 (Q180*)* were performed to precisely determine per-base substitution rates for individually cloned or isolated TP-DNAP1 variants. Large-scale p1 fluctuation tests are performed similarly to small-scale p1 fluctuation tests, with several modifications. First, large-scale p1 fluctuation tests were typically performed with 36-48 replicates. For highly error-prone TP-DNAP1s obtained from later rounds of screening, fewer replicates (3-16) were used, which is

sufficient for similar precision [11]. Second, p1 copy number was determined by the flow cytometry method, described below. Third, cell titers were measured for each fluctuation test to estimate the average number of cells per culture. Cell resuspensions were diluted and plated on solid SC-UH medium, and colony counts were determined after incubation at 30 °C for 2-3 days. Alternatively, cell resuspensions were diluted and subject to an event-count measurement via flow cytometry. Fourth, inoculums of highly error-prone TP-DNAP1s occasionally contained preexisting mutants, despite the 1:10,000 dilution, so mutant frequencies were estimated by plating precultures on solid SC-UHLW, pH 5.8 medium. Plates were incubated for 2-3 days, in parallel with cultures grown for fluctuation tests. Preculture mutant titers were counted to estimate the number of replicates in the fluctuation test expected to contain preexisting mutants (n). Revertants were counted from all replicates of the fluctuation tests, counts were sorted, and n replicates with the highest counts were omitted from calculations.

To calculate per-base substitution rates, large-scale fluctuation data were analyzed by the maximum likelihood method, implemented using newton.LD.plating in rSalvador 1.7 [12]. Phenotypic mutation rates were calculated by normalizing to the average number of cells per culture. Phenotypic mutation rates were divided by the measured p1 copy number and by the number of ways *leu2 (Q180*)* can revert to *LEU2* (2.33 for the ochre codon) to yield per-base substitution rates. 95% confidence intervals were similarly scaled by these factors. All data related to large-scale p1 fluctuation tests are provided in **Appendix E**.

p1 copy number assay. A calibration curve was established to correlate p1 copy number, determined via quantitative PCR (qPCR), with fluorescence of p1-encoded *mKate2* (**Table 3.1**). Five TP-DNAP1 variants representing a wide range of copy numbers were

transformed into OR-Y24 and passaged until WT p1 was displaced. The five OR-Y24 strains were grown to saturation, diluted 1:10,000 to mimic p1 fluctuation tests, and grown in triplicate 100 µL cultures and duplicate 40 mL cultures. 100 µL saturated cultures were subject to fluorescence measurement of mKate2 (ex/em = 561 nm/620 nm, bandwidth = 15 nm) on a flow cytometer (Invitrogen Attune NxT). Whole-cell DNA extracts were prepared from 40 mL cultures and used as templates for qPCR measurement of p1-encoded *leu2 (Q180*)*, as described in **2.4 Methods**. p1-encoded *LEU2* was PCR amplified with qPCR-Leu2F and qPCR-Leu2R; and genomic *LEU3* was PCR amplified with qPCR-Leu3F and qPCR-Leu3R using SyBR Green (Fisher Scientific) master mix (see **2.4 Methods** for oligo sequences). The correlation of mKate2 fluorescence and qPCR-determined p1 copy number had a strong linear fit across p1 copy numbers ranging from 9-90, and had low background ($y = 0.048x + 0.206$, $r^2 = 0.954$).

To assay p1 copy number for large-scale p1 fluctuation tests, additional replicates of OR-Y24 strains were grown from the 1:10,000 dilution in triplicate and saturated cultures were subject to fluorescence measurement via flow cytometry. Fluorescence measurements were converted to copy number with the linear calibration curve.

To assay relative p1 copy number for small-scale p1 fluctuation tests, an additional 200 µL culture of OR-Y24 strains were grown from the 1:10,000 dilution. 200 µL saturated cultures were subject to an OD_{600} measurement and fluorescence measurement of mKate2 (ex/em = 588 nm/633 nm), using a microplate reader (TECAN Infinite M200 PRO). A linear relationship was assumed between p1 copy number and OD_{600}-normalized mKate2 fluorescence. Copy numbers were calculated from normalization to a WT TP-DNAP1 control.

For a few large-scale p1 fluctuation tests, p1 copy number was assayed by the method of small-scale p1 fluctuation tests (i.e. microplate reader measurement). For these experiments, mutant TP-DNAP1 p1 copy numbers were calculated by normalization to a WT TP-DNAP1 control. The copy number of this control was assumed to be the average WT TP-DNAP1 copy number from all large-scale p1 fluctuation tests that used flow cytometry measurements.

Characterization of high activity TP-DNAP1s. From mKate2 measurements of the 13,625 clones screened with small-scale p1 fluctuation tests, 283 clones exhibiting high fluorescence were chosen for additional characterization. TP-DNAP1 plasmids were isolated from these strains and subject to Sanger sequencing. 210 unique variants were identified and the corresponding plasmids were re-transformed into OR-Y24. Transformed strains were passaged in SC-UH until WT p1 was displaced, and subject to p1 copy number measurement in triplicate. Four high copy variants were directly subject to qPCR measurements for validation (unpublished results). To test the suppressor activity of the G410H mutation, which yields increased p1 copy number, this amino acid change was added to several low-activity TP-DNAP1 variants. p1 copy number was similarly assayed for these strains.

Characterization of mutational preferences. Two additional fluctuation tests, coupled with sequencing, were used to determine mutational preferences of TP-DNAP1s, as described in Chapter 2. In one experiment, strain AR-Y302 was used for fluctuation tests. AR-Y302 contains recombinant p1 encoding *leu2 (Q180*)*, which contains two substitutions (538C>T and 540A>G), which create an amber nonsense mutation (TAG) in *LEU2*.

Fluctuation tests can be used to determine the frequency of reversion from *leu2 (Q180*)* to *LEU2*. Single point mutations that restore *LEU2* from *leu2 (Q180*)* include T:A→A:T, T:A→C:G, T:A→G:C, G:C→T:A and G:C→C:G. This leaves the G:C→A:T mutation unrepresented. To measure the rate of this final mutation, we used strain AR-Y401, which contains recombinant p1 encoding *ura3 (K93R)*. This disabled allele contains a 278A>G substitution and the only way to restore *URA3* is via the G:C→A:T mutation [13, 14].

Strains of AR-Y302 containing TP-DNAP1 variants were subject to large-scale p1 fluctuation tests. p1 plasmids were extracted from 50-70 revertant colonies, each from an independent replicate of the fluctuation test. The restored *LEU2* alleles were PCR amplified and subject to Sanger sequencing. To calculate individual per-base pair mutation rates, the corresponding per-base substitution rate (measured from fluctuation tests of the corresponding TP-DNAP1 in OR-Y24) was used. Per-base substitution rates calculated from reversion of *leu2 (Q180*)* is more specifically, the sum of the mutation rates of T:A→A:T, T:A→C:G, and T:A→G:C, weighted in proportion to the number of possible sites at which each mutation can occur and be detected. The relative preferences for these mutations, determined from Sanger sequencing, was used to calculate per-base pair mutation rates.

The G:C→A:T mutation rate is equal to the per-base mutation rate calculated from fluctuation tests of TP-DNAP1s expressed in AR-Y401. Fluctuation tests were performed and analyzed by the large-scale protocol described above. The G:C→A:T mutation rate was proportionally added to the full substitution spectra.

Genomic orthogonality measurements. Fluctuation analyses on the genomic *URA3* gene were performed to determine genomic per-base substitution rates, as previously described [15]. Strains were grown to saturation in SC media lacking uracil, histidine and tryptophan,

in order to select for genomically expressed URA3 (to purge preculture mutants), HIS3 expressed from the *CEN6/ARS4* plasmid encoding TP-DNAP1 mutants, and p1-expressed TRP1 (to ensure that p1 is replicated). Saturated strains were diluted 1:5,000 into SC media lacking histidine and tryptophan, and split into replicate 200μL cultures in 96-well trays. Cultures were grown for 2-2.5 days until saturation. Cultures were washed with 400μL 0.9% NaCl and resuspended in 210μL 0.9% NaCl. Four cultures from each strain were pooled, diluted, and titered on solid SC media lacking histidine and tryptophan. 200μL of all remaining cultures were spot plated onto solid SC media lacking histidine and tryptophan and supplemented with 5-FOA (1g/L), and spots were allowed to dry before incubation. Plates were incubated at 30°C. Colonies were counted from titer plates after 2 days, and from spot plates after 5 days.

Genomic fluctuation data were analyzed by the maximum likelihood method, as described for large-scale p1 fluctuation tests. Phenotypic mutation rates were divided by the target size for loss of function of *URA3* via base pair substitution [15], to yield per-base substitution rates. 95% confidence intervals were similarly scaled by these factors. All data related to genomic fluctuation tests are fully described in **Table 3.4**.

3.4 References

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

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

[3] 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.

[4] 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.

[5] 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.

[6] 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

[7] Burker, D.J., Amberg, D.C. and Strathern, J.N. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*, 2005 edition, CSHL Press, New York.

[8] 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.

[9] 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.

[10] 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.

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

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

[13] Smiley, J.A and Jones, M.E. (1992). A unique catalytic and inhibitor-binding role for Lys93 of yeast orotidylate decarboxylase. Biochemistry *31,* 12162-12168.

[14] Miller, B.G. and Wolfenden, R. (2002). Catalytic proficiency: the unusual case of OMP decarboxylase. Annu. Rev. Biochem*. 71,* 847-885.

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

Chapter 4

Crossing mutation-induced genomic error thresholds with OrthoRep

4.1 Crossing the haploid extinction threshold

OrthoRep can access and sustain mutation rates that untargeted genome mutagenesis cannot. There is a theoretically predicted and empirically observed inverse relationship between the length of an information-encoding polymer, such as a gene or genome, and the tolerable error rate of replication [1-4]. At sufficiently high mutation rates, essential genetic information is destroyed every generation, guaranteeing extinction, and even moderately elevated mutation rates can erode fitness [5, 6]. Continuous directed evolution systems fundamentally work by targeting mutagenesis to desired genes in order to bypass the low error thresholds of large cellular genomes, but existing systems still elevate genome-wide mutation rates of cells or phages, falling short of a complete bypass [7-14]. Since OrthoRep is fully orthogonal to genomic replication, it achieves the complete bypass of genomic error thresholds for genes of interest, which should result in the ability to run *in vivo* continuous evolution for indefinitely large numbers of generation at mutation rates that are exclusively limited by the thresholds of user-selected genes.

In order to demonstrate the limitations of genomic error thresholds on continuous evolution, we experimentally applied high mutation rates to the host genome. This was done by transplanting previously discovered mutations that increase the substitution mutation rate of POL3 [5], the primary yeast lagging strand DNAP, into WT or mismatch repair-deficient (*Δmsh6*) versions of AH22, the parent of OrthoRep strains. Mutator phenotypes, verified by fluctuation tests at a genomic locus, were accompanied by severe

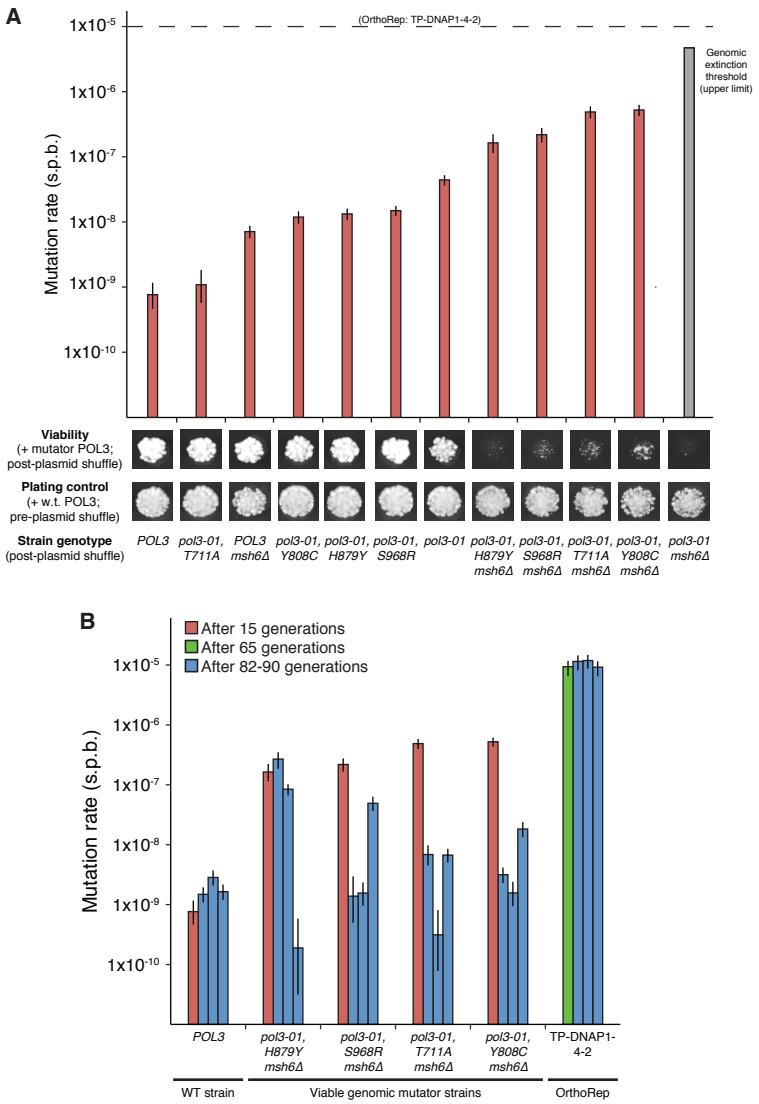


Figure 4.1 | OrthoRep crosses mutation-induced error thresholds of the yeast genome. **(A)** A series of yeast genomic mutator strains spanning mutation rates from WT to the extinction threshold (upper limit). OrthoRep's parent strain, AH22, was modified to express its genomic WT *POL3* from a plasmid, and *POL3* variants were introduced into WT or mismatch repair-deficient (*Δmsh6*) versions of this strain via plasmid shuffle. WT *POL3* is retained in pre-plasmid shuffle plating controls. Genomic mutation rates were measured ~15 generations after plasmid shuffle. The projected mutation rate of the inviable *pol3-01*, *Δmsh6* strain was calculated as the mutational increases due to *pol3-01* (58-fold) and *Δmsh6* mutations (106-fold, averaged across genotypes). The proofreading deficient *pol3-01* allele encodes *POL3 (D321A, E323A)*. T711A, Y808C, H879Y, and S968R are suppressor mutations that reduce the error-rate of *pol3-01*. **(B)** Mutational stability of viable genomic mutator strains versus OrthoRep. Strains harboring *POL3* variants or OrthoRep with TP-DNAP1-4-2 were passaged in triplicate for 82 or 90 generations, respectively. Afterward, genomic or OrthoRep substitution mutation rates were measured at the genomic *CAN1* locus or with p1-encoded *leu2 (Q180*)*, respectively. Data shown in **(A)** and **(B)** are individual measurements with 95% confidence intervals shown as error bars.

growth defects and led to immediate extinction in the case of *pol3-01, Δmsh6* AH22 (**Figure 4.1A**). In agreement with a previous estimate [5], the projected mutation rate imposed in this nonviable AH22 strain was 4.72×10^{-6} s.p.b., calculated from the individual contribution of *pol3-01* and the average effect of *MSH6* loss. This mutation rate is presumed to exceed the haploid yeast error-induced extinction threshold thereby killing the host cell [5]. Since replication of p1 by TP-DNAP1-4-2 occurs at a higher mutation rate than 4.72×10^{-6} s.p.b., we conclude that OrthoRep can stably exceed categorical mutation rate limits on replicating cellular genomes.

4.2 Crossing intermediate genomic error thresholds

We also asked whether viable genomic mutator strains could sustain mutagenesis. Four AH22 strains with mutation rates of 1.64×10^{-7}–5.24×10^{-7} s.p.b. were propagated for 82 generations in triplicate, and afterward, a clone from each was subject to genomic mutation rate measurements via fluctuation tests (**Figure 4.1B**). Across replicates, the mutation rate dropped an average of 284-fold (**Figure 4.1B**), likely due to suppressor mutants that alleviated deleterious genome mutagenesis and overtook the population [5]. In contrast, mutagenesis on p1 remained constant (**Figure 4.1B**). This indicates that in durations relevant to directed evolution experiments, even moderate genome mutagenesis is unsustainable whereas continuous mutagenesis in OrthoRep is sustainable.

4.3 Methods

Yeast strains. All *Saccharomyces cerevisiae* parent strains used in this chapter are listed in **Appendix G**. Parent strains do not include AR-Y432-derived strains containing POL3

variants. Strains AH22 and F102-2 are described previously [15]. All genetic modifications that were made during strain construction were verified by sequencing and phenotyping.

To construct *POL3* mutator strains, plasmids 19-25 (see **DNA cloning** section below and **Appendix G**) were transformed into AR-Y432 and AR-Y445. Transformants were expanded in selective SC medium and spot plated on selective SC medium or selective SC medium supplemented with 5-FOA (1 g/L) for plasmid shuffle of plasmid 18 (see **DNA cloning** section below and **Appendix G**) via *URA3* counter-selection (16).

DNA cloning. Plasmids used in this study are listed in **Appendix G**. *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 were assembled by the Gibson method [17].

Yeast transformation. All transformations were performed using the Yeastmaker Yeast Transformation System 2 kit (Clontech). For integrations into p1, 10μg of plasmid was linearized by digestion with ScaI, which generated blunt ends containing homologous regions as per the design of our integration cassettes. DNA fragments were transformed directly into p1-containing cells. Transformants were selected on the appropriate selective solid SC media. The appearance of colonies following integration typically took 4-6 days of growth at 30°C. Genomic modifications were made using a CRISPR-Cas9 system for *S. cerevisiae* [18].

Yeast DNA extraction. Whole-cell DNA extractions followed the yeast DNA miniprep procedure as summarized in *Methods in Yeast Genetics* [19]. We note the following modifications: (1) cells were washed in 0.9% NaCl prior to treatment with Zymolyase (US

Biological); (2) 200μg/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 treatments. Small-scale preparations, sufficient for culture volumes between 1-10mL, were performed by scaling down the large-scale protocol by 32-fold. Extracted DNA was still enough to be observed by agarose gel electrophoresis.

Genomic fluctuation tests of *POL3* mutator strains. Fluctuation tests of the genomic *CAN1* gene were performed to determine genomic per-base substitution rates. To minimize propagation of *POL3* mutator strains, fluctuation tests were performed directly on colonies from plasmid shuffle plates. For each strain, 48 colonies from plasmid shuffle plates were individually scraped and resuspended in 120 μL 0.9% NaCl. 10 μL from each resuspension was diluted and subject to an event counts measurement via flow cytometry. This was done to identify colonies of similar cell count, because fluctuation tests are only appropriate when final population sizes for all replicates are similar. 24 resuspensions with similar event counts were used for fluctuation tests. 90 μL from the 24 resuspensions were mixed and plated onto SC medium lacking arginine and supplemented with 10X canavanine (0.6 g/L). 10 μL from four of the 24 resuspensions was pooled, diluted and titered on solid SC medium. Plates were incubated at 30 °C. Colonies were counted from titer plates after 2-4 days and from spot plates after 3-6 days. Based on titer counts, the average number of generations that occurred during colony formation was ~15.

To calculate per-base substitution rates, fluctuation data were analyzed by the maximum likelihood method, implemented using newton.LD.plating in rSalvador 1.7 [20]. Phenotypic mutation rates were calculated by normalizing to the average number of cells per culture. Phenotypic mutation rates were divided by the target size for loss of function

of *CAN1* via base pair substitution [21], to yield per-base substitution rates. 95% confidence intervals were similarly scaled by these factors. The Fenton approximation [22] was used to calculate the predicted rate of the extinct mutator strain.

Stability tests of *POL3* mutator strains. To test the stability of mutator phenotypes, three colonies from each plasmid shuffle plate were passaged ten times at 1:100 dilutions (67 generations). A single clone from each final population was isolated and subject to *CAN1* fluctuation tests, as described above.

Large-scale p1 fluctuation tests. All individual p1 mutation rate measurements, confidence intervals, and related information are provided in **Appendix C** and were measured according to the protocols detailed in **3.3 Methods**.

4.4 References

[1] 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.

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

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

[4] 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.

[5] 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.

[6] 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.

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

[8] 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.

[9] 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.

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

[11] 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.

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

[13] 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.

[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Burker, D.J., Amberg, D.C. and Strathern, J.N. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*, 2005 edition, CSHL Press, New York.

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

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

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

Chapter 5

High-throughput directed evolution of *Plasmodium falciparum* DHFR

5.1 Directed evolution of *Pf*DHFR drug resistance using OrthoRep

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 [1-6]. 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.

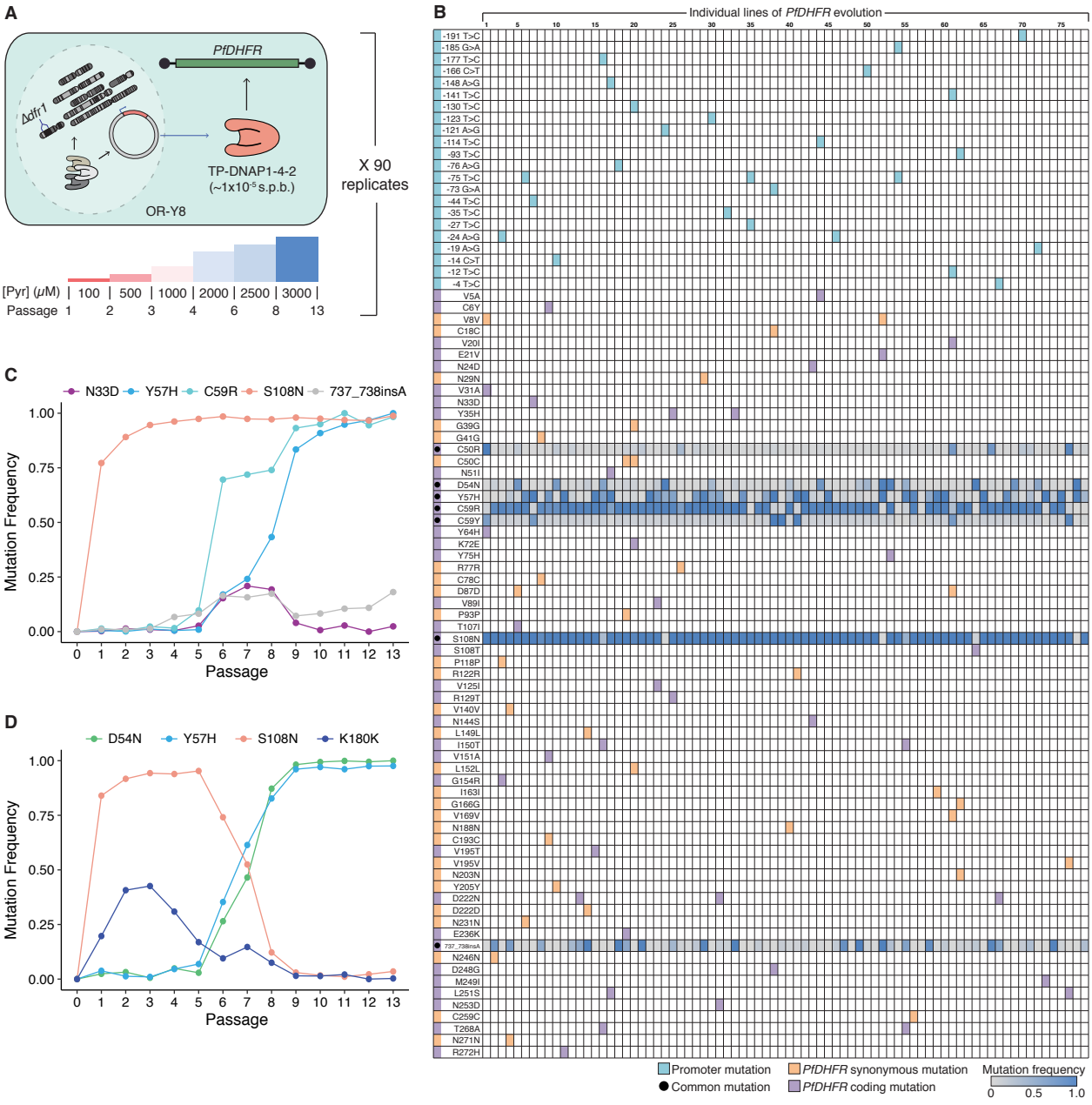
We used OrthoRep to evolve *Pf*DHFR resistance to pyrimethamine in 90 independent 0.5 mL cultures (**Figure 5.1A**). Based on a well-established yeast model of *Pf*DHFR, we constructed transgenic yeast strains that lack endogenous DHFR and depend on p1-encoded *Pf*DHFR. These strains acquired sensitivity to pyrimethamine and in pilot studies, evolved resistance by accumulating mutations in *Pf*DHFR (**Figure 5.2**). We found that resistance arose more commonly and successfully as the mutation rate of p1 was increased (**Figure 5.2**), 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 5.1A**). 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 5.1**).) From Sanger sequencing analysis of bulk adapted populations (**Figure 5.1B**; see **5.3 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 [1-3, 7]. 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 [8].

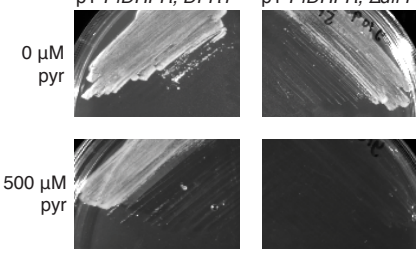
5.2 New adaptive pathways to resistance

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 5.1B**), including 737_738insA, which creates an adaptive C-terminal truncation (**Figure 5.3A**). The two most common mutations, C59R and S108N, occur together in 62/78 adapted populations

Figure 5.1 | High-throughput directed 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 hr. 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 **5.3 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)**.


A
PfDHFR
Δdfr1
TP-DNAP1-4-2
(~1x10-5 s.p.b.)
OR-Y8
X 90
replicates
[Pyr] (μM) | 100 | 500 | 1000 | 2000 | 2500 | 3000 |
Passage 1 2 3 4 6 8 13
B
Individual lines of PfDHFR evolution
-191 T>C
-185 G>A
-177 T>C
-166 C>T
-148 A>G
-141 T>C
-130 T>C
-123 T>C
-121 A>G
-114 T>C
-93 T>C
-76 A>G
-75 T>C
-73 G>A
-44 T>C
-35 T>C
-27 T>C
-24 A>G
-19 A>G
-14 C>T
-12 T>C
-4 T>C
V5A
C6Y
V8V
C18C
V20I
E21V
N24D
N29N
V31A
N33D
Y35H
G39G
G41G
C50R
C50C
N51I
D54N
Y57H
C59R
C59Y
Y64H
K72E
Y75H
R77R
C78C
D87D
V89I
P93P
T107I
S108N
S108T
P118P
R122R
V125I
R129T
V140V
N144S
L149L
I150T
V151A
L152L
G154R
I163I
G166G
V169V
N188N
C193C
V195T
V195V
N203N
Y205Y
D222N
D222D
N231N
E236K
737_738insA
N246N
D248G
M249I
L251S
N253D
C259C
T268A
N271N
R272H
Promoter mutation
PfDHFR synonymous mutation
Mutation frequency
Common mutation
PfDHFR coding mutation
0 0.5 1.0
C
N33D Y57H C59R S108N 737_738insA
Mutation Frequency
Passage
D
D54N Y57H S108N K180K
Mutation Frequency
Passage

A



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 5.2 | Pilot studies of p1-*Pf*DHFR evolution.
(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 *PfDHFRs* 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. OD_{600} 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.

Index of extinct population in 96-well tray		
Original experiment	Revival experiment 1	Revival experiment 2
1	51	39
9	63	51
32	66	63
33	84	88
34		90
40		92
42		
49		
50		
55		
57		
90		

Table 5.1 | Stochastic extinction in revival experiments of *Pf*DHFR evolution.
Indices of populations that went extinct during the 90-replicate *Pf*DHFR evolution experiment and during revival experiments. In the revival experiments, cultures were inoculated from glycerol stocks of passage 5, at which point all 90 populations grew robustly. Cultures were revived in SC media supplemented with 2.5 mM pyrimethamine and passaged with the same protocol used for the original evolution experiment.

(**Figure 5.1B**). Although these mutations are present in qm-wild, only one population accumulated a third mutation from the qm-wild peak (N51I; replicate 17 in **Figure 5.1B**). 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 5.3A**). 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 5.3A**). 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

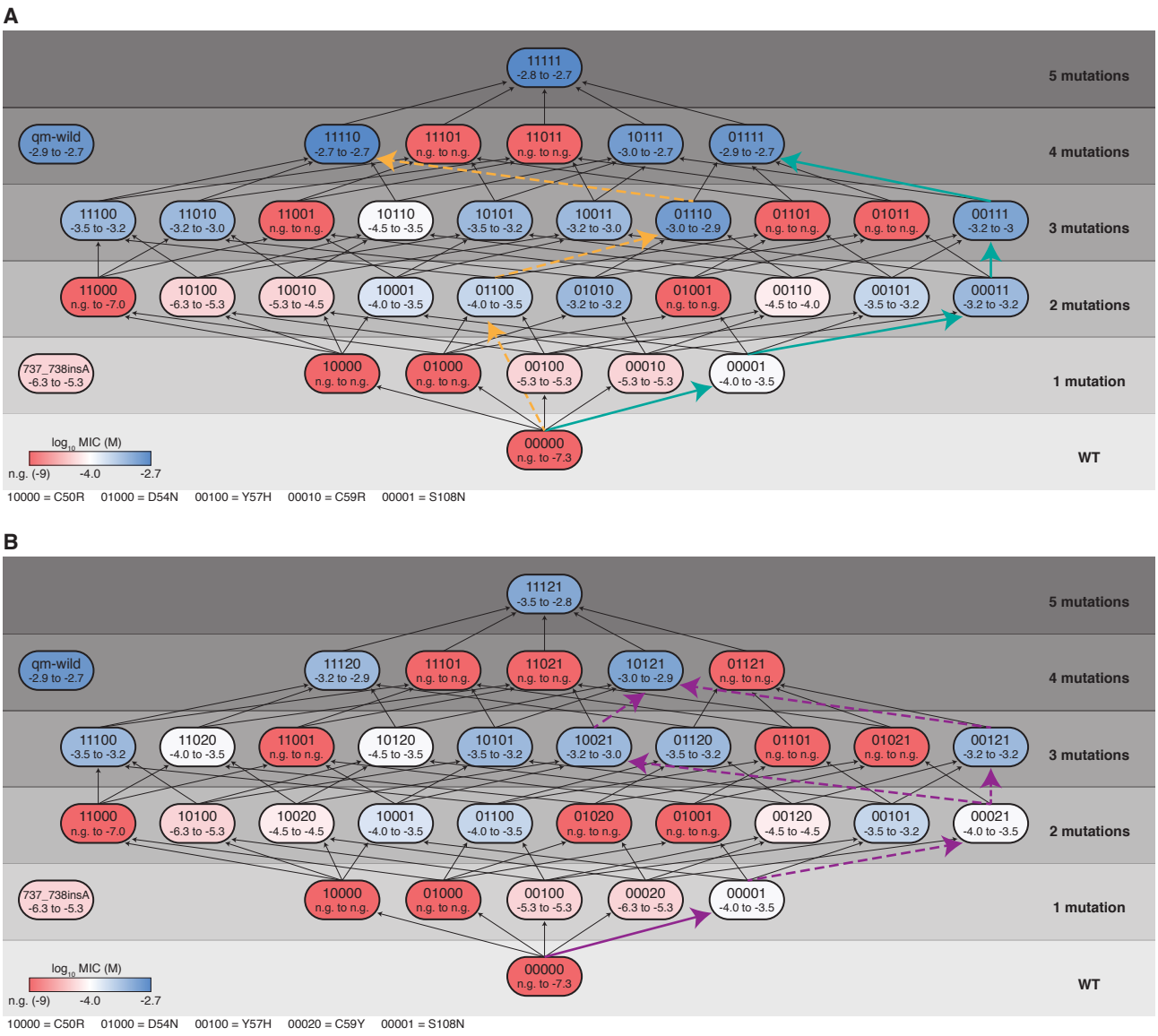


Figure 5.3 | Five-mutation *Pf*DHFR fitness landscapes.
(A) A fitness map of a five-mutation *Pf*DHFR landscape defined by C50R, D54N, Y57H, C59R, and S108N. Black arrows show all theoretically possible single-mutation steps in this landscape. The common S108N-dependent pathway is highlighted with a 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.

populations (**Figures 5.1C, 5.4**), and blocked access to the 96 out of 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 both present (**Figure 5.3A**). (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 5.3A**). 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 5.3A**). Three populations in our experiment avoid mutation at S108 (**Figure 5.1B**) 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 5.1D**). One of these replicates additionally fixed C59R to reach 01110, the triple mutant with the highest MIC (replicate 24 in **Figure 5.1B**). Stronger selection for pyrimethamine resistance, if feasible, should also fix C50R and lead to 11110 (**Figure 5.3A**).

Since 11110 is suppressed by rapid fixation of S108N, weaker early selection or greater population structure [9, 10], 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

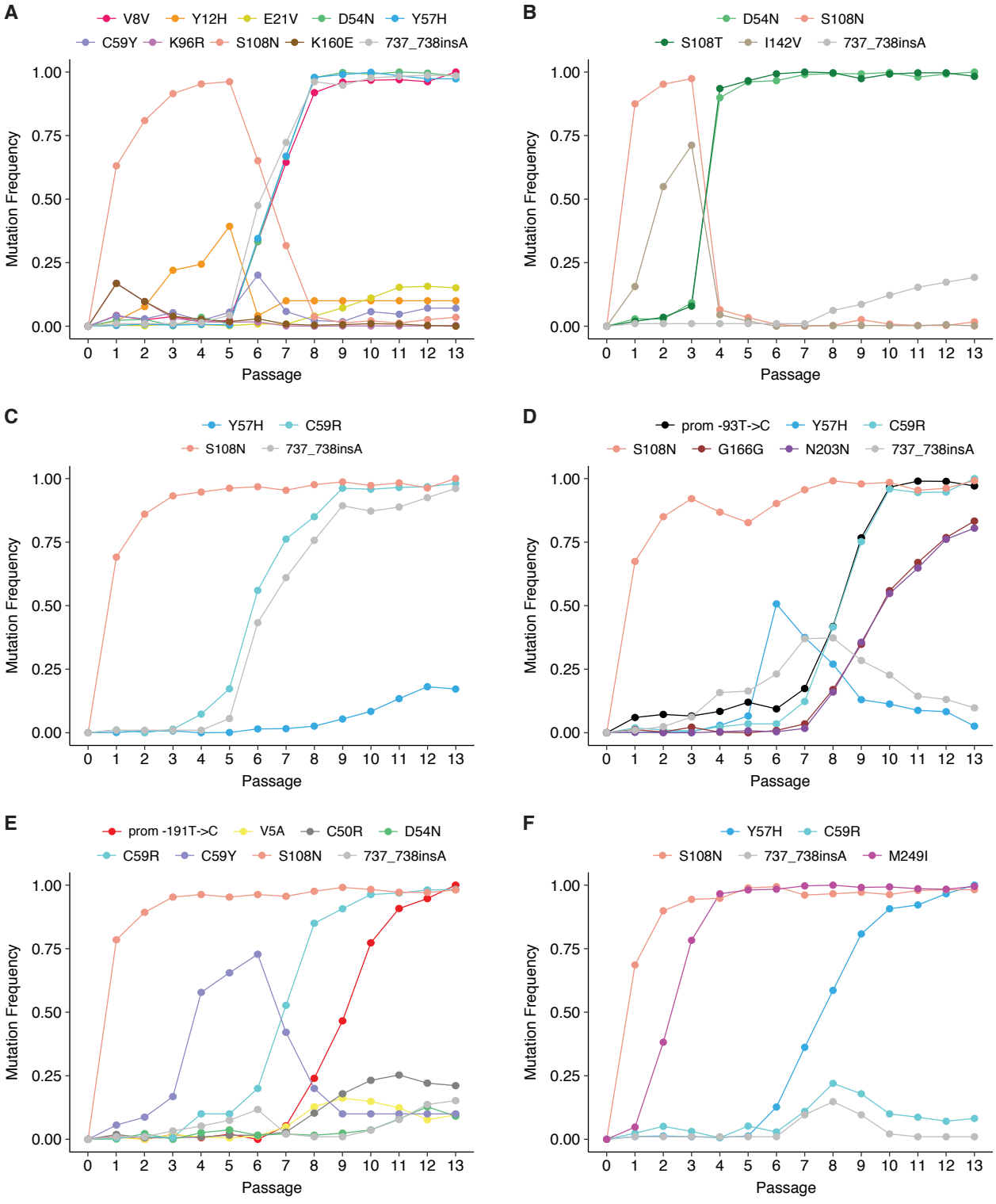
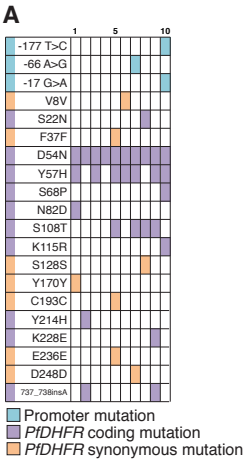


Figure 5.4 | Dynamics of *Pf*DHFR evolution in six representative populations. **(A-F)** Mutation frequencies tracked across all 13 passages from lines 52, 64, 74, 62, 70, and 73 shown in **Figure 5.3**, respectively. Populations from each passage were revived from glycerol stocks in the same media condition that they were initially grown in. Mutation frequencies were calculated from Sanger sequencing of revived populations.

along new trajectories [11]. We examined this latter possibility by repeating evolution from a variant of WT *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 5.5**). 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.

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 5.3B**) at which populations are occasionally trapped (**Figure 5.1B**). In replicate 64, D54N fixes with S108T and avoids negative epistasis with S108N. Future analysis will include less frequent candidate



B

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

Figure 5.5 | 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 **5.3 Methods** for SNP analysis details. **(B)** Percentage of replicates encoding commonly observed mutations, from two *Pf*DHFR evolution experiments. Evolution was started from the WT *PfDHFR* sequence, or from a recoded allele encoding a synonymous codon at S108.

adaptive mutations that occur in multiple replicates (*e.g.* Y35H, I150T, D222N, L251S, T268A from **Figure 5.1B**) or fix independently in time (*e.g.* M249I from **Figure 5.4F**). 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.

5.3 Methods

Yeast strains. All *Saccharomyces cerevisiae* parent strains used in this chapter are listed in **Appendix G**. Strains AH22 and F102-2 are described previously [12]. All genetic modifications that were made during strain construction were verified by sequencing and phenotyping.

DNA cloning. Plasmids used in this Chapter are listed in **Appendix G**. *E. coli* strain TG1 (Lucigen) was used for all of the DNA cloning steps. All primers used in this Chapter were purchased from IDT. All enzymes for PCR and cloning were obtained from NEB. All individually cloned plasmids were assembled by the Gibson method [13]. To clone plasmid 29, a DNA fragment encoding the open-reading frame of *PfDHFR* (819 bp) was obtained from IDT.

Yeast transformation. All transformations were performed using the Yeastmaker Yeast Transformation System 2 kit (Clontech). For integrations into p1, 10μg of plasmid was

linearized by digestion with ScaI, which generated blunt ends containing homologous regions as per the design of our integration cassettes. DNA fragments were transformed directly into p1-containing cells. Transformants were selected on the appropriate selective solid SC media. The appearance of colonies following integration typically took 4-6 days of growth at 30°C. Genomic modifications were made using a CRISPR-Cas9 system for *S. cerevisiae* [14].

Yeast DNA extraction. Whole-cell DNA extractions followed the yeast DNA miniprep procedure as summarized in *Methods in Yeast Genetics* [15]. We note the following modifications: (1) cells were washed in 0.9% NaCl prior to treatment with Zymolyase (US Biological); (2) 200µg/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 treatments. Small-scale preparations, sufficient for culture volumes between 1-10mL, were performed by scaling down the large-scale protocol by 32-fold. Extracted DNA was still enough to be observed by agarose gel electrophoresis.

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 5.2** 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) [16] 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 [17]. Insertion frequencies at base 737 in *PfDHFR* were calculated from trace files using TIDE [18]. Additional details are provided below in the **Quantification of *PfDHFR* mutation frequencies** section.

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 [17] and TIDE [18]. Additional details are provided below in the **Quantification of *PfDHFR* mutation frequencies** section. (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 (**Appendix G**). 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) [16] 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 [6]. Plasmids 23-70 in **Appendix G**, are yeast centromeric plasmids that express *Pf*DHFR variants from the weak *DFR1* promoter. Plasmids 23-70 were transformed

into YH5 [6] 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. Additional details of MIC analysis are provided below in the ***Pf*DHFR MIC analysis** section.

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 [17]. The *PfDHFR* trace file of GA-Y229 served as the template of the homozygous WT base, at all five positions. Insertion frequencies at base 737 in *PfDHFR* were calculated from trace files using TIDE (default settings) [18]. 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 [17] 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 [19]. Merged reads containing insertions or deletions were removed from analysis. Mutation frequencies match closely with results obtained from QSVanalyzer [17]. From one of the replicates, mutation frequencies calculated for the C50R, D54N, Y57H, C59R, C59Y and S108N mutations by QSVanalyzer [17] 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 [17] 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.

5.4 References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

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

[9] 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.

[10] Szendro, I.G., Franke, J., de Visser, J.A.G.M., and Krug, J. (2013). Predictability of evolution depends nonmotonically on population size. Proc. Natl. Acad. Sci. USA *110*, 571-576.

[11] 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.

[12] 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.

[13] 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.

[14] 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.

[15] Burker, D.J., Amberg, D.C. and Strathern, J.N. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*, 2005 edition, CSHL Press, New York.

[16] 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.

[17] 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.

[18] 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.

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

Chapter 6

Discussion

6.1 OrthoRep as a new paradigm for directed evolution

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 (**Appendix A**) [1-6]. 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 [7, 8], protein-protein interactions (PPIs) or PPI inhibitors via yeast two-hybrid systems [9], 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 [10, 11] or to improve enzymes from difficult-to-transform hosts like plants [12]. 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; selections utilizing cell-based technologies such as fluorescence-activated cell sorting (FACS), continuous culturing devices [13], and droplet screening systems; and other selections that have been developed for rewiring the specificities of biosensors, GPCRs, or transcription factors [14-16]. 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 22 kb of DNA, allowing for the continuous evolution of both individual genes and multi-gene pathways (manuscript 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 (**Appendix A**) [4, 17-25]. 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 [25], reverse transcriptase recognition sequences [19], gRNA target sequences [22-24], or target sequences for specific DNA-binding domains [21]). 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 the critical distinction of enforced continuous mutagenesis, OrthoRep is unique in a number of 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 [26], this ability to control OrthoRep's mutation rate should be valuable in directed evolution. Ongoing engineering of TP-DNAP1, 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 (see **6.2 Future OrthoRep technology development**). 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 evolution, because it can sustain large population sizes with fast generation times, and the availability of yeast mating should allow for *in vivo* recombination of genes being evolved on OrthoRep (see **6.2 Future OrthoRep technology development**), 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.

6.2 Future OrthoRep technology development

A number of approaches can be taken to increase the speed of evolution with OrthoRep. Ongoing efforts are focused on engineering TP-DNAP1 for higher substitution mutation rates using an error-rate selection, rather than a screen. The selection relies on reversion of an internal stop codon in the counter-selectable yeast marker, *URA3*, which allows pre-culture revertants to be suppressed with 5-FOA. This should narrow the long tail of the Luria-Delbrück distribution of mutants and center the mean closer to the actual number of mutations [27]. Selection should permit much larger TP-DNAP1 library sizes than screening. We anticipate that finding TP-DNAP1 variants with mutation rates of 10^{-4} – 10^{-3} will require extensive shuffling libraries that sample more amino acid variation at fidelity-

determining residues, cross inter-motif mutations, and include the activity-boosting mutations that were found from the large screen (see **3.1 Screening single amino acid TP-DNAP1 mutants**; **Table 3.2**). The latter will likely be important, since the hits from the final shuffling library failed to incorporate the highest error-rate exonuclease mutations, which generally have lower activity (**Table 6.1**). DNAP processivity can also be improved through fusion to DNA binding domains [28, 29], although initial attempts to do this with TP-DNAP1 mutants were unsuccessful (**Figure 6.1**). Additional TP-DNAP1 engineering will seek to create variants with custom mutational spectra or with high in/del rates for specialized experiments, such as the evolutionary optimization of loop regions in protein scaffolds. All of these efforts will be informed by the *in vitro* characterization and structure determination of TP-DNAP1 currently underway in our lab. We have also explored the possibility of increasing OrthoRep mutation rates through untargeted approaches such as disabling host repair machinery (data not shown), adding chemical mutagens (**Figure 6.2**) or creating imbalanced dNTP pools (**Figure 6.3**). Transferring OrthoRep to host strains with fast doubling times (*e.g. Kluyveromyces maxianus*) [30], or limiting clonal interference by engineering hosts with lower genomic mutation rates [31], can accelerate OrthoRep evolution as well.

Since OrthoRep consists of a plasmid-DNAP pair mechanistically and spatially separated from host nuclear DNA replication, we aim to create future versions that allow *in vivo* control over a number of parameters of gene diversification. For example, gene dosage can be titrated through the expression of a catalytically inactive TP-DNAP1 that can still compete for initiation of p1 replication (**Figure 6.4**). In addition, our lab has shown that p1 plasmids within the same cell undergo homologous recombination at high frequencies, and

that yeast mating allows intercellular transfer of cytoplasmic plasmids, meaning that evolution with OrthoRep can be expanded to include sexual recombination. This process can even be targeted to OrthoRep through the use of yeast karyogamy mutants that exchange cytoplasmic but not nuclear content during mating. Finally, we note that p2 can serve as a second, mutually orthogonal DNA replication system with tunable error rates for *in vivo* accelerated evolution of different genes or sets of genes at distinct, custom mutation rates [32].

More broadly, OrthoRep may have an important role beyond laboratory evolution, in the long-term effort to control all aspects of gene replication and expression *in vivo* [33]. At the level of replication, OrthoRep enables a variety of applications including continuous barcoding, molecular recording of non-biological information, and implementation of new genetic alphabets or XNAs. Beyond replication, OrthoRep also serves as an orthogonal transcription system, as it uses a special p2-encoded RNA polymerase to recognize unique cytoplasmic promoters (**Figure 6.5**). Based on these, we have created a collection of evolved promoters, as well as genetically-encoded poly(A) tails, that can be combined to tune the expression level of OrthoRep-encoded genes over a large range, reaching at least ~40% of the strength of the genomic *TDH3* promoter [34]. Therefore, our system, more fully described, is an orthogonal DNA replication-transcription system, which frees two of the three key processes in the central dogma for *in vivo* engineering. Assembly of OrthoRep with other orthogonalized central dogma components (*e.g.* tRNAs, aminoacyl-tRNA synthetases, ribosomes, and ribosome-specific mRNAs) could one day yield an entirely orthogonal central dogma *in vivo* [33].

6.3 Methods

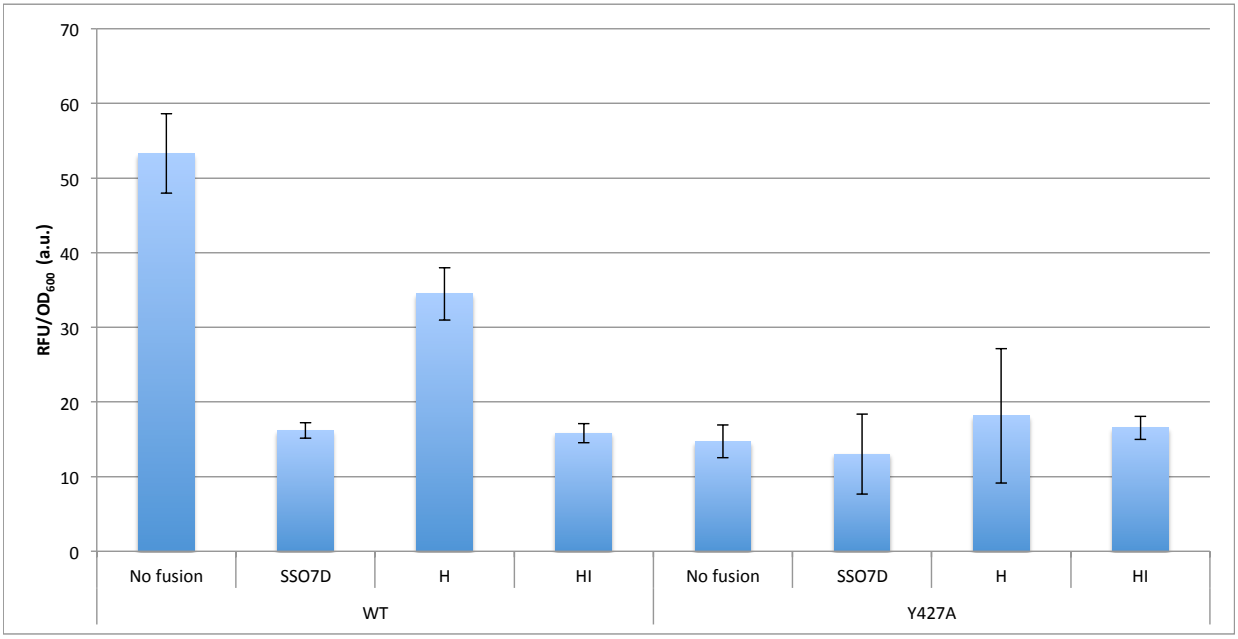


Figure 6.1 | Activity of TP-DNAP1 fusions to double-stranded DNA binding domains. WT TP-DNAP1 or the low-activity TP-DNAP1 (Y427A) mutant were fused C-terminally to three double-stranded DNA binding proteins: Sso7d from *Sulfolobus solfataricus*, Topoisomerase V domain H (residues 696-751) from *Methanopyrus kandleri*, or Topoisomerase V domains H and I (residues 696-802) from *M. kandleri*. TP-DNAP1s were expressed in a strain encoding mKate2 on p1. Strains were passaged until p1 copy number stabilized, after which OD_{600} and mKate2 fluorescence were measured. Data shown are the mean OD_{600}-normalized mKate2 fluorescence ± SD (measured in arbitrary units [a.u.]) for biological triplicates.

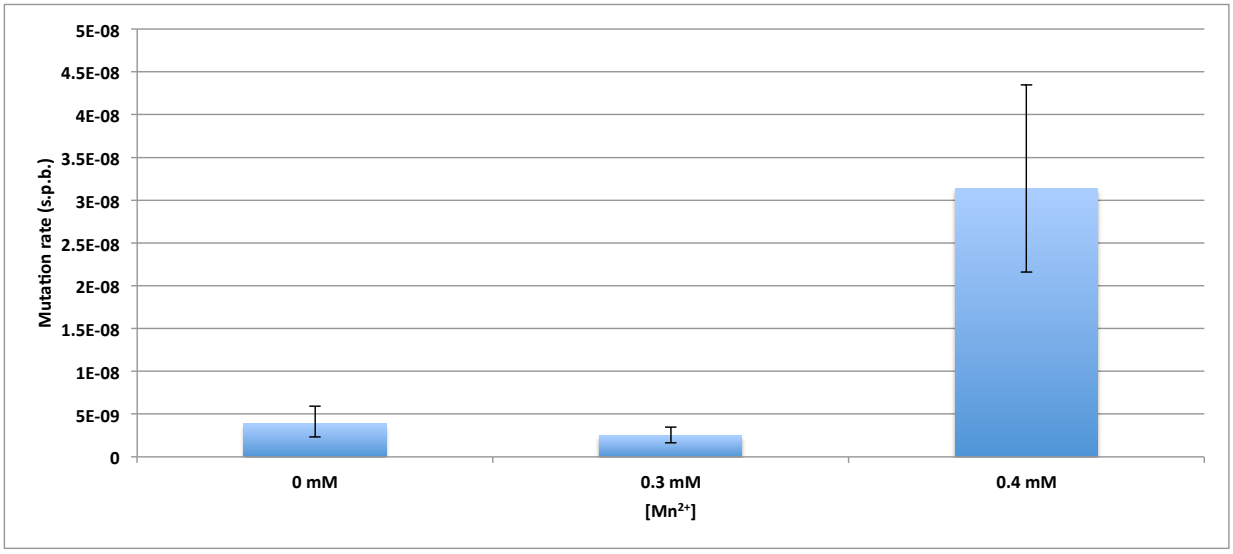


Figure 6.2 | Mn^{2+} increases WT p1 mutation rates.
OrthoRep mutation rates with WT TP-DNAP1 in the presence of various concentrations of $MnCl_2$. p1 mutation rates were measured with fluctuation tests using p1-encoded *leu2 (Q180*)*. 95% confidence intervals are shown as error bars.

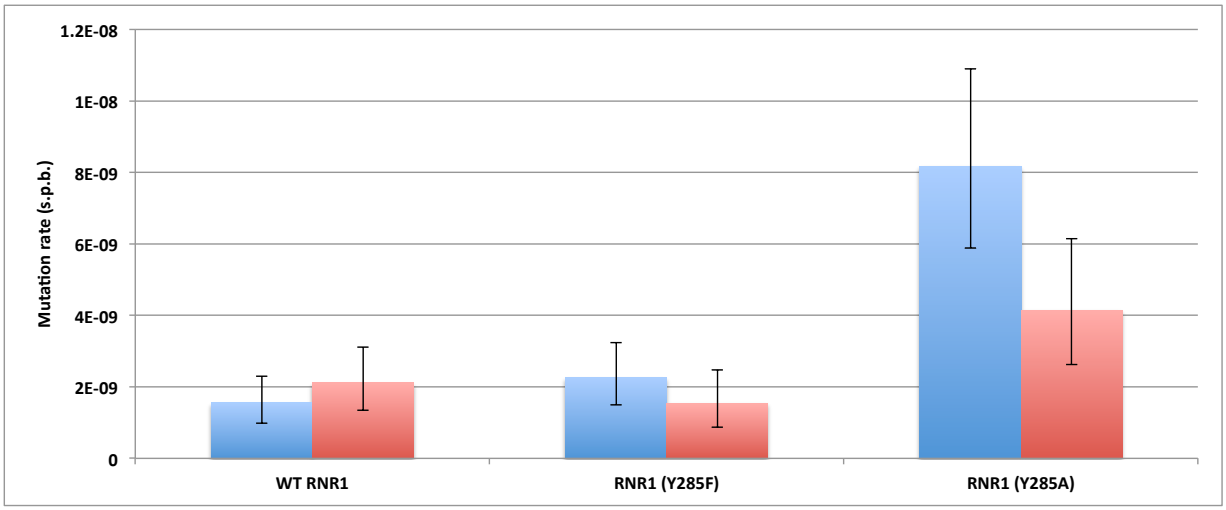


Figure 6.3 | RNR1 mutants known to create imbalanced dNTP pools increase WT p1 mutation rates.

Overexpression of RNR1 variants Y285F and Y285A from the high-copy 2μ plasmid was previously shown to severely imbalance dNTP pools in yeast. p1 mutation rates were measured in yeast containing these plasmids. Mutation rates were measured with fluctuation tests using p1-encoded *leu2 (Q180*)*. 95% confidence intervals are shown as error bars.

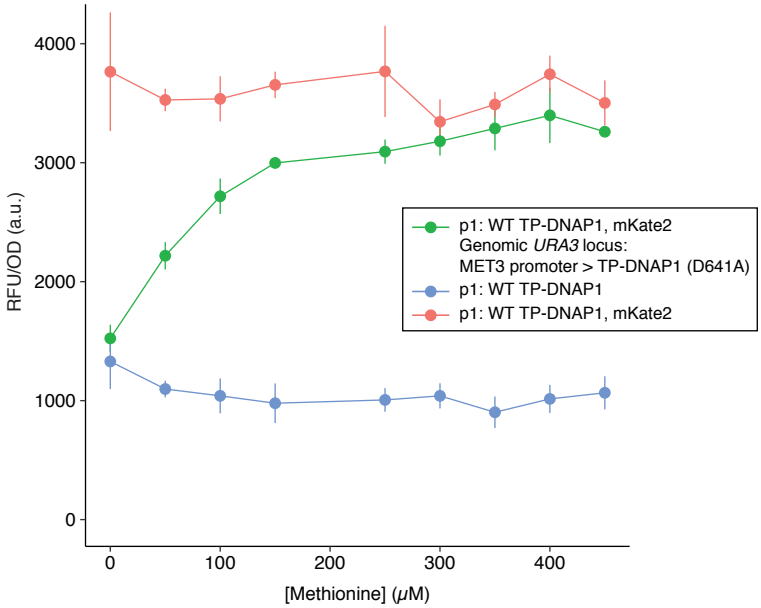
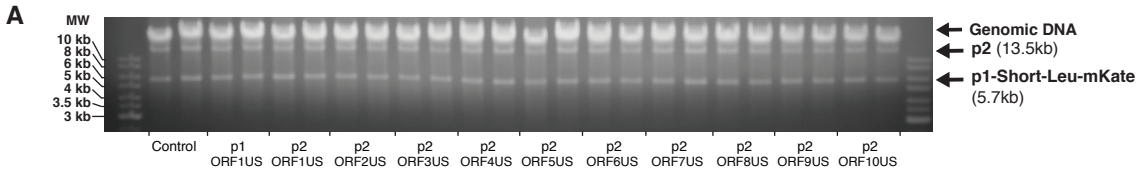
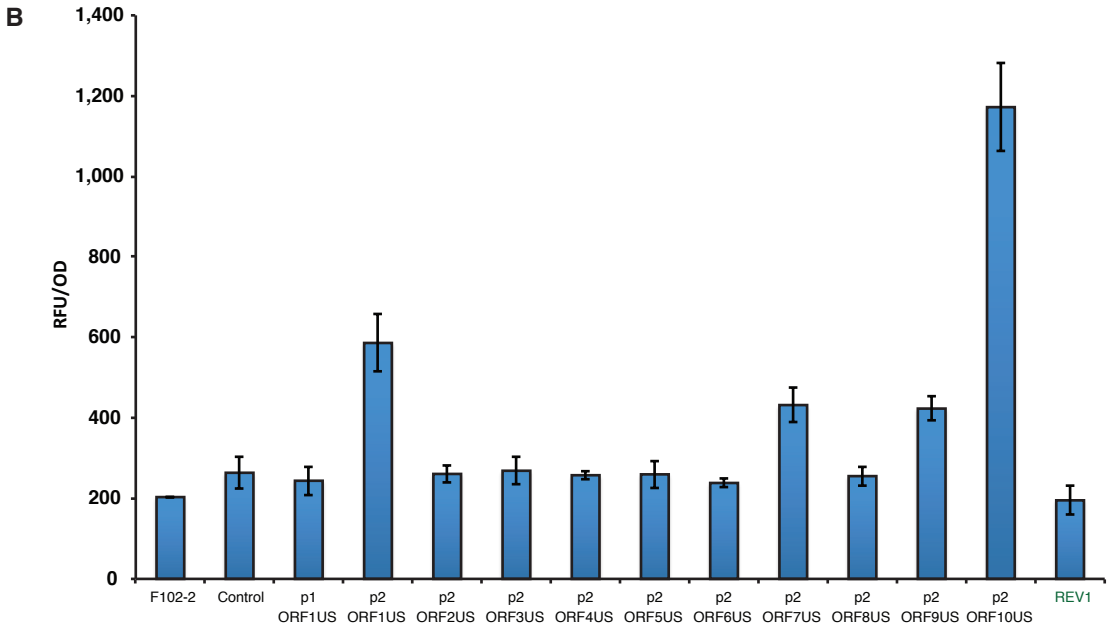
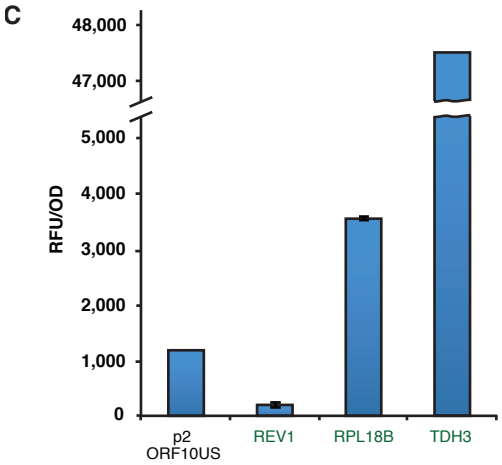


Figure 6.4 | Titratable control of p1 copy number.
TP-DNAP1 encodes a highly conserved catalytic residue at D641. The TP-DNAP1 (D641A) variant is unable to polymerize DNA, but can still compete with fully functional TP-DNAp1 for replication initiation at the TP origin. TP-DNAP1 (D641A) was expressed under the control of the repressible *MET3* promoter in a strain expressing WT TP-DNAP1 and mKate2 from p1. Strains were grown in SC media supplemented with methionine for 2 days, after which OD_{600} and mKate2 fluorescence were measured. Data shown are the mean OD_{600}-normalized mKate2 fluorescence ± SD (measured in arbitrary units [a.u.]) for biological triplicates.

Figure 6.5 | Characterization of promoter strengths of the p1/2 transcription system. 100bp upstream sequences (USs) of each ORF of the p1/2 system shown were inserted directly upstream of an mKate2 coding sequence. Corresponding integration cassettes were used to generate p1-Short-Leu-mKate cytoplamic plasmids, each containing mKate2 driven by a different US driving mKate2 expression. Control contains mKate2 without an US. **(A)** DNA gel analyses (1% agarose, EtBr stained) showing cytoplasmic plasmids extracted from representative replicate clones after passaging for several days in SC without leucine. The correct size p1-Short-Leu-mKate2 cytoplasmic plasmids are observed in all cases. **(B)** Mean expression of mKate2 driven by different USs (n = 6, error-bars are ± s.d.). Relative expression was determined by mKate2 fluorescence normalized to OD_{600}. F102-2 are cells free of any mKate2-containing plasmids. REV1 corresponds to an mKate2 coding sequence driven by the *REV1* nuclear promoter, which was cloned upstream of mKate2 in a nuclear *CEN6/ARS4* plasmid and transformed into F102-2. **(C)** mKate2 expression from nuclear *CEN6/ARS4* plasmids using various yeast nuclear promoters (green) in comparison to mKate2 expression from the p1 cytoplasmic plasmid using the p2 ORF10US promoter. Relative expression was determined by mKate2 fluorescence normalized to OD_{600} (n = 6, error-bars are ± s.d.).


A
MW
10 kb
8 kb
6 kb
5 kb
4 kb
3.5 kb
3 kb
Genomic DNA
p2 (13.5kb)
p1-Short-Leu-mKate
(5.7kb)
Control
p1 ORF1US
p2 ORF1US
p2 ORF2US
p2 ORF3US
p2 ORF4US
p2 ORF5US
p2 ORF6US
p2 ORF7US
p2 ORF8US
p2 ORF9US
p2 ORF10US


B
RFU/OD
1,400
1,200
1,000
800
600
400
200
0
F102-2
Control
p1 ORF1US
p2 ORF1US
p2 ORF2US
p2 ORF3US
p2 ORF4US
p2 ORF5US
p2 ORF6US
p2 ORF7US
p2 ORF8US
p2 ORF9US
p2 ORF10US
REV1


C
RFU/OD
48,000
47,000
5,000
4,000
3,000
2,000
1,000
0
p2 ORF10US
REV1
RPL18B
TDH3

Yeast strains. All *Saccharomyces cerevisiae* parent strains used in this chapter are listed in **Appendix G**. Strains AH22 and F102-2 are described previously [35]. All genetic modifications that were made during strain construction were verified by sequencing and phenotyping.

DNA cloning. Plasmids used in this Chapter are listed in **Appendix G**. *E. coli* strain TG1 (Lucigen) was used for all of the DNA cloning steps. All primers used in this Chapter were purchased from IDT. All enzymes for PCR and cloning were obtained from NEB. All individually cloned plasmids were assembled by the Gibson method [36].

Yeast transformation. All transformations were performed using the Yeastmaker Yeast Transformation System 2 kit (Clontech). For integrations into p1, 10μg of plasmid was linearized by digestion with ScaI, which generated blunt ends containing homologous regions as per the design of our integration cassettes. DNA fragments were transformed directly into p1-containing cells. Transformants were selected on the appropriate selective solid SC media. The appearance of colonies following integration typically took 4-6 days of growth at 30°C. Genomic modifications were made using a CRISPR-Cas9 system for *S. cerevisiae* [37].

Yeast DNA extraction. Whole-cell DNA extractions followed the yeast DNA miniprep procedure as summarized in *Methods in Yeast Genetics* [38]. We note the following modifications: (1) cells were washed in 0.9% NaCl prior to treatment with Zymolyase (US Biological); (2) 200μg/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 treatments. Small-scale preparations, sufficient for culture volumes between 1-10mL, were

performed by scaling down the large-scale protocol by 32-fold. Extracted DNA was still enough to be observed by agarose gel electrophoresis.

Improving activity of TP-DNAP1 with fusion to DNA binding domains. TP-DNAP1 fusions to DNA binding domains were expressed in strain OR-Y24. p1 mutation rate measurements were made according to the large-scale protocol detailed in **3.3 Methods**.

Increasing p1 mutation rate with chemical mutagens. Plasmid 2, encoding WT TP-DNAP1, was transformed into OR-Y24 to measure WT p1 mutation rates in the presence of various concentrations of $MnCl_2$ supplemented in the SC media. p1 mutation rate measurements were made according to the large-scale protocol detailed in **3.3 Methods**.

Increasing p1 mutation rate with RNR1 overexpression. Plasmids 115-117, encoding RNR1 mutants, were individually co-transformed, with plasmid 2, encoding WT TP-DNAP1, into OR-Y24. The resulting strains were passaged to stabilize p1 copy number and p1 mutation rate was measured according to the large-scale protocol detailed in **3.3 Methods**.

p1 copy number control. To titrate p1 copy number, the catalytically inactive TP-DNAP1 (D641A) variant was placed under the control of the repressible *MET3* promoter in AR-Y062, which expresses WT TP-DNAP1 and the mKate2 reporter from p1. Strains were grown in SC media containing methionine, ranging in concentration from 0μM to 450 μM, for 2 days at 30 °C. Afterward, mKate2 fluorescence was measured to assay p1 copy number, as described in **3.3 Methods**.

6.4 References

[1] 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.

[2] 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.

[3] 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.

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

[5] 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.

[6] 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.

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

[8] 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.

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

[10] 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.

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

[12] Sun *et al*. (2018). Parts-prospecting for a high-efficiency thiamin thiazole biosynthesis pathway. Plant Physiol. DOI: https://doi.org/10.1104/pp.18.01085.

[13] 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.

[14] 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.

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

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

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

[18] 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.

[19] 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.

[20] 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.

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

[22] 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.

[23] Hess, G.T., Frećard, 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.

[24] 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.

[25] 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.

[26] 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.

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

[28] de Vega, M., Lázaro, J.M., Mencía, M., Blanco, L., and Salas, M. (2010). Improvements of φ29 DNA polymerase amplification performance by fusion of DNA binding motifs. Proc. Natl. Acad. Sci. USA. *107,* 16506-16511.

[29] Wang, Y., Prosen, D.E., Mei, L., Sullivan, J.C., Finney, M., and Vander Horn, P.B. (2004). A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. Nucleic Acids Res. *32*, 1197-1207.

[30] Groeneveld, P., Stouthamer, A.H., and Westerhoff, H.V. (2009). Super life—how and why 'cell selection' leads to the fastest-growing eukaryote. FEBS J. *276*, 254-270.

[31] Deatherage, D.E., Leon, D., Rodriguez, Á.E., Omar, S.K., and Barrick, J.E. (2018). Directed evolution of *Escherichia coli* with lower-than-natural plasmid mutation rates. Nucleic Acids Res. *46*, 9236-9250.

[32] Arzumanyan, G.A., Gabriel, K.N., Ravikumar, A., Javapour, A.A., and Liu, C.C. (2018). Mutually orthogonal DNA replication systems *in vivo*. ACS Synth. Biol. *7*, 1722-1729.

[33] Liu, C.C., Jewett, M.C., Chin, J.W., and Voigt, C.A. (2018). Toward an orthogonal central dogma. Nat. Chem. Biol. *14*, 103-106.

[34] Zhong, Z., Ravikumar, A., and Liu, C.C. (2018). Tunable expression systems for orthogonal DNA replication. ACS Synth. Biol. DOI: https://doi.org/10.1021/acssynbio.8b00400

[35] 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.

[36] 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.

[37] 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.

[38] Burker, D.J., Amberg, D.C. and Strathern, J.N. *Methods in Yeast Genetics: A Cold Spring Harbor Laboratory Course Manual*, 2005 edition, CSHL Press, New York.

Appendix A | Comparison of directed evolution systems

Approach	*Orthogonal DNA replication*	*Repurposing of the bacteriophage life cycle*	*Targeted in vivo mutagenesis*	*Rounds of multiplexed genome mutagenesis*	*Rounds of ex vivo mutagenesis and in vivo selection*
Systems	OrthoRep	Phage assisted continuous evolution (PACE)	CRISPR-guided DNA polymerases*, Pol I/ColE1-based systems**, In vivo continuous evolution (ICE), cytidine deaminase-based systems, TaGTEAM	MAGE-based systems, Cas9-based multiplexed genome editing systems	Traditional directed evolution methods, compartmentalized partenered replication (CPR)***
Durability of mutagenesis	Indefinitely continuous. Mutagenesis is enforced because it is coupled to replication of the target gene that is being selected for during evolution. It has been implemented for up to 300 generations without any sign of erosion (unpublished results).	In principle, indefinitely continuous. However, it can be tedious to implement for long periods of time due to the constant influx of mutagneic *E. coli* required.	These systems risk breaking down due to the possibility of mutations in the *cis* elements that recruit mutagenesis machinery. Several systems cease mutagenesis within a few generations.	These are primarily genome engineering methods but they can be used for step-wise directed evolution as well. They are limited by the number of labor-intensive rounds that can be performed.	These are step-wise methodologies and are limited by the number of labor-intensive rounds that can be performed.
Mutagenesis rates achieved	Genes of interest are mutated *in vivo* at ~10^{-5} substitutions per-base. Future DNAP engineering should yield variants with error rates up to ~10^{-3} and higher.	Genes of interest are mutated in *E. coli* at ~10^{-3} substitutions per-base.	Mutation rates vary across systems, but are generally high (>10^{-5} substitutions per-base).*CRISPR-guided DNA polymerases induce exceptionally high mutation rates (10^{-4} - 10^{-3} substitutions per base) for short target regions.	Mutagenesis can be controlled via oligo synthesis, so mutation rates can be as high as desired, but the sequence space that can be accessed is bottlenecked by DNA transformation efficiency. Furthermore, mutations can't accumulate close to each other because oligos need preprogrammed constant binding sites.	Mutagenesis is typically performed via error-prone PCR or site saturation mutagenesis, so mutation rates can be as high as desired, but the sequence space that can be accessed is bottlenecked by DNA transformation efficiency.
Targeting	Completely orthogonal (at least 100,000-fold mutational targeting).	Mutations exclusively accumulate in the genes encoded in the bacteriophage genome, because the host *E. coli* cells are continuously diluted away.	These systems suffer from moderate to severe off-target mutagenesis.	Mutations are mostly targeted to genes of interest, but off-target mutations may occasionally occur, especially in MAGE strains that disable mismatch repair.	Complete mutational targeting to genes of interest.
Range of available selections	All growth-based selections and non-growth based selections like FACS, droplet sorting, cell surface antibody display, etc.	Selectable phenotypes must be coupled to gene expression.	All growth-based selections and non-growth based selections like FACS, droplet sorting, cell surface antibody display, etc.	All growth-based selections and non-growth based selections like FACS, droplet sorting, cell surface antibody display, etc.	All growth-based selections and non-growth based selections like FACS, droplet sorting, cell surface antibody display, etc.
Scalability	Can be used for hundreds of parallel evolution experiments.	Generally limited to <10 parallel experiments.	Can be used for hundreds of parallel evolution experiments.	Generally limited to <10 parallel experiments.	Generally limited to <10 parallel experiments.
Ease of use	Requires only a simple strain construction step, followed by serial passaging under selective conditions.	Requires integration of genes into the bacteriohpage genome, followed by continuous evolution in a turbidostat setup.	Requires only a simple strain construction step, followed by serial passaging under selective conditions.	Requires step-wise rounds of mutation, transformation, selection, and DNA extraction.	Requires step-wise rounds of mutation, transformation, selection, and DNA extraction. CPR additionally requires an emulsification step.
Number of genes that can be simultatenously evolved	<10 genes	<10 genes	Depending on the system, up to hundreds of genes	Up to hundreds of genes	<10 genes
Host organisms used so far	Currently in yeast. Should be extensible to bacteria and mammalian systems using related protein-priming DNAPs.	Currently bacteriophage M13 in *E. coli*. Should be extensible to mammalian cells using non-integrative viruses (*e.g.* adenovirus).	Depending on the system, these methods have been established in *E. coli*, yeast, and mammalian cells.	Depending on the system, these methods have been established in *E. coli* and yeast.	*E. coli*, yeast, and mammalian cells.
Mutational spectrum	TP-DNAP1-4-2 strongly favors transition mutations. This can be readily improved through DNAP engineering.	Fairly unbiased spectrum.	Error-prone Pol I mutates ColE1 plasmids with a bias towards transition mutations. EvolvR generates substitutions of all four nucleotide types, in a relatively unbiased manner. If needed, this can be improved through DNAP engineering. TaGTEAM generates a broad spectrum of both transitions and transversions. In addition, 25% of mutations are single base deletions. In ICE, there is a 1:1 ratio between transitions and transversions. AID generates point mutations rather than insertions and deletions, and it favors transitions over transversions.	Mutational spectrum can be user-defined.	Mutational spectrum can be user-defined.
Additional notes			**Although the Pol I/ColE1-based systems couple mutagenesis to DNA replication, they quickly cease mutagenesis due to high levels of genome-wide mutagenesis.		***CPR requires selection to be coupled to gene expression.
References	Ravikumar et al., 2014; Ravikumar et al., 2018	Badran and Liu, 2015; Esvelt et al., 2011	Camps et al., 2003; Crook et al., 2016; 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	Barbieri et al., 2017; DiCarlo et al., 2013; Jakočiūnas et al., 2018; Wang et al., 2009	Ellefson et al., 2014; Packer and Liu, 2015;

Appendix B | List of plasmids used in Chapter 2.

	Name	Source	Selection Marker (yeast, bacterial)	Notes	Internal #
1	pCCL-GKL1RC	This work	N/A, Ampicillin	p1 integration cassette with multiple cloning sites	3, 4
2	pST39	[18]	N/A, Ampicillin	Low-copy mutant ColE1 origin of replication with multiple cloning sites	2
3	pCCL-GKL1RC-LEUg1	This work	N/A, Ampicillin	Integration cassette to create p1-Leu	5
4	pCCL-GKL1RC-URAg2	This work	N/A, Ampicillin	Integration cassette to create p1-Ura	20
5	pCCL-GKL1RC-LEUg1URAg2	This work	N/A, Ampicillin	Integration cassette to create p1-LU	13
6	pCCL-GKL1RC-LEUg1URAg2-ScaI	This work	N/A, Ampicillin		14
7	pCCL-GKL1RC-LEU*g1URAg2	This work	N/A, Ampicillin	Integration cassette to create p1-L*U	15
8	pCCL-GKL1RC-LEU*g1URAg2-NsiI	This work	N/A, Ampicillin		18
9	pCCL-GKL1RC-URAg1LEU*g2	This work	N/A, Ampicillin	Integration cassette to create p1-UL*	22
10	pAH005	Courtesy of J. Dueber	LEU2, Ampicillin	CEN6/ARS4 plasmid, REV1 promoter>*mKate*	N/A
11	pAH007	Courtesy of J. Dueber	LEU2, Ampicillin	CEN6/ARS4 plasmid, RPL18B promoter>*mKate*	N/A
12	pAH056	Courtesy of J. Dueber	LEU2, Ampicillin	CEN6/ARS4 plasmid, TDH3 promoter>*mKate*	N/A
13	pCCL-GKL1RC-LEUg1mKateg2	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate	27
14	pCCL-GKL1PolRC-URAg1LEU*g2	This work	N/A, Ampicillin	Integration cassette with larger p1 homology region	44
15	pCCL-GKL1PolRC-ExoISAla-UL*	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoISAla-UL*	65
16	pCCL-GKL1PolRC-ExoISIle-UL*	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoISIle-UL*	66
17	pCCL-GKL1PolRC-ExoISThr-UL*	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoISThr-UL*	67
18	pCCL-GKL1PolRC-ExoIINAla	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoIINAla-UL*	68
19	pCCL-GKL1PolRC-ExoIINAsp	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoIINAsp-UL*	69
20	pCCL-GKL1PolRC-ExoIIFAla	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoIIFAla-UL*	70
21	pCCL-GKL1PolRC-ExoIIFSer	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoIIFSer-UL*	71
22	pCCL-GKL1PolRC-	This work	N/A,	Integration cassette to	72

	ExoIIFTyr		Ampicillin	create p1-MutExoIIFTyr-UL*	
23	pCCL-GKL1PolRC-ExoII2NAla	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoII2NAla-UL*	73
24	pCCL-GKL1PolRC-ExoII2NAsp	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoII2NAsp-UL*	74
25	pCCL-GKL1PolRC-ExoII2YAla	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoII2YAla-UL*	75
26	pCCL-GKL1PolRC-ExoII2YSer	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoII2YSer-UL*	76
27	pCCL-GKL1PolRC-ExoII2YPhe	This work	N/A, Ampicillin	Integration cassette to create p1-MutExoII2YPhe-UL*	77
28	pML225	Courtesy of J. Dueber	URA3, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter, ADH1 terminator	N/A
29	pCCL-C/A-UL	This work	URA3, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter>*LEU2*	45
30	pCCL-C/A-UL*Glu	This work	URA3, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter> Q180E *LEU2*	78
31	pCCL-C/A-UL*Lys	This work	URA3, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter> Q180K *LEU2*	79
32	pCCL-C/A-UL*Ser	This work	URA3, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter> Q180S *LEU2*	80
33	pCCL-C/A-UL*Leu	This work	URA3, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter> Q180L *LEU2*	81
34	pCCL-C/A-UL*Tyr	This work	URA3, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter> Q180Y *LEU2*	82
35	pWCD-0432	Courtesy of J. Dueber	URA3, Ampicillin	CEN6/ARS4 plasmid	35
36	pWCD-0432-URA3*	This work	N/A, Ampicillin	CEN6/ARS4 plasmid	36
37	pCCL-GKL1RC-FullDelPol-UL*	This work	N/A, Ampicillin	Integration cassette to create p1-delPol-UL*	244
38	pCCL-GKL1RC-ShortMutPRC-UL*	This work	N/A, Ampicillin	Integration cassette to create p1-Short-UL*	192
39	pCCL-GKL1RC-KanMXg2	This work	N/A, Ampicillin	Integration cassette to create p1-KanMX	193
40	pCCL-GKL1RC-LEUg1mKateg2-NsiI	This work	N/A, Ampicillin		41
41	pCCL-GKL1RC-LEUg1mKateg2CS	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with no US>*mKate*	48
42	pCCL-GKL1RC-LEUg1mKateg2-1ORF1	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p1 ORF1 US >*mKate*	52
43	pCCL-GKL1RC-LEUg1mKateg2-2ORF1	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF1 US>*mKate*	53
44	pCCL-GKL1RC-	This work	N/A,	Integration cassette to	54

	LEUg1mKateg2-2ORF2		Ampicillin	create p1-Leu-mKate with p2 ORF2 US>*mKate*	
45	pCCL-GKL1RC-LEUg1mKateg2-2ORF3	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF3 US>*mKate*	55
46	pCCL-GKL1RC-LEUg1mKateg2-2ORF4	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF4 US>*mKate*	56
47	pCCL-GKL1RC-LEUg1mKateg2-2ORF5	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF5 US>*mKate*	57
48	pCCL-GKL1RC-LEUg1mKateg2-2ORF6	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF6 US>*mKate*	58
49	pCCL-GKL1RC-LEUg1mKateg2-2ORF7	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF7 US>*mKate*	59
50	pCCL-GKL1RC-LEUg1mKateg2-2ORF8	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF8 US>*mKate*	60
51	pCCL-GKL1RC-LEUg1mKateg2-2ORF9	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF9 US>*mKate*	61
52	pCCL-GKL1RC-LEUg1mKateg2-2ORF10	This work	N/A, Ampicillin	Integration cassette to create p1-Leu-mKate with p2 ORF10 US>*mKate*	62
53	pCCL-GKL1RC-Short-LEUg1mKateg2CS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with no US>*mKate*	231
54	pCCL-GKL1RC-Short-LEUg1mKateg2GKL1ORF1UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p1 ORF1 US >*mKate*	232
55	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF1UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF1 US>*mKate*	233
56	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF2UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF2 US>*mKate*	234
57	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF3UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF3 US>*mKate*	235
58	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF4UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF4 US>*mKate*	236
59	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF5UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF5 US>*mKate*	237
60	pCCL-GKL1RC-	This work	N/A,	Integration cassette to	238

	Short-LEUg1mKateg2GKL2ORF6UCS		Ampicillin	create p1-Short-Leu-mKate with p2 ORF6 US>*mKate*	
61	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF7UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF7 US>*mKate*	239
62	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF8UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF8 US>*mKate*	240
63	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF9UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF9 US>*mKate*	241
64	pCCL-GKL1RC-Short-LEUg1mKateg2GKL2ORF10UCS	This work	N/A, Ampicillin	Integration cassette to create p1-Short-Leu-mKate with p2 ORF10 US>*mKate*	242
65	pCCL-C/A-UL*	This work	URA3, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter>*leu2**	46
66	pCCL-C/A-HL*	This work	HIS4, Kanamycin	CEN6/ARS4 plasmid, TDH3 promoter>*leu2**	124
67	pCCL-C/A-MCS	This work	HIS4, Kanamycin	CEN6/ARS4 plasmid with multiple cloning site	138
68	pWCD376	Courtesy of J. Dueber	URA3, Ampicillin	CEN6/ARS4 plasmid, REV1 promoter, ADH1 terminator	N/A
69	pCCL-C/A-DNAP-REV1	This work	URA3, Ampicillin	CEN6/ARS4 plasmid, REV1 promoter>WT *TP-DNAP1* (corresponding to the annotation of p1 ORF1 in GenBank (X01095.1))	122
70	pCCL-C/A-MCS-DNAP-REV1	This work	HIS4, Kanamycin	CEN6/ARS4 plasmid, REV1 promoter>WT *TP-DNAP1* (corresponding to the annotation of p1 ORF1 in GenBank (X01095.1))	147
71	pCCL-C/A-MCS-DNAPa-REV1	This work	HIS4, Kanamycin	CEN6/ARS4 plasmid, REV1 promoter> WT *TP-DNAP1*	204
72	pCCL-ReDNAP	This work	HIS4, Kanamycin	CEN6/ARS4 plasmid, REV1 promoter>Recoded WT *TP-DNAP1*	265
73	pCCL-ReDNAP(Y427A)	This work	HIS4, Kanamycin	CEN6/ARS4 plasmid, REV1 promoter>Recoded *TP-DNAP1(Y427A)*	268
74	pUG6	Courtesy of J. Dueber	N/A, Ampicillin	Contains *KanMX* gene	N/A
75	pYES-LEU2	Courtesy of A. Arkin	LEU2, Ampicillin	Contains the *LEU2* gene	N/A
76	pCCL-GKL1ShortMutPRC-Ug1Lg2starTGA	This work	URA3, Ampicillin	Integration cassette to create p1-Short-UL*(TGA)	278

77	pCCL-FullDelPol-UL	This work	N/A, Ampicillin	Integration cassette to create p1-delPol-UL	203
78	pCCL-FullDelPol-UK93Rg1Lg2	This work	N/A, Ampicillin	Integration cassette to create p1-delPol-U*L	281
79	pCCL-GKL1ShortMutPRC-Ug1K93RLg2	This work	LEU2, Ampicillin	Integration cassette to create p1-Short-U*L	280

Appendix C | List of primers used in Chapter 2.

	Name	Sequence
1	CCLGKL1RCtopSTPCR1	CATGCTCGAGAGTACTTATAATAATTTTGAAGAA
2	CCLGKL1RCtopSTPCR2	CATGTCTAGAAGTACTTCTAAACAAAGTAA
3	CCLpSTforGKL1RCPCR1	CATGTCTAGAAGATCCGGCTGCTAACAAAG
4	CCLpSTforGKL1RCPCR2	CATGCTCGAGGGTGCCAACACCTACATCTG
5	CCLLeu2g1PCR1	CATGATGCATGTTCCTAAGAAGATCGTCGT
6	CCLLeu2g1PCR2	GATCGAGCTCCGTTTCTATTATGAATTTCAT
7	CCLUrag2PCR1	GGCCCATATGTCGAAAGCTACATATAA
8	CCLUrag2PCR2	GGCCGAGCTCGGGTAATAACTGATATAATTAAA
9	CCLUra-ScaPCR1	5'-Phos-ATACTGCAATTTGACTGTATTA
10	CCLUra-ScaPCR2	5'-Phos-TCTGCGGGTGTATACAGAAT
11	CCLLeug2*PCR1	5'-Phos-ATAGGGCCATGAAAGCGGCCA
12	CCLLeug2*PCR2	5'-Phos-AACATGAGCCACCATTGCCT
13	CCLUra-NsiIPCR1	5'-Phos-CTGCATTTACTTATAATACAGTTT
14	CCLUra-NsiIPCR2	5'-Phos-GTATACTAAACTCACAAATTAGAGC
15	CCLmKateg2PCR1	GGCCCATATGGTGAGCGAGCTGATTAA
16	CCLmKateg2PCR2	GGCCGAGCTCGAGCGACCTCATGCTATACC
17	CCLPolExoISAlaPCR1	5'-Phos-GCTTATTTTGATCCAGAAAAAGAATCTAATC
18	CCLPolExoISIlePCR1	5'-Phos-ATTTATTTTGATCCAGAAAAAGAATCTAATC
19	CCLPolExoISThrPCR1	5'-Phos-ACTTATTTTGATCCAGAAAAAGAATCTAATC
20	CCLPolExoISPCR2	5'-Phos-TTCTATATCAAAACATACTGTTATATTTTT
21	CCLPolExoIINAlaPCR1	5'-Phos-GCTATTGTAGATTTTGAAGGAAGAGATTG
22	CCLPolExoIINAspPCR1	5'-Phos-GATATTGTAGATTTTGAAGGAAGAGATTG
23	CCLPolExoIINPCR2	5'-Phos-TCCTATGACTTTATTATATATTATAGATGCA
24	CCLPolExoIIFAlaPCR1	5'-Phos-GCTGAAGGAAGAGATTGTGTAGCTCAA
25	CCLPolExoIIFSerPCR1	5'-Phos-TCTGAAGGAAGAGATTGTGTAGCTCAA
26	CCLPolExoIIFTyrPCR1	5'-Phos-TACGAAGGAAGAGATTGTGTAGCTCAA
27	CCLPolExoIIFPCR2	5'-Phos-GTCTACAATATTTCCTATGACTT
28	CCLPolExoII2NAlaPCR1	5'-Phos-GCTGGTGGAGGTTATGATTTTCA
29	CCLPolExoII2NAspPCR1	5'-Phos-GATGGTGGAGGTTATGATTTTCA
30	CCLPolExoII2NPCR2	5'-Phos-ATGTGCAATTAGTTCCACTGAAG
31	CCLPolExoII2YAlaPCR1	5'-Phos-GCTGATTTTCATTATATTTTAAGTAGTATGTATAAT
32	CCLPolExoII2YSerPCR1	5'-Phos-TCTGATTTTCATTATATTTTAAGTAGTATGTATAAT
33	CCLPolExoII2YPhePCR1	5'-Phos-TTTGATTTTCATTATATTTTAAGTAGTATGTATAAT
34	CCLPolExoII2YPCR2	5'-Phos-ACCTCCACCATTATGTGCAA
35	CCLLeu*IntPCR1	CCAAAAGATCTATGCATGTTCCTAAGAAGAT
36	CCLLeu*IntPCR2	TTACTCGAGTTAAGCAAGGATTTTCTTAA
37	CCLLeuMisPCR1	5'-Phos-TAGGGCCATGAAAGCGGCCATTCTTGTGATTCTTTG
38	CCLLeuMisGluPCR2	5'-Phos-GAACATGAGCCACCATTGCCTATTTGGT
39	CCLLeuMisLysPCR2	5'-Phos-AAACATGAGCCACCATTGCCTATTTGGT
40	CCLLeuMisSerPCR2	5'-Phos-TCACATGAGCCACCATTGCCTATTTGGT
41	CCLLeuMisLeuPCR2	5'-Phos-TTACATGAGCCACCATTGCCTATTTGGT
42	CCLLeuMisTyrPCR2	5'-Phos-TACCATGAGCCACCATTGCCTATTTGGT
43	CCLUra3*PCR1	5'-Phos-TAAGAAGTAACAAAGGAACCT
44	CCLUra3*PCR2	5'-Phos-TGCCGCCTGCTTCAAACCGC
45	CCLFullDelPol-iPCR1	5'-Phos-ATATTATAGTTGGAAAATGTGTATAATAAAATGAC
46	CCLFullDelPol-iPCR2	5'-Phos-TATATGAAAGTTTTTATAATAATTATAAAGAGCTCC
47	CCLShortPRCPCR1	AAAGATCCTCAAGGCAATAGAATAACAGATTTTGATAGTACT TCTAGAAGATCCGGCTGC
48	CCLShortPRCPCR2	GAAACACCAATCTGAGTCTGAGGGGATCCGATCTTGTATAGA TAAAAAATTACGTATAT

49	CCLGKL1RCKanMXPCR1	GCAACATATGGGTAAGGAAAAGACTCACGTTT
50	CCLGKL1RCKanMXPCR2	TCGAGAGCTCGTTTTCGACACT
51	CCLUra3*RecPCR1	CGAAAGCTACATATAAGGAA
52	CCLUra3*RecPCR2	CTGGCCGCATCTTCTCAAAT
53	CCLLg1mKg2-NsiIPCR1	5'-Phos-CGTTCCTAAGAAGATCGTCGT
54	CCLLg1mKg2-NsiIPCR2a	5'-Phos-TGCATTTTATAATTATTATAAAAACTTTC
55	CCLLg1mKg2CSPCR1b	CCGGAATACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
56	CCLLg1mKg2CSPCR2	ATGCATGTGAGCGAGCTGATTAAGGAGAA
57	CCL1ORF1iPCR1	CTTAAAATATAAAAAAATAGACTTGTTATACAGAGTTGATACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
58	CCL1ORF1iPCR2	TCTATTTTACACTTTTGACCTATAAGTCATTTTATTATACACATTTTCCAACTATAATATATGCATGTGAGCGAGCTGAT
59	CCL2ORF1iPCR1	TTATATAGAAAAGTATGACCTAATTGACTCCGGCGAAAAAACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
60	CCL2ORF1iPCR2	TCCAAATCCCAAAAATCAATTGAATGATTCTTAATATGATTTAATAGTTTATGATTATAAATGCATGTGAGCGAGCTGAT
61	CCL2ORF2iPCR1	GTATTCATGGTTCTATTTATTTTCATATGCTTCCAATTCTACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
62	CCL2ORF2iPCR2	CGGAAGTGCACAAGAAATAGAATTATCACGCATGTATTCTTATCATTTCTTTTCTAAAAAATGCATGTGAGCGAGCTGAT
63	CCL2ORF3iPCR1	ACTATCTGATTATAACAATCAGATTAAAAATGAAAGTATTACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
64	CCL2ORF3iPCR2	AACATCTCACCCTTTTTAATACCATATTTAGAAATTATATATCTAAGGTAAACCTTTTGCATGCATGTGAGCGAGCTGAT
65	CCL2ORF4iPCR1	TCTCTATTGCATCTATCCATTTAAGATTATACGAAACTATACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
66	CCL2ORF4iPCR2	AATTCGATCTAAAGGATCCAGACATGCAAAAGGTTTACCTTAGATATATAATTTCTAAATATGCATGTGAGCGAGCTGAT
67	CCL2ORF5iPCR1	TATTGTTTTCACAATTAAGATCTTTAGCTATCTCACTAGTACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
68	CCL2ORF5iPCR2	TTGAAAGTATAATAAATACATTAAAAGAACAAAATAGATATTTTGACAAACAAATTGAATATGCATGTGAGCGAGCTGAT
69	CCL2ORF6iPCR1	ACAACTTCCTTCAAGTAAACTTGTAGAGTAGTCTTTTCCGACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
70	CCL2ORF6iPCR2	TGTTACTAAGATTATTAAGAATAATTTAGAAGTTGAATTTGACAAACTATCATTAGCTATATGCATGTGAGCGAGCTGAT
71	CCL2ORF7iPCR1	TCTAGCCACCTTATAGGAGGTAGTATATAAGGTCCTTCTTACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
72	CCL2ORF7iPCR2	TAGTTTTGAAAAGAATGCTGAATTTACAATTTTATGTGAAGTTGATGATATAAAGTAAAAATGCATGTGAGCGAGCTGAT
73	CCL2ORF8iPCR1	TAAAAAGAATAAGATTGTACAAAAAAGTTACATGAATAGAACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
74	CCL2ORF8iPCR2	AATCTTAATATTTCCCATTGTTTATCTAATTTAAAATTATCTGTCCATTGATATTCGTCTATGCATGTGAGCGAGCTGAT
75	CCL2ORF9iPCR1	CGTAATATATTCTTATTTAATGAATTCTACTTGTAGTAATACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
76	CCL2ORF9iPCR2	TCTTTGGAAACATTATTTGAATAGTATTCGACAAATGAAACATTGGTTATCTTATACATTATGCATGTGAGCGAGCTGAT
77	CCL2ORF10iPCR1	ATCTTTTTGTTTCTATAGATCTTCCTATGTATAGGTCATCACGCGTTAATATTAATAAATCTAAAAACTTTGATAATTTA
78	CCL2ORF10iPCR2	TAATAACTTTCAAATATCAGAAAAATGTAGAAATATATGATAAGCTCATAGACATGTAAAATGCATGTGAGCGAGCTGAT

79	CCLSRCStitchPCR1	CAGTACGCGTCCTCAGACTCAGATTGGTGTAAAGATCCTCAAGGCAATAGAA
80	CCLSRCStitchPCR2	ATCTTCTAGAAGTACTATCAAAATCTGTTATTCTATTGCCTTGAGGATCT
81	AR-His4gib-His4PCR1a	GCAGGAAACGAAGATAAATCATGGTTTTGCCGATTCTACC
82	AR-His4gib-His4PCR2a	CATGCATTTACTTATAATACAGTTTTCTACTGGAAATCCTTTGGGATC
83	CCLCAULBBGibPCR1	AAAACTGTATTATAAGTAAATGCATGTATACTAAACTCACA
84	AR-His4gib-UraPCR2a	GATTTATCTTCGTTTCCTGCAC
85	CCLCAMCSPCR1	5'-Phos-ACGCGTGGTACCGGCCGCGTTGCTGGCGTTTTTCCATA
86	CCLCAMCSPCR2	5'-Phos-CCATGGACTAGTGAGCTCCCTGATGCGGTATTTTCTCC
87	ARpGKL1-DNAP-PCR1	CCAAAAGATCTATGAATTACATTATTAATTTAAAAATGGATTACAAAGA
88	ARpGKL1-DNAP-PCR2	GCTAGCTCGAGTTAGGATCCTTTATAATTATTATAAAAACTTTCATATATTAAGTAGCTT
89	CCL-DNAPREV1PCR1	CGGAGAGCTCGTGTTGTTATCCGATACAACCG
90	CCL-DNAPREV1PCR2	GCACACTAGTGAGCGACCTCATGCTATACCT
91	ARDNAPaPCR1	5'-Phos-ATGGATTACAAAGATAAGGC
92	ARDNAPaPCR2	5'-Phos-AGATCTCGCTGGATATGCCTAGAA
93	CCLRecodedDNAP-Gib-fwd	CATGGAGATCTATGGACTACAAAGACAA
94	CCLRecodedDNAP-Gib-rvs	GGTCATCCTCGAGTTAGGTGGC
95	CCLReEII2YAPCR1	GGTGGTGGTGCTGATTTCCACTACATCTTGTCATCCA
96	CCLReEII2YAPCR2	TGGAAATCAGCACCACCACCGTTATGGGCGA
97	qPCR-Leu2F	GCTAATGTTTTGGCCTCTTC
98	qPCR-Leu2R	ATTTAGGTGGGTTGGGTTCT
99	qPCR-Leu3F	CAGCAACTAAGGACAAGG
100	qPCR-Leu3R	GGTCGTTAATGAGCTTCC
101	CCLGKL1PolRCPCR1	CGTCCTCGAGAGTACTAAATTTAGTCCTTTGTCAA
102	CCLGKL1PolRCPCR2	CAGTATGCATTTTATAATTATTATAAAAACTTTCATATA
103	CCLGKL1PolBBPCR1	ATAAAATGCATAAAGCTACATATAAGGA
104	CCLGKL1PolBBPCR2	AGTACTCTCGAGGGTGCCAA
105	CCLGKL1RCseq2	GGATCAGAAGTAGGACAATTAGAAT
106	ARLeug2*TGAPCR1	GGCTCATGTCATAGGGCCATGAAAGCGGCCA
107	ARLeug2*TGAPCR2	TGGCCCTATGACATGAGCCACCATTGCCTA
108	ARUraK93RPCR1	AAGACAGAAGATTTGCTGACATTGGTAATACAGT
109	ARUraK93RPCR2	TCAGCAAATCTTCTGTCTTCGAAGAGTAAAAAATT

Appendix D | 99 TP-DNAP1 homologs generated via protein BLAST

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

Appendix E | All TP-DNAP1 variants characterized by fluctuation tests in Chapter 3.
All independent measurements of mutation rate are shown, with corresponding 95% confidence intervals and the number of replicates performed for each fluctuation test listed. The number of replicates assayed for determination of p1 copy number is shown as (n).

TP-DNAP1	p1 copy number (n)	Mutation rate (s.p.b.)	Lower 95% C.I. (s.p.b.)	Upper 95% C.I. (s.p.b.)	Number of replicates	Notes
K302G	87.8 (1)	2.27×10^{-9}	1.57×10^{-9}	3.15×10^{-9}	36	From saturation mutagenesis library
T352E	31.8 (1)	4.58×10^{-9}	3.12×10^{-9}	3.12×10^{-9}	34	From homology study
C354F	25.1 (1)	3.86×10^{-9}	2.46×10^{-9}	2.46×10^{-9}	36	From homology study
F355M	88.8 (1)	4.63×10^{-9}	3.22×10^{-9}	6.35×10^{-9}	36	From saturation mutagenesis library
Y360F	47 (1)	3.76×10^{-9}	2.65×10^{-9}	2.65×10^{-9}	36	From homology study
N371A	16.6 (1)	6.03×10^{-8}	4.69×10^{-8}	7.54×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
N371C	26.9 (1)	1.08×10^{-8}	7.85×10^{-9}	1.42×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
N371M	23.3 (1)	1.56×10^{-8}	1.14×10^{-8}	2.06×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
	27 (1)	1.10×10^{-8}	7.53×10^{-9}	1.53×10^{-8}	36	
C376I	58.1 (1)	1.73×10^{-9}	1.13×10^{-9}	1.13×10^{-9}	33	From homology study
K384D	91.1 (1)	2.84×10^{-9}	1.88×10^{-9}	4.06×10^{-9}	36	From saturation mutagenesis library
V405I	68.3 (1)	2.34×10^{-9}	1.63×10^{-9}	1.63×10^{-9}	36	From homology study
D407Q	176 (1)	1.50×10^{-9}	8.05×10^{-10}	2.50×10^{-9}	48	From saturation mutagenesis library
	160 (1)	1.31×10^{-9}	6.33×10^{-10}	2.35×10^{-9}	48	
G410H, N423Q	55.4 (1)	4.77×10^{-8}	3.94×10^{-8}	5.62×10^{-8}	24	From saturation mutagenesis library; Rd2 mutant
	60.3 (1)	3.91×10^{-8}	3.21×10^{-8}	4.63×10^{-8}	24	
	61.7 (1)	3.50×10^{-8}	2.97×10^{-8}	4.05×10^{-8}	36	
G410Y	138 (1)	1.06×10^{-9}	6.42×10^{-10}	1.61×10^{-9}	24	From saturation mutagenesis library
	132 (1)	3.08×10^{-10}	1.23×10^{-10}	6.20×10^{-10}	24	
E411T, G426C	31.7 (1)	3.42×10^{-9}	2.05×10^{-9}	5.28×10^{-9}	24	From saturation mutagenesis library
	42.2 (1)	3.10×10^{-9}	1.82×10^{-9}	4.81×10^{-9}	24	
	46 (1)	1.84×10^{-9}	9.68×10^{-10}	3.10×10^{-9}	24	
	39.8 (1)	1.58×10^{-9}	8.04×10^{-10}	2.72×10^{-9}	24	
N413I	25.6 (1)	7.89×10^{-9}	3.62×10^{-9}	1.46×10^{-8}	48	From saturation mutagenesis library
	32.3 (1)	4.47×10^{-9}	1.78×10^{-9}	9.04×10^{-9}	48	
	41.4 (1)	4.14×10^{-9}	1.78×10^{-9}	7.99×10^{-9}	48	
	33.4 (1)	2.30×10^{-9}	5.73×10^{-10}	5.96×10^{-9}	45	
I414H	81.4 (1)	1.87×10^{-9}	1.16×10^{-9}	2.79×10^{-9}	24	From saturation mutagenesis library
	94 (1)	8.43×10^{-10}	4.34×10^{-10}	1.44×10^{-9}	24	
I420Y	3.52 (1)	2.45×10^{-9}	4.08×10^{-10}	4.08×10^{-10}	35	From homology study

A421N	5.73 (1)	$9.76x10^{-8}$	$5.94x10^{-8}$	$1.49x10^{-7}$	46	From saturation mutagenesis library; Rd2 mutant
	7.37 (1)	$7.80x10^{-8}$	$4.62x10^{-8}$	$1.21x10^{-7}$	45	
A421S	33.6 (1)	$7.31x10^{-9}$	$5.34x10^{-9}$	$9.65x10^{-9}$	36	From saturation mutagenesis library; Rd2 mutant
N423R	36.5 (1)	$3.67x10^{-8}$	$3.02x10^{-8}$	$4.36x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	37.1 (1)	$3.27x10^{-8}$	$2.67x10^{-8}$	$3.92x10^{-8}$	36	
	37.6 (1)	$2.57x10^{-8}$	$1.99x10^{-8}$	$3.19x10^{-8}$	24	
	43.3 (1)	$2.07x10^{-8}$	$1.59x10^{-8}$	$2.60x10^{-8}$	24	
	43.9 (1)	$1.97x10^{-8}$	$1.52x10^{-8}$	$2.47x10^{-8}$	24	
	69.5 (1)	$1.55x10^{-8}$	$1.20x10^{-8}$	$1.93x10^{-8}$	23	
N423D	4.61 (1)	$5.14x10^{-7}$	$3.89x10^{-7}$	$6.58x10^{-7}$	48	From Ravikumar et al., 2014; Rd1 mutant
	4.84 (1)	$4.03x10^{-7}$	$2.99x10^{-7}$	$5.24x10^{-7}$	48	
	10.6 (3)	$1.74x10^{-7}$	$1.49x10^{-7}$	$2.01x10^{-7}$	48	
	9.48 (3)	$1.82x10^{-7}$	$1.52x10^{-7}$	$2.14x10^{-7}$	44	
	7.46 (3)	$2.49x10^{-7}$	$2.07x10^{-7}$	$2.94x10^{-7}$	45	
N423E	12.6 (1)	$2.67x10^{-7}$	$2.04x10^{-7}$	$3.39x10^{-7}$	48	From saturation mutagenesis library; Rd2 mutant
	17.5 (1)	$1.81x10^{-7}$	$1.27x10^{-7}$	$2.48x10^{-7}$	42	
N423Q	36.2 (1)	$6.43x10^{-8}$	$5.34x10^{-8}$	$7.59x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	20.3 (1)	$6.19x10^{-8}$	$4.94x10^{-8}$	$7.52x10^{-8}$	24	
	25.8 (1)	$5.89x10^{-8}$	$4.78x10^{-8}$	$7.07x10^{-8}$	36	
	18.1 (1)	$5.26x10^{-8}$	$4.11x10^{-8}$	$6.52x10^{-8}$	24	
	23.7 (1)	$4.86x10^{-8}$	$3.82x10^{-8}$	$5.98x10^{-8}$	24	
	46.8 (1)	$3.10x10^{-8}$	$2.49x10^{-8}$	$3.74x10^{-8}$	24	
N423W	22.5 (1)	$4.67x10^{-9}$	$2.96x10^{-9}$	$2.96x10^{-9}$	35	From homology study
N423Y	4.65 (1)	$1.29x10^{-9}$	$7.36x10^{-11}$	$7.36x10^{-11}$	34	From homology study
G425S	42.9 (1)	$3.04x10^{-9}$	$2.05x10^{-9}$	$2.05x10^{-9}$	36	From homology study
G426A	52.2 (1)	$3.35x10^{-9}$	$2.35x10^{-9}$	$2.35x10^{-9}$	36	From homology study
G426K	106 (1)	$1.31x10^{-9}$	$7.67x10^{-10}$	$2.04x10^{-9}$	24	From saturation mutagenesis library
	153 (1)	$1.25x10^{-9}$	$7.87x10^{-10}$	$1.85x10^{-9}$	24	
G426P	20.2 (1)	$1.52x10^{-8}$	$1.09x10^{-8}$	$2.02x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
Y427A	6.36 (3)	$2.14x10^{-8}$	$5.36x10^{-9}$	$5.51x10^{-8}$	24	From Ravikumar et al., 2014; Rd1 mutant
Y427N	2.28 (1)	$1.65x10^{-7}$	$1.25x10^{-7}$	$1.25x10^{-7}$	36	From homology study; Rd1 mutant
Y427D	1.49 (1)	$7.16x10^{-8}$	$4.41x10^{-8}$	$4.41x10^{-8}$	36	From homology study; Rd1 mutant
Y427G	2.92 (1)	$7.31x10^{-8}$	$5.20x10^{-8}$	$5.20x10^{-8}$	36	From homology study; Rd1 mutant
Y427H	15.4 (1)	$5.93x10^{-8}$	$4.14x10^{-8}$	$8.11x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
Y427K	7.09 (1)	$3.89x10^{-8}$	$2.88x10^{-8}$	$2.88x10^{-8}$	36	From homology study; Rd1 mutant

H430Q	1.85 (1)	$3.84x10^{-8}$	$2.07x10^{-8}$	$2.07x10^{-8}$	36	From homology study; Rd1 mutant
Y431H	31.7 (1)	$3.82x10^{-8}$	$3.10x10^{-8}$	$4.59x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
P439A	1.74 (1)	$2.14x10^{-8}$	$9.21x10^{-9}$	$9.21x10^{-9}$	36	From homology study; Rd1 mutant
N450P	19.2 (1)	$4.19x10^{-8}$	$3.29x10^{-8}$	$5.17x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
L473Y	19.1 (1)	$1.90x10^{-8}$	$1.32x10^{-8}$	$2.61x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
L474F	17.2 (1)	$5.65x10^{-8}$	$4.42x10^{-8}$	$7.00x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
L474W	33.4 (1)	$1.02x10^{-8}$	$7.51x10^{-9}$	$1.34x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	34.3 (1)	$9.92x10^{-9}$	$7.16x10^{-9}$	$1.32x10^{-8}$	36	
L477I	63.1 (1)	$4.33x10^{-9}$	$2.83x10^{-9}$	$6.24x10^{-9}$	36	From saturation mutagenesis library
L477T	16 (1)	$1.08x10^{-7}$	$8.86x10^{-8}$	$1.29x10^{-7}$	36	From saturation mutagenesis library; Rd2 mutant
L477V	36.4 (1)	$2.06x10^{-8}$	$1.62x10^{-8}$	$2.54x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	46.2 (1)	$1.64x10^{-8}$	$1.22x10^{-8}$	$2.14x10^{-8}$	36	
N479P	23.3 (1)	$8.02x10^{-9}$	$5.37x10^{-9}$	$1.13x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
S481F	25.5 (1)	$1.44x10^{-8}$	$1.04x10^{-8}$	$1.91x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
K492T	18.3 (1)	$4.94x10^{-8}$	$3.94x10^{-8}$	$6.04x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	20 (1)	$4.40x10^{-8}$	$3.48x10^{-8}$	$5.41x10^{-8}$	36	

K492V	21.2 (1)	$2.69x10^{-8}$	$2.05x10^{-8}$	$3.43x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	27.3 (1)	$2.03x10^{-8}$	$1.56x10^{-8}$	$2.57x10^{-8}$	36	
T493P	16.2 (1)	$4.95x10^{-8}$	$3.89x10^{-8}$	$6.13x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	16.8 (1)	$3.75x10^{-8}$	$2.39x10^{-8}$	$5.50x10^{-8}$	36	
H497N	15.7 (1)	$5.37x10^{-8}$	$4.03x10^{-8}$	$6.91x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	15 (1)	$3.25x10^{-8}$	$2.43x10^{-8}$	$4.20x10^{-8}$	36	
K511H	60.6 (1)	$1.49x10^{-9}$	$8.81x10^{-10}$	$2.32x10^{-9}$	36	From saturation mutagenesis library
S514R	64.3 (1)	$3.57x10^{-9}$	$2.57x10^{-9}$	$4.77x10^{-9}$	36	From saturation mutagenesis library
S533I	48.9 (1)	$1.86x10^{-9}$	$1.10x10^{-9}$	$2.89x10^{-9}$	36	From saturation mutagenesis library
I549L	52.7 (1)	$3.75x10^{-9}$	$2.63x10^{-9}$	$2.63x10^{-9}$	35	From homology study
E550D	56.3 (1)	$2.36x10^{-9}$	$1.60x10^{-9}$	$1.60x10^{-9}$	36	From homology study
C556G	32.6 (1)	$1.36x10^{-8}$	$9.85x10^{-9}$	$1.81x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
	33.3 (1)	$9.37x10^{-9}$	$6.41x10^{-9}$	$1.30x10^{-8}$	36	
	35.4 (1)	$7.27x10^{-9}$	$5.21x10^{-9}$	$9.72x10^{-9}$	36	
C556T	16.9 (1)	$4.10x10^{-8}$	$3.04x10^{-8}$	$5.34x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
R557A	41.4 (1)	$1.90x10^{-9}$	$1.06x10^{-9}$	$3.06x10^{-9}$	36	From saturation mutagenesis library
R557C	43.8 (1)	$2.65x10^{-9}$	$1.73x10^{-9}$	$1.73x10^{-9}$	35	From homology study
N558S	14.1 (1)	$7.90x10^{-9}$	$4.39x10^{-9}$	$1.29x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
L561I	54.1 (1)	$2.84x10^{-9}$	$1.94x10^{-9}$	$1.94x10^{-9}$	36	From homology study
S564A	53.7 (1)	$2.67x10^{-9}$	$1.84x10^{-9}$	$1.84x10^{-9}$	36	From homology study
E569A	52 (1)	$2.51x10^{-9}$	$1.69x10^{-9}$	$1.69x10^{-9}$	36	From homology study
E569K	48.7 (1)	$2.72x10^{-9}$	$1.82x10^{-9}$	$1.82x10^{-9}$	36	From homology study
E569P	3.72 (1)	$3.20x10^{-8}$	$1.90x10^{-8}$	$1.90x10^{-8}$	36	From homology study; Rd1 mutant
A573T	32.9 (1)	$5.24x10^{-9}$	$3.37x10^{-9}$	$7.65x10^{-9}$	36	From saturation mutagenesis library; Rd2 mutant
	34.8 (1)	$4.82x10^{-9}$	$3.25x10^{-9}$	$6.77x10^{-9}$	36	
V574F	20.9 (1)	$1.04x10^{-8}$	$7.17x10^{-9}$	$1.43x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
V574L	31.2 (1)	$9.34x10^{-9}$	$6.35x10^{-9}$	$1.30x10^{-8}$	36	From saturation mutagenesis library; Rd2 mutant
E575L	35.5 (1)	$3.72x10^{-9}$	$2.45x10^{-9}$	$5.32x10^{-9}$	36	From saturation mutagenesis library
	22.8 (1)	$6.37x10^{-9}$	$4.12x10^{-9}$	$9.26x10^{-9}$	36	From saturation mutagenesis

F578I	22.7 (1)	5.82×10^{-9}	3.54×10^{-9}	8.85×10^{-9}	36	From saturation mutagenesis library; Rd2 mutant
	26.6 (1)	4.38×10^{-9}	2.75×10^{-9}	6.51×10^{-9}	36	
L592M	51.3 (1)	1.28×10^{-9}	7.43×10^{-10}	7.43×10^{-10}	36	From homology study
A599S, C639T	49.4 (3)	3.33×10^{-9}	2.47×10^{-9}	4.35×10^{-9}	48	From saturation mutagenesis library
	54.7 (1)	2.53×10^{-9}	1.61×10^{-9}	3.71×10^{-9}	36	
D614L	50.5 (3)	3.22×10^{-9}	1.48×10^{-9}	5.98×10^{-9}	45	From homology study
M618V	99.8 (1)	2.20×10^{-9}	1.58×10^{-9}	1.58×10^{-9}	35	From homology study
E620D	20.8 (1)	3.68×10^{-9}	2.11×10^{-9}	2.11×10^{-9}	36	From homology study
A621K	6.3 (1)	2.35×10^{-9}	5.85×10^{-10}	5.85×10^{-10}	36	From homology study
A621S	63.5 (1)	4.02×10^{-9}	2.94×10^{-9}	2.94×10^{-9}	35	From homology study
L622I, C639I	31.3 (3)	4.78×10^{-8}	4.12×10^{-8}	5.46×10^{-8}	48	From saturation mutagenesis library; Rd2 mutant
	42.9 (1)	3.70×10^{-8}	3.06×10^{-8}	4.38×10^{-8}	36	
C627V	42.7 (1)	4.73×10^{-9}	3.35×10^{-9}	3.35×10^{-9}	36	From homology study
V630K	57.3 (1)	3.26×10^{-9}	2.28×10^{-9}	2.28×10^{-9}	36	From homology study
V630Y	50.7 (1)	2.48×10^{-9}	1.59×10^{-9}	1.59×10^{-9}	35	From homology study
N631A	42.3 (1)	1.87×10^{-9}	1.12×10^{-9}	2.90×10^{-9}	36	From saturation mutagenesis library
C639V	52.8 (1)	2.54×10^{-9}	1.73×10^{-9}	1.73×10^{-9}	36	From homology study
C639Y	67.6 (3)	2.90×10^{-9}	1.16×10^{-9}	5.88×10^{-9}	48	From homology study
L640A	4.26 (3)	2.47×10^{-7}	1.54×10^{-7}	3.71×10^{-7}	45	From homology study; Rd1 mutant
L640N	5.62 (1)	6.60×10^{-9}	2.84×10^{-9}	2.84×10^{-9}	36	From homology study; Rd1 mutant
L640D	4.51 (1)	1.05×10^{-9}	6.00×10^{-11}	6.00×10^{-11}	36	From homology study
L640G	30 (1)	7.47×10^{-9}	5.27×10^{-9}	5.27×10^{-9}	36	From homology study; Rd1 mutant
	38.8 (1)	5.25×10^{-9}	3.67×10^{-9}	7.16×10^{-9}	36	From saturation mutagenesis library; Rd1 mutant
L640K	5.98 (1)	1.75×10^{-9}	2.92×10^{-10}	2.92×10^{-10}	35	From homology study
L640F	19.7 (3)	1.10×10^{-8}	5.52×10^{-9}	1.91×10^{-8}	47	From homology study; Rd1 mutant
L640W	51.6 (1)	2.25×10^{-9}	1.51×10^{-9}	1.51×10^{-9}	36	From homology study
L640Y	59.3 (3)	6.51×10^{-9}	3.85×10^{-9}	1.01×10^{-8}	46	From homology study; Rd1 mutant

K643N	37.4 (1)	$2.69x10^{-9}$	$1.68x10^{-9}$	$1.68x10^{-9}$	36	From homology study
K643S	3.34 (1)	$2.42x10^{-9}$	$8.52x10^{-10}$	$8.52x10^{-10}$	35	From homology study
K643V	7.31 (1)	$5.92x10^{-10}$	$1.47x10^{-10}$	$1.47x10^{-10}$	36	From homology study
L645N	12.1 (3)	$6.21x10^{-9}$	$2.36x10^{-9}$	$1.13x10^{-8}$	48	From homology study
L645M	69.1 (3)	$1.52x10^{-8}$	$8.75x10^{-9}$	$2.39x10^{-8}$	21	From homology study; Rd1 mutant
Y646A	3.46 (1)	$2.25x10^{-9}$	$3.74x10^{-10}$	$3.74x10^{-10}$	36	From homology study
Y646F	22.3 (1)	$4.42x10^{-9}$	$2.78x10^{-9}$	$2.78x10^{-9}$	36	From homology study
A648S	30.2 (1)	$1.80x10^{-9}$	$9.99x10^{-10}$	$9.99x10^{-10}$	35	From homology study
S649A	46.1 (1)	$3.88x10^{-9}$	$2.73x10^{-9}$	$2.73x10^{-9}$	35	From homology study
F652G	58.5 (1)	$1.62x10^{-9}$	$9.06x10^{-10}$	$2.64x10^{-9}$	24	From saturation mutagenesis library
	56 (1)	$1.29x10^{-9}$	$6.33x10^{-10}$	$2.27x10^{-9}$	23	
F652L	2.45 (1)	$1.62x10^{-8}$	$7.00x10^{-9}$	$7.00x10^{-9}$	36	From homology study; Rd1 mutant
Y653R	26.1 (1)	$3.08x10^{-9}$	$1.83x10^{-9}$	$1.83x10^{-9}$	35	From homology study
Y653H	40.6 (1)	$2.79x10^{-9}$	$1.76x10^{-9}$	$1.76x10^{-9}$	36	From homology study
Y653L	2.81 (1)	$1.42x10^{-8}$	$6.13x10^{-9}$	$6.13x10^{-9}$	36	From homology study; Rd1 mutant
Q655H	37.9 (1)	$2.08x10^{-9}$	$1.26x10^{-9}$	$1.26x10^{-9}$	36	From homology study
P656A	24.5 (1)	$6.17x10^{-9}$	$4.18x10^{-9}$	$4.18x10^{-9}$	36	From homology study; Rd1 mutant
R662T	106 (1)	$1.83x10^{-9}$	$1.21x10^{-9}$	$2.61x10^{-9}$	36	From saturation mutagenesis library
	109 (1)	$1.26x10^{-9}$	$8.34x10^{-10}$	$1.81x10^{-9}$	36	
S664H	56 (1)	$1.57x10^{-9}$	$8.58x10^{-10}$	$2.58x10^{-9}$	36	From saturation mutagenesis library
	63.7 (1)	$1.27x10^{-9}$	$6.69x10^{-10}$	$2.12x10^{-9}$	36	
I708Q	122 (1)	$9.31x10^{-10}$	$6.05x10^{-10}$	$1.35x10^{-9}$	36	From saturation mutagenesis library
N725R	54.6 (1)	$2.70x10^{-9}$	$1.80x10^{-9}$	$3.83x10^{-9}$	36	From saturation mutagenesis library
K769G	49.9 (1)	$3.74x10^{-9}$	$2.57x10^{-9}$	$5.17x10^{-9}$	36	From saturation mutagenesis library
N773Q	46 (1)	$4.80x10^{-9}$	$3.46x10^{-9}$	$3.46x10^{-9}$	36	From homology study
V774A	52.8 (1)	$2.41x10^{-9}$	$1.59x10^{-9}$	$1.59x10^{-9}$	36	From homology study
V774I	22.7 (1)	$9.02x10^{-9}$	$6.46x10^{-9}$	$6.46x10^{-9}$	35	From homology study; Rd1 mutant
I775A	16.1 (3)	$8.66x10^{-9}$	$4.09x10^{-9}$	$1.44x10^{-8}$	47	From homology study; Rd1 mutant
	45.7 (1)	$3.05x10^{-9}$	$2.06x10^{-9}$	$2.06x10^{-9}$	36	
I777A	28.5 (1)	$8.63x10^{-9}$	$6.40x10^{-9}$	$6.40x10^{-9}$	36	From homology study; Rd1 mutant
I777K	19.6 (3)	$2.39x10^{-8}$	$1.95x10^{-8}$	$2.88x10^{-8}$	48	From saturation mutagenesis library; Rd2 mutant
	46.6 (1)	$8.06x10^{-9}$	$6.15x10^{-9}$	$1.02x10^{-8}$	36	
	49.8 (1)	$7.12x10^{-9}$	$5.21x10^{-9}$	$9.36x10^{-9}$	36	
	28.9 (3)	$1.68x10^{-8}$	$1.35x10^{-8}$	$2.05x10^{-8}$	45	
	24.8 (3)	$2.05x10^{-8}$	$1.65x10^{-8}$	$2.49x10^{-8}$	45	
I777K, W814N	31.5 (3)	$1.86x10^{-8}$	$1.53x10^{-8}$	$2.23x10^{-8}$	48	From saturation mutagenesis library; Rd2 mutant
	39.6 (1)	$1.29x10^{-8}$	$9.96x10^{-9}$	$1.63x10^{-8}$	36	
I777V	46.3 (1)	$4.71x10^{-9}$	$3.39x10^{-9}$	$3.39x10^{-9}$	36	From homology study

I778A	3.41 (3)	$5.13x10^{-9}$	$7.22x10^{-10}$	$1.22x10^{-8}$	48	From homology study
M779L	43.2 (3)	$7.22x10^{-9}$	$4.34x10^{-9}$	$1.11x10^{-8}$	46	From homology study; Rd1 mutant
M779S	3.58 (1)	$4.04x10^{-9}$	$1.01x10^{-9}$	$1.01x10^{-9}$	36	From homology study
S781G, L782G, W783Y	23.7 (3)	$1.69x10^{-8}$	$8.83x10^{-9}$	$2.89x10^{-8}$	48	From homology study; Rd1 mutant
K785R	5.93 (1)	$5.04x10^{-9}$	$2.17x10^{-9}$	$2.17x10^{-9}$	36	From homology study
K785S	3.6 (1)	$2.23x10^{-9}$	$3.70x10^{-10}$	$3.70x10^{-10}$	36	From homology study
A787V	49.9 (3)	$3.99x10^{-9}$	$1.72x10^{-9}$	$7.69x10^{-9}$	48	From homology study
W790P	4.81 (3)	$1.03x10^{-8}$	$1.72x10^{-9}$	$3.19x10^{-8}$	48	From homology study
V791H	50.9 (1)	$1.46x10^{-9}$	$8.37x10^{-10}$	$2.33x10^{-9}$	36	From saturation mutagenesis library
D804M	86.1 (1)	$1.25x10^{-9}$	$7.69x10^{-10}$	$1.88x10^{-9}$	36	From saturation mutagenesis library
I824V	75 (1)	$1.59x10^{-9}$	$1.02x10^{-9}$	$2.32x10^{-9}$	36	From saturation mutagenesis library
P833C	22 (1)	$6.44x10^{-9}$	$4.11x10^{-9}$	$9.47x10^{-9}$	36	From saturation mutagenesis library; Rd2 mutant
M848N	25.5 (1)	$3.31x10^{-9}$	$2.07x10^{-9}$	$2.07x10^{-9}$	35	From homology study
M848D	6.65 (1)	$2.51x10^{-10}$	$3.38x10^{-11}$	$3.38x10^{-11}$	36	From homology study
M848H	4.95 (1)	$3.48x10^{-9}$	$1.08x10^{-9}$	$1.08x10^{-9}$	36	From homology study
M848P	7.73 (1)	$2.45x10^{-9}$	$7.62x10^{-10}$	$7.62x10^{-10}$	35	From homology study
K849H, K857S	25.9 (3)	$2.79x10^{-9}$	$1.80x10^{-9}$	$4.07x10^{-9}$	48	From saturation mutagenesis library
	44 (1)	$2.12x10^{-9}$	$1.23x10^{-9}$	$3.35x10^{-9}$	36	
I851L	63.1 (1)	$4.68x10^{-9}$	$3.43x10^{-9}$	$6.14x10^{-9}$	36	From saturation mutagenesis library
E858H	63.4 (1)	$1.90x10^{-9}$	$1.23x10^{-9}$	$2.75x10^{-9}$	36	From saturation mutagenesis library
E861R	66 (1)	$1.69x10^{-9}$	$1.08x10^{-9}$	$1.08x10^{-9}$	36	From homology study
C862I	4.18 (1)	$4.82x10^{-10}$	$6.23x10^{-11}$	$6.23x10^{-11}$	35	From homology study
S865G	46.3 (1)	$2.03x10^{-9}$	$1.27x10^{-9}$	$1.27x10^{-9}$	35	From homology study
D866N	3.46 (1)	$2.54x10^{-9}$	$4.23x10^{-10}$	$4.23x10^{-10}$	36	From homology study
S869T	4.28 (1)	$1.74x10^{-9}$	$6.09x10^{-10}$	$6.09x10^{-10}$	36	From homology study
F871I	48.3 (1)	$1.24x10^{-9}$	$7.03x10^{-10}$	$2.00x10^{-9}$	36	From saturation mutagenesis library
F871Y	28 (3)	$3.92x10^{-8}$	$3.27x10^{-8}$	$4.61x10^{-8}$	42	From saturation mutagenesis library; Rd2 mutant
	29.3 (3)	$4.29x10^{-8}$	$3.62x10^{-8}$	$5.00x10^{-8}$	43	
	35.5 (1)	$3.12x10^{-8}$	$2.50x10^{-8}$	$3.81x10^{-8}$	36	
V872I	5.18 (3)	$1.68x10^{-8}$	$5.21x10^{-9}$	$3.89x10^{-8}$	48	From homology study; Rd1 mutant
V872L	60.4 (1)	$2.17x10^{-9}$	$1.46x10^{-9}$	$1.46x10^{-9}$	36	From homology study
H873R	52.9 (1)	$2.75x10^{-9}$	$1.81x10^{-9}$	$1.81x10^{-9}$	35	From homology study
H873T	49.5 (1)	$3.16x10^{-9}$	$2.16x10^{-9}$	$2.16x10^{-9}$	36	From homology study
K874P	87.5 (1)	$3.01x10^{-9}$	$2.19x10^{-9}$	$3.98x10^{-9}$	36	From saturation mutagenesis library
L900S	55.9 (1)	$5.24x10^{-9}$	$3.72x10^{-9}$	$7.05x10^{-9}$	36	From saturation mutagenesis

L909S	68.8 (1)	4.77×10^{-9}	3.50×10^{-9}	6.26×10^{-9}	36	From saturation mutagenesis library; Rd2 mutant
L909F	58.2 (1)	6.99×10^{-9}	5.34×10^{-9}	8.87×10^{-9}	36	From saturation mutagenesis library; Rd2 mutant
K934F	44.8 (1)	2.23×10^{-9}	1.36×10^{-9}	3.41×10^{-9}	36	From saturation mutagenesis library
K934W	34 (1)	1.20×10^{-8}	9.01×10^{-9}	1.55×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
S955W, K967C	12.6 (3)	5.69×10^{-9}	3.61×10^{-9}	8.41×10^{-9}	48	From saturation mutagenesis library
	21.9 (1)	4.55×10^{-9}	2.46×10^{-9}	7.55×10^{-9}	36	
D958A	88 (1)	1.98×10^{-9}	1.34×10^{-9}	2.78×10^{-9}	36	From saturation mutagenesis library
F968C	14.4 (1)	1.23×10^{-8}	7.42×10^{-9}	1.88×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
F968T	12.8 (1)	1.98×10^{-8}	1.24×10^{-8}	2.96×10^{-8}	36	From saturation mutagenesis library; Rd2 mutant
L622I, C639I, I775A	16.1 (3)	3.26×10^{-8}	2.56×10^{-8}	4.06×10^{-8}	48	From shuffling library (A X B)
L622I, C639I, I777A	7.74 (3)	1.11×10^{-7}	9.04×10^{-8}	1.33×10^{-7}	48	From shuffling library (A X B)
L622I, C639I, I777K	7.59 (3)	3.06×10^{-7}	2.66×10^{-7}	3.47×10^{-7}	48	From shuffling library (A X B)
L622I, C639I, I777K, W814N	10.4 (3)	1.44×10^{-7}	1.21×10^{-7}	1.69×10^{-7}	46	From shuffling library (A X B)
L622I, C639I, S781G, L782G, W783Y	11.4 (3)	4.41×10^{-10}	2.52×10^{-11}	1.94×10^{-9}	48	From shuffling library (A X B)
C639Y, L640A, I777A	18.9 (3)	7.16×10^{-8}	6.09×10^{-8}	8.29×10^{-8}	48	From shuffling library (A X B)
L640A, I775A	58.6 (3)	1.68×10^{-8}	1.40×10^{-8}	1.98×10^{-8}	48	From shuffling library (A X B)
L640A, I777A	20.5 (3)	9.06×10^{-8}	7.77×10^{-8}	1.04×10^{-7}	46	From shuffling library (A X B)
L640A, I777K	10.8 (3)	1.83×10^{-7}	1.57×10^{-7}	2.09×10^{-7}	48	From shuffling library (A X B)

L640A, I777K, W814N	10.2 (3)	$5.41x10^{-7}$	$4.80x10^{-7}$	$6.02x10^{-7}$	48	From shuffling library (A X B)
L640A, M779L	40.1 (3)	$2.79x10^{-9}$	$1.81x10^{-9}$	$4.07x10^{-9}$	47	From shuffling library (A X B)
L640N, I777A	12.4 (3)	$3.04x10^{-8}$	$2.23x10^{-8}$	$3.98x10^{-8}$	46	From shuffling library (A X B)
L640G, I775A	32.5 (3)	$1.81x10^{-8}$	$1.44x10^{-8}$	$2.21x10^{-8}$	48	From shuffling library (A X B)
L640G, I777A	19.5 (3)	$9.28x10^{-8}$	$7.97x10^{-8}$	$1.06x10^{-7}$	47	From shuffling library (A X B)
L640G, I777K, W814N	5.21 (3)	$2.63x10^{-7}$	$2.19x10^{-7}$	$3.09x10^{-7}$	48	From shuffling library (A X B)
L640Y, I777A	43.1 (3)	$5.70x10^{-8}$	$4.96x10^{-8}$	$6.45x10^{-8}$	45	From shuffling library (A X B)
L640Y, I777K, W814N	13.8 (3)	$9.70x10^{-7}$	$7.97x10^{-7}$	$1.14x10^{-6}$	16	From shuffling library (A X B); Rd3 mutant
	2.56 (3)	$3.36x10^{-6}$	$2.23x10^{-6}$	$4.54x10^{-6}$	6	
L645M, I777K	13.1 (3)	$9.19x10^{-9}$	$5.72x10^{-9}$	$1.38x10^{-8}$	46	From shuffling library (A X B)
L645M, M779L	8.71 (3)	$4.99x10^{-9}$	$2.15x10^{-9}$	$9.63x10^{-9}$	47	From shuffling library (A X B)
F652L, I775A	14.4 (3)	$6.08x10^{-9}$	$3.60x10^{-9}$	$9.45x10^{-9}$	45	From shuffling library (A X B)
Y653L, I775A	7.64 (3)	$9.70x10^{-9}$	$5.22x10^{-9}$	$1.62x10^{-8}$	48	From shuffling library (A X B)
Y653L, I777A	8.52 (3)	$2.02x10^{-8}$	$1.31x10^{-8}$	$2.96x10^{-8}$	45	From shuffling library (A X B)
Y653L, I777K	13.6 (3)	$9.59x10^{-8}$	$7.82x10^{-8}$	$1.15x10^{-7}$	43	From shuffling library (A X B)
I775A, F871Y	39.4 (3)	$4.73x10^{-8}$	$4.09x10^{-8}$	$5.39x10^{-8}$	48	From shuffling library (B X C)
I775A, L900S	39.1 (3)	$2.53x10^{-8}$	$2.12x10^{-8}$	$2.98x10^{-8}$	48	From shuffling library (B X C)
I775A, L909F	31.8 (3)	$2.46x10^{-8}$	$2.03x10^{-8}$	$2.92x10^{-8}$	48	From shuffling library (B X C)
I775A, K934W	28.8 (3)	$3.59x10^{-8}$	$3.00x10^{-8}$	$4.23x10^{-8}$	48	From shuffling library (B X C)
I775A, F968T	6.62 (3)	$7.29x10^{-8}$	$5.52x10^{-8}$	$9.35x10^{-8}$	45	From shuffling library (B X C)
I777A, F871Y	12.9 (3)	$2.57x10^{-7}$	$2.24x10^{-7}$	$2.91x10^{-7}$	45	From shuffling library (B X C)

I777A, V872I	11.8 (3)	$1.34x10^{-10}$	$3.18x10^{-10}$	$1.89x10^{-11}$	48	From shuffling library (B X C)
I777A, L900S	26.9 (3)	$1.36x10^{-7}$	$1.20x10^{-7}$	$1.51x10^{-7}$	48	From shuffling library (B X C)
I777A, L909F	13 (3)	$1.07x10^{-7}$	$9.02x10^{-8}$	$1.24x10^{-7}$	46	From shuffling library (B X C)
I777A, K934W	6.04 (3)	$1.86x10^{-7}$	$1.55x10^{-7}$	$2.18x10^{-7}$	48	From shuffling library (B X C)
I777K, W790L, K934W	8.23 (3)	$2.02x10^{-7}$	$1.73x10^{-7}$	$2.33x10^{-7}$	47	From shuffling library (B X C)
I777K, F871Y	11.2 (3)	$1.35x10^{-6}$	$1.14x10^{-6}$	$1.56x10^{-6}$	16	From shuffling library (B X C); Rd3 mutant
I777K, V872I	33 (3)	$2.59x10^{-8}$	$2.15x10^{-8}$	$3.06x10^{-8}$	47	From shuffling library (B X C)
I777K, L900S	20.6 (3)	$9.22x10^{-7}$	$7.82x10^{-7}$	$1.05x10^{-6}$	16	From shuffling library (B X C); Rd3 mutant
	12.4 (3)	$2.87x10^{-6}$	$2.07x10^{-6}$	$3.55x10^{-6}$	4	
	15 (3)	$2.18x10^{-6}$	$1.55x10^{-6}$	$2.72x10^{-6}$	4	
I777K, L909F	9.15 (3)	$2.22x10^{-7}$	$1.91x10^{-7}$	$2.55x10^{-7}$	46	From shuffling library (B X C)
I777K, K934W	8.27 (3)	$2.71x10^{-7}$	$2.33x10^{-7}$	$3.11x10^{-7}$	48	From shuffling library (B X C)
I777K, F968C	6.72 (3)	$6.67x10^{-7}$	$5.30x10^{-7}$	$8.07x10^{-7}$	16	From shuffling library (B X C)
I777K, F968T	7.11 (3)	$3.19x10^{-7}$	$2.69x10^{-7}$	$3.71x10^{-7}$	45	From shuffling library (B X C)
M779L, K934W	23.1 (3)	$2.85x10^{-9}$	$1.62x10^{-9}$	$4.57x10^{-9}$	48	From shuffling library (B X C)
M779L, F968C	6.57 (3)	$3.13x10^{-10}$	$7.42x10^{-10}$	$4.40x10^{-11}$	48	From shuffling library (B X C)
V774I, F871Y	6.57 (3)	$1.64x10^{-7}$	$1.33x10^{-7}$	$1.97x10^{-7}$	43	From shuffling library (B X C)
V774I, L900S	20.7 (3)	$1.10x10^{-7}$	$9.54x10^{-8}$	$1.24x10^{-7}$	46	From shuffling library (B X C)
V774I, F968C	8.91 (3)	$1.14x10^{-7}$	$9.11x10^{-8}$	$1.39x10^{-7}$	46	From shuffling library (B X C)
Y431H, L640Y, I777K, W814N	9.69 (3)	$3.65x10^{-6}$	$2.65x10^{-6}$	$4.71x10^{-6}$	11	Rd4 mutant; ***Entries 4-6 for this TP-DNAP1 variant are measurements taken after 90 generations.***
	17.3 (3)	$1.65x10^{-6}$	$1.30x10^{-6}$	$1.99x10^{-6}$	11	
	7.97 (3)	$2.67x10^{-6}$	$1.48x10^{-6}$	$3.89x10^{-6}$	3	
	12.3 (3)	***$2.33x10^{-6}$***	***$1.70x10^{-6}$***	***$2.90x10^{-6}$***	***5***	
	7.65 (3)	***$4.29x10^{-6}$***	***$2.93x10^{-6}$***	***$5.66x10^{-6}$***	***6***	
	4.83 (3)	***$3.24x10^{-6}$***	***$2.25x10^{-6}$***	***$4.17x10^{-6}$***	***5***	
L474W, L640Y,	9.56 (3)	$4.03x10^{-6}$	$2.85x10^{-6}$	$5.13x10^{-6}$	5	Rd4 mutant; ***Entries 3-4 for this TP-DNAP1 variant are***
	3.55 (3)	$2.74x10^{-6}$	$1.84x10^{-6}$	$3.64x10^{-6}$	5	

I777K, W814N	***3.82 (3)***	***$2.20x10^{-6}$***	***$1.49x10^{-6}$***	***$2.89x10^{-6}$***	***5***	***measurements taken after 90 generations.***
	2.38 (3)	***$1.04x10^{-6}$***	***$6.08x10^{-7}$***	***$1.54x10^{-6}$***	***6***	
V574F, I777K, L900S	7.8 (3)	$4.86x10^{-6}$	$3.77x10^{-6}$	$5.94x10^{-6}$	11	Rd4 mutant; TP-DNAP1-4-1 ***Entries 4-6 for this TP-DNAP1 variant are measurements taken after 90 generations.***
	5.95 (3)	$8.53x10^{-6}$	$6.26x10^{-6}$	$1.09x10^{-5}$	10	
	5.09 (3)	$3.78x10^{-6}$	$2.58x10^{-6}$	$4.86x10^{-6}$	4	
	6.02 (3)	***$7.58x10^{-6}$***	***$5.50x10^{-6}$***	***$9.47x10^{-6}$***	***5***	
	4.08 (3)	***$6.87x10^{-6}$***	***$4.76x10^{-6}$***	***$8.89x10^{-6}$***	***5***	
	3.39 (3)	***$7.71x10^{-6}$***	***$5.20x10^{-6}$***	***$1.00x10^{-5}$***	***4***	
L477V, L640Y, I777K, W814N	11.8 (3)	$7.19x10^{-6}$	$5.45x10^{-6}$	$8.68x10^{-6}$	5	Rd4 mutant; TP-DNAP1-4-2; ***Entries 4-6 for this TP-DNAP1 variant are measurements taken after 90 generations.***
	6.09 (3)	$1.88x10^{-5}$	$1.40x10^{-5}$	$2.30x10^{-5}$	5	
	9.67 (3)	$9.41x10^{-6}$	$6.53x10^{-6}$	$1.17x10^{-5}$	3	
	5.05 (3)	***$1.15x10^{-5}$***	***$8.14x10^{-6}$***	***$1.45x10^{-5}$***	***4***	
	7.45 (3)	***$1.19x10^{-5}$***	***$8.70x10^{-6}$***	***$1.48x10^{-5}$***	***5***	
	7.82 (3)	***$9.22x10^{-6}$***	***$6.48x10^{-6}$***	***$1.16x10^{-5}$***	***4***	

WT	59.8 (1)	$2.49x10^{-9}$	$1.77x10^{-9}$	$3.37x10^{-9}$	48	***Entries 24-27 for this TP-DNAP1 variant are measurements taken after 90 generations.***
	47.8 (3)	$4.21x10^{-9}$	$1.51x10^{-9}$	$9.03x10^{-9}$	23	
	59.8 (1)	$2.63x10^{-9}$	$1.89x10^{-9}$	$1.89x10^{-9}$	36	
	72.6 (1)	$4.37x10^{-9}$	$2.42x10^{-9}$	$7.16x10^{-9}$	48	
	46.9 (1)	$3.83x10^{-9}$	$1.65x10^{-9}$	$7.39x10^{-9}$	48	
	53.9 (1)	$2.69x10^{-9}$	$1.79x10^{-9}$	$3.83x10^{-9}$	36	
	55.8 (1)	$2.38x10^{-9}$	$1.39x10^{-9}$	$3.73x10^{-9}$	24	
	57.3 (1)	$2.36x10^{-9}$	$1.46x10^{-9}$	$3.55x10^{-9}$	36	
	78.6 (1)	$2.26x10^{-9}$	$1.44x10^{-9}$	$3.30x10^{-9}$	24	
	55.3 (1)	$2.23x10^{-9}$	$1.41x10^{-9}$	$3.28x10^{-9}$	36	
	58.9 (1)	$2.06x10^{-9}$	$1.36x10^{-9}$	$2.95x10^{-9}$	36	
	54.3 (1)	$1.96x10^{-9}$	$1.14x10^{-9}$	$3.09x10^{-9}$	36	
	52.7 (1)	$1.95x10^{-9}$	$1.10x10^{-9}$	$3.14x10^{-9}$	36	
	59.8 (1)	$1.84x10^{-9}$	$9.75x10^{-10}$	$3.07x10^{-9}$	19	
	68.3 (1)	$1.73x10^{-9}$	$1.11x10^{-9}$	$2.52x10^{-9}$	36	
	71.5 (1)	$1.65x10^{-9}$	$1.08x10^{-9}$	$2.39x10^{-9}$	36	
	59.5 (1)	$1.62x10^{-9}$	$1.02x10^{-9}$	$2.39x10^{-9}$	36	
	71.8 (1)	$1.37x10^{-9}$	$8.77x10^{-10}$	$2.01x10^{-9}$	36	
	65.9 (1)	$1.22x10^{-9}$	$7.30x10^{-10}$	$1.89x10^{-9}$	36	
	58.8 (1)	$1.16x10^{-9}$	$6.05x10^{-10}$	$1.96x10^{-9}$	24	
	53.1 (3)	$3.10x10^{-9}$	$2.18x10^{-9}$	$4.24x10^{-9}$	48	
	51.8 (3)	$2.23x10^{-9}$	$1.50x10^{-9}$	$3.15x10^{-9}$	45	
	47.7 (3)	$2.49x10^{-9}$	$1.74x10^{-9}$	$3.42x10^{-9}$	45	
	54.2 (3)	***$1.85x10^{-9}$***	***$1.22x10^{-9}$***	***$2.65x10^{-9}$***	***45***	
	46.9 (3)	***$2.95x10^{-9}$***	***$2.05x10^{-9}$***	***$4.06x10^{-9}$***	***45***	
	45.2 (3)	***$1.84x10^{-9}$***	***$1.15x10^{-9}$***	***$2.74x10^{-9}$***	***45***	
	51.4 (3)	***$4.28x10^{-9}$***	***$3.05x10^{-9}$***	***$5.77x10^{-9}$***	***45***	

Appendix F | 210 TP-DNAP1 variants that replicate p1 at a higher copy number than WT TP-DNAP1.

From p1 copy number measurements of 13,625 yeast clones screened in small-scale p1 fluctuation tests, 210 unique variants exhibited elevated copy numbers. Variants were re-transformed into OR-Y24 and subject to additional p1 copy number measurements for verification. Data shown are fold-change mean and standard deviation (calculated using equation (5).2 of Frishman, 1975) of biological triplicate measurements of each mutant and 15 measurements of WT TP-DNAP1. High activity of four TP-DNAP1 variants was independently validated with qPCR measurements of p1 (unpublished results).

Mutation	Mean fold-increase over WT	Standard deviation
I287M	2.39	0.482
S289K	2.02	0.306
N291K	1.66	0.341
N291M	2.28	0.263
N291W	2.26	0.456
L295Q	2.48	0.285
Y296K	2.28	0.408
E298R	2.6	0.66
E298G	2.64	0.365
E298G, S305F	2.06	0.411
E298H	2.57	0.284
E298I	2.34	0.413
E298F	1.74	0.305
E298P	2.57	0.286
E298S	2.43	0.385
E298Y	2.17	0.357
E298V	2.64	0.387
I301A	2.41	0.251
I301C	2.46	0.457
I301E	2.57	0.32
I301L	2.94	0.981
I301K	2.71	0.34
I301M	2.76	0.326
I301S	2.29	0.408
I301Y	3.28	1.39
I301V	2.53	0.317
K302A	1.88	0.436
T303H	2.3	0.317
T303L	2.41	0.332
T303M	2.5	0.298
T303W	2.96	1.74
F304A	2.14	0.234
F304R	3.49	1.85
F304Y	2.41	0.451

S305N	2.3	0.278
S305G	2.53	0.369
I307L	2.76	0.391
D308Q	2.15	0.264
N309C	2.25	0.282
N309K, I311G	2.2	0.256
T310N	2.83	0.353
T310D	2.64	0.325
T310E	2.37	0.466
T310H	2.57	0.361
T310F	2.66	0.35
T310W	2.48	0.255
I311A	2.19	0.324
I311N	2.55	0.572
I311D	1.86	0.595
I311G	2.05	0.484
I311M	2.26	0.232
T312E	2.29	0.279
T312Q	2.85	0.486
Y313H	3.08	1.45
Y313M	1.94	0.373
Y313F	2.11	0.429
Y313W	2.15	0.563
Y313V	2.48	0.977
Y316R	2.46	0.778
I327Q	2.15	0.546
S330R	2.21	0.592
D333A	1.57	0.242
D333N	2.07	0.455
D333N, K344R	2.07	0.398
D333M	2.6	0.342
D333T	2.41	0.279
D333V	2.18	0.26
K365N	1.89	0.198
I372C	1.73	0.316
Y382A	1.6	0.18
Y382V	1.62	0.238
V385W	1.41	0.234
G387H	1.57	0.173
R395P	1.91	0.263
V406C	2.34	0.36
V406Q	2.16	0.231
D407V, S415T	1.95	0.216
D407A	2.25	0.305
D407C	2.26	0.408
D407G	2.53	0.281
D407H	2.57	0.314
D407M	2.71	0.283

D407F	2.28	0.286
D407T	3.95	1.87
G410R	2.87	0.46
G410E	2.37	0.284
G410Q	2.2	0.229
G410H	2.19	0.334
G410I	2.57	0.321
G410L	2.57	0.31
G410K	2.53	0.729
G410M	2.55	0.516
G410F	2.66	0.291
G410S	2.27	0.323
G410T	2.57	0.304
G410W	2.5	0.363
G410Y	2.73	1.36
E411A	2.08	0.407
E411R	2.07	0.278
E411N	1.9	0.309
E411D	2.32	0.28
E411G	1.54	0.357
E411L	2.22	0.335
E411K	2.39	0.307
E411S	1.52	0.162
L412V	2.21	0.365
N413E	2.62	0.366
N413H	2.2	0.327
N413T	1.76	0.228
I414F	1.29	0.186
S415V	1.44	0.223
G426T	1.22	0.141
D460G	1.63	0.221
G461K	2.15	1
N611Q	1.92	0.249
N611P, V897V	2.04	0.259
K612A	1.94	0.347
K612Q	2.04	0.249
K612S	1.75	0.531
K612V	1.85	0.355
E613D	1.82	0.196
F616E	2.2	0.249
E620R	2.29	0.744
R662A	1.75	0.196
R662C	1.7	0.227
D669G, T698M	1.65	0.274
E670M, S693G	1.55	0.368
I673V, E703D	1.88	0.229
Y675H	2.14	0.259
Y675L	1.96	0.244

R677Q	1.84	0.189
R677L	1.68	0.386
R677M	1.97	0.211
R677S	1.87	0.268
T679A	1.82	0.256
T679R	1.73	0.346
T679E	1.7	0.176
T679Q	1.71	0.369
N681Q, R692I	1.74	0.225
N681F	1.75	0.263
R682D	1.74	0.214
R682P	1.75	0.246
N683A	1.93	0.703
N683H	1.68	0.242
N683T	1.57	0.199
N684D	1.54	0.173
N687G	1.65	0.233
R692C	1.69	0.34
R692I	1.99	0.409
R692K	1.88	0.24
R692F	2.6	1.29
R692F, T706E	1.89	0.257
R692W	1.97	0.237
R692V	2.29	0.487
S693R	2.09	0.23
S693Q	2.22	0.382
H694I	1.78	0.264
H694T	1.92	0.311
N695E	2.34	0.328
N695Q	1.68	0.622
N695H	1.76	0.183
N695F	1.44	0.265
N695S	1.82	0.365
K696C	1.87	0.228
K696G	2.11	0.231
K696M	1.71	0.248
K696S, E704V	1.86	0.253
K696T	2.12	0.257
K696Y	1.49	0.159
T698A	2.02	0.228
T698L	2	0.266
T698M	1.75	0.239
E703R	1.84	0.274
E703H	1.53	0.164
E703W	1.72	0.213
E704N	1.85	0.376
E704D	1.89	0.366
E704G	1.75	0.215

E704I	1.35	0.202
E704K	1.8	0.422
E704M	2.02	0.228
S705R	1.84	0.277
S705L	1.98	0.308
S705F	1.76	0.204
T706R	2	0.332
T706Q	2.92	1.5
T706G	1.72	0.226
T706P	1.73	0.235
T706W	1.73	0.297
I708A	2.55	0.276
I708E	2.41	0.264
I708L	2.16	0.284
I708M	2.32	0.35
I708T	2.41	0.57
I708V	2.01	0.6
A709N	2.01	0.522
I729F	1.7	0.285
I729W	1.97	0.406
I729V	1.87	0.254
S733R	1.98	0.207
S733N	1.86	0.266
S733E	2.06	0.354
S733Q	1.86	0.191
I824E	2.12	0.327
Y829I	1.86	0.202
Y829F	2.46	0.787
S831T	1.71	0.182
K959A	1.55	0.166
K959M	2.02	0.288
K962Q	1.56	0.171

Appendix G | A list of all plasmids and yeast strains used in Chapter 3-6.

#	Name	Source	Parent plasmid or strain	Description (plasmid entries are split into three columns denoting: (1) origin of replication (yeast, bacteria), (2) selection marker (yeast, bacteria), (3) notes)		
1	AR-Ec265	Previous work (Ravikumar et al., 2014)	See previous work	*CEN6/ARS4, ColE1*	*HIS4, KanR*	*REV1 promoter > Recoded WT TP-DNAP1*
2	AR-Ec318	This work	AR-Ec265	*CEN6/ARS4, ColE1*	*HIS3, KanR*	*REV1 promoter > Recoded WT TP-DNAP1 with HIS3 instead of HIS4*
3	AR-Ec244	Previous work (Ravikumar et al., 2014)	See previous work	N/A, *ColE1*	*URA3 (p1), AmpR*	p1 recombination cassette that integrates URA3 and leu2 (538C>T) in place of WT TP-DNAP1 with to create p1-FulldelPol-U-I*(TAA)
4	AR-Ec354	This work	AR-Ec244	N/A, *ColE1*	*URA3 (p1), AmpR*	p1 recombination cassette that integrates mKate2, URA3, and leu2 (538C>T) in place of WT TP-DNAP1 to create p1-FulldelPol-mK-U-I*(TAA)
5	GA-Ec78	This work	AR-Ec354	N/A, *ColE1*	*TRP1 (p1), AmpR*	p1 recombination cassette that integrates mKate2 and TRP1 in place of WT TP-DNAP1 to create p1-FulldelPol-mK-W
6	AR-Ec507	This work	AR-Ec354	N/A, *ColE1*	*URA3 (p1), AmpR*	p1 recombination cassette that integrates mKate2, URA3, and leu2 (538C>T, 540A>G) in place of WT TP-DNAP1 to create p1-FulldelPol-mK-U-I*(TAG)

7	AR-Ec281	Previous work (Ravikumar et al., 2014)	See previous work	N/A, *ColE1*	*LEU2 (p1)*, *AmpR*	p1 recombination cassette that integrates LEU2 and ura3 (278A>G) in place of WT TP-DNAP1 to create p1-FulldelPol-L-u(K93R)
8	AR-Ec596	Gift from A. Herr (pGL310)	N/A	*CEN4/ARS1, ColE1*	*URA3*, *AmpR*	*SUP11 and POL3 promoter > WT POL3*
9	AR-Ec600	Gift from A. Herr (YCplac111)	N/A	*CEN4/ARS1, ColE1*	*LEU2, AmpR*	Shuttle vector
10	AR-Ec601	Gift from A. Herr (YCplac111 *POL3*)	N/A	*CEN4/ARS1, ColE1*	*LEU2, AmpR*	*POL3 promoter > WT POL3*
11	AR-Ec602	Gift from A. Herr (YCplac111 *pol3-01*)	N/A	*CEN4/ARS1, ColE1*	*LEU2, AmpR*	*POL3* promoter > *pol3-01*
12	AR-Ec605	Gift from A. Herr (YCplac111 *pol3-01+T711A*)	N/A	*CEN4/ARS1, ColE1*	*LEU2, AmpR*	*POL3* promoter > *pol3-01+T711A*
13	AR-Ec606	Gift from A. Herr (YCplac111 *pol3-01+Y808C*)	N/A	*CEN4/ARS1, ColE1*	*LEU2, AmpR*	*POL3* promoter > *pol3-01+Y808C*
14	AR-Ec607	Gift from A. Herr (YCplac111 *pol3-01+H879Y*)	N/A	*CEN4/ARS1, ColE1*	*LEU2, AmpR*	*POL3* promoter > *pol3-01+H879Y*
15	AR-Ec609	Gift from A. Herr (YCplac111 *pol3-01+S968R*)	N/A	*CEN4/ARS1, ColE1*	*LEU2, AmpR*	*POL3* promoter > *pol3-01+S968R*
16	GA-Ec64	This work	GA-Ec78	N/A, *ColE1*	*TRP1 (p1)*, *AmpR*	p1 recombination cassette that integrates TRP1, mKate2 PfDHFR, and leu2 (538C>T) in place of WT TP-DNAP1 to create p1-FulldelPol-W-mK-PfDHFR-l*(TAA)

17	GA-Ec51	This work	GA-Ec50	*CEN6/ARS4, ColE1*	*URA3*, *AmpR*	*DFR1* promoter > *DFR1*
18	GA-Ec52	This work	GA-Ec49	*CEN6/ARS4, ColE1*	*LEU2, AmpR*	*DFR1* promoter > *PfDHFR*
19	GA-Ec49	Gift from R. Jajoo (pRS415)	N/A	*CEN6/ARS4, ColE1*	*LEU2, AmpR*	Shuttle vector
20	GA-Ec50	Gift from R. Jajoo (pRS416)	N/A	*CEN6/ARS4, ColE1*	*URA3, AmpR*	Shuttle vector
21	GA-Ec119	This work	GA-Ec52	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR* and *TEF1* promoter > *KanMX*
22	GA-Ec121	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (S108N)* and *TEF1* promoter > *KanMX*
23	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA2	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (C50R)* and *TEF1* promoter > *KanMX*
24	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA3	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (D54N)* and *TEF1* promoter > *KanMX*
25	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA4	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (Y57H)* and *TEF1* promoter > *KanMX*
26	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA5	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (C59R)* and *TEF1* promoter > *KanMX*

27	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA6	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C59Y)* and *TEF1* promoter > *KanMX*
28	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA8	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N)* and *TEF1* promoter > *KanMX*
29	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA9	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, Y57H)* and *TEF1* promoter > *KanMX*
30	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA10	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, C59R)* and *TEF1* promoter > *KanMX*
31	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA11	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, C59Y)* and *TEF1* promoter > *KanMX*
32	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pA12	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, S108N)* and *TEF1* promoter > *KanMX*
33	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB1	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, Y57H)* and *TEF1* promoter > *KanMX*
34	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB2	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, C59R)* and *TEF1* promoter > *KanMX*
35	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB3	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, C59Y)* and *TEF1* promoter > *KanMX*
36	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB4	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, S108N)* and *TEF1* promoter > *KanMX*

37	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB5	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (Y57H, C59R)* and *TEF1* promoter > *KanMX*
38	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB6	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (Y57H, C59Y)* and *TEF1* promoter > *KanMX*
39	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB7	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (Y57H, S108N)* and *TEF1* promoter > *KanMX*
40	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB8	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C59R, S108N)* and *TEF1* promoter > *KanMX*
41	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB9	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C59Y, S108N)* and *TEF1* promoter > *KanMX*
42	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB10	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, Y57H)* and *TEF1* promoter > *KanMX*
43	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB11	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, C59R)* and *TEF1* promoter > *KanMX*
44	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pB12	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, C59Y)* and *TEF1* promoter > *KanMX*
45	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC1	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, S108N)* and *TEF1* promoter > *KanMX*
46	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC2	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, Y57H, C59R)* and *TEF1* promoter > *KanMX*

47	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC3	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, Y57H, C59Y)* and *TEF1* promoter > *KanMX*
48	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC4	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, Y57H, S108N)* and *TEF1* promoter > *KanMX*
49	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC5	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, C59R, S108N)* and *TEF1* promoter > *KanMX*
50	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC6	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, C59Y, S108N)* and *TEF1* promoter > *KanMX*
51	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC7	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, Y57H, C59R)* and *TEF1* promoter > *KanMX*
52	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC8	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, Y57H, C59Y)* and *TEF1* promoter > *KanMX*
53	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC9	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, Y57H, S108N)* and *TEF1* promoter > *KanMX*
54	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC10	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, C59R, S108N)* and *TEF1* promoter > *KanMX*
55	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC11	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, C59Y, S108N)* and *TEF1* promoter > *KanMX*
56	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pC12	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (Y57H, C59R, S108N)* and *TEF1* promoter > *KanMX*

57	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD1	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (Y57H, C59Y, S108N)* and *TEF1* promoter > *KanMX*
58	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD2	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, Y57H, C59R)* and *TEF1* promoter > *KanMX*
59	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD3	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, Y57H, C59Y)* and *TEF1* promoter > *KanMX*
60	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD4	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, Y57H, S108N)* and *TEF1* promoter > *KanMX*
61	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD5	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, C59R, S108N)* and *TEF1* promoter > *KanMX*
62	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD6	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, C59Y, S108N)* and *TEF1* promoter > *KanMX*
63	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD7	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, Y57H, C59R, S108N)* and *TEF1* promoter > *KanMX*
64	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD8	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (C50R, Y57H, C59Y, S108N)* and *TEF1* promoter > *KanMX*
65	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD9	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, Y57H, C59R, S108N)* and *TEF1* promoter > *KanMX*
66	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD10	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX, AmpR*	*DFR1* promoter > *PfDHFR (D54N, Y57H, C59Y, S108N)* and *TEF1* promoter > *KanMX*

67	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD11	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, Y57H, C59R, S108N)* and *TEF1* promoter > *KanMX*
68	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pD12	This work	GA-Ec121	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (C50R, D54N, Y57H, C59Y, S108N)* and *TEF1* promoter > *KanMX*
69	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pE2	This work	GA-Ec119	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (737insA)* and *TEF1* promoter > *KanMX*
70	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pE6	This work	[AR Tray: 11.22.17 DHFR alleles (E. coli)]-pE3	*CEN6/ARS4, ColE1*	*LEU2 and KanMX*, *AmpR*	*DFR1* promoter > *PfDHFR (N51I, C59R, S108N, I164L)* and *TEF1* promoter > *KanMX*
71	AR-Ec451	Addgene (#60847)	N/A	*2μ*, *ColE1*	*KanMX*, *KanR*	*tRNA TYR* promoter > *HDV* ribozyme + *sgRNA* > *SNR52* terminator and *RNR2* promoter > *SpCAS9-SV40NLS-8XHIS* > *CYC1* terminator
72	AR-Ec448	This work	AR-Ec451	*2μ*, *ColE1*	*KanMX*, *KanR*	*tRNA TYR* promoter > *HDV* ribozyme + *LEU2*-targeting sgRNA > *SNR52* terminator and *RNR2* promoter > *SpCAS9-SV40NLS-8XHIS* > *CYC1* terminator
73	AR-Ec634	This work	AR-Ec451	*2μ*, *ColE1*	*KanMX*, *KanR*	*tRNA TYR* promoter > *HDV* ribozyme + *URA3*-targeting sgRNA > *SNR52* terminator and *RNR2* promoter > *SpCAS9-SV40NLS-8XHIS* > *CYC1* terminator

74	AR-Ec556	This work	AR-Ec451	*2μ*, *ColE1*	*KanMX*, *KanR*	*tRNA TYR* promoter > *HDV* ribozyme + *TRP1*-targeting sgRNA > *SNR52* terminator and *RNR2* promoter > *SpCAS9-SV40NLS-8XHIS* > *CYC1* terminator
75	AR-Ec454	This work	AR-Ec451	*2μ*, *ColE1*	*KanMX*, *KanR*	*tRNA TYR* promoter > *HDV* ribozyme + *HIS3*-targeting sgRNA > *SNR52* terminator and *RNR2* promoter > *SpCAS9-SV40NLS-8XHIS* > *CYC1* terminator
76	AR-Ec242	Previous work (Ravikumar et al., 2014)	See previous work	*N/A, ColE1*	*LEU2 (p1), AmpR*	p1 recombination cassette that integrates LEU2 and mKate2 adjacent 3' to WT TP-DNAP1 to create p1-ShortPol-L-m
77	AR-Ec193	Previous work (Ravikumar et al., 2014)	See previous work	N/A, *ColE1*	*KanMX (p1), AmpR*	p1 recombination cassette that integrates KanMX adjacent 3' to WT TP-DNAP1 to create p1-ShortPol-KanMX
78	AR-Ec298	This work	AR-Ec318	N/A, *ColE1*	*KanMX (genomic integration), AmpR*	Genomic recombination cassette that integrates *KanMX* and *MET3(-1 to -593)-TP-DNAP1 (D641A)* at the genomic *URA3* locus
79	AR-Ec649	This work	GA-Ec64	N/A, *ColE1*	*TRP1 (p1)*, *AmpR*	p1 recombination cassette that integrates TRP1, mKate2 PfDHFR(S108TCA), and leu2 (538C>T) in place of WT TP-DNAP1 to create p1-FulldelPol-W-mK-PfDHFR(S108TCA)-l*(TAA)

80	AH22	ATCC (Catalog #38626)	See Gunge et al., 1981	*MAT***a** *can1 his4-519 leu2-3*, *112*
81	F102-2	ATCC (Catalog #200585)	See Gunge et al., 1981	*MAT***a** *can1 his4-519 leu2-3*, *112* ρ^0 + p1 + p2
82	AR-Y246	This work	F102-2	*MAT***a** *can1 his4-519 leu2-2, 112 ura3Δ0* ρ^0 + p1 + p2
83	AR-Y258	This work	AR-Y246	*MAT***a** *can1 his4-519 leu2Δ0 ura3Δ0* ρ^0 + p1 + p2
84	AR-Y288	This work	AR-Y258	*MAT***a** *can1 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1 + p2
85	AR-Y292	This work	AR-Y288	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1 + p2
86	AR-Y293 (Referred to interchange-ably with AR-Y285 as OR-Y24)	This work	AR-Y292	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1 + p1-FullDelPol-U-mK-I*(TAA) [*tp-dnap1::URA3/mKate2/leu2*(*538C>T*)*]* + p2
87	AR-Y285 (Referred to interchange-ably with AR-Y293 as OR-Y24)	This work	AR-Y258	*MAT***a** *can1 his4-519 leu2Δ0 ura3Δ0* ρ^0 + p1 + p1-FullDelPol-U-mK-I*(TAA) [*tp-dnap1::URA3/mKate2/leu2*(*538C>T*)*]* + p2
88	AR-Y391	This work	AH22	*MAT***a** *can1 leu2-3*, *112 HIS4*

89	AR-Y383	This work	AR-Y292	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4* + p1 + p2
90	AR-Y436	This work	AR-Y383	*MAT***a** *can1 his3 leu2Δ0 trp1 HIS4 URA3* + p1 + p2
91	AR-Y443	This work	AR-Y436	*MAT***a** *can1 his3 leu2Δ0 trp1 HIS4 URA3* + p1 + p1-FulldelPol-mK-W [*tp-dnap1::TRP1/mKate2*] + p2
92	AR-Y302	This work	AR-Y292	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1 + p1-FulldelPol-mK-U-I*(TAG) [*tp-dnap1::URA3/mKate2/leu2*(*538C>T, 540A>G*)] + p2
93	AR-Y401	This work	AR-Y292	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 HIS4* ρ^0 + p1 + p1-FulldelPol-L-u(K93R) [*tp-dnap1::LEU2 /ura3*(278A>G)] + p2
94	AR-Y402	This work	AR-Y391	*MAT***a** *can1 his3 leu2-3*, *112 HIS4*
95	AR-Y404	This work	AR-Y402	*MAT***a** *can1 his3 leu2Δ0 HIS4*
96	AR-Y408	This work	AR-Y404	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 HIS4*
97	AR-Y412	This work	AR-Y408	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4*
98	AR-Y416	This work	AR-Y412	*MAT***a** *his3 leu2Δ0 ura3Δ0 trp1 HIS4 can1::URA3*

99	AR-Y424	This work	AR-Y416	*MAT***a** *his3 leu2Δ0 ura3Δ0 trp1 HIS4 CAN1*
100	AR-Y428	This work	AR-Y424	*MAT***a** *his3 leu2Δ0 ura3Δ0 trp1 HIS4 CAN1* + pAR-Ec596 [*URA3/POL3*]
101	AR-Y432	This work	AR-Y428	*MAT***a** *his3 leu2Δ0 ura3Δ0 trp1 HIS4 CAN1 pol3::TRP1* + pAR-Ec596 [*URA3/POL3*]
102	AR-Y445	This work	AR-Y432	*MAT***a** *his3 leu2Δ0 ura3Δ0 trp1 HIS4 CAN1 pol3::TRP1 msh6::KanMX* + pAR-Ec596 [*URA3/POL3*]
103	GA-Y102	This work	AR-Y383	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0 + pGA-Ec51 [*URA3/DFR1*] + p1 + p2
104	GA-Y109	This work	GA-Y102	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0 + pGA-Ec51 [*URA3/DFR1*] + p1 + p1-FulldelPol-W-mK-*Pf*DHFR-I*(TAA) [*tp-dnap1::TRP1/mKate2/PfDHFR/ leu2*(*538C>T*)] + p2
105	GA-Y149	This work	GA-Y109	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0 + pAR-Ec318 [*HIS3/TP-DNAP1]* + p1-FulldelPol-W-mK-*Pf*DHFR-I*(TAA) [*tp-dnap1::TRP1/mKate2/PfDHFR/ leu2*(*538C>T*)] + p2
106	GA-Y151	This work	GA-Y109	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0 + pAR-Ec519 [*HIS3/TP-DNAP1 (N423E)]* + p1-FulldelPol-W-mK-*Pf*DHFR-I*(TAA) [*tp-dnap1::TRP1/mKate2/PfDHFR/ leu2*(*538C>T*)] + p2
107	GA-Y155	This work	GA-Y109	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0 + pAR-Ec611 [*HIS3/TP-DNAP1 (I777K, L900S)]* + p1-FulldelPol-W-mK-*Pf*DHFR-I*(TAA) [*tp-dnap1::TRP1/mKate2/PfDHFR/ leu2*(*538C>T*)] + p2
108	GA-Y077	This work	AR-Y292	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 HIS4 dfr1::KanMX* ρ^0 + pGA-Ec52 [*URA3/PfDHFR*] + p1 + p2

109	GA-Y229 (Referred to as OR-Y8)	This work	GA-Y109	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0 + pAR-Ec633 [*HIS3/TP-DNAP1 (L477V, L640Y, I777K, W814N)]* + p1-FulldelPol-W-mK-*Pf*DHFR-I*(TAA) [*tp-dnap1::TRP1/mKate2/PfDHFR/ leu2*(*538C>T*)] + p2		
110	YH5	Gift from C. Sibley	See Wooden et al., 1997	*MAT***a** *ura3-52 leu2-3, 112 trp1 tup dfr1::URA3*		
111	AR-Y146	Previous work	See Ravikumar et al., 2014	*MAT***a** *can1 his4-519 leu2-3*, *112* ρ^0 + p1-ShortPol-L-m [*TP-DNAP1*::*LEU2/mKate2*] + p2		
112	AR-Y062	This work	AR-Y146	*MAT***a** *can1 his4-519 leu2-3*, *112 ura3::KanMX-MET3(-1 to -593)-TP-DNAP1(D641A)* ρ^0 + p1-ShortPol-L-m [*TP-DNAP1*::*LEU2/mKate2*] + p2		
113	AR-Y463	This work	GA-Y102	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0 + pGA-Ec51 [*URA3/DFR1*] + p1 + p1-FulldelPol-W-mK-*Pf*DHFR(S108TCA)-I*(TAA) [*tp-dnap1::TRP1/mKate2/PfDHFR/ leu2*(*538C>T*)] + p2		
114	AR-Y470	This work	AR-Y463	*MAT***a** *can1 his3 leu2Δ0 ura3Δ0 trp1 HIS4 dfr1::KanMX* ρ^0 + pAR-Ec633 [*HIS3/TP-DNAP1 (L477V, L640Y, I777K, W814N)]* + p1-FulldelPol-W-mK-*Pf*DHFR(S108TCA)-I*(TAA) [*tp-dnap1::TRP1/mKate2/PfDHFR/ leu2*(*538C>T*)] + p2		
115	AR-Ec558	Gift from R. Jajoo (pRJ167)	N/A	*2μ, ColE1*	*TRP1, AmpR*	*RNR1 promoter > RNR1*
116	EJ-Ec1	This work	AR-Ec558	*2μ, ColE1*	*TRP1, AmpR*	*RNR1 promoter > RNR1 (Y285A)*
117	EJ-Ec2	This work	AR-Ec558	*2μ, ColE1*	*TRP1, AmpR*	*RNR1 promoter > RNR1 (Y285F)*