Title
CRISPR-guided DNA polymerases for targeted genetic diversification

Author
Halperin, Schaked

Publication Date
2018

Peer reviewed|Thesis/dissertation

CRISPR-guided DNA polymerases for targeted genetic diversification

By

Schaked Omer Halperin

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Joint Doctor of Philosophy
with the University of California, San Francisco

in

Bioengineering

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor David Schaffer, Co-Chair
Professor John Dueber, Co-Chair
Professor David Savage
Professor Wendell Lim

Fall, 2018

Abstract

CRISPR-guided DNA polymerases for targeted genetic diversification

by Schaked Omer Halperin

Joint Doctor of Philosophy in Bioengineering with the University of California, San Francisco and University of California, Berkeley

Professor David Schaffer, Co-Chair; Professor John Dueber, Co-Chair

The capacity to diversify genetic codes advances our ability to understand and engineer biological systems. A method for continuously diversifying user-defined regions of a genome would enable forward genetic approaches in systems that are not amenable to efficient homology-directed oligonucleotide integration. It would also facilitate the rapid evolution of biotechnologically useful phenotypes through accelerated and parallelized rounds of mutagenesis and selection, as well as cell-lineage tracking through barcode mutagenesis. Programmable nucleases, such as CRISPR/Cas9, have revolutionized our ability to easily disable targeted genetic elements; however, substituting nucleotides in user-defined regions of a genome using programmable nucleases remains inefficient in many contexts due to the need for homology-directed repair (HDR), donor templates, cytotoxic double-stranded breaks, and competition with non-homologous end-joining (NHEJ). Here I will present EvolvR, the first system that can continuously diversify all nucleotides within a tunable window length at user-defined loci without relying on HDR, donor templates, double-stranded breaks, or NHEJ. This is achieved by enzymatically generating mutations using engineered DNA polymerases targeted to loci via CRISPR-guided nickases. I identified nickase and polymerase variants that offer a range of targeted mutation rates that are up to 7,770,000-fold greater than rates seen in wild-type cells, and editing windows with lengths of up to 350 nucleotides. I used EvolvR to identify novel ribosomal mutations that confer resistance to the antibiotic spectinomycin and adapted EvolvR for use in human cells. My results demonstrate that CRISPR-guided DNA polymerases enable multiplexed and continuous diversification of user-defined genomic loci, which will be useful for a broad range of basic and biotechnological applications.

This dissertation is dedicated to Yossi Halperin for inspiring me to question, learn, and build and Milene Halperin for persistently and lovingly promoting passion, ambition, and balance.

Table of Contents

Acknowledgements iii

1. Chapter 1: Introduction 4

1.1. Problem statement 1

1.2. Approach 1

2. Chapter 2: Background 4

2.1. Broader context 4

2.2. Targeted continuous directed evolution 5

2.3. Motivation 5

2.4. Previous diversification techniques 7

3. Chapter 3: Engineering a CRISPR-guided DNA polymerase architecture for targeted diversification 10

3.1. Identify CRISPR-guided architectures that offer targeted mutagenesis 10

3.2. Measure targeted and global mutation rates 13

4. Chapter 4: Optimizing, tuning, and characterizing CRISPR-guided DNA polymerases 17

4.1. Tuning the targeted mutation rate 18

4.1.1. Decreasing nCas9 non-specific DNA affinity 18

4.1.2. Screening novel DNA polymerase mutants 20

4.2. Tuning the editing window length 25

4.2.1. Incorporation of processive DNA polymerases 26

4.2.2. Incorporation of processivity-enhancing domains 27

4.3. Combinatorial mutations 29

4.4. Decreasing off-target mutagenesis 31

4.5. Targeting chromosomal and essential loci 33

4.6. Comparing to previously developed continuous mutagenesis tools 35

4.7. Multiplexing targets 39

5. Chapter 5: CRISPR-guided DNA polymerases in eukaryotic cells 41

5.1. Measure targeted and global mutagenesis in human cells 43

6. Chapter 6: Using CRISPR-guided DNA polymerases for directed evolution 49

6.1. Coupling CRISPR-guided DNA polymerase-mediated mutagenesis to non-selectable genetic screens 50

6.2. Continuous directed evolution enables adaptation to two distinct selective pressures 53

6.3. Identifying novel spectinomycin resistance mutations 55

7. Chapter 7: Further Work 64

7.1. Evolving efficient utilization of D-galacturonic acid in yeast 65
7.2. Utilizing CRISPR-guided DNA polymerases for crop trait engineering 66

8. Conclusions 67

9. Methods 68
9.1. Plasmid Construction 69
9.2. High-throughput sequencing of pTarget sample preparation 69
9.3. High-throughput sequencing data analysis 69
9.4. Fluctuation analysis assay 70
9.5. Calculation of mutation rate and statistics 70
9.6. Fluorescence-activated cell sorting of EvolvR libraries 70
9.7. Continuous evolution of E. coli resistant to both spectinomycin and streptomycin 70
9.8. High-throughput sequencing of spectinomycin resistant E. coli 71
9.9. Oligonucleotide recombination 71
9.10. Characterization of spectinomycin resistance 71
9.11. Transfection of HEK293-BFP cells 71
9.12. Flow cytometry and fluorescence assisted cell sorting of HEK293-BFP cells 72
9.13. Data availability 72

10. Appendix 72
10.1. Computer Code for variant analysis 72
10.2. Oligonucleotides, gRNAs, plasmids, and EvolvR component amino acid sequences used in this study 75

11. Bibliography 94

Acknowledgements

I thank Shana McDevitt at the University of California, Berkeley Vincent J. Coates Genomics Sequencing Laboratory for assistance with high-throughput sequencing, the Arkin laboratory for supplying *E. coli* strain RE1000, W. DeLoache for helping edit our manuscript, and the Innovative Genomics Institute and the National Science Foundation for funding.

1. Chapter 1: Introduction

1.1. Problem statement

Natural biological systems evolve astounding functionality by sampling immense genetic diversity under selective pressures. Forward genetics emulates this process to help us understand naturally evolved biological phenomena and to direct the evolution of biotechnologically useful material by applying an artificial selection pressure or screen to libraries of genetic variants. New forward genetic approaches would be enabled by a targeted mutator capable of continuously diversifying all nucleotides within user-defined regions of a genome. However, previous targeted continuous-diversification techniques are confined to either evolving specific loci within particular cells under stringent culture conditions (Esvelt, Carlson, and Liu 2011; Camps et al. 2003) or mutating particular types of nucleotides in a narrow, user-defined window (Ma et al. 2016; Hess et al. 2016). Conversely, current techniques capable of diversifying all nucleotides within user-defined loci remain discrete and inefficient, owing to their requirement for efficient integration of oligonucleotide libraries at the target site (H. H. Wang et al. 2009; Costantino and Court 2003). Therefore, no method currently exists to continuously diversify all nucleotides within user-defined regions of a genome (Table 1).

1.2. Approach

DNA polymerases have the capacity to create all 12 substitutions, as well as deletions (Troll et al. 2011; de Boer and Ripley 1988). These enzymes vary in processivity (average number of nucleotides incorporated after each binding event), fidelity (misincorporation rate) and substitution bias (nucleotide bias during misincorporation). In particular, nick-translating DNA polymerases are able to initiate synthesis from a single-stranded break in double-stranded DNA while displacing the downstream nucleotides, and their flap endonuclease domain subsequently degrades the displaced nucleotides, leaving a ligatable nick. I hypothesized that recruiting an error-prone, nick-translating DNA polymerase with a nicking variant of Cas9 (Jinek et al. 2012) (nCas9) could offer an ideal targeted mutagenesis tool that is independent of homology-directed repair, and which I term EvolvR (Figure 1). The specificity of the polymerase initiation site created by the nCas9 specifies the start site of the editing window, and the mutagenesis window length, mutation rate and substitution bias are controlled by the processivity, fidelity and misincorporation bias of the polymerase variant, respectively.

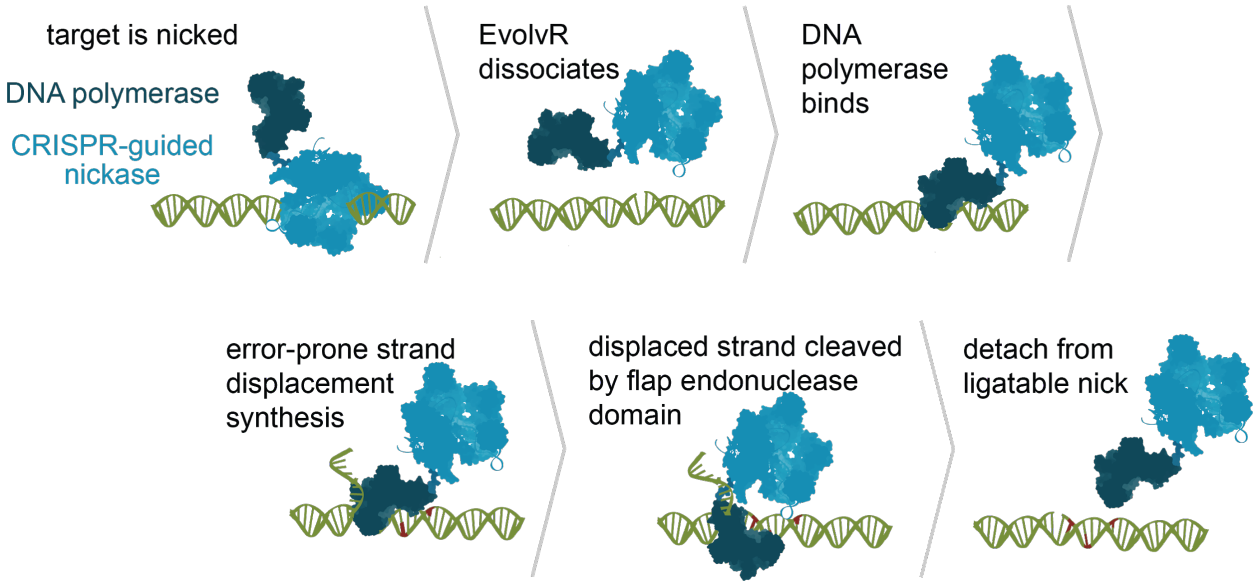

Figure 1 | The EvolvR system consists of a CRISPR-guided nickase that nicks the target locus and a fused DNA polymerase that performs error-prone nick translation.

Table 1 | Comparison of *E. coli* diversification methods

***E. coli* Diversification Method**	**Host/culture requirements**	**Targetability**	**Ease of use**	***In vivo*/ Continuous?**	**When to use**
XL1-Red (Jinek et al. 2012; Greener, Callahan, and Jerpseth 1997)	Specific Strain	None	Transformation	Yes	Continuous whole genome evolution; target is unknown
MP6 (Badran and Liu 2015)	None	None	Transformation	Yes	Continuous whole genome evolution of any strain; target is unknown
Orthogonal polymerase/plasmid (Camps et al. 2003)	Specific strain and culture requirements	Plasmid	Transformation	Yes	Continuous plasmid evolution; target must be located next to the origin of replication of a specific plasmid
PACE (Esvelt, Carlson, and Liu 2011)	Specific strain and culture requirements	Phage genome	Custom turbidostat operation, bacteriophage propogation	Yes	Continuous engineered phage genome evolution; target must be inserted within phage genome, and target activity must be coupled to phage propagation
MAGE (H. H. Wang et al. 2009)	Specific strain	User-defined targets	Recombination machinery induction and high-efficiency transformation of oligonucleotide library	No	Generating rationally designed, discrete, user-defined libraries of recombineering strains
EvolvR (Halperin et al. 2018)	None	User-defined targets	Transformation	Yes	Continuous diversification of user-defined genomic loci in any strain

2. Chapter 2: Background

2.1. Broader context

The efficiency and diversity of functions that have sustained living systems for billions of years make biology an attractive engineering substrate. One approach to programming biology to perform useful functions is rationally modifying or assembling characterized genetic parts. However, such rational design is often ineffective due to our incomplete understanding of how sequence determines function within a specific context. Directed evolution, a complementary approach, relies less on prior knowledge of a system's sequence to function relationship but, instead, isolates the desired functionality from a library of variants using an artificial selection pressure or screen. Because the degree of improvement is dependent upon the number of genetic variants explored, the success of directed evolution is limited by the maximum library size that can be synthesized and transformed, as well as the time, material, and labor cost of each round of evolution. Furthermore, directed evolution has been confined to only a few model organisms with sufficient transformation efficiency to generate large libraries. For example, targeted genome diversification of crop plants that are difficult to transform has relied on an expensive, cumbersome, and imprecise process of chemical or radioactivity-induced global mutagenesis followed by high-throughput sequencing to identify clones that received mutations in the target locus (Tsai et al. 2011). These limitations can *theoretically* be overcome by programming a cell to autonomously generate diversity in the sequence targeted for evolution. Although a few species-specific techniques are able to provide localized mutagenesis at specific engineered loci, no method currently exists for species-independent targeting of unbiased mutagenesis to long stretches of user-defined sequences.

An ideal targeted mutagenesis tool would offer: 1) species-independent functionality, 2) user-defined single base resolution of the mutagenesis window, 3) equal distribution of the 12 types of substitutions, 4) a mutagenesis window length that can be tuned between 1 to 10,000 nucleotides, and 5) a mutation rate that can be tuned - ideally to a rate as high as one mutation per cellular generation.

DNA polymerases have the potential to offer unbiased long-range mutagenesis due to their inherent processivity (ability to synthesize multiple nucleotides from a single binding event) as well as their range of fidelities (misincorporation rates). However, the ability to target synthesis by an error-prone DNA polymerase to a user-defined locus in the genome has yet to be achieved. Nick-translating DNA polymerases are able to initiate synthesis from a single stranded break in double stranded DNA while displacing the downstream nucleotides. A flap endonuclease domain subsequently degrades the displaced nucleotides, leaving a ligatable nick. I hypothesized that recruiting an error-prone nick-translating DNA polymerase with a variant of Cas9 (nCas9) that nicks its target sequence could offer an ideal targeted mutagenesis tool. The specificity of the polymerase initiation site created by the nCas9 offers single-base resolution for the start site of the editing window, while the mutagenesis window length, mutation rate, and substitution bias are controlled by the polymerase processivity, fidelity, and misincorporation bias, respectively. Thus, it should be feasible to optimize each of these characteristics to desired specifications through the choice of polymerase variant and mutagenesis. Because this genetic device would not rely on host machinery for the mutagenesis, I also anticipate it will be readily

adaptable to any transformation-competent organism. I term the proposed system "EvolvR", reflecting how evolution often involves replicative errors.

2.2. Targeted continuous directed evolution

Natural evolution, a mechanism of iterated mutation and selection, has yielded organisms and biomolecules with astoundingly diverse and complex behaviors. Since these naturally evolved behaviors often do not offer a specific desirable function, people have applied an artificial selection such as animal domestication and crop selective breeding for the past several millennia (Driscoll, Macdonald, and O'Brien 2009; Wright et al. 2005). Directed evolution is a more recent laboratory technique to generate organisms and biomolecules with behavior not found in nature through direct genetic manipulation, which has produced many of our medicines, biofuels, biosensors, and industrial enzymes (Tsai et al. 2011; Cobb, Sun, and Zhao 2013).

In general, directed evolution involves 1. Mutating the information carrier (most commonly DNA), 2. translating the evolving molecule (most commonly a protein) from the information carrier, 3. selecting/screening the evolving molecule for a desired behavior, and 4. replicating the information carrier. Traditionally, each of these four steps is performed manually and discretely, which is an extremely time consuming, costly, and labor-intensive process. Continuous directed evolution methods integrate these four steps into a sustained cycle requiring minimal human intervention, thereby allowing many evolution cycles in a single day as well as a typically unattainable exploration of the possible sequence space in search for optimal variants. While increased random global mutations can be generated continuously using mutator strains, researchers typically would like to target the increased mutagenesis to a sequence of interest.

2.3. Motivation

The evolution of living organisms has produced a vast array of molecules with useful behaviors. Unfortunately, evolution cannot efficiently produce molecules with new desirable functions due to the slow and untargeted mechanism of natural mutation.

A laboratory technique called directed evolution traditionally overcomes these limitations by manually discretizing the steps of evolution, beginning with synthetic construction of a library of variants followed by screening or selection of a desired behavior. While directed evolution has led to the discovery of many medicines, catalysts, and other molecular technologies used today, the process is extremely expensive, labor-intensive and slow.

Therefore, several systems have been developed to achieve continuous directed evolution in which integrated rounds of semi-targeted evolution are cycled continuously, significantly increasing the sequence space that can be explored. However, these systems suffer from complex hardware operation, low rates of mutation, a lack of specificity in target sequence, and an inability to target chromosomal DNA.

The proposed investigation assesses the ability to program a cell to increase the rate of mutation of a defined region of DNA using a novel error-prone DNA polymerase/Cas9 system, which I termed EvolvR. If successful, the proposed system will offer the following advantages over previous methods of traditional and continuous directed evolution: 1. screen a greater

library size in a shorter amount of time, 2. maintain all previously generated variants in the growing library, 3. eliminate hardware, 4. easily target multiple loci, 5. easily target chromosomal DNA and 6. designate mutagenesis start site with single-base resolution. In conclusion, this tool may provide a means of accelerating the discovery and production of new medicines, industrial catalysts, and other biomolecular technologies.

Once an EvolvR toolkit offering a range of mutation rates and mutagenesis window lengths has been developed and generalized to organisms of interest, it has the potential to advance many areas of current interest.

Developing new systems for gene editing: While CRISPR-Cas9 has shown promise as a gene-editing therapeutic, off-target activity has raised safety concerns for clinical applications (Schaefer et al. 2017). Higher-fidelity Cas9 mutants may be isolated through continuous diversification of the Cas9 coding sequence in *E. coli* under a dual selection pressure. Specifically, Cas9 mutants with higher fidelity may be enriched by requiring on-target cleavage of a lethal bacteriophage and no cleavage of essential genes targeted with sgRNA's containing mismatches.

Advanced breeding tools: Crop productivity is in part limited by the rate of CO_2 fixation with RuBisCO (Zhu, Long, and Ort 2010). This consideration has motivated decades of rational engineering attempts to improve the RuBisCO's catalytic properties (Whitney, Houtz, and Alonso 2011). Directed evolution of RuBisCO has been attempted by coupling heterologous RuBisCO activity to growth in *E. coli* but mostly selected for variants with improved solubility rather than catalysis. I hypothesize that EvolvR can be used to simultaneously target all RuBisCO catalytic subunits within cyanobacteria, whose growth under high temperature and low CO_2 conditions will enrich for enzymatic variants with improved catalytic efficiency and specificity. Since previous work has shown that plant's native RuBisCO enzyme can be replaced by homologs derived from cyanobacteria and archaea (Lin et al. 2014; Wilson, Alonso, and Whitney 2016), introducing these improved variants into crops is a potential route to breed crops with higher yield and temperature resistance.

Stress response and disease resistance in developing world crops and woody species: Global mutagenesis of plant seeds through x-ray radiation followed by improved trait screening and selective backcrossing revolutionized food production in the mid-20th century. But, in the 21st century, we are equipped with a better understanding of the genetic determinants of crop traits. For instance, R genes are involved in plant innate immune response, and modifications to a plant's R genes have conferred disease resistance (Tai et al. 1999). Targeting EvolvR to various R loci or stress-responsive promoters in the germline of crops would offer a more targeted random-mutagenesis approach to screening for disease and stress resistant crops.

Discovering new pathways and systems: Actinomycetes produce many known, pharmaceutically useful, antimicrobial metabolites. Because many biosynthetic gene clusters are not expressed under laboratory conditions, natural product discovery has often resorted to integration of heterologous constitutive promoters (Rutledge and Challis 2015; Zhang et al. 2017). A useful approach to generate cells with altered transcriptional profiles has been to transform a library of mutated global transcription factors (Alper et al. 2006). However, the difficulty of transforming actinomycetes makes the construction of libraries with altered transcriptional profiles incredibly laborious. A single actinomycete cell transformed with EvolvR targeting endogenous global transcription factors can create a library of cells with diversified transcriptional profiles. Plating this culture on a lawn of multidrug resistant *E. coli* and screening

for a zone of inhibition will identify mutants with altered transcriptional profiles that activate novel antibiotic biosynthesis.

Industrial biotechnology applications: Error-prone PCR libraries of global transcription factors have been used to obtain laboratory strains of microbes that are resistant to various abiotic stressors (Alper et al. 2006). Unfortunately, many industrially relevant strains are difficult to transform and therefore not amenable to this library approach. EvolvR would be able to generate transcription factor diversity in these microbes during continuous growth under a desired selective condition to improve a variety of traits (e.g., increased tolerance to high or low pH, butanol, or temperature).

2.4. Previous diversification techniques

In one instance, an *Escherichia coli* (*E. coli*) ColEI plasmid was able to be selectively mutated with an error-prone mutant of *E. coli* DNA polymerase I (Pol I) (Camps et al. 2003). However, this method suffered from a high level of global mutation due to Pol I's role in chromosomal replication. Another method recruited a mutagenic protein to a telomere region of a yeast chromosome. However, the 20 kb region of increased mutations offers minimal control of sequence diversification (Finney-Manchester and Maheshri 2013). A recent orthogonal plasmid - polymerase system in yeast has allowed continuous evolution of an extranuclear plasmid, but suffers from poor expression and low levels of mutation (Ravikumar, Arrieta, and Liu 2014). Crook *et al.* engineered the native retroelement in yeast to diversify heterologous sequences, but was limited to diversification of a single open reading frame (Crook et al. 2016). Finally, Wang *et al* showed an elevated rate of mutagenesis within heterologous sequences integrated near the antibody regions of B cells (L. Wang et al. 2004). Lastly, phage assisted continuous evolution coupled the production of an essential bacteriophage coat protein to an activity of interest and has allowed a greater than 100-fold reduction in evolution cycle time compared to non-continuous methods (Esvelt, Carlson, and Liu 2011). Unfortunately, the complex turbidostat operation, low rates of mutation, and requirement for phage-associable genes has not allowed this tool to gain widespread usage.

Importantly, all these techniques are highly limited to the particular locus and species for which they were developed. This means that defined regions of DNA, such as a promoter or the coding sequence of a single gene still cannot benefit from the advantages of continuous directed evolution. Furthermore, multiple loci and chromosomal DNA cannot be targeted, which would be a significant tool for metabolic engineering in which an entire pathway of genes could be co-evolved for production of a small molecule of interest. Species-independent targeted mutagenesis would enable library generation in organisms with low transformation efficiency such as *Streptomyces*, human stem-cells, and plants. Additionally, the ability for a user to co-target multiple regions for elevated mutagenesis would enable diversification of endogenous genomic sequences in a manner that allows the leveraging of genetic drift to discover epistatic relationships.

Investigation and engineering of specific regions of a genome would benefit from the ability to efficiently generate libraries of cells with locally diversified genomes. This goal would be enabled by a targeted mutator ideally capable of directly generating all nucleotide substitutions, as well as deletions or insertions, within a defined region of a genome without the

requirement for inserting libraries of nucleic acid sequences at the target locus. The advent of programmable nucleases offers potential solutions to this challenge of targeted genome modification. These systems, such as CRISPR/Cas9, can create double-stranded breaks at user-defined loci. Precise alterations can then be introduced at these break sites by integrating donor constructs via homology-directed repair (HDR). Unfortunately, HDR is extremely inefficient compared to non-homologous end-joining (NHEJ) in many organisms, which results in insertion or deletion of a variable number of nucleotides at the break site rather than the intended modification (Mao et al. 2008a, [b] 2008). Additionally, the double-stranded breaks generated by these nucleases can be toxic, often leading to induced cell death and genomic rearrangements (Aguirre et al. 2016; Choi and Meyerson 2014; Frock et al. 2014). Therefore, while programmable nucleases have been useful for creating NHEJ-mediated insertions and deletions that disrupt genetic elements, the requirement for integrating synthesized oligonucleotides at the site of toxic double-stranded breaks to generate substitutions limits programmable nucleases' utility in genome diversification. Several species-specific, *in vivo* random mutagenesis techniques can provide localized substitutions at particular engineered loci without transforming nucleic acids or generating double-stranded breaks (Ravikumar, Arrieta, and Liu 2014; Esvelt, Carlson, and Liu 2011; L. Wang et al. 2004; Camps et al. 2003). However, no HDR-independent and organism-agnostic method currently exists to introduce all types of mutations at user-defined genomic loci.

The advent of CRISPR systems for specifically targeting enzymatic activities in the genome offers potential solutions to the challenge of mutagenesis of user-defined loci that can be generalized to many organisms. Cas9 is a CRISPR endonuclease that can create a double stranded break in a defined sequence determined by an engineered single guide RNA (sgRNA). Two separate studies recently recruited Activation Induced Deaminase (AID) to specific sequences within the mammalian genome using Cas9 variants (Sohail et al. 2003; Hess et al. 2016; Ma et al. 2016). AID converts cytosine to uracil as part of somatic hypermutation in the immune system, forming an aberrant U:G base-pairing that can lead to mutagenesis during the repair process (Sohail et al. 2003). Unfortunately, most mutations were confined to a narrow 10 base pair window, and in order to target a longer region, they needed 7 sgRNA's per 100 base-pairs. By extrapolation, mutagenizing even a single protein (1,000 nucleotides on average) would require at least 70 sgRNAs. The construction and transfection of this magnitude of sgRNAs is prohibitive, and optimizing an entire pathway would be even more impractical. Additionally, AID-driven mutagenesis only led to functional mutations of cytosine or guanine, highlighting the expected strong mutation bias of this system. Somatic hypermutation overcomes the bias of AID mutagenesis by employing a strong codon bias in the immunoglobin locus of lymphocytes, which would be difficult to reconstitute at user-defined loci (N.-Y. Zheng et al. 2005). For directed evolution of protein coding sequences, mutating a codon's cytosine and guanine can lead to an average of six out of the nineteen missense mutations (Figure 2). This tool's short mutagenesis window length and strong substitution bias indicate that state-of-the-art CRISPR-guided mutagenesis is currently limited to specific single-nucleotide-polymorphism engineering within very narrow DNA regions, precluding its usefulness in directed evolution.

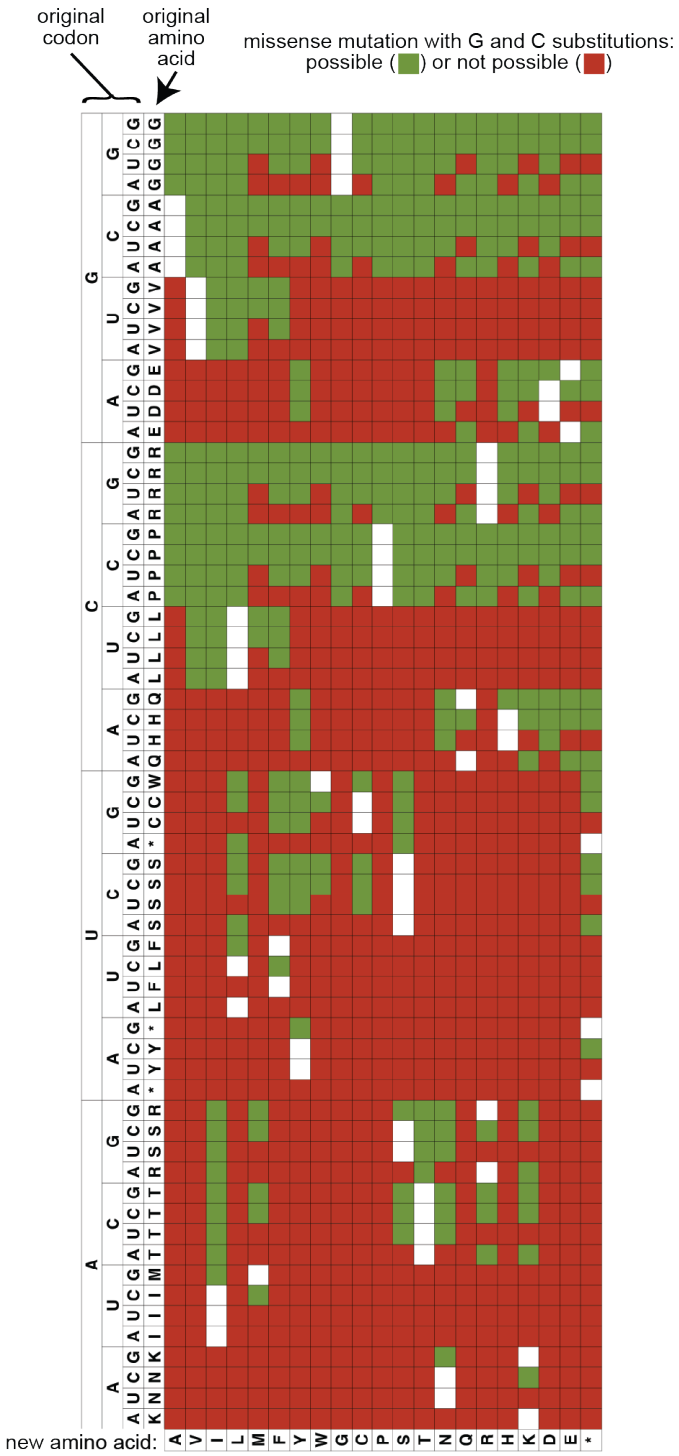

Figure 2 | Previous tools enabling diversification of user-defined loci by substituting cytosines and guanines limit the protein coding space that can be explored. This chart shows which amino acids can (green) and cannot (red) be reached by mutating cytosines and guanines to any other base for each of the 64 codons, highlighting that only 32% of missense mutations are achievable with targeted cytidine deaminases. The white area depicts the original amino acid identity.

3. Chapter 3: Engineering a CRISPR-guided DNA polymerase architecture for targeted diversification

3.1. Identify CRISPR-guided architectures that offer targeted mutagenesis

In the initial design, nCas9 (*Streptococcus pyogenes* Cas9 containing a D10A mutation) was fused to the N terminus of a fidelity-reduced variant of *Escherichia coli* DNA polymerase I (PolI) harboring the mutations D424A, I709N and A759R (PolI3M) (Camps et al. 2003). A plasmid (pEvolvR) expressing the nCas9–PolI3M and a guide RNA (gRNA) in *E. coli* was tested for its ability to mutate a second plasmid targeted by the gRNA over 24 h of propagation. High-throughput targeted amplicon sequencing revealed that the target plasmid accrued substitutions in an approximately 17-nucleotide window 3′ of the nick site (Figure 3), consistent with the established 15–20 nucleotide processivity of PolI (Bambara and Choi 1973). Although the sequencing results are probably under-sampling the total diversity generated, I observed substitutions of all four nucleotide types (Figure 4). The presence of low-frequency substitutions 5′ of the nick site may be due to endogenous 3′-to-5′ exonucleases removing a few nucleotides 5′ of the nick prior to the polymerase initiating synthesis. Controls expressing an unfused nCas9 and PolI3M with the on-target guide only yielded one low-frequency substitution, whereas expressing nCas9–PolI3M with an off-target guide, as well as expressing nCas9 alone, did not yield any substitutions at a frequency above the detection threshold.

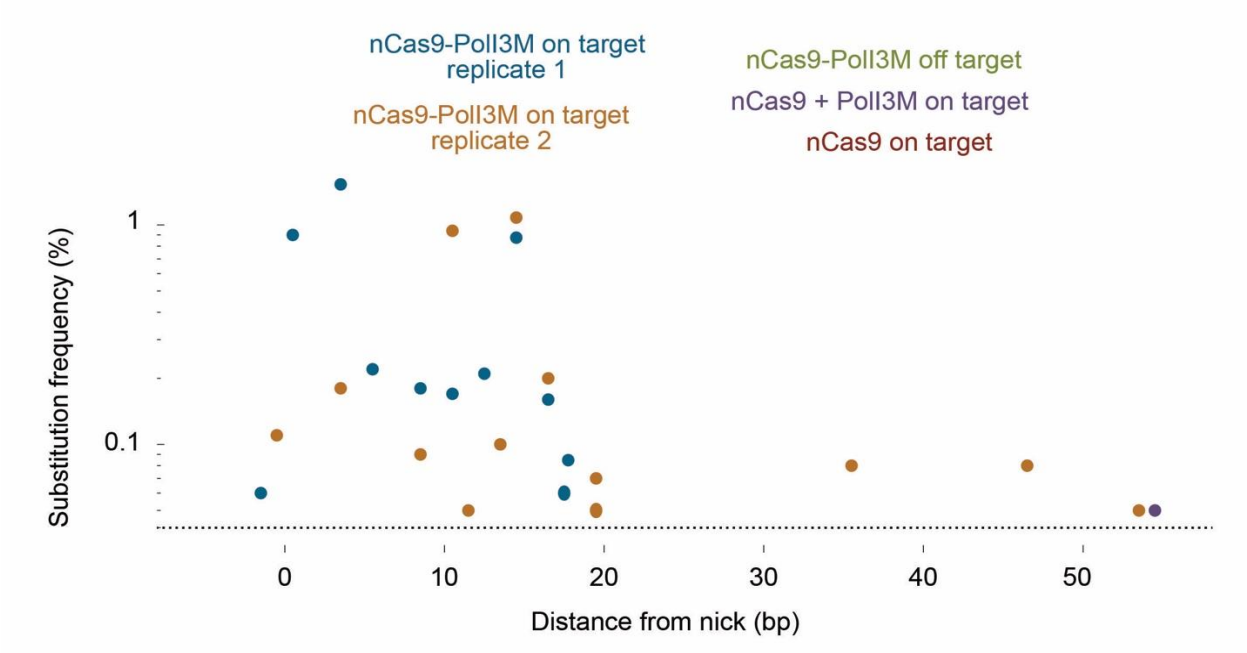

Figure 3 | High-throughput sequencing shows that fusing nCas9 to PolI3M resulted in substitutions across an approximately 17-nucleotide window 3′ from the nick. Expressing nCas9–PolI3M with an off-target guide did not show substitutions at a frequency above our detection threshold (dotted line, see Methods section 'High-throughput sequencing data analysis'), whereas an unfused nCas9 and PolI3M yielded only one substitution and at low-frequency.

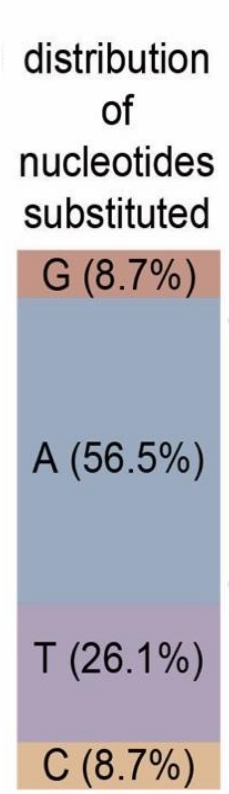

Figure 4 | Distribution of the substituted nucleotides; all four nucleotides were substituted by nCas9–PolI3M.

3.2. Measure targeted and global mutation rates

To sensitively quantify the mutation rate and mutagenesis window length of EvolvR variants, I designed a fluctuation analysis (Sarkar, Ma, and Sandri 1992). For this assay, the pEvolvR plasmid was co-transformed into *E. coli* with a plasmid (pTarget) containing the *aadA* spectinomycin resistance gene disabled by a nonsense mutation (Figure 5). After 16 hours of growth, the cultures were plated on the antibiotic, and the mutation rates were determined from the number of resistant colony forming units (CFU). As shown in Figure 6, fluctuation analysis estimated the mutation rate of wild-type *E. coli* to be approximately 10^{-10} mutations per nucleotide per generation, similar to a previously reported 5.4×10^{-10} mutations per nucleotide per generation (Drake 1991). The global mutation rate (the mutation rate of the untargeted genome in cells expressing EvolvR) was determined by measuring the antibiotic-resistance reversion rate of cells carrying a gRNA targeting *dbpA*, a fitness-neutral RNA helicase gene in the *E. coli* genome (Jagessar and Jain 2010).

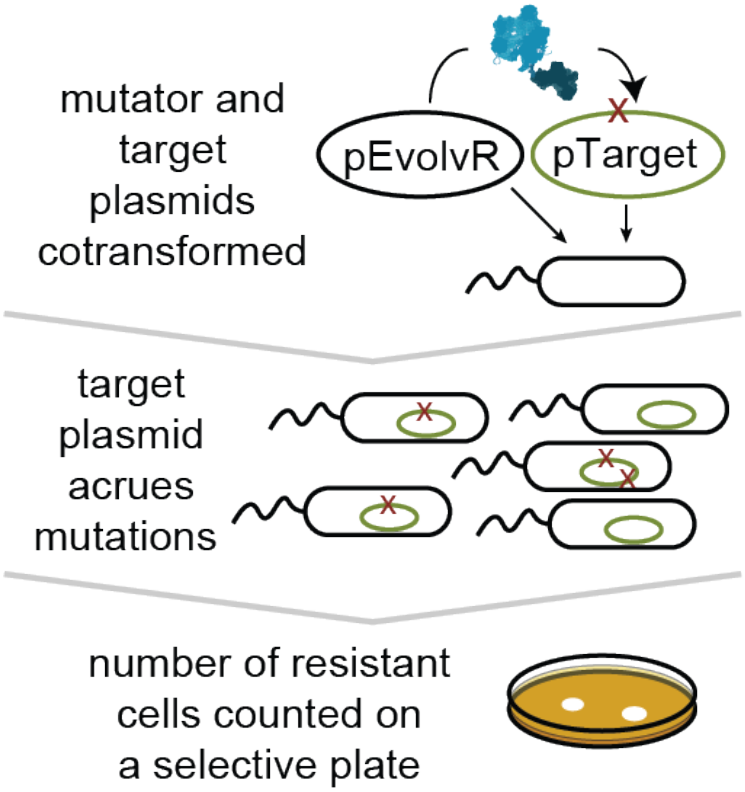

Figure 5 | Schematic of the fluctuation analysis workflow used to sensitively quantify targeted and global mutation rates. The pEvolvR plasmid was cotransformed into *E. coli* with a plasmid (pTarget) containing the *aadA* spectinomycin resistance gene disabled by a nonsense mutation. After 16 hours of growth, the cultures were plated on the antibiotic, and the mutation rates were determined from the number of resistant colony forming units (CFU).

Expressing nCas9–PolI3M markedly increased the mutation rate at the targeted locus 24,500-fold over the wild type while increasing the global mutation rate 120-fold over the wild type (Figure 6), a global mutation rate comparable to that of previous targeted mutagenesis techniques in *E. coli (Camps et al. 2003; H. H. Wang et al. 2009)*. By comparison, expressing nCas9 and PolI3M as separate proteins, PolI3M alone, nCas9 alone or a catalytically inactive Cas9 (dCas9) fused to PolI3M, showed significantly lower targeted mutation rates ($P < 0.0001$; two-sided student's *t*-test). These results suggest that both PolI3M and the nick created by nCas9 are essential for EvolvR-mediated mutagenesis. Expressing nCas9 and PolI3M as separate proteins or PolI3M alone showed a 98-fold or 554-fold increase in global mutation rates compared to wild-type *E. coli*, respectively.

Finally, by replacing the D10A nCas9, which nicks the strand complementary to the gRNA, with the H840A nickase, which nicks the strand non-complementary to the gRNA, I found that the direction of EvolvR-mediated mutagenesis relative to the target site of the gRNA is dependent on which strand is nicked (Figure 7).

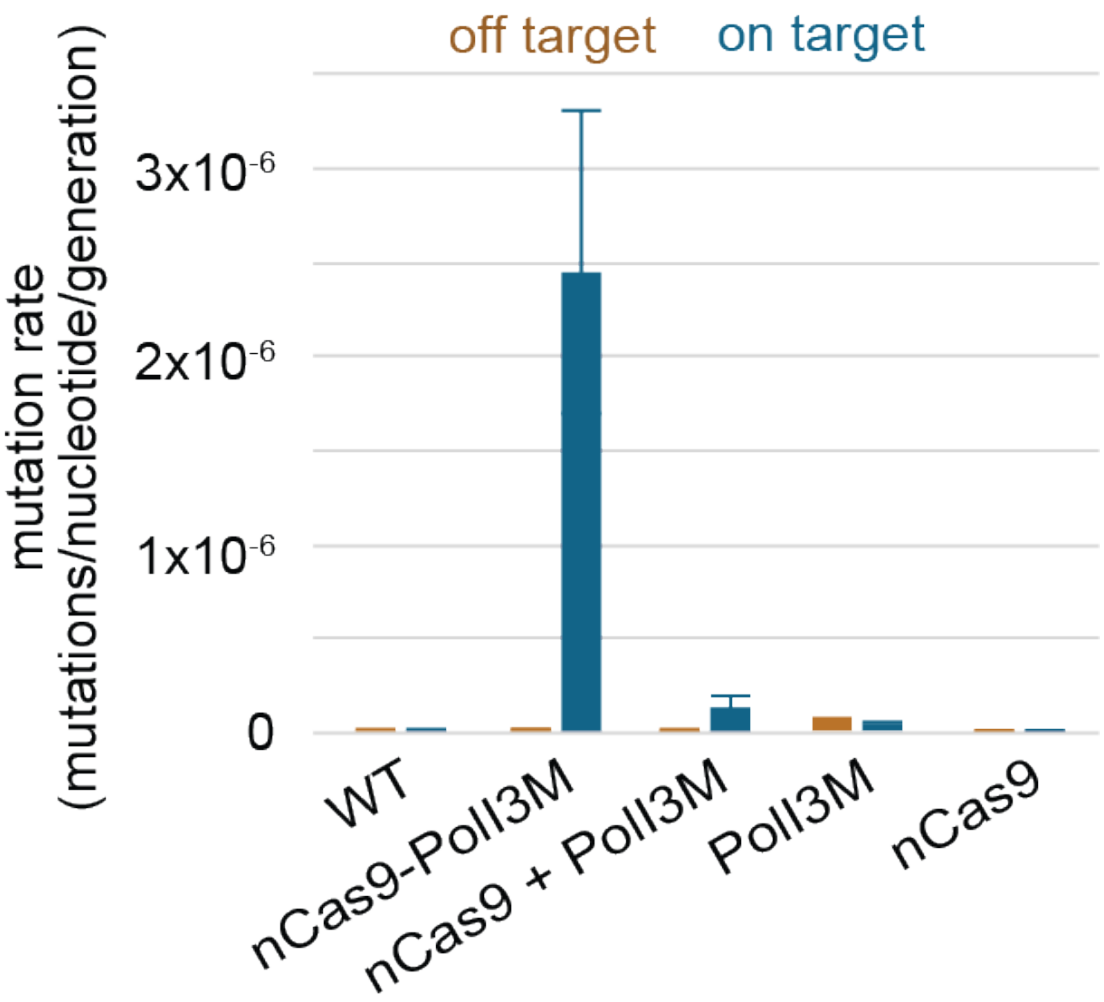

Figure 6 | The global and targeted mutation rates of wild-type (WT) *E. coli*, nCas9–PolI3M, unfused nCas9 and PolI3M, PolI3M alone, and nCas9 alone were determined by fluctuation analysis. For all figures, 'on target' mutation rates were determined by expressing a gRNA that nicks 11 nucleotides 5' of the nonsense mutation unless labelled otherwise, whereas the 'off target' mutation rates were determined by expressing a gRNA targeting *dbpA*, a fitness-neutral RNA helicase gene in the *E. coli* genome. Data are the mean of ten biologically independent samples and the error bars indicate 95% confidence intervals. Mutation rates throughout are mutations per nucleotide per generation.

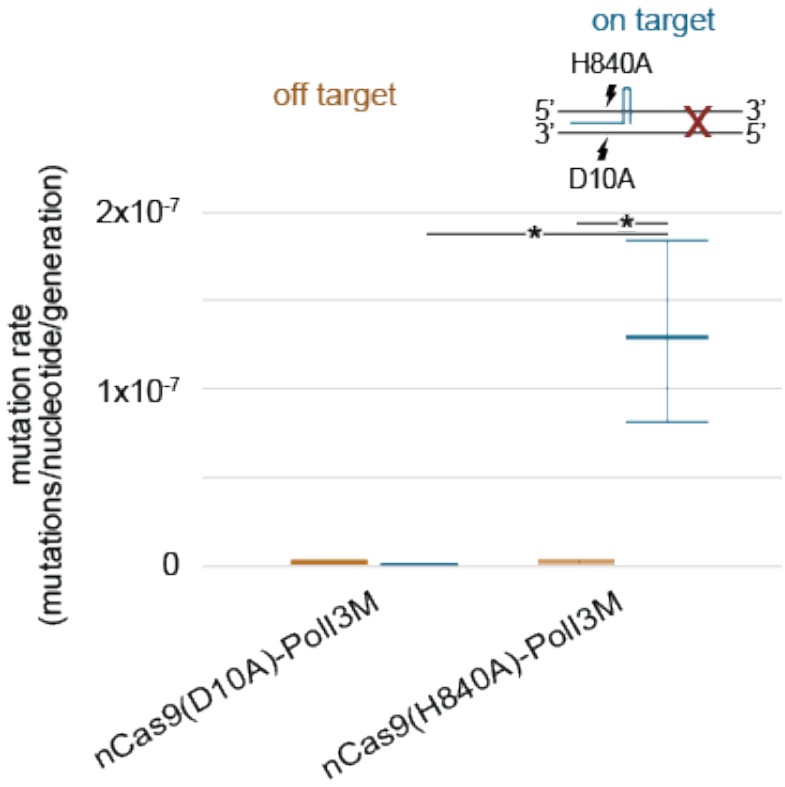

Figure 7 | The previous fluctuation analysis in Figure 6 demonstrated that nCas9(D10A)–PolI3M mutates a window 3′ of the nick site. Here I directly tested whether mutations are generated 5′ of the nick site using a different gRNA. Because DNA polymerases synthesize in the 5′-to-3′ direction, I anticipated that nCas9(D10A)–PolI3M would not provide an elevated mutation rate 5′ of the nick site. I indeed found that expressing a guide RNA which targeted nCas9(D10A)–PolI3M to nick 16 nucleotides 3′ from the nonsense mutation (indicated by a red cross) did not show targeted mutagenesis. I hypothesized that I could induce targeted mutagenesis using the same gRNA by using a Cas9 variant harboring the H840A mutation, which nicks the DNA strand non-complementary to the gRNA, rather than the D10A mutation, which nicks the strand complementary to the gRNA. nCas9(H840A)–PolI3M increased the mutation rate 16 nucleotides 3′ from the nick by 52-fold compared to the global mutation rate of cells expressing an off-target gRNA. I used the D10A nCas9 variant for all subsequent experiments. Data are mean ± 95% confidence intervals from ten biologically independent samples. *$P < 0.0001$; two-sided Student's *t*-test.

4. Chapter 4: Optimizing, tuning, and characterizing CRISPR-guided DNA polymerases

4.1. Tuning the targeted mutation rate

Ideally, the mutation rate at target loci can be tuned depending on a user's needs. Tuning this mutation rate can be achieved by modifying the nuclease and DNA polymerase components of EvolvR.

4.1.1. Decreasing nCas9 non-specific DNA affinity

I hypothesized that the targeted mutation rate could be further increased by promoting the dissociation of nCas9 from DNA after nicking the target locus. Therefore, three mutations (K848A, K1003A, R1060A) that have previously been suggested to lower the non-specific DNA affinity of Cas9 (Slaymaker et al. 2015) were introduced into the fused nCas9. The resulting enhanced nCas9 (enCas9) fused to PolI3M increased the global mutation rate 223-fold compared to wild-type cells (1.9–fold greater than nCas9–PolI3M), yet elevated the mutation rate at the targeted locus by 212,000-fold (8.7-fold greater than nCas9–PolI3M) (Figure 8).

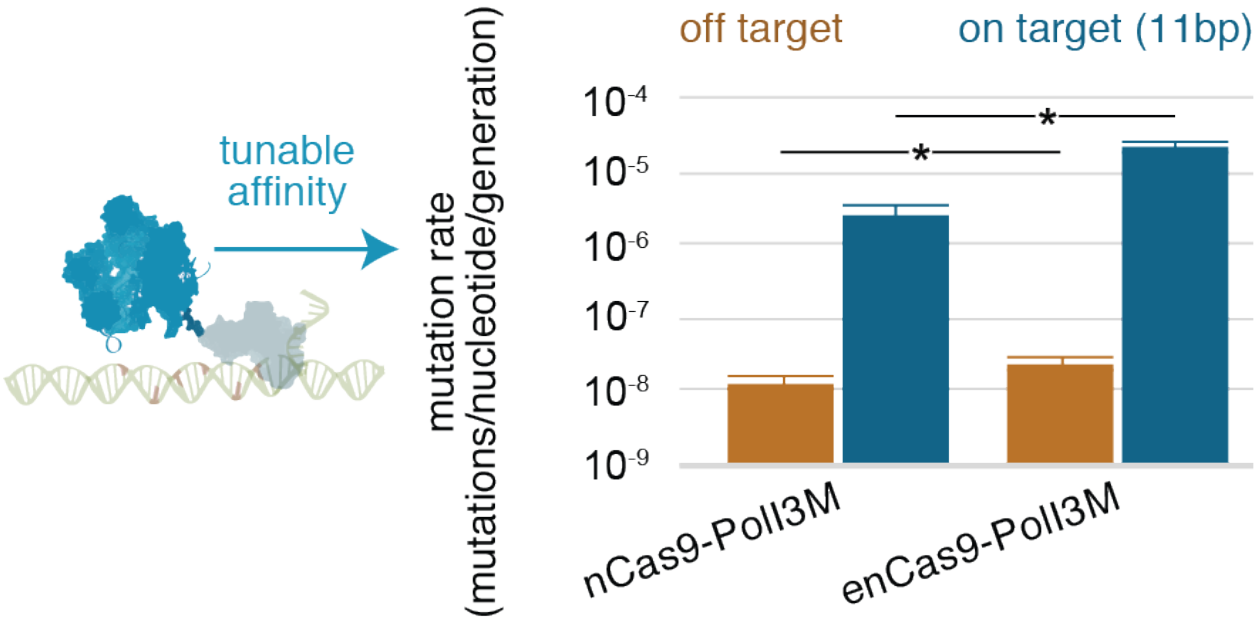

Figure 8 | Introducing mutations suggested to lower non-specific DNA affinity into the fused nCas9 (producing enCas9) (Slaymaker et al. 2015) increased the global mutation rate 223-fold compared to the wild-type mutation rate (enCas9–PolI3M 1.9-fold greater than nCas9–PolI3M), and increased the targeted mutation rate by 212,038-fold over the wild type (enCas9–PolI3M 8.7-fold greater than nCas9–PolI3M).

4.1.2. Screening novel DNA polymerase mutants

PolI3M was initially chosen because it was the most error-prone variant of PolI previously characterized. However, the modularity of EvolvR enables tuning of the mutation rate by using polymerases with different fidelities. First, I confirmed that the fidelity of the polymerase determines mutation rates by comparing enCas9 fused to PolI variants, in decreasing order of fidelity: PolI1M (D424A), PolI2M (D424A, I709N), and PolI3M (D424A, I709N, A759R) (Figure 9).

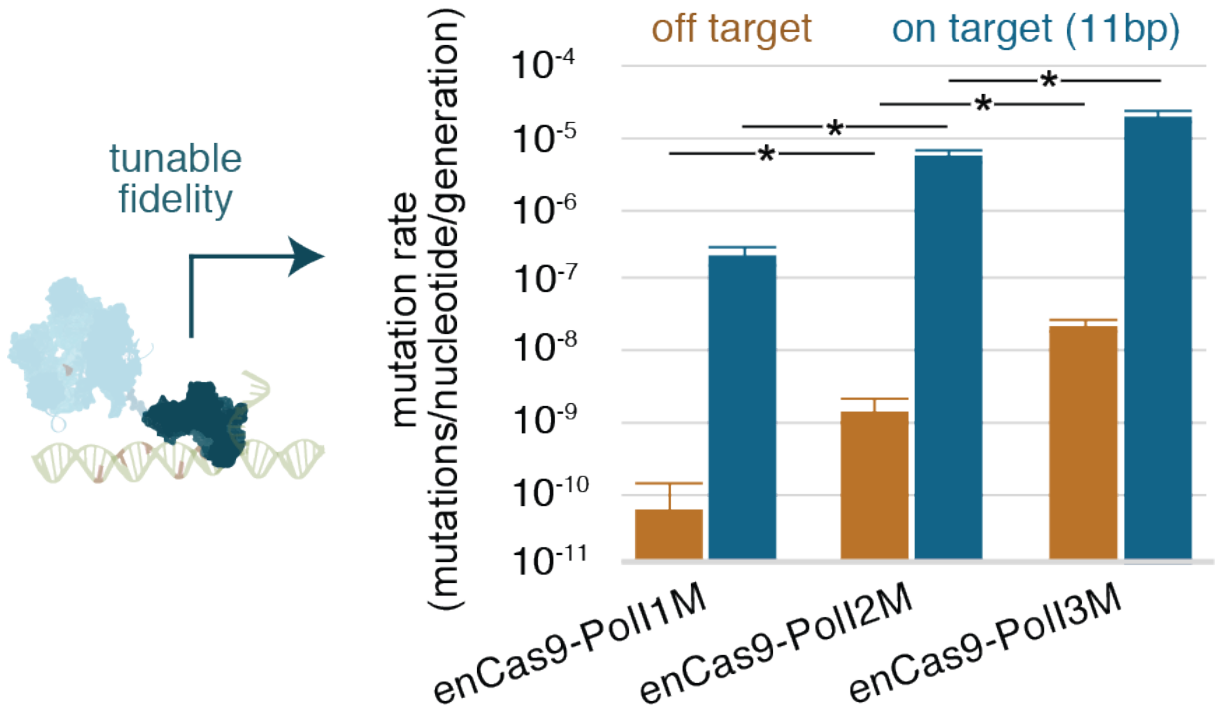

Figure 9 | Mutagenesis rates were dependent on the fidelity of the polymerase. PolI with a D424A mutation (PolI1M) was less mutagenic than PolI with both D424A and I709N mutations (PolI2M), and PolI3M (D424A, I709N, A759R) was the most mutagenic.

Next, to further increase the targeted mutation rate of EvolvR, I screened several additional mutations previously shown to individually decrease wild-type PolI fidelity (Camps et al. 2003; Minnick et al. 1999; Loh, Salk, and Loeb 2010) (Figure 10). Although several of the additional mutations yielded low-activity variants, PolI3M with the additional mutations F742Y and P796H (PolI5M) displayed a mutation rate one nucleotide from the nick that was 7,770,000-fold higher than wild-type cells, and 33-fold higher than PolI3M, making it the most error-prone PolI mutant ever reported.

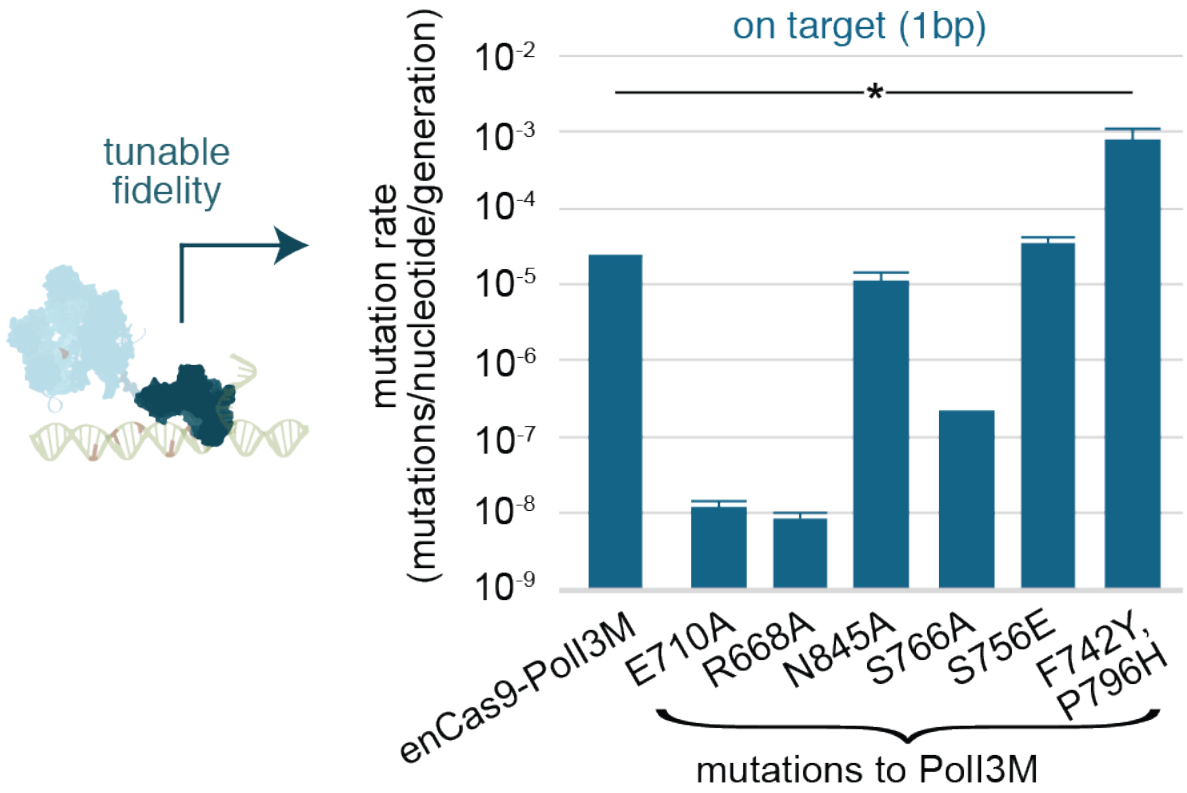

Figure 10 | Screening mutations in PolI3M previously shown to decrease wild-type PolI fidelity revealed that PolI3M with additional mutations F742Y and P796H (PolI5M) had a mutation rate 7,770,000-fold greater than wild-type cells one nucleotide from the nick.

Notably, enCas9–PolI5M did not exhibit either a higher global rate of mutagenesis than enCas9–PolI3M or higher mutation rates than enCas9–PolI3M 11 nucleotides from the nick (Figure 11).

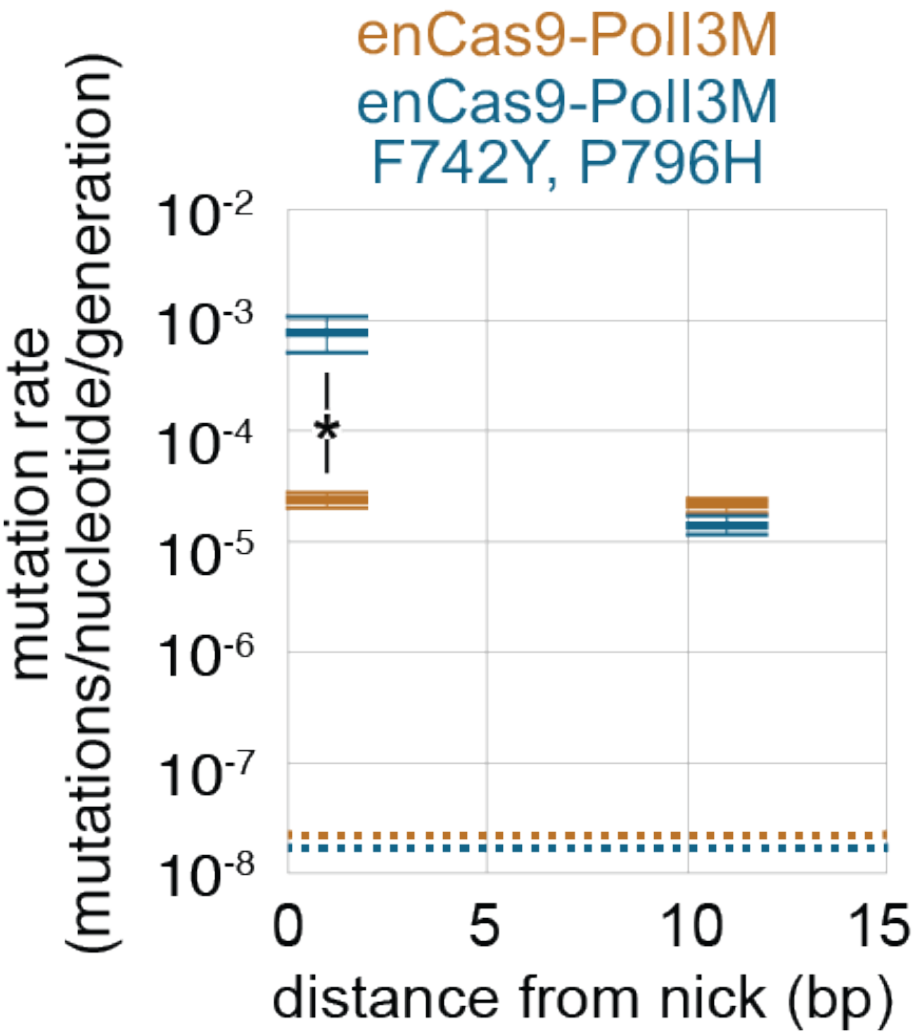

Figure 11 | PolI3M with additional F742Y and P796H mutations (PolI5M) elevates the mutation rate 33-fold 1 nucleotide from the nick compared to PolI3M. PolI5M did not have a higher mutation rate than PolI3M 11 nucleotides from the nick. Data are mean ± 95% confidence intervals from ten biologically independent samples. *$P < 0.0001$; two-sided Student's *t*-test.

4.2. Tuning the editing window length

Ideally, the editing window length can be tuned depending on a user's needs. Tuning this editing window length can be achieved by modifying the DNA polymerase component of EvolvR.

4.2.1. Incorporation of processive DNA polymerases

A more processive DNA polymerase could potentially increase the length of the editing window, so PolI5M was exchanged for the more processive bacteriophage Phi29 DNA polymerase (Phi29). Expression of Phi29 variants with previously reported fidelity-reducing and thermostabilizing mutations in combination with the Phi29 single-stranded binding protein showed targeted mutagenesis 56 and 347 nucleotides from the nick site (Figure 12). However, the mutation rate at these distances was not nearly as high as that achieved with PolI3M at shorter distances.

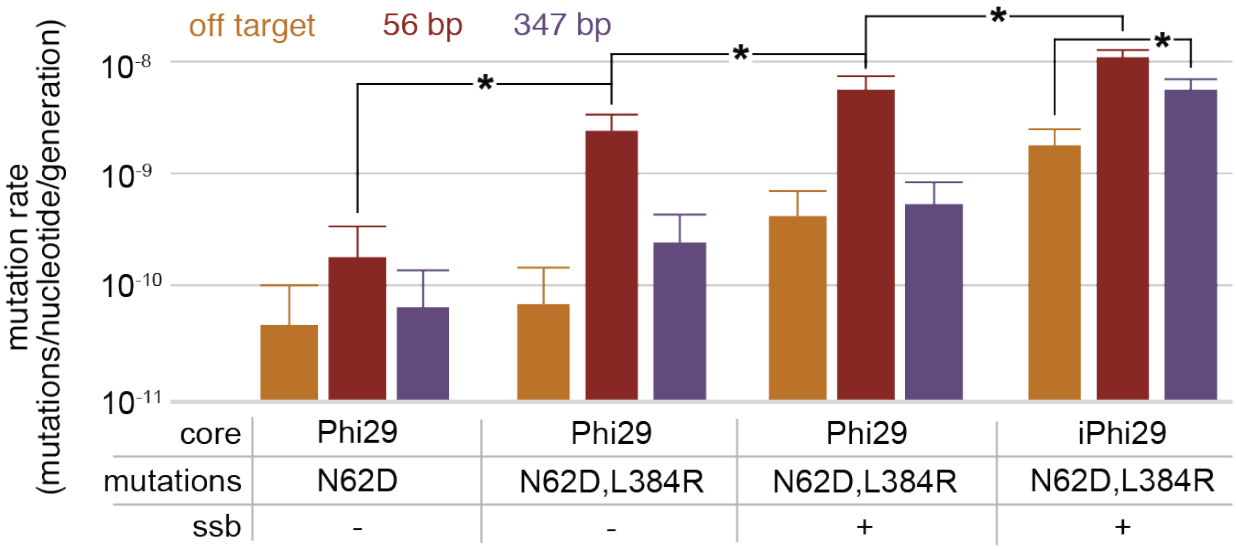

Figure 12 | PolI was exchanged for a more processive and higher-fidelity bacteriophage Phi29 DNA polymerase (Phi29). Owing to Phi29 not having a flap endonuclease, residues 1–325 of PolI were inserted between enCas9 and Phi29. Using gRNAs targeting different distances from the nonsense mutation, I found that Phi29 with two previously reported fidelity-reducing mutations (N62D and L384R) elevated the mutation rate 56 nucleotides from the nick compared to the global mutation rate (Truniger et al. 2003; de Vega et al. 1996). When I expressed Phi29's single-stranded binding protein (ssb), which is known to improve the activity of Phi29, I observed an elevation in the targeted mutation rate (Ducani, Bernardinelli, and Högberg 2014). Finally, because the activity of Phi29 is known to decrease at temperatures above 30 °C and the fluctuation analysis was performed at 37 °C, I added mutations previously reported to improve the thermostability of Phi29 (iPhi29) and observed a targeted mutation rate 347 nucleotides from the nick site that was significantly greater than the global mutation rate (Povilaitis et al. 2016). Unfortunately, mutations decreasing Phi29's fidelity are known to decrease its processivity explaining our inability to identify Phi29 variants that retain high processivity while offering as high of a mutation rate as PolI3M (Truniger et al. 2003). Data are mean ± 95% confidence intervals from ten biologically independent samples. *$P < 0.0001$; two-sided Student's *t*-test.

4.2.2. Incorporation of processivity-enhancing domains

An alternative method to increase the length of the editing window and retain high mutation rates would be to increase the processivity of PolI. Previous work has shown that inserting the thioredoxin-binding domain (TBD) of bacteriophage T7 DNA polymerase into the thumb domain of PolI increases the processivity of the polymerase in the presence of thioredoxin from *E. coli* (Y. Wang et al. 2004). Figure 13 shows that whereas the original enCas9–PolI3M did not show a difference between global and targeted mutation rates 56 nucleotides from the nick, incorporation of the TBD into the PolI3M EvolvR gene (enCas9–PolI3M–TBD) produced a 555-fold increase over the global mutation rate at this range.

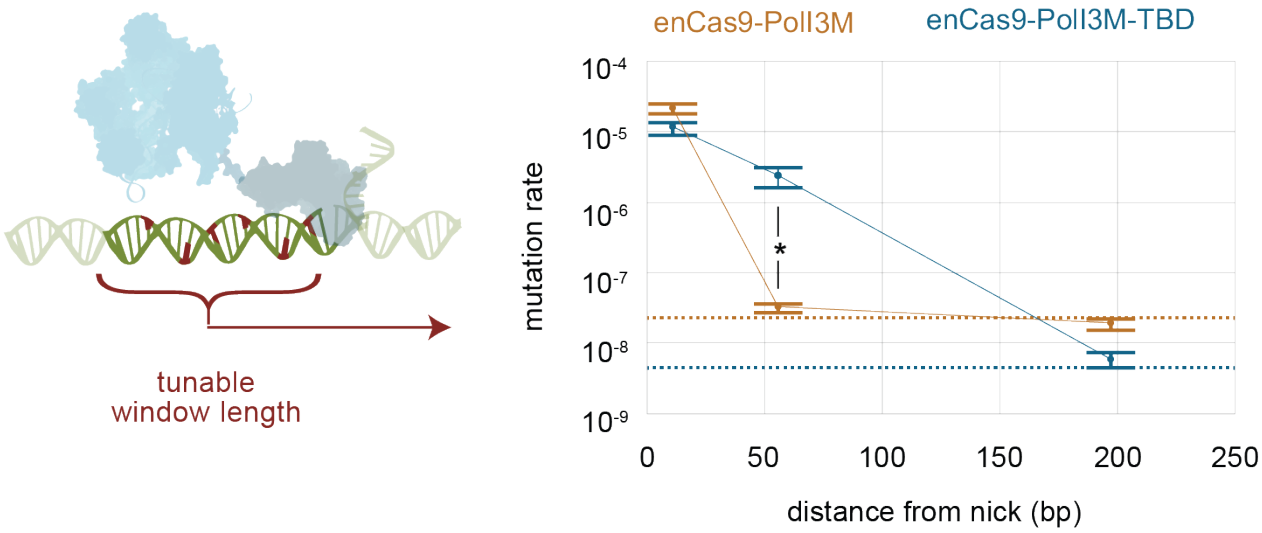

Figure 13 | The editing-window length was increased by incorporating TBD into PolI3M. enCas9–PolI3M–TBD provided a targeted mutation rate 56 nucleotides from the nick that was 555-fold above the global mutation rate, whereas enCas9–PolI3M showed no targeted mutagenesis 56 nucleotides from the nick. Mutation rate data are mean±95% confidence intervals from ten biologically independent samples. *P<0.05; two-sided student's t-test.

4.3. Combinatorial mutations

Leveraging this increased editing window length, I targeted enCas9–PolI3M–TBD to a plasmid (pTarget2) containing two nonsense mutations (11 and 37 nucleotides from the nick) in the antibiotic-resistance gene, and thereby showed the ability of EvolvR to generate combinations of multiple mutations with a single gRNA (Figure 14).

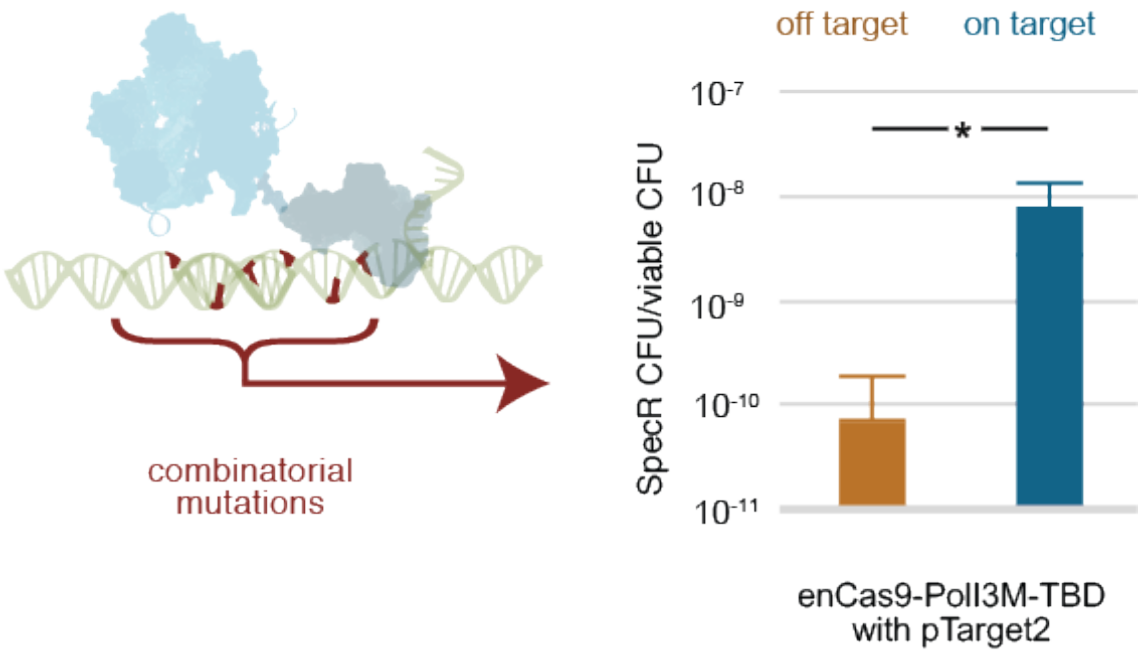

Figure 14 | enCas9–PolI3M–TBD targeted to a plasmid containing two nonsense mutations in the spectinomycin resistance gene (pTarget2) showed that EvolvR is able to generate combinations of multiple mutations. Mutation rate data are mean±95% confidence intervals from ten biologically independent samples. *P<0.05; two-sided student's t-test.

4.4. Decreasing off-target mutagenesis

I hypothesized that unintended translation products consisting of functional DNA polymerase not fused to a functional CRISPR-guided protein contributed to undesirable off-target mutagenesis. Therefore, I codon-optimized the EvolvR coding sequence (enCas9–PolI3M–TBD-CO) to remove three strong internal ribosome binding sites identified using the RBS Calculator (Salis 2011). I found that the off-target mutation rate decreased 4.14-fold when expressing enCas9–PolI3M–TBD-CO compared to enCas9–PolI3M–TBD while the on-target mutation rate only decreased 1.23-fold (Figure 15).

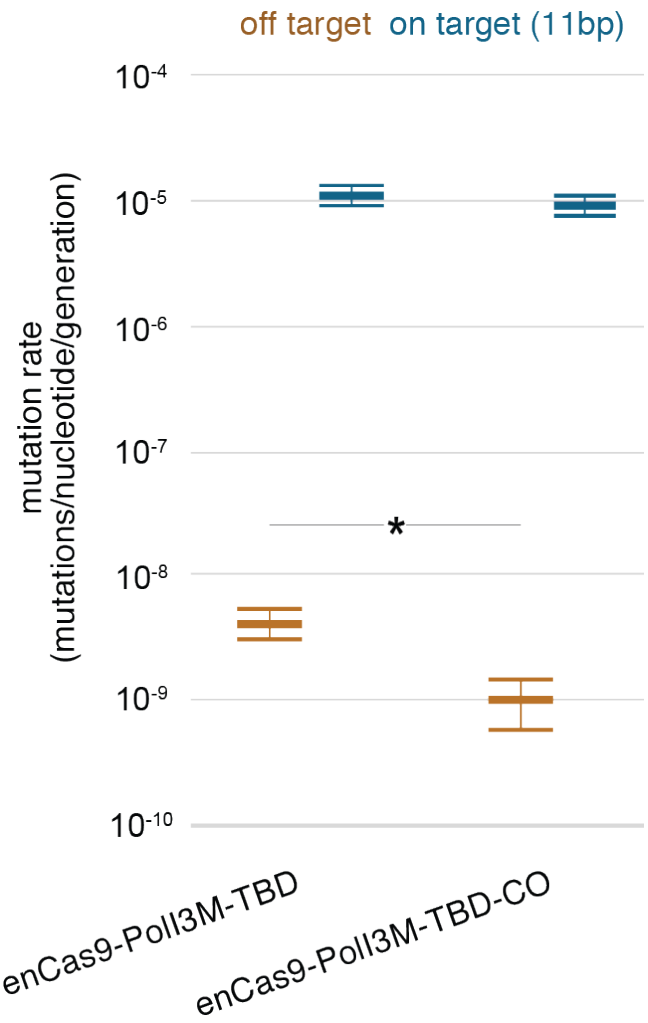

Figure 15 | Removing internal ribosome binding sequences decreases EvolvR-mediated off-target mutagenesis. enCas9–PolI3M– TBD was codon optimized to remove strong ribosome binding sites in the EvolvR coding sequence that were predicted to produce an untethered DNA polymerase. The off-target mutation rate decreased 4.14-fold when expressing enCas9–PolI3M–TBD-CO compared to enCas9–PolI3M–TBD (P=0.000482) whereas the on-target mutation rate only decreased 1.23- fold. Data are mean±95% confidence intervals from ten biologically independent samples. *P < 0.0001; two-sided student's t-test.

4.5. Targeting chromosomal and essential loci

Importantly, EvolvR also showed the capacity to diversify chromosomal loci by increasing the fraction of the population resistant to spectinomycin approximately 16,000-fold after targeting enCas9–PolI3M to the endogenous ribosomal protein subunit 5 gene of *E. coli* (*rpsE*), which has mutations that are known to confer resistance to spectinomycin (Funatsu, Schiltz, and Wittmann 1972) (Figure 16).

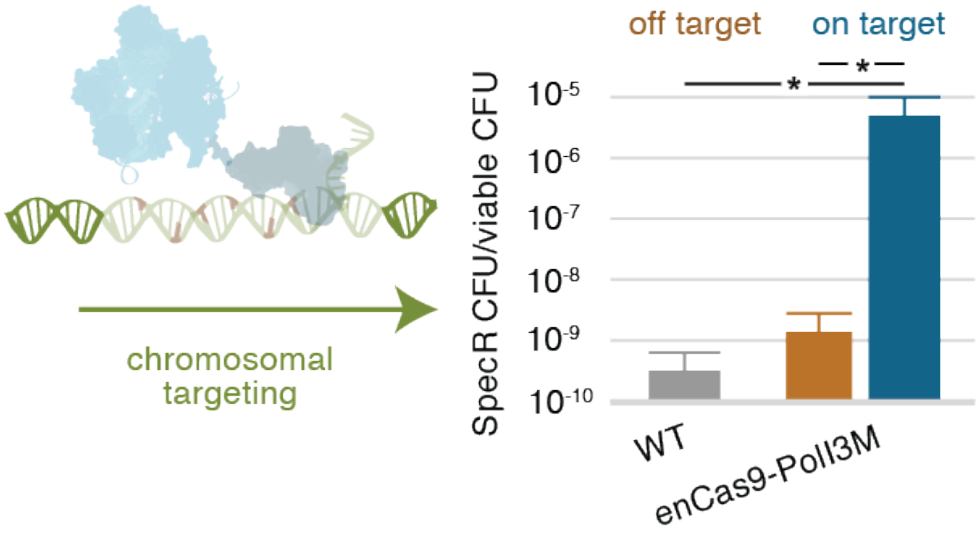

Figure 16 | enCas9–PolI3M targeted to *E. coli rpsE* generated approximately 16,000-fold more spectinomycin-resistant CFUs (SpecR CFUs) than when targeted to the *dbpA* locus. Resistant CFUs/viable CFUs data are mean±s.d. from ten biologically independent samples. *P < 0.05; two-sided student's t-test.

4.6. Comparing to previously developed continuous mutagenesis tools

Next, I tested whether EvolvR avoids the toxicity associated with non-targeted mutagenesis systems (X. Zheng, Xing, and Zhang 2017). I found that, unlike two previously developed non-targeted continuous-mutagenesis systems, EvolvR does not impede cell viability or growth rate (Figure 17, Figure 18). Additionally, targeting EvolvR to the *rpsE* gene evolved more spectinomycin-resistant CFUs per ml compared to these previous non-targeted mutagenesis systems (Figure 19).

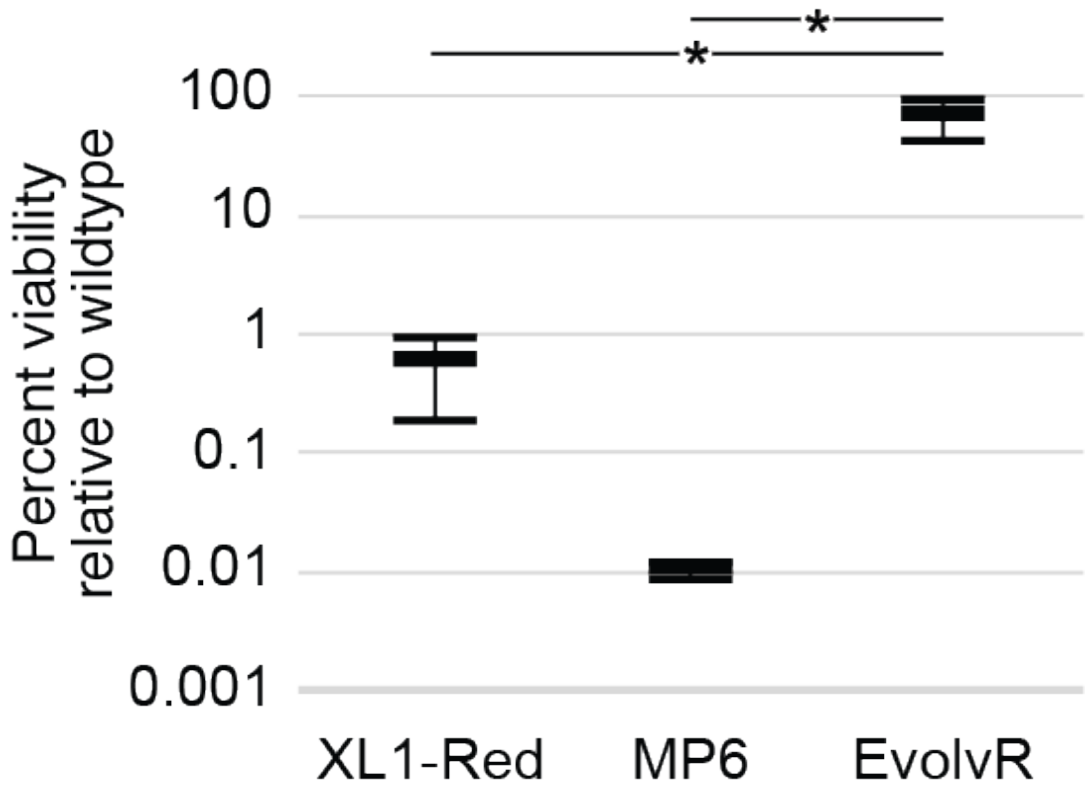

Figure 17 | The viability of TG1 *E. coli* expressing EvolvR targeted to the essential *rpsE* gene was significantly higher than TG1 *E. coli* transformed with the MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose (a previously developed plasmid for continuous non-targeted mutagenesis (Badran and Liu 2015), P=0.0108) as well as XL1-Red *E. coli* (a previously developed strain for continuous nontargeted mutagenesis (Greener, Callahan, and Jerpseth 1997), P=0.0105). Viability was measured relative to TG1 *E. coli* transformed with an empty control plasmid. Data are mean±s.d. from three biologically independent samples. *P < 0.05; two-sided student's t-test.

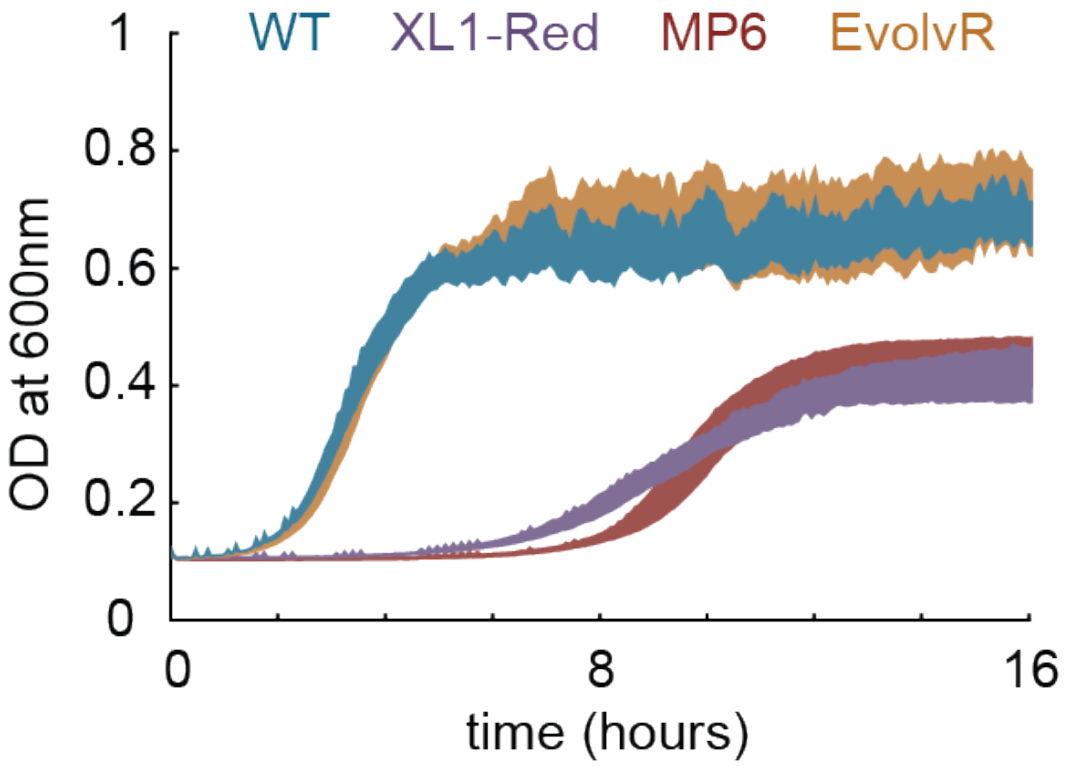

Figure 18 | TG1 *E. coli* transformed with an empty control plasmid and TG1 *E. coli* transformed with pEvolvR targeting the *rpsE* gene resulted in similar growth curves whereas XL1-Red *E. coli* and TG1 *E. coli* transformed with MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose grew much slower and saturated at lower final optical densities. Shaded area represents mean±s.d. from three biologically independent samples.

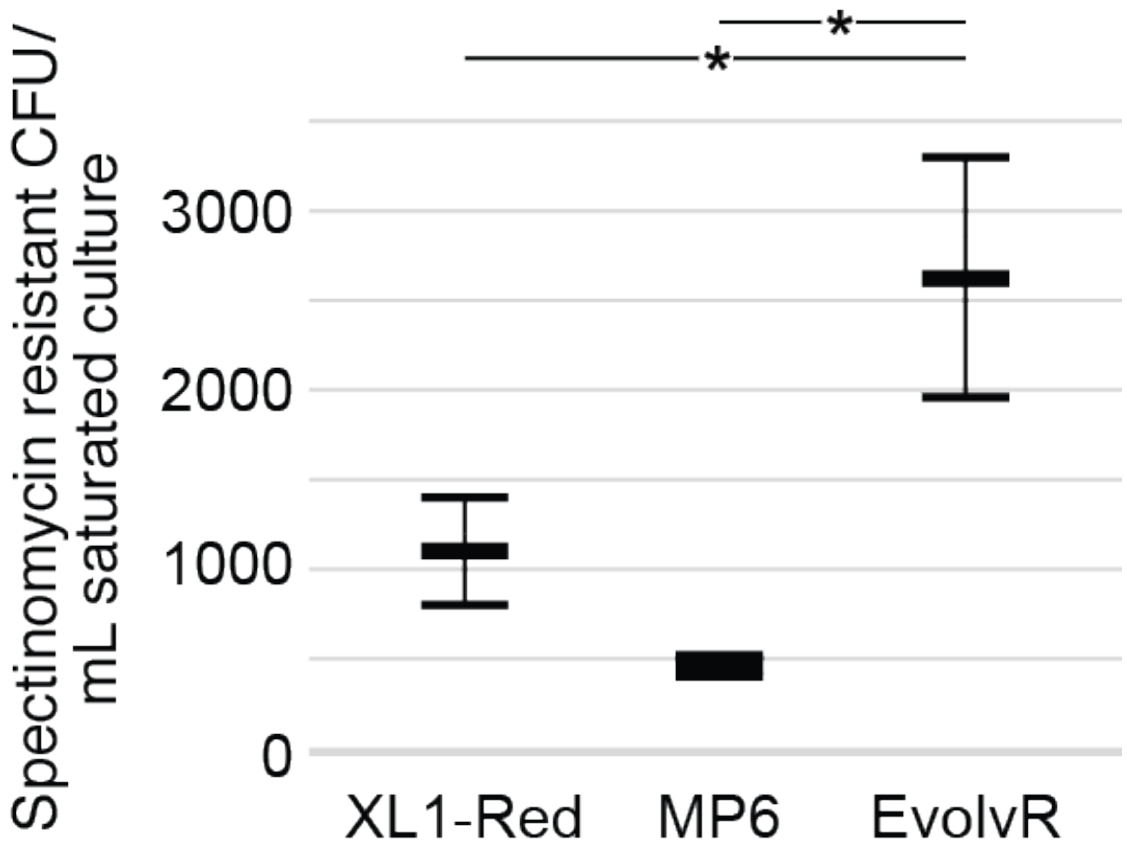

Figure 19 |The spectinomycin-resistant CFUs per ml saturated culture of TG1 *E. coli* targeting EvolvR to the *rpsE* gene was significantly higher than XL1-Red *E. coli* (P=0.022) and TG1 *E. coli* transformed with MP6 plasmid and induced with 25 mM arabinose and 25 mM glucose (P=0.0049). Data are mean±s.d. from three biologically independent samples. *P < 0.05; two-sided student's t-test.

4.7. Multiplexing targets

EvolvR could enable simultaneous diversification of distant genomic loci through coexpression of multiple gRNAs. Expression of a gRNA targeting enCas9–PolI3M–TBD to *rpsL*, a ribosomal protein subunit gene capable of acquiring mutations that confer streptomycin resistance (Timms et al. 1992), increased the rate of acquiring streptomycin resistance compared to wild-type cells, without altering sensitivity to spectinomycin (Figure 20). By comparison, coexpression of the gRNAs targeting *rpsE* and *rpsL* generated approximately the same number of respective spectinomycin- and streptomycin-resistant CFUs as observed for individual expression of the *rpsE* gRNA ($P = 0.0752$; two-sided Student's *t*-test) and *rpsL* gRNA ($P = 0.885$; two-sided Student's *t*-test). This capacity to simultaneously diversify multiple loci will be useful for identifying epistatic interactions. I also note that the expression of two gRNAs that nick separate strands at genomic loci separated by 100 bp was lethal, whereas nicking the same strand at this 100-bp distance was not lethal. Therefore, if multiple gRNAs are to be used to increase the length of the target region, I recommend targeting the same strand.

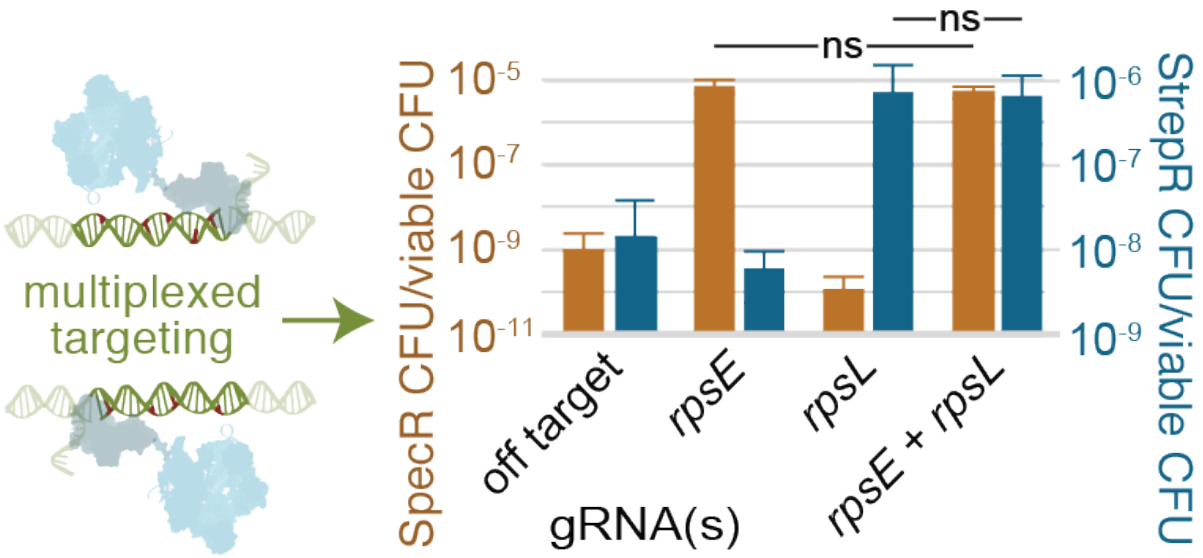

Figure 20 | enCas9–PolI3M–TBD targeted to *rpsL* increased the rate of acquiring streptomycin resistance without increasing the rate of acquiring spectinomycin resistance. Coexpression of both *rpsL* and *rpsE* gRNAs increased both spectinomycin and streptomycin resistant CFUs. Resistant CFUs/viable CFUs data are mean ± s.d. from ten biologically independent samples.

5. Chapter 5: CRISPR-guided DNA polymerases in eukaryotic cells

Diversification of the human genetic code for phenotypic selection, screening, and tracking enables our efforts to both understand our biology and treat our diseases. Programmable nucleases, such as CRISPR/Cas9, have revolutionized our ability to easily disable targeted genetic elements; however, substituting nucleotides in user-defined regions of the human genome using programmable nucleases remains inefficient due to the need for homology-directed repair (HDR), donor templates, cytotoxic double-stranded breaks, and competition with non-homologous end-joining (NHEJ) (Mao et al. 2008a; Aguirre et al. 2016; Choi and Meyerson 2014; Frock et al. 2014; Mao et al. 2008b). Here, I develop the first HDR-independent method to diversify all nucleotide types at user-defined loci in the human genome using EvolvR. My results demonstrate that CRISPR-guided DNA polymerases enable programmable diversification of user-defined loci in the human genome that will be useful for forward genetic approaches to understanding pathologies and engineering therapies.

Investigation and engineering of human cells and proteins would benefit from the ability to generate libraries of cells genetically diversified within user-defined regions of their genomes. This goal would be enabled by a programmable mutator ideally capable of generating all nucleotide substitutions, as well as deletions or insertions, within a targeted region of a genome without the requirement for inserting libraries of nucleic acid sequences at the target locus. The advent of programmable nucleases offers potential solutions to this challenge of targeted genome modification. These systems, such as CRISPR/Cas9, can create double-stranded breaks at user-defined loci. Precise alterations can then be introduced at these break sites by integrating donor constructs via homology-directed repair (HDR). Unfortunately, HDR is extremely inefficient compared to non-homologous end-joining (NHEJ) in many eukaryotic organisms, which results in insertion or deletion of a variable number of nucleotides at the break site rather than the intended modification (Mao et al. 2008a, [b] 2008). Additionally, the double-stranded breaks generated by these nucleases can be toxic, often leading to induced cell death and genomic rearrangements (Aguirre et al. 2016; Choi and Meyerson 2014; Frock et al. 2014). Therefore, while programmable nucleases have been useful for creating NHEJ-mediated insertions and deletions that disrupt genetic elements, the requirement for integrating synthesized oligonucleotides at the site of toxic double-stranded breaks to generate substitutions limits programmable nucleases' utility in genome diversification. Therefore, no HDR-independent method currently exists to diversify all nucleotide types at user-defined genomic loci in human cells.

Two studies recently diversified user-defined loci in mammalian cells without generating double-stranded breaks via targeting a cytidine deaminase with programmable nucleases (Ma et al. 2016; Hess et al. 2016). The positions most frequently mutated using these tools clustered within a narrow 10 base pair window, and one report accordingly found that evenly mutating a 100 base-pair window required 7 guide RNAs (gRNAs) (Hess et al. 2016). Additionally, cytosine deaminase-driven diversification only generated functional mutations of cytosine and guanine. For directed evolution of protein coding sequences, mutating a codon's cytosine and guanine can lead to an average of six out of the 19 possible missense mutations. While this tool

represented an advance, the short mutagenesis window length and strong substitution bias limit its utility in diversifying genomes. Therefore, I sought to adapt EvolvR for use in human cells to offer the first method for diversification of all nucleotides at user-defined positions without generating double stranded breaks or relying on HDR.

5.1. Measure targeted and global mutagenesis in human cells

I acquired a HEK293T cell line with a genomically-integrated, constitutively-expressed blue fluorescent protein (BFP) gene (HEK293T-BFP) (Richardson et al. 2016). BFP can undergo a particular single-nucleotide substitution causing an H67Y missense mutation to become green fluorescent protein (GFP) (Glaser, McColl, and Vadolas 2016) (Figure 21).

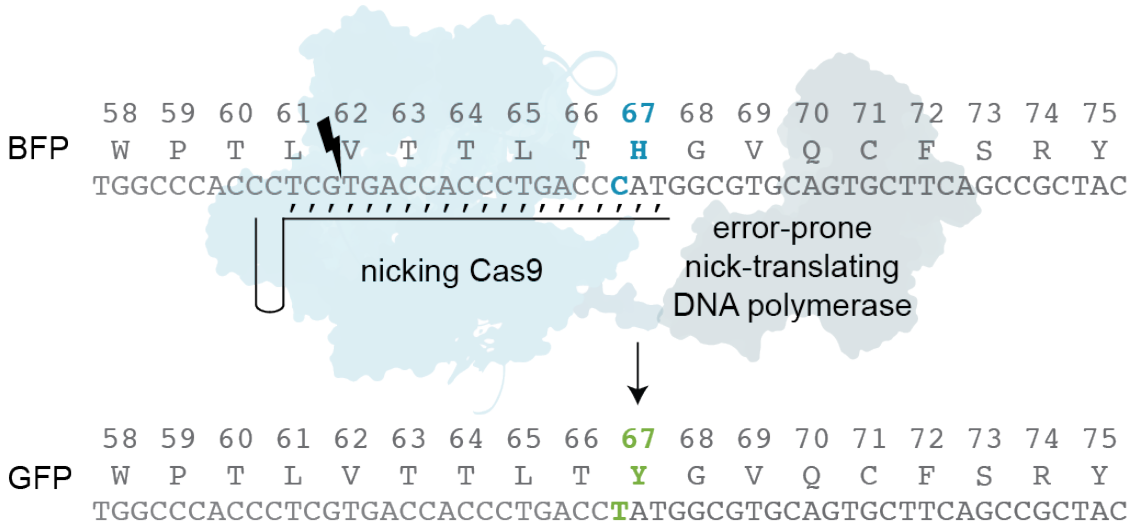

Figure 21 | The blue fluorescent protein (BFP) gene can become a green fluorescent protein (GFP) gene by undergoing a particular single-nucleotide substitution causing an H67Y missense mutation. Therefore, the frequency of cells expressing GFP after targeting EvolvR to a BFP gene integrated into the genome of HEK293T cells can be used as a relative proxy for EvolvR's targeted mutation rate in human cells.

Therefore, I used the frequency of GFP positive cells in an EvolvR-expressing population as a relative proxy for the cell's targeted mutation rate. I constructed an EvolvR expression plasmid the can be transiently transfected into human cells (pEvolvR-HT). The plasmid consists of a single guide RNA expression cassette driven by the human U6 promoter as well as a CMV promoter-driven enCas9-PolI5M gene tagged with two SV40 nuclear localization sequences (NLSs) and an mCherry fluorescent reporter (Figure 22).

Two days after transiently transfecting EvolvR into HEK293T-BFP cells, I enriched for transfectants by sorting mCherry positive cells using Fluorescence Assisted Cell Sorting (FACS). Following five days of expansion, I analyzed the frequency of GFP-expressing cells using flow cytometry (Figure 23).

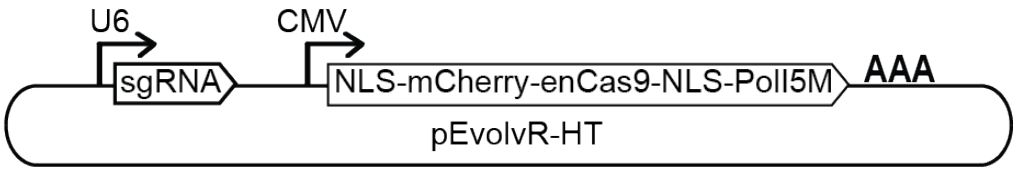

Figure 22 | The plasmid used for expression of EvolvR in human cells following transient transfection (pEvolvR-HT) consists of a single guide RNA expression cassette driven by the human U6 promoter as well as a CMV promoter-driven enCas9-PolI5M gene tagged with two SV40 nuclear localization sequences (NLSs) and an mCherry fluorescent reporter.

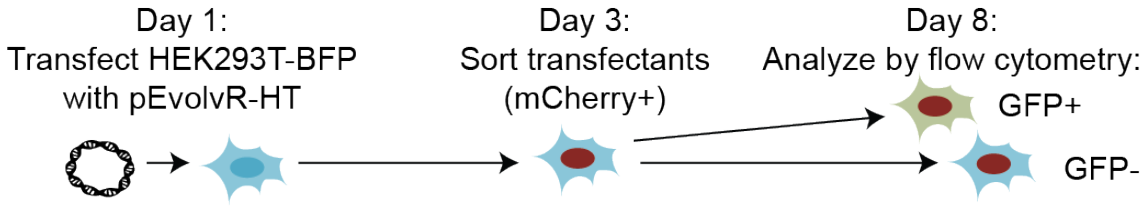

Figure 23 | Two days after transiently transfecting EvolvR-HT into HEK293T-BFP cells, transfectants were enriched by sorting mCherry positive cells using Fluorescence Assisted Cell Sorting (FACS). Following five days of expansion, the frequency of GFP-expressing cells was analyzed using flow cytometry.

While a population not expressing EvolvR and a population expressing EvolvR with an off-target gRNA targeting the aavs1 locus did not produce any GFP-expressing cells, a population expressing EvolvR with a gRNA that nicks 15 base-pairs away from the H67Y mutation showed 0.05% GFP-expressing cells (Figure 24). Estimating that the population expressed EvolvR for six generations, the observed mutation rate is approximately 2.5×10^{-4} mutations per nucleotide per generation. This mutation rate is comparable to the mutation rate I observed in *E. coli*.

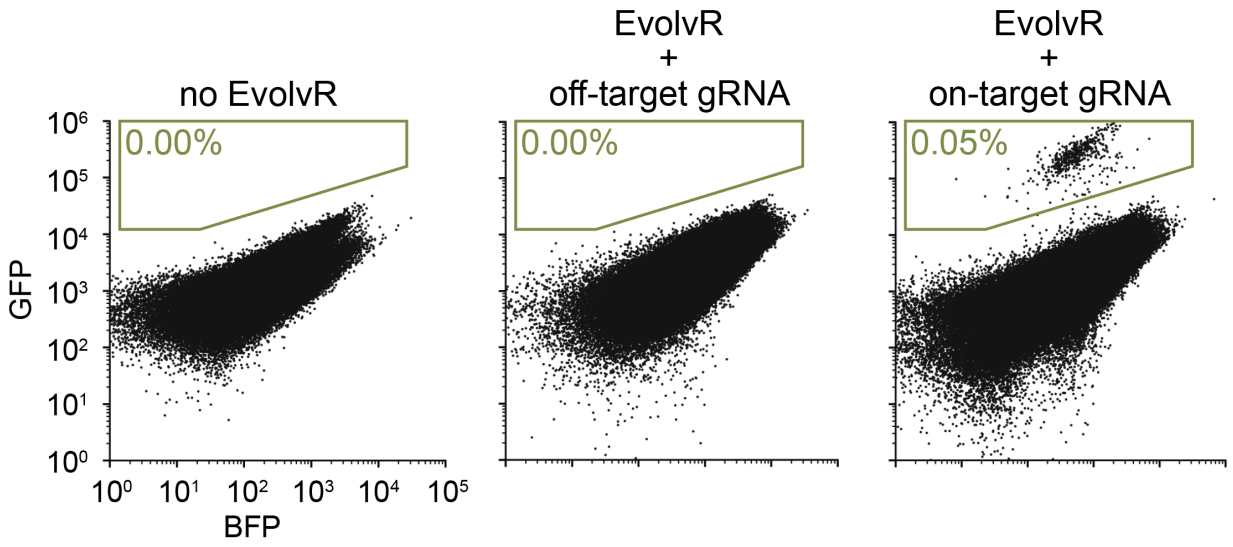

Figure 24 | A population not expressing EvolvR and a population expressing EvolvR with a gRNA targeting the aavs1 locus (off-target gRNA) did not produce any GFP-expressing cells. A population expressing EvolvR with a gRNA that nicks 15 base-pairs away from the H67Y mutation showed 0.05% GFP-expressing cells.

6. Chapter 6: Using CRISPR-guided DNA polymerases for directed evolution

6.1. Coupling CRISPR-guided DNA polymerase-mediated mutagenesis to non-selectable genetic screens

The ability to couple EvolvR-mediated mutagenesis to a genetic screen of a non-selectable phenotype would considerably broaden the utility of EvolvR. I found that after targeting EvolvR to a plasmid containing a green fluorescent protein (GFP) cassette with an early termination codon, 0.06% and 0.07% of the population expressed GFP, whereas no cells expressed GFP when an off-target gRNA was used (Figure 26 and 27).

Finally, targeting EvolvR to a GFP cassette containing a nonsense mutation resulted in 28 times more GFP-positive cells than when using the most recently developed non-targeted continuous mutagenesis technique (Figure 28).

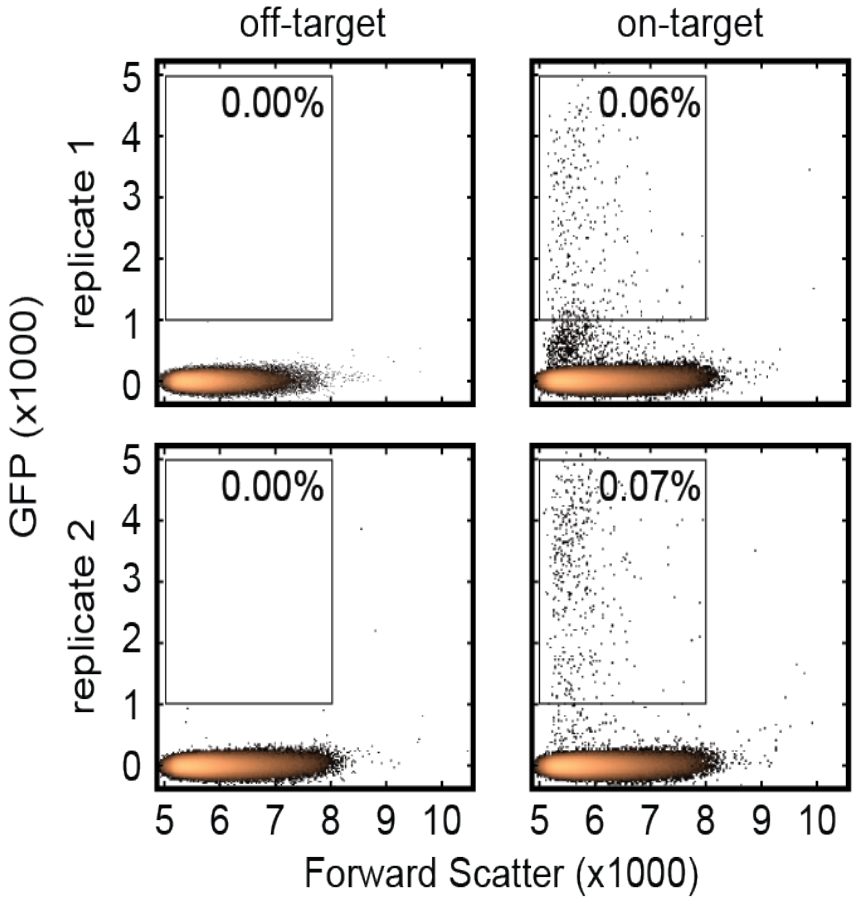

Figure 26 | To test the capability for coupling EvolvR-mediated mutagenesis with a non-selectable genetic screen, I designed a target plasmid containing a GFP cassette with an early termination codon in the GFP coding sequence (pTarget-GFP*). After co-transforming pEvolvR with pTarget-GFP* and growing for 24 h, I analysed and sorted the GFP-positive fraction. In the two replicates expressing an off-target gRNA, I did not detect or sort any GFP cells. By contrast, for the two replicates expressing a gRNA nicking four nucleotides away from the chain-terminating mutation in the coding sequence of GFP, I found that 0.06% and 0.07% of the total cells were GFP positive. These results agree with sequencing outcomes from Figure 3, which showed that expressing nCas9–PolI3M for 24 h produces substitutions in the target region at frequencies between 0.5% to 1%.

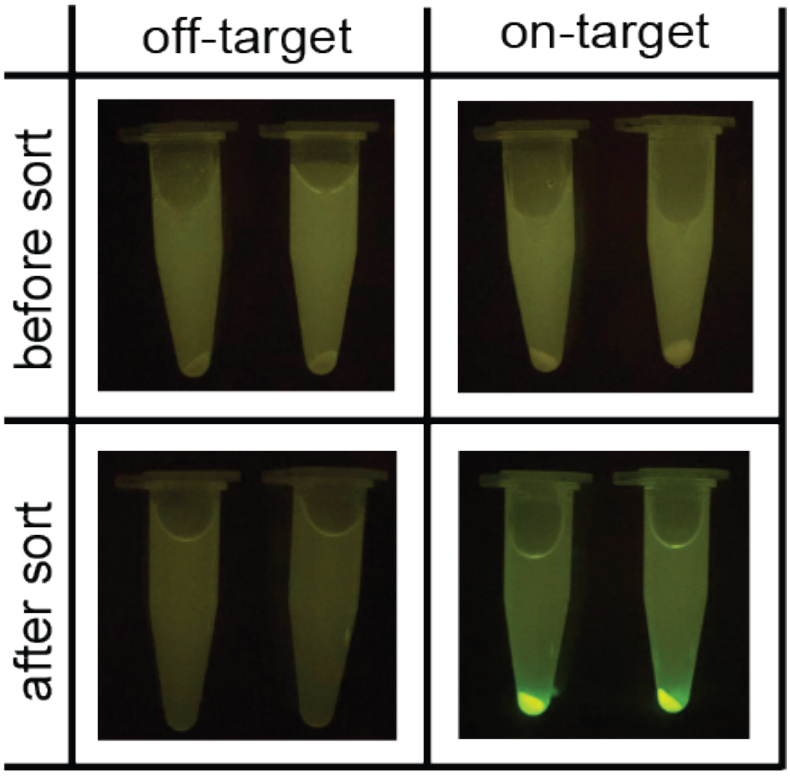

Figure 27 | After culturing the sorted populations, both replicates expressing an off-target gRNA did not show growth, whereas both replicates expressing the on-target gRNA grew bright green.

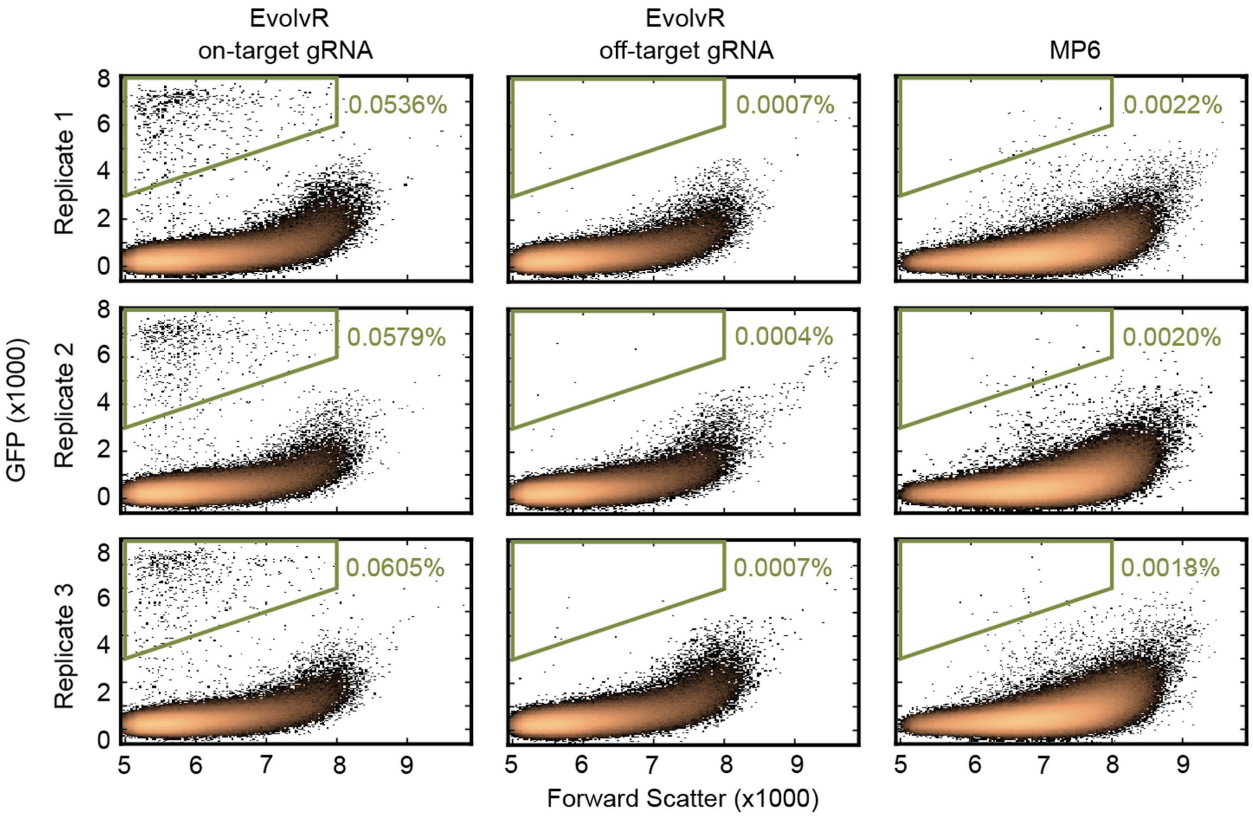

Figure 18 | EvolvR-mediated mutagenesis performs better than a previous non-targeted diversification technique. To compare the performance of EvolvR and the previously developed non-targeted mutagenesis plasmid MP6 in screen-based directed evolution applications, I co-transformed pEvolvR (enCas9–PolI3M–TBD) or MP6 with a target plasmid containing a GFP cassette with an early termination codon in the GFP coding sequence (pTarget-GFP*). The cultures expressing EvolvR were grown for 24 h and the MP6 cultures followed a two day growth–induction protocol as previously described. Flow cytometry revealed that cultures expressing EvolvR and an on-target gRNA resulted in 28-fold more GFP-positive cells than MP6 cultures.

6.2. Continuous directed evolution enables adaptation to two distinct selective pressures

To evolve resistance to both spectinomycin and streptomycin, I used the continuous diversity generation of EvolvR for continuous directed evolution (in which mutagenesis, selection and amplification occur simultaneously) to allow adaptation to modulated selection pressures with minimal researcher intervention. Cultures expressing enCas9–PolI3M–TBD and either the *rpsL* gRNA or both *rpsE* and *rpsL* gRNAs grew in liquid medium supplemented with streptomycin, whereas cultures expressing an off-target gRNA or the *rpsE* gRNA did not (Figure 29). After the cultures were diluted 1,000-fold into liquid medium supplemented with both spectinomycin and streptomycin, only cultures expressing both *rpsE* and *rpsL* gRNAs grew.

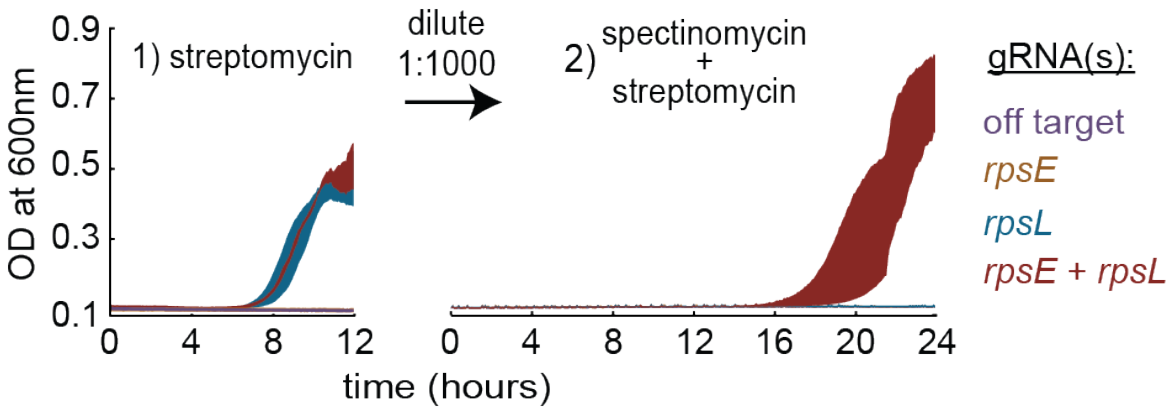

Figure 29 | Cultures expressing enCas9–PolI3M–TBD and either the *rpsL* gRNA or both *rpsE* and *rpsL* gRNAs grew in streptomycin-supplemented medium, whereas cultures expressing an off-target gRNA or the *rpsE* gRNA did not. After back-dilution into spectinomycin- and streptomycin-supplemented media, only cultures expressing both *rpsE* and *rpsL* gRNAs grew. The shaded region of OD600 nm indicates mean ± s.d. from three biologically independent samples.

6.3. Identifying novel spectinomycin resistance mutations

The clinical utility of spectinomycin as a broad-spectrum antibiotic has motivated previous efforts to characterize genomic mutations conferring spectinomycin resistance (Brocklehurst and Peter 2002). I used the capacity of EvolvR to diversify the genomic *rpsE* gene to identify novel mutations that confer spectinomycin resistance by disrupting the spectinomycin-binding pocket of the 30S ribosome (Figure 30).

First, I targeted enCas9–PolI3M–TBD to five dispersed loci in the endogenous *rpsE* gene using gRNAs that nick after the 119th, 187th, 320th, 403rd or 492nd base pair within the 504-bp *rpsE* coding sequence (Figure 31).

Then, I challenged the cell populations for growth on agar plates supplemented with varying concentrations of spectinomycin and observed that resistance was highest with the gRNAs targeted to the domain of the ribosomal subunit protein that is proposed to interact with spectinomycin (Figure 32).

After selection, high-throughput sequencing of the resistant cells containing gRNAs A, B and C revealed that all 12 types of substitutions, as well as deletions, were generated (Figure 33).

For functional analysis, I introduced five of the candidate mutations not previously described as providing spectinomycin resistance into a different strain of *E. coli* (RE1000) using oligonucleotide-mediated recombination. Growth curves in varying concentrations of spectinomycin confirmed that each of the five mutations (Δ17–19; K23N, Δ24; Δ24; Δ26; G27D) provided varying levels of spectinomycin resistance, but reduced fitness in the absence of selection (Figure 34).

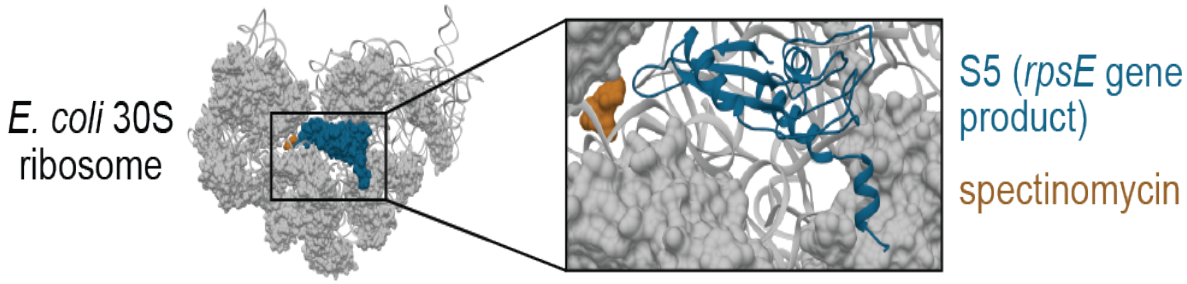

Figure 30 | Spectinomycin inhibits protein synthesis through interactions with the 30S ribosome.

Figure 31 | enCas9–PolI3M–TBD was targeted to five dispersed loci in the endogenous *rpsE* gene using gRNAs that nick after the 119th, 187th, 320th, 403rd or 492nd base pair of the 504-bp *rpsE* coding sequence. The locations of the previously identified *rpsE* mutations that provide spectinomycin resistance are colored orange, and the region where I identified new spectinomycin-resistance mutations is highlighted in red.

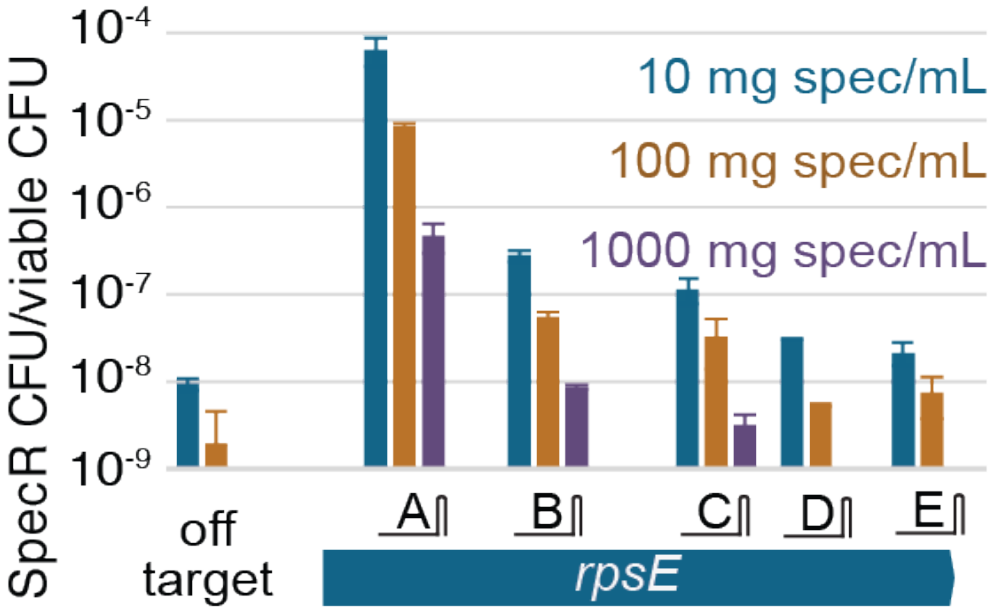

Figure 32 | enCas9–PolI3M–TBD targeted to different parts of the endogenous *rpsE* gene with five gRNAs showed higher rates of spectinomycin (spec) resistance than targeting *dbpA* (off-target). Data are mean ± s.d. from three biologically independent samples.

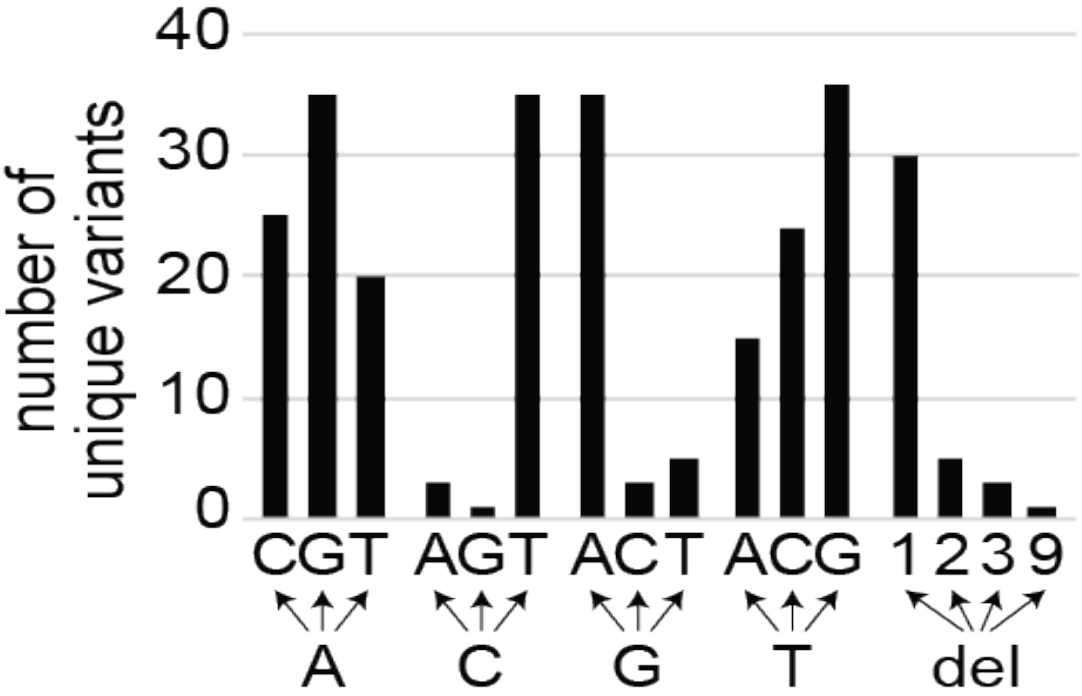

Figure 33 | After selection, high-throughput sequencing of the resistant cells containing gRNAs A, B and C revealed that all 12 types of substitutions as well as deletions were generated.

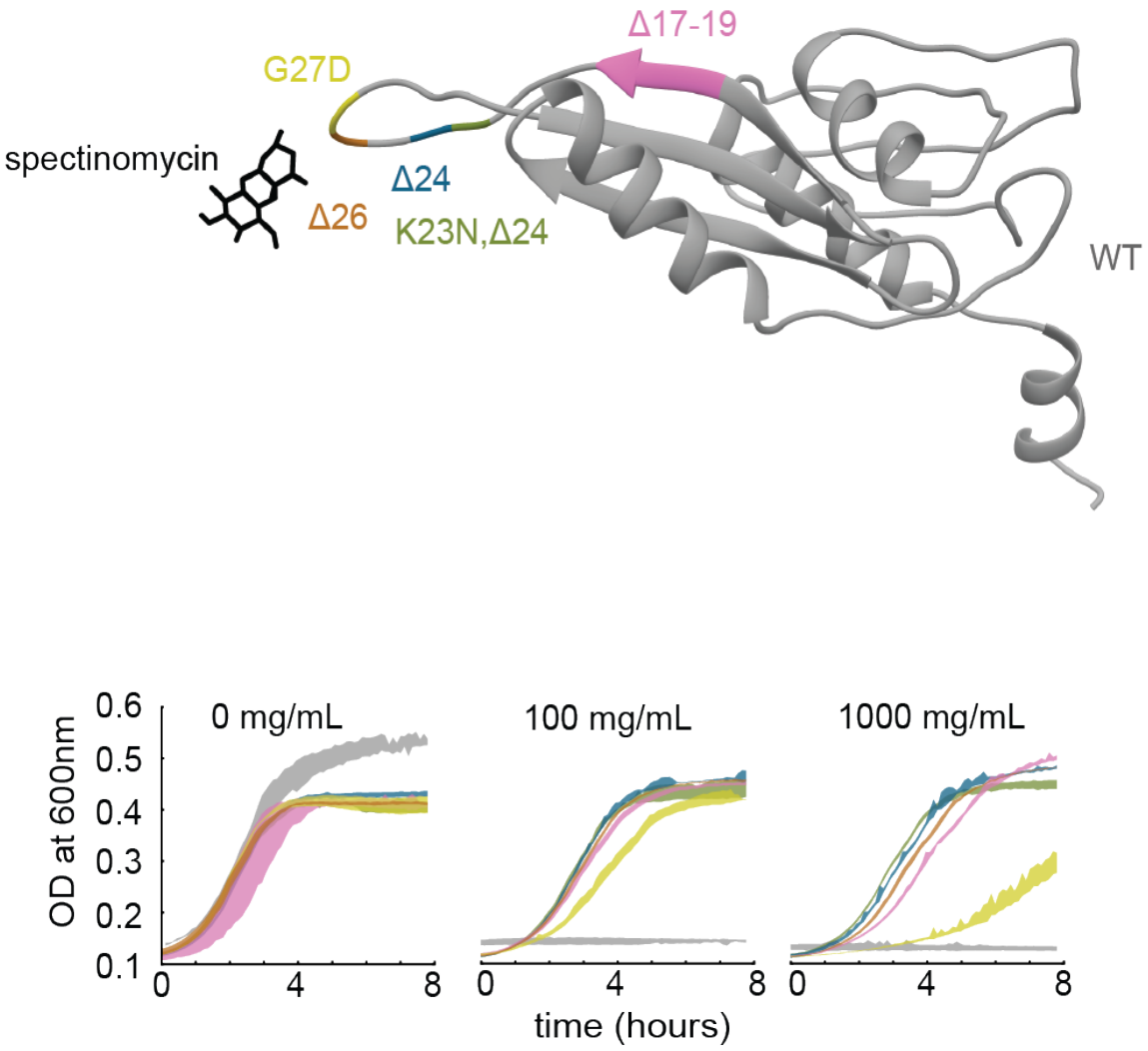

Figure 34 | Top, five mutations not previously described as conferring spectinomycin resistance were regenerated in a new strain of *E. coli* (RE1000). Bottom, growth curves in varying concentrations of spectinomycin confirmed that the mutations provide spectinomycin resistance. Shaded area represents mean ± s.d. from three biologically independent samples.

These mutations are located between the 48th and 81st base pair of the *rpsE* coding sequence, suggesting that enCas9-PolI3M-TBD provides elevated mutation rates more than 71 nucleotides away from the nick site. The mutations would be expected to move Lys26, which is predicted to hydrogen bond with spectinomycin, relative to the spectinomycin binding pocket (Figure 35).

On the basis of these mutations, I hypothesized that mutations that move Lys26 relative to the spectinomycin-binding pocket confer resistance to spectinomycin by removing a hydrogen bond that stabilizes the interaction of spectinomycin with the ribosome. Therefore, I tested an array of deletions that I predicted would move Lys26 and discovered additional novel mutations that confer spectinomycin resistance (Figure 36). This rapid method for discovering genotypes conferring antibiotic resistance will be generally useful for improving the effective use of antibiotics.

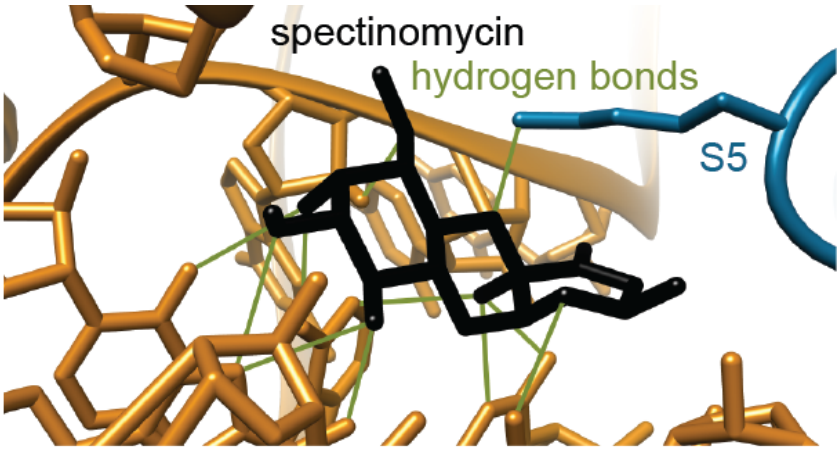

Figure 35 | The mutations that I discovered confer spectinomycin resistance would be expected to move Lys26 (which is predicted to hydrogen bond with spectinomycin) relative to the spectinomycin-binding pocket. I hypothesized that mutations that move Lys26 relative to the spectinomycin-binding pocket remove that hydrogen bond and destabilize the interaction of spectinomycin with the ribosome, thereby conferring spectinomycin resistance.

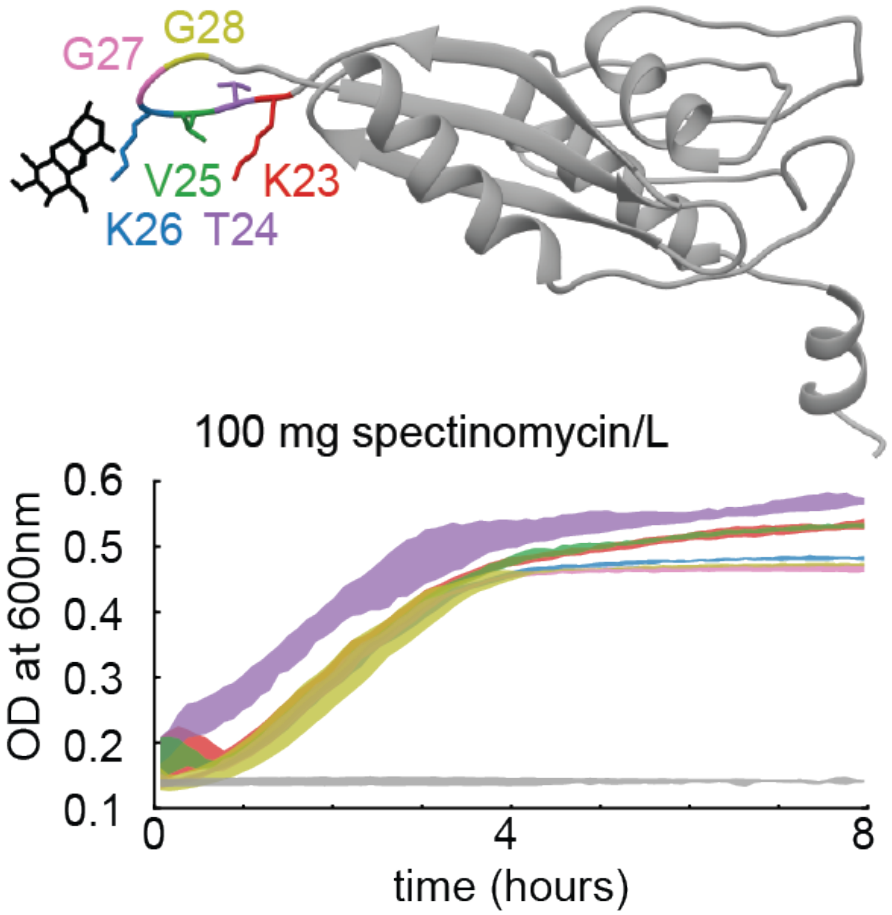

Figure 36 | The effects of deleting any single amino acid between residues 16 and 35 were determined. I found that deleting residues 23, 24, 25, 26, 27 or 28 provides spectinomycin resistance whereas deleting any of the residues between 16 and 22 or 29 and 35 does not. These results support the hypothesis that one mechanism of resistance to spectinomycin is disruption of the interaction between Lys26 and spectinomycin. Data are mean ± s.d. from three biologically independent samples.

7. Chapter 7: Further Work

7.1. Evolving efficient utilization of D-galacturonic acid in yeast

To expand the utility of EvolvR to targeted mutagenesis in other eukaryotic organisms, we can adapt EvolvR to function in the model eukaryote *S. cerevisiae*. One concern for using EvolvR in a eukaryotic host is whether the mutagenesis will be affected by the presence of chromatin structures. CRISPR-Cas9 and Cpf1 systems have shown robust activity in eukaryotic cells. Additionally, *E. coli* PolI has been widely used for *in situ* DNA-labeling in mammalian cells, suggesting the polymerase will display enzymatic activity despite chromatin structures (Cremer et al. 2012). However, if polymerase processivity is affected in eukaryotes, we can express orthogonal processivity-enhancing replisome machinery. For example, while T7 bacteriophage's DNA polymerase alone synthesizes hundreds of nucleotides per binding event, when complexed with the T7 helicase it synthesizes approximately 16,000 nucleotides per binding event (Kulczyk et al. 2012). For targeting several regions simultaneously in *S. cerevisiae*, which does not possess the CRISPR-array processing machinery, we can use a previously developed sgRNA multiplexing strategy (Xie, Minkenberg, and Yang 2015). The mutation rate and editing window length of different designs can be determined with a fluctuation analysis measuring the reversion of a nonsense mutation in an uracil auxotrophic marker. The mutagenesis rates and editing window lengths of final designs will be characterized with targeted amplicon sequencing.

As an initial proof-of-concept directed evolution application, we can engineer a strain of *S. cerevisiae* for efficient utilization of D-galacturonic acid, which is abundant in pectin-rich biomass. One million tons of dried orange pulp and 7 million tons of dried sugar beet pulp are produced yearly in the United States as byproducts of the orange juice and sugar industries. These and other agricultural wastes, which contain large fractions of pectin, are disposed by drying and selling cheaply as cattle feedstock. Accordingly, they represent an inexpensive feedstock for microbial fermentations. Pectin, unlike woody lignocellulosic polymers, can be efficiently and cheaply hydrolyzed enzymatically into monomers of its backbone sugar, >70% D-galacturonic acid. Engineering a microbial strain to convert pectin hydrolysates into commodity chemicals poses a number of challenges the Dueber lab has already tackled. First, the low pH of high galacturonic acid concentrates was addressed by the choice of *Saccharomyces cerevisiae* as the production host, which is naturally robust to low pH. Second, the ability to import galacturonic acid in the presence of the high glucose concentrations, also present in pectin hydrolysate, by isolating a transporter that can efficiently import GalA without inhibition by glucose when heterologously expressed in *S. cerevisiae*.

This system provides identified bottlenecks that can be targeted under a constant selection (i.e., growth on D-galacturonic acid as the sole carbon source). Previous adaptations of a strain expressing the pathway enzymes and transporter improved the doubling time from 80 to 20 hours by increasing the copy number of the pathway genes. Thus, flux through the pathway is empirically limited by the activity of at least one of the enzymes. The Dueber lab has performed biochemical analysis on each of the four enzymes required for integrating into central metabolism and determined that three have activities higher than the lowest activity glycolytic

enzyme. The fourth enzyme, L-galactonate dehydratase (LGD1), has considerably lower activity and therefore is a promising directed evolution target. We are thus well positioned to continuously mutate LGD1 with EvolvR while growing on D-galacturonic acid as the sole carbon source to enrich for improved enzyme properties, stability, and kinetics. Subsequently, the entire pathway (four catabolic enzymes and transporter), can be co-targeted for increased growth on D-galacturonic acid resulting from improved expression balancing or further increased enzyme stability and kinetics. The most frequent mutations in the adapted population will be characterized with targeted amplicon sequencing. Combinations of these mutations will be integrated into the wild-type cell to isolate their effect on D-galacturonic acid utilization. This effort will further establish EvolvR's broad utility for industrial microbiology and provide a platform for other applications.

7.2. Utilizing CRISPR-guided DNA polymerases for crop trait engineering

Humans have been breeding plants with more desirable characteristics for 10,000 years. Over the past 70 years, breeders leveraged a new understanding of genetic variability and heritability to accelerate this process by using physical and chemical agents to induce mutations and increase the genetic variability that can be explored. These mutagenesis program have resulted in over 3200 new crop varieties in over 200 plant species. Despite this success, the major limitation of chemical and physical mutagenesis is that mutations are randomly generated across the genome, requiring enormous, expensive, and slow screening efforts followed by backcrossing with parental strains. Crop trait engineering efforts would benefit from a method to efficiently diversify regions of plant genomes known to be responsible for agriculturally valuable traits.

Currently, the most popular method for exploring diversity in a specific region of a plant genome is Targeting Induced Local Lesions In Genomes (TILLING). This process generally consists of randomly mutating germplasm with chemical mutagens and then employing DNA-screening techniques to segregate seeds containing mutations in the target region. Unfortunately, this process is cumbersome, expensive, inefficient, and imprecise (Jacob, Avni, and Bendahmane 2018). Furthermore, the segregated seeds contain thousands of off target mutations, which necessitates extensive backcrossing with parental strains.

The development of programmable nucleases such as CRISPR/Cas9 have enabled precision editing of plant genomes through homology directed repair (HDR). However, the process of a plant repairing a genomic break with foreign homologous DNA remains inefficient or even impossible in many crop plants (Svitashev et al. 2015; Zhao et al. 2016). Since HDR editing efficiency is often below 1% in plants, this tool has not been used to generate large libraries of variants for forward genetic studies in crops. Additionally, multiplexed HDR in plants is often unsuccessful because generating multiple double-stranded breaks can result in large deletions and rearrangements rather than the intended changes. CRISPR-deaminase base editors have enabled high efficiency precision editing in crop plants without relying on the integration of oligonucleotides or double-stranded breaks at the target locus. Unfortunately, these tools are restricted to generating certain types of mutations in a narrow editing window (Shimatani et al. 2017).

EvolvR, our CRISPR-guided DNA polymerase system, has already demonstrated continuous and multiplexable targeted diversification of all nucleotides with tunable mutation rates and tunable editing window lengths in multiple organisms. Integrating EvolvR into a crop trait development pipeline would offer several advantages over current technologies: (i) targetability alleviates the need for both segregating seeds with mutations in the target locus and backcrossing progeny with parental strains to remove off-target mutations, (ii) EvolvR would not rely on the plant to repair a toxic double-stranded break with foreign oligonucleotides, and (iii) EvolvR generates all types of mutations within a tunable editing window length. EvolvR-mediated diversification could be used to generate targeted libraries of germplasm that can be screened for improved yield, disease resistance, herbicide tolerance, extreme weather adaptation, and grain composition.

In our work to adapt EvolvR for use in crop trait engineering, the first milestone will be to establish EvolvR's ability to generate targeted variation in the model plant *Nicotiana Benthamiana*. To achieve this, we will target EvolvR to *N. Benthamiana*'s Phytoene Desaturase gene (PDS). Homozygous knockout of PDS generates a visible albino phenotype that we can use as a proxy for targeted mutagenesis. GFP-tagged EvolvR variant enCas9-PolI5M and a sgRNA will be delivered by agrobacterium infiltration of leaf tissue followed by callus regeneration in selective media. A similar construct delivering nCas9 not fused to a DNA polymerase will be used as a control. Expression of EvolvR will be confirmed by GFP imaging. After 60 days, leaves will be inspected for albinism. We will then extract genomes from leaf tissue and perform targeted amplicon sequencing to determine the location, frequency, and types of mutations generated.

The next milestone will be to establish EvolvR's ability to generate targeted variation in an agriculturally important dicot crop plant as well as EvolvR's ability to simultaneously diversify multiple sites in plants. Similar to the previous experiment, we will target EvolvR to the PDS gene of *Solanum lycopersicum* (tomato) strain M82. GFP-tagged EvolvR variant enCas9-PolI5M and the sgRNA will be delivered by agrobacterium infiltration of leaf tissue followed by callus regeneration in selective media. After 90 days, leaves will be inspected for albinism and genomes from leaf tissue will be sequenced to determine the location, frequency, and types of mutations generated. Simultaneously, we will target two exons within the *N. Benthamiana* PDS gene using two sgRNAs and determine whether EvolvR is able to diversify both regions while avoiding large deletion events.

The final milestone will be to establish EvolvR's ability to (i) generate targeted variation in two monocot crops, (ii) provide tunable mutation rates and editing window lengths and (iii) install a heritable trait. We will determine the types and frequencies of mutations generated by delivering the EvolvR gene to rice and wheat tissue through bombardment transformation. We will simultaneously introduce EvolvR variants possessing different processivities and error rates into tomato and determine EvolvR's ability to offer tunable editing window lengths and mutation rates in plants. Finally, we will target the acetolactate synthase (ALS) gene in rice and determine EvolvR's ability to generate a heritable mutation that confers tolerance to the herbicide imazamox over multiple generations. If successful, this tool could be used to generate libraries of seeds harboring targeted diversity that can be screened for useful crop traits.

8. Conclusions

EvolvR offers the first example of continuous targeted diversification of all nucleotides at user-defined loci, which will be useful for evolving protein structure and function, mapping protein–protein and protein–drug interactions, investigating the non-coding genome, engineering industrially and therapeutically valuable cells and tracking the lineage of cell populations. As a guiding principle for using this tool, our data suggest that 1 µl saturated *E. coli* culture expressing enCas9–PolI3M–TBD for 16 h contains all single substitutions in the 60-nucleotide window with more than tenfold coverage. Future work towards adapting EvolvR for use in cells possessing low transformation efficiency, as well as increasing the mutation rate and window length of EvolvR mutagenesis, would enable new forward genetic applications.

9. Methods

No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

9.1. Plasmid Construction

All plasmids were constructed using a modular Golden Gate strategy. pEvolvR consisted of EvolvR and gRNA expression cassettes, a pBR322 origin of replication and a kanamycin resistance cassette. pTarget consisted of a p15a origin of replication carrying both a functional trimethoprim resistance cassette for selection and a disabled spectinomycin resistance gene (*aadA*) harbouring a L106X nonsense mutation. pTarget2 is identical to pTarget except that the *aadA* gene now carried both Q98X and L106X mutations. The full plasmid sequences are provided in Supplementary Table 1.

9.2. High-throughput sequencing of pTarget sample preparation

A pTarget and pEvolvR plasmid were cotransformed into 50 µl chemically competent TG1 *E. coli* prepared by a TSS/KCM method. Cells were allowed to recover in the TSS/LB solution for 1 h, before 4 µl transformation mix was inoculated into 2 ml LB containing 25 µg/ml kanamycin and 15 µg/ml trimethoprim. The cultures were grown for 24 h at 37 °C while shaking at 750 rpm. A 1.5-ml sample of each culture was miniprepped using a Zippy Plasmid Prep kit (Zymo Research). The oligonucleotides pTarget-F and pTarget-R were used to amplify the target region in a 20-cycle PCR reaction using 100 ng miniprepped DNA as the template. A second PCR reaction added Illumina sequencing adapters and indices to the previous PCR product over 10 thermocycles. A Qubit fluorimeter was used to quantify the DNA before pooling samples. The sample pool was submitted to the University of California, Berkeley Vincent J. Coates Genomics Sequencing Laboratory for quality control and sequencing. Quality control consisted of fragment analysis (Advanced Analytical) and concentration measurement of the sequenceable fraction by quantitative PCR (Kapa Biosystems). The pooled sample was mixed with Illumina PhiX sequencing control library at 10% molarity, diluted to 14 pM, denatured, and run on an Illumina MiSeq using a 150-bp paired-end read MiSeq Reagent Kit v2. Resulting basecall files where converted into demultiplexed fastq format using Illumina bcl2fastq v.2.17.

9.3. High-throughput sequencing data analysis

Perfectly complementary paired reads were filtered, and the five randomized nucleotides, amplification primer sequences, and first and last five nucleotides were trimmed using a custom Python script. Bwa and samtools were used to generate alignment files using the wild-type *aadA* gene sequence as a reference. VarScan2 was used for variant calling with the parameters: min-coverage 1; min-reads2 1; variants 1; min-var.-freq 0.0005; p-value 0.99 (Koboldt et al. 2012). The limit of detection was determined by sequencing a culture transformed

with an empty vector as a control. The highest frequency variant was 0.04% so all substitutions with a frequency under 0.05% were discarded.

9.4. Fluctuation analysis assay

A 50-µl sample of chemically competent TG1 *E. coli* were contransformed with pEvolvR and pTarget or pTarget2. After 1 h of recovery at 37 °C, 4 µl was inoculated into a 1.996 ml LB containing 25 µg/ml kanamycin and 15 µg/ml trimethoprim. After shaking at 37 °C for 16 h, 1 ml and 1 µl culture were plated on separate LB agar plates containing 50 µg/ml spectinomycin. For viable CFU counting, 300 µl of 1:50,000,000 diluted culture was plated on LB agar plates. After 24 h of incubation at 37 °C, spectinomycin-resistant CFUs and viable CFUs were counted. Ten replicates were used for each condition.

9.5. Calculation of mutation rate and statistics

The Ma–Sandri–Sarkar Maximum Likelihood Estimator was used to determine mutation rates as it is the most accurate and valid for all mutation rates (Sarkar, Ma, and Sandri 1992). Falcor was used to calculate the mutation rates by inputing the viable and resistant CFU counts for the ten replicates (Hall et al. 2009). A two-tailed Student's *t*-test was carried out to determine *P* values as previously described (Rosche and Foster 2000).

9.6. Fluorescence-activated cell sorting of EvolvR libraries

pEvolvR expressing either an on- or off-target gRNA was contransformed with pTarget-GFP* and shaken at 37 °C for 24 h. For each sample, the GFP positive fraction of a million events was sorted with a Cell Sorter SH800 (Sony) using a 488-nm laser and a 525/50-nm emission filter.

9.7. Continuous evolution of *E. coli* resistant to both spectinomycin and streptomycin

pEvolvR expressing enCas9–PolI3M–TBD and either the off-target gRNA (targeting *dbpA*), *rpsL* gRNA, *rpsE* gRNA or both *rpsL* and *rpsE* gRNAs was transformed into TG1 *E. coli* as previously described. After recovering for one hour, 4 µl of transformation mix was inoculated into 2 ml of LB supplemented with 25 µg/ml kanamycin and cultures were propagated over 16 h at 37 °C. For each culture 2 µl of culture was re-inoculated into 198 µl of LB supplemented with 50 µg/ml of streptomycin. A Tecan M1000 Pro spectrophotometer was used to measure the optical density of each well over 12 h of growth at 37 °C. Each well was then diluted 1,000-fold into LB supplemented with 50 µg/ml of streptomycin and 25 µg/ml of spectinomycin and the optical density of 200 µl of culture was again measured with a Tecan M1000 Pro spectrophotometer over 24 h of growth at 37 °C. Three biological replicates for each gRNA were characterized.

9.8. High-throughput sequencing of spectinomycin resistant *E. coli*

A pEvolvR plasmid expressing enCas9–PolI3M–TBD with *rpsE* gRNA A, B, C, D or E was transformed into chemically competent TG1 *E. coli*. Cells were allowed to recover for 1 h before innoculating 4 µl transformation mix into 1.996 ml LB supplemented with 25 µg/ml kanamycin. The cultures were grown for 16 h at 37 °C while shaking. One millilitre and one microlitre of each culture were plated on separate LB agar plates containing 10, 100, or 1,000 µg/ml spectinomycin. Resistant CFUs were counted in the same manner as the fluctuation assays. The colonies from each plate were then pooled into separate cultures containing 2 ml of LB supplemented with 50 µg/ml spectinomycin and grown for 16 h at 37 °C. Genomic DNA was purified using the Wizard Genomic DNA Purification Kit (Promega). One hundred nanograms of purified genome was then processed and sequenced in the same manner as already described for the sequencing analysis of pTarget, with the one alteration that the oligonucleotides *rpsE*-F and *rpsE*-R were used for the first round of PCR.

9.9. Oligonucleotide recombination

Re-introduction of *rpsE* mutations was performed using RE1000 *E. coli* (MG1655 λ-Red::bioA/bioB ilvG+ pTet2:gam-bet-exo-dam pN25:tetR dnaG.Q576A lacIQ1 Pcp8-araE ΔaraBAD pConst-araC ΔrecJ ΔxonA) developed for recombineering. Electro-competent cells were prepared fresh from overnight cultures of bacteria. The saturated culture was back-diluted 1:70 into 5 ml LB with 100 ng/µl anhydrous tetracycline and shaken at 37 °C until the optical density reached 0.5. Cultures were then transferred to an ice-water bath and swirled for approximately 30 s before being chilled on ice for 10 min. Chilled cultures were centrifuged at 9,800 *g* for 1 min. The supernatant was aspirated and the pellet was resuspended in 1 ml ice-chilled 10% glycerol. Washing with glycerol was repeated twice. The final pellet was resuspended in 70 µl chilled 10% glycerol for each transformation. 1 µg of oligonucleotide was electroporated into the cells. The cells were recovered for 1 h at 37 °C in 1 ml LB and streaked out on LB agar plates containing 50 µg/ml spectinomycin. Successful recombination was verified by Sanger sequencing a PCR amplification of the genomic *rpsE* gene.

9.10. Characterization of spectinomycin resistance

Single colonies of sequence-verified *rpsE* mutants were grown overnight in LB media and then back-diluted 1:200 into LB containing 0, 100 or 1,000 µg/ml spectinomycin. A Tecan M1000 Pro spectrophotometer was used to measure the optical density of each well over 8 h of growth at 37 °C. Three biological replicates of each mutant at each spectinomycin concentration were characterized.

9.11. Transfection of HEK293-BFP cells

HEK293-BFP cells were cultured on polystyrene plates in growth media (Dulbecco's Modified Eagle Medium (DMEM) with 10% FBS (FBS) and 1% penicillin/streptomycin) at 37°C and 5% CO2. 4.5 mL growth media was added to each well of 6-well plates. When cells reached 90% confluency, growth media was aspirated from plates and cells were washed with 5 mL PBS. Cells were

trypsinized with 2 mL trypsin (0.25% EDTA) and incubated for 5 minutes at 37°C. Cells were further dislodged and trypsin was inactivated by rinsing the plate with 8 mL growth media. A 10 uL sample of the 10 mL solution of cells was mixed with an equal volume of trypan blue. Cells were counted with a hemocytometer and determined to be >95% viable. Cells were diluted to a density of $6x10^5$ cells/mL and 500 uL (300,000 cells) was dispensed into each well of 6-well plates. Cells were incubated overnight at 37°C, 5% CO2. After two days, cells were transfected with Mirus LT1 transfection reagent as recommended by the manufacturer's protocol (3 uL transfection reagent : 1 ug DNA ratio; total 2.5 ug DNA; 500 ng pEvolvR-HT, 2 ug filler DNA).

9.12. Flow cytometry and fluorescence assisted cell sorting of HEK293-BFP cells

Media was aspirated from each well within the 6-well plate. Cells were trypsinized with 250 uL trypsin EDTA (0.25%) for 5 minutes at 37°C. Cells were resuspended and rinsed with 750 uL growth media and transferred to separate 1.5 mL eppendorf tubes. Cells were centrifuged at 300 rcf for 5 minutes. Supernatant was aspirated from each tube and the pellets were resuspended in 1% BSA, 1% EDTA PBS. Cell suspensions were dispensed into FACS tubes through cell-strainer lids. Cells from desired gate condition were sorted into catch tubes containing growth media using a Sony SH800Z flow cytometer. Cells were plated in 10 cm dishes and allowed to grow to confluence.

9.13. Data availability

The data that support the findings of this study are available from the corresponding authors upon request. High-throughput sequencing data have been deposited as a NCBI BioProject under accession number PRJNA472658. Plasmids encoding enCas9-PolI3M-TBD and enCas9-PolI5M are available from Addgene (plasmids 113077 and 113078).

10. Appendix

10.1. Computer Code for variant analysis

```
import pyfaidx
        from Bio import SeqIO
        import os
        import csv
        import matplotlib
        matplotlib.use('TkAgg')
        import matplotlib.pyplot as plt


        filenames =
        ["JDHS03B_S2_L001_","JDHS03D_S4_L001_","JDHS03E_S5_L001_","JDHS03C_S3_L001_","JDHS03A
        _S1_L001_"]
```

```
# Place fastq files into subfolder
# Modify these according to your primers and target site relative to your amplicon
FwdPrimer = "TCACTGTGTGGCTTCAGG"
RevPrimer = "TCGACCCAGCTGTCGGAG"
TargetStart = 28
TargetEnd = 95


# Cycle through files
for i in range(len(filenames)):


# Trim off primers from R1
    trimmed = (rec[TargetStart:TargetEnd] for rec in
SeqIO.parse(filenames[i]+"R1.fastq", "fastq"))
    count = SeqIO.write(trimmed, filenames[i]+"R1_trimmed.fastq", "fastq")


# Make a reverse complement R1 file
    trimmed_RC = (rec.reverse_complement() for rec in
SeqIO.parse(filenames[i]+"R1_trimmed.fastq", "fastq"))
    count = SeqIO.write(trimmed_RC, filenames[i]+"R1_trimmed_RC.fastq", "fastq")
    print("Found %i trimmed RC sequences" % count)


# Make a file of the reads in R1 whose reverse complement matches perfectly with part
of R2. This removes illumina base-calling errors.
    count = 0
    matches = []
    rcrecords = list(SeqIO.parse(filenames[i]+"R1_trimmed_RC.fastq", "fastq"))
    for rec in SeqIO.parse(filenames[i]+"R2.fastq", "fastq"):
        if rcrecords[count].seq in rec.seq:
            matches.append(rec)
        count += 1
    count = SeqIO.write(matches, filenames[i]+"R1_matches.fastq", "fastq")
    print("Found %i perfectly matching sequences" % count)


# make file containing R1 reads that have either the forward or reverse primer (to
remove non target sequences)
    primer_reads = []
    for rec in SeqIO.parse(filenames[i]+"R1_matches.fastq", "fastq"):
        if FwdPrimer in rec.seq:
                primer_reads.append(rec)
        elif RevPrimer in rec.seq:
                primer_reads.append(rec)
    count = SeqIO.write(primer_reads, filenames[i]+"R1_primed_matches.fastq",
"fastq")
    print("Found %i matched sequences with primers" % count)


# make file with all reads oriented in same direction and with the primers trimmed
    count=0
    oriented = []
    for rec in SeqIO.parse(filenames[i]+"R1_matches.fastq", "fastq"):
        if FwdPrimer in rec.seq:
                oriented.append(rec[TargetStart:TargetEnd])
        elif RevPrimer in rec.seq:
                rec = rec[TargetStart:TargetEnd].reverse_complement()
                oriented.append(rec)
```

```
    count = SeqIO.write(oriented,
filenames[i]+"R1_primed_matches_oriented_trimmed.fastq", "fastq")
    print("Found %i matched sequences with primers_oriented" % count)


# alignment
    os.system("bwa index 'pSH0201.fa'")
    os.system("bwa mem -R '@RG\tID:foo\tSM:bar\tLB:library1' 'pSH0201.fa'
"+filenames[i]+"'R1_primed_matches_oriented_trimmed.fastq' > "+filenames[i]+"'.sam'")
    os.system("samtools fixmate -O bam "+filenames[i]+"'.sam'
"+filenames[i]+"'_fixmate.bam'")
    os.system("samtools sort -O bam -o "+filenames[i]+"'_sorted.bam' -T
'/tmp/201_651_temp' "+filenames[i]+"'_fixmate.bam'")
    os.system("samtools index "+filenames[i]+"'_sorted.bam'")
    os.system("samtools mpileup -f 'pSH0201.fa' -q 1 -B "+filenames[i]+"'_sorted.bam'
> "+filenames[i]+"'intermediate.mpileup'")


# variant analysis
    os.system("java -jar VarScan.v2.3.9.jar  mpileup2snp
"+filenames[i]+"'intermediate.mpileup' --pileup 1 \
    --min-coverage 1 \
    --min-reads2 1 \
    --variants 1 \
    --p-value 0.99 \
    --min-var-freq 0.0003 \
    --output-vcf 1 > "+filenames[i]+"'snp3.vcf'")
    os.system("bcftools query -f '%REF\t%ALT[\t%FREQ]\t%POS\n'
"+filenames[i]+"'snp3.vcf' > "+filenames[i]+"'snp3.txt'")
    with open(filenames[i]+'snp3.txt', 'r') as infile,
open(filenames[i]+'snp3formatted.txt', 'w') as outfile:
        temp = infile.read().replace("%", "")
        outfile.write(temp)


#plot
freq1=[]
pos1=[]
f = open('JDHS03A_S1_L001_snp3formatted.txt')
reader = csv.reader(f,delimiter='\t')
for row in reader:
       freq1.append(row[2])
       pos1.append(row[3])
freq2=[]
pos2=[]
f = open('JDHS03B_S2_L001_snp3formatted.txt')
reader = csv.reader(f,delimiter='\t')
for row in reader:
       freq2.append(row[2])
       pos2.append(row[3])
freq3=[]
pos3=[]
f = open('JDHS03C_S3_L001_snp3formatted.txt')
reader = csv.reader(f,delimiter='\t')
for row in reader:
       freq3.append(row[2])
       pos3.append(row[3])
freq4=[]
pos4=[]
f = open('JDHS03D_S4_L001_snp3formatted.txt')
```

```
reader = csv.reader(f,delimiter='\t')
for row in reader:
        freq4.append(row[2])
        pos4.append(row[3])
freq5=[]
pos5=[]
f = open('JDHS03E_S5_L001_snp3formatted.txt')
reader = csv.reader(f,delimiter='\t')
for row in reader:
        freq5.append(row[2])
        pos5.append(row[3])
```

```
pos1=map(int,pos1)
pos1[:]=[-1*(x - 583.5) for x in pos1]
pos2=map(int,pos2)
pos2[:]=[-1*(x - 583.5) for x in pos2]
pos3=map(int,pos3)
pos3[:]=[-1*(x - 583.5) for x in pos3]
pos4=map(int,pos4)
pos4[:]=[-1*(x - 583.5) for x in pos4]
pos5=map(int,pos5)
pos5[:]=[-1*(x - 583.5) for x in pos5]


fig1, ax = plt.subplots(figsize=(9,2.7), dpi=100)


fig1=matplotlib.pyplot.scatter(pos2,freq2, color='blue',label = r'$\bf{nCas9-PolI3M
on target replicate 1}$')
fig1=matplotlib.pyplot.scatter(pos4,freq4, color='orange', label = r'$\bf{nCas9-
PolI3M on target replicate 2}$')
fig1=matplotlib.pyplot.scatter(pos5,freq5, color='purple', label = r'$\bf{nCas9-
PolI3M off target}$')
fig1=matplotlib.pyplot.scatter(pos3,freq3, color='green', label = r'$\bf{nCas9 +
PolI3M on target}$')
fig1=matplotlib.pyplot.scatter(pos1,freq1, color='red', label = r'$\bf{nCas9 on
target}$')


ax.set_yscale('log')
plt.ylim([0.04,2])
plt.xlim([-8,59]) #this removes the first and last 5 nucleotides to remove the low
fidelity region of the reads from the analysis
plt.tight_layout()
# matplotlib.pyplot.show()
plt.savefig("NGSfig.eps", format="eps")
```

10.2. Oligonucleotides, gRNAs, plasmids, and EvolvR component amino acid sequences used in this study

Oligonucleotides used in this study

pTarget-F	GCTCTTCCGATCTNNNNNTCGACCCAGCTGTCGGAG
pTarget-R	GCTCTTCCGATCTNNNNNTCACTGTGTGGCTTCAGG
rpsE-F	GCTCTTCCGATCTNNNNNGCTGGCCTTCAGTTCTAAGGTAG
rpsE-R	GCTCTTCCGATCTNNNNNCGTTAATCATATTGCGACGGGC
rpsE-Δ26	GTAACCAAAACCAACGCGACCGTTACCATCGCCAACTACAGTCAGAGCT GTGAAGGAGAAAATACGACCACCTaCGGTTTTAGATACGCG
rpsE-G27D	ACCGTAACCAAAACCAACGCGACCGTTACCATCGCCAACTACAGTCAGA GCTGTGAAGGAGAAAATACGACCAtCTTTAACGGTTTTAGA
rpsE-K23N, Δ24	AACCAACGCGACCGTTACCATCGCCAACTACAGTCAGAGCTGTGAAGGA GAAAATACGACCACCTTTAACGtTAGATACGCGGTTTACCG
rpsE-Δ24	AACCAACGCGACCGTTACCATCGCCAACTACAGTCAGAGCTGTGAAGGA GAAAATACGACCACCTTTAACtTTAGATACGCGGTTTACCG
rpsE-Δ17-19	GTGAAGGAGAAAATACGACCACCTTTAACGGTTTTAGATACGCGGaTCA GCTTTTCCTGCAGTTCGCCAGCTTGTTTTTCGATGTGAGCC
gRNA protospacers used in this study	
nicks 1 nucleotide from nonsense mutation in pTarget	ttgctggccgtacattaata
nicks 11 nucleotides from nonsense mutation in pTarget	acattaatacggctccgcag
nicks 56 nucleotides from nonsense mutation in pTarget	tgatattgatttgctggtta
nicks 347 nucleotides from nonsense mutation in pTarget	tgccttggtaggtccagcgg
Off-target gRNA (targets *dbpA*)	gcatggaaacagttacaggg
Nicks after	GTATCTAAAACCGTTAAAGG

nucleotide position 77 of the genomic *rpsE* gene	
Nicks after nucleotide position 293 of the genomic *rpsL* gene	GTTCGTTACCACACCGTACG
rpsE-gRNA-A	GCTCTGACTGTAGTTGGCGA
rpsE-gRNA-B	AGCAGCGATCCAGAAAGCGA
rpsE-gRNA-C	GGTACCGGTATCATCGCCGG
rpsE-gRNA-D	TTCCACCAACCCGATCAACG
rpsE-gRNA-E	AATCCGTTGAAGAAATTCTG
Plasmids used in this study	
pTarget	aacggatttacggctagctcagtcctaggtacaatgctagcgctagcaa agaggagaaaagatctatgcgctcacgcaactggtcccgcaccttgacc gaacgcagcggtggtaacggcgcagtggcggttttcatggcttgttatg actgtttttttggggtacagtctatgcctcgcgcatccaagcagcaagc gcgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgtt acgcagcagggcagtcgccctaaaacaaagttaaacattatgcgcgaag cggtgatcgccgaagtatcgacccagctgtcggaggtggtgggggtcat tgaacgtcacctggaaccgaccttgctggccgtacattaatacggctcc gcagtggatggcggcctgaagccacacagtgatattgatttgctggtta cggtgaccgtacgccttgatgaaacaacgcgccgcgctttgatcaacga ccttttggaaacttcggcttcccctggagagagcgagattctccgcgct gtagaagtcaccattgttgtgcacgacgacatcattccgtggcgttatc cagctaagcgcgaactgcaatttggagaatggcagcgcaatgacattct tgcgggtatcttcgagccagccacgatcgacattgatctggctatcttg ctgacaaaagcacgcgaacatagcgttgccttggtaggtccagcggcgg aggaactctttgatccggttcctgaacaggatctgtttgaggcgctgaa tgaaaccttaacgctgtggaactcgccgcccgactgggctggcgatgag cgcaatgtagtgcttacgttgtcccgcatttggtacagcgcagtaaccg gcaaaatcgcgccgaaggatgtcgctgccgactgggcaatggagcgcct gccggcccagtatcagcccgtcattcttgaagctcgccaggcttatctt ggacaagaagaagatcgcttggcctcgcgcgcagatcagttggaagaat ttgtccactacgtgaaaggcgagatcaccaaggtagtcggcaaataagg atcctaactcgagctcggtaccaaattccagaaaagaggcctcccgaaa gggggggccttttttcgttttggtccgctgccaatgagacgacggggtca tcacggctcatcatgcgccaaacaaatgtgtgcaatacacgctcggatg

	actgcatgatgaccgcactgactggggacagcagatccacctaagcctg tgagagaagcagacacccgacagatcaaggcagttaactagtgcactgc agtacagcggccgctagcggagtgtatactggcttactatgttggcact gatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgc accggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcg ctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatgg cttacgaacggggcggagatttcctggaagatgccaggaagatacttaa cagggaagtgagagggccgcggcaaagccgttttccataggctccgcc cccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaa cccgacaggactataaagataccaggcgtttccccctggcggctccctc gtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgt tatggccgcgtttgtctcattccacgcctgacactcagttccgggtagg cagttcgctccaagctggactgtatgcacgaacccccgttcagtccga ccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaaga catgcaaaagcaccactggcagcagccactggtaattgatttagaggag ttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagtt ttggtgactgcgctcctccaagccagttacctcggttcaaagagttggt agctcagagaaccttcgaaaaaccgccctgcaaggcggtttttttcgttt tcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatc ttattaatcagataaaatatttttgacagctagctcagtcctaggtata atgctagcgttaataaggagaaggaaaagtatgggtcaaagtagcgatg aagccaacgctcccgttgcagggcagtttgcgcttcccctgagtgccac ctttggcttaggggatcgcgtacgcaagaaatctggtgccgcttggcag ggtcaagtcgtcggttggtattgcacaaaactcactcctgaaggctatg cggtagagtccgaatcccacccaggctcagtgcaaatttatcctgtggc tgcacttgaacgtgtggcctaaactgctcccggtcacactagaaaataa aaacagccgtgcctctttctggcaacggctgttttatttattcaacct taattaaggacagctattgtagataagtggggatactacgggtataaaa aaacccgccggggcgggtttttttacgttgcttcagagcggccgcccct gaattcgcatctagatggtagagccacaaacagccggtacaagcaacga tctccaggaccatctgaatcatgcgcggatgacacgaactcacgacggc gatcacagacattaacccacagtacagacactgcgacaacgtggcaatt cgtcgcaataccgtctcactgaactggccgataattgcagacg BBa_J23110 promoter-*aadA* with a chain-terminating codon at position 106-L3S2P21 terminator-p15a origin of replication-Trimethoprim resistance cassette
pTarget2	aacggatttacggctagctcagtcctaggtacaatgctagcgctagcaa agaggagaaaagatctatgcgctcacgcaactggtcccgcaccttgacc gaacgcagcggtggtaacggcgcagtggcggttttcatggcttgttatg actgtttttttggggtacagtctatgcctcgcgcatccaagcagcaagc gcgttacgccgtgggtcgatgtttgatgttatggagcagcaacgatgtt acgcagcagggcagtcgccctaaaacaaagttaaacattatgcgcgaag cggtgatcgccgaagtatcgacccagctgtcggaggtggtggggtcat tgaacgtcacctgtaaccgaccttgctggccgtacattaatacggctcc gcagtggatggcggcctgaagccacacagtgatattgatttgctggtta cggtgaccgtacgccttgatgaaacaacgcgccgcgctttgatcaacga ccttttggaaacttcggcttcccctggagagagcgagattctccgcgct

	gtagaagtcaccattgttgtgcacgacgacatcattccgtggcgttatc cagctaagcgcgaactgcaatttggagaatggcagcgcaatgacattct tgcgggtatcttcgagccagccacgatcgacattgatctggctatcttg ctgacaaaagcacgcgaacatagcgttgccttggtaggtccagcggcgg aggaactctttgatccggttcctgaacaggatctgtttgaggcgctgaa tgaaaccttaacgctgtggaactcgccgcccgactgggctggcgatgag cgcaatgtagtgcttacgttgtcccgcatttggtacagcgcagtaaccg gcaaaatcgcgccgaaggatgtcgctgccgactgggcaatggagcgcct gccggcccagtatcagcccgtcattcttgaagctcgccaggcttatctt ggacaagaagaagatcgcttggcctcgcgcgcagatcagttggaagaat ttgtccactacgtgaaaggcgagatcaccaaggtagtcggcaaataagg atcctaactcgagctcggtaccaaattccagaaaagaggcctcccgaaa gggggggccttttttcgttttggtccgctgccaatgagacgacggggtca tcacggctcatcatgcgccaaacaaatgtgtgcaatacacgctcggatg actgcatgatgaccgcactgactggggacagcagatccacctaagcctg tgagagaagcagacacccgacagatcaaggcagttaactagtgcactgc agtacagcggccgctagcggagtgtatactggcttactatgttggcact gatgagggtgtcagtgaagtgcttcatgtggcaggagaaaaaaggctgc accggtgcgtcagcagaatatgtgatacaggatatattccgcttcctcg ctcactgactcgctacgctcggtcgttcgactgcggcgagcggaaatgg cttacgaacggggcggagatttcctggaagatgccaggaagatacttaa cagggaagtgagagggccgcggcaaagccgttttttccataggctccgcc cccctgacaagcatcacgaaatctgacgctcaaatcagtggtggcgaaa cccgacaggactataaagataccaggcgtttccccctggcggctccctc gtgcgctctcctgttcctgcctttcggtttaccggtgtcattccgctgt tatggccgcgtttgtctcattccacgcctgacactcagttccgggtagg cagttcgctccaagctggactgtatgcacgaaccccccgttcagtccga ccgctgcgccttatccggtaactatcgtcttgagtccaacccggaaaga catgcaaaagcaccactggcagcagccactggtaattgatttagaggag ttagtcttgaagtcatgcgccggttaaggctaaactgaaaggacaagtt ttggtgactgcgctcctccaagccagttacctcggttcaaagagttggt agctcagagaaccttcgaaaaaccgccctgcaaggcggtttttttcgttt tcagagcaagagattacgcgcagaccaaaacgatctcaagaagatcatc ttattaatcagataaaatatttttgacagctagctcagtcctaggtata atgctagcgttaataaggagaaggaaaagtatgggtcaaagtagcgatg aagccaacgctcccgttgcagggcagtttgcgcttcccctgagtgccac ctttggcttaggggatcgcgtacgcaagaaatctggtgccgcttggcag ggtcaagtcgtcggttggtattgcacaaaactcactcctgaaggctatg cggtagagtccgaatcccacccaggctcagtgcaaatttatcctgtggc tgcacttgaacgtgtggcctaaactgctcccggtcacactagaaaataa aaacagccgtgcctctttctggcaacggctgttttatttattcaacct taattaaggacagctattgtagataagtggggatactacgggtataaaa aaacccgccggggcgggtttttttacgttgcttcagagcggccgccct gaattcgcatctagatggtagagccacaaacagccggtacaagcaacga tctccaggaccatctgaatcatgcgcggatgacacgaactcacgacggc gatcacagacattaacccacagtacagacactgcgacaacgtggcaatt cgtcgcaataccgtctcactgaactggccgataattgcagacg

	BBa_J23110 promoter-*aadA* with a chain-terminating codon at positions 98 and 106-L3S2P21 terminator-p15a origin of replication-Trimethoprim resistance cassette
pEvolvR	tttatagctagctcagcccttggtacaatgctagc [insert 20 nucleotide gRNA protospacer] gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaa cttgaaaaagtggcaccgagtcggtgcttttttt gaagcttgggcccga acaaaaactcatctcagaagaggatctgaatagcgccgtcgaccatcat catcatcatcattgagtttaaacggactccagcttggctgttttggcgg atgagagaagattttcagcctgatacagattaaatcagaacgcagaagc ggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccac ctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtag tgtggggactccccatgcgagagtagggaactgccaggcatcaaataaa acgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttg tcggtgaacgacagg gctgccaaaccagatgtcaacacagctacaacgt taagacccactttcacatttaagttgtttttctaatccgcatatgatca attcaaggccgaataagaaggctggctctgcaccttggtgatcaaataa ttcgatagcttgtcgtaataatggcggcatactatcagtagtaggtgtt tccctttcttctttagcgacttgatgctcttgatcttccaatacgcaac ctaaagtaaaatgccccacagcgctgagtgcatataatgcattctctag tgaaaaaccttgttggcataaaaaggctaattgattttcgagagtttca tactgtttttctgtaggccgtgtacctaaatgtacttttgctccatcgc gatgacttagtaaagcacatctaaaacttttagcgttattacgtaaaaa atcttgccagctttccccttctaaagggcaaaagtgagtatggtgccta tctaacatctcaatggctaaggcgtcgagcaaagcccgcttatttttta catgccaatacaatgtaggctgctctacacctagcttctgggcgagttt acgggttgttaaaccttcgattccgacctcattaagcagctctaatgcg ctgttaatcactttacttttatctaatctagacatcattaattcctaat ttttgttgacactctatcgttgatagagttatttaccactccctatca gtgatagagaaaagaattcaaaagatctaaagaggagaaaagatctatg [Insert Cas9 variant] tggttctagtgaaaccccgggaacaagtgagtcggccacccctgaaggt ggatcaggggtagcggatcc [Insert DNA Polymerase variant] taataatggcctcggtaccaaagacgaacaataagacgctgaaaagcgt ctttttttcgttttggtccgctgagcagttacagagatgttacgaaccac tagtgcactgcagtacagtgttacaaccaattaaccaattctgattaga aaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggatt atcaataccatattttgaaaaagccgtttctgtaatgaaggagaaaac tcaccgaggcagttccataggatggcaagatcctggtatcggtctgcga ttccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaa aataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggt gagaatggcaaaagcttatgcatttctttccagacttgttcaacaggcc agccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttatt cattcgtgattgcgcctgagcgaggcgaaatacgcgatcgctgttaaaa ggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgcca gcgcatcaacaatattttcacctgaatcaggatattcttctaatacctg

	gaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatca ggagtacggataaaatgcttgatggtcggaagaggcataaattccgtca gccagtttagtctgaccatctcatctgtaacatcattggcaacgctacc tttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaat cgatagattgtcgcacctgattgcccgacattatcgcgagcccatttat acccatataaatcagcatccatgttggaatttaatcgcggcctggagca agacgtttcccgttgaatatggctcataacacccttgtattactgttt atgtaagcagacagttttattgttcatgatgatatattttatcttgtg caatgtaacatcagagattttgagacacaacgtggctttgttgaataaa tcgaactttgctgagttgaaggatcaggagttcatgaccaaaatccct taacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatca aaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgca aacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagag ctaccaactctttttccgaaggtaactggcttcagcagagcgcagatac caaatactgtccttctagtgtagccgtagttaggccaccacttcaagaa ctctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtg gctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagac gatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtg cacacagcccagcttggagcgaacgacctacaccgaactgagataccta cagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcgg acaggtatccggtaagcggcagggtcggaacaggagagcgcacgaggga gcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgc cacctctgacttgagcgtcgattttgtgatgctcgtcaggggggcgga gcctatggaaaaacgccagcaacgcggccttttacggttcctggcctt ttgctggccttttgctcacatgttctttcctgcgttatcccctgattct gtggataaccgtattaccgcctttgagtgagctgataccgctcgccgca gccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcg cctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgc aatggtgcactctcagtacaatctgctctgatgccgcatagttaagcca gtatacactccgctatcgctacgtgactgggtcatggctgcgccccgac acccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggc atccgcttacagacaagctgtgaccttctccgggagctgcatgtgtcag aggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagct catcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgc gtccagctcgttgagtttctccagaagcgttaatgtctggcttctgata aagcgggccatgttaagggcggtttttcctgtttggtcactgatgcct ccgtgtaaggggatttctgttcatccctgaattcgcatctagactgaa ctggccgataattgcagacgaacggatttactttacagctagctcagtc ctaggtattatgctagcgctagcaaagaggagaaaagatctatggagaa aaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaa gaacattttgaggcatttcagtcagttgctcaatgtacctataaccaga ccgttcagctggatattacggccttttaaagaccgtaaagaaaaataa gcacaagttttaaccggcctttattcacattcttgcccgcctgatgaat gctcatccggaatttcgtatggcaatgaaagacggtgagctggtgatat gggatagtgttcacccttgttacaccgttttccatgagcaaactgaaac gttttcatcgctctggagtgaataccacgacgatttccggcagtttcta cacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatt tccctaaagggtttattgagaatatgttttcgtttcagccaatccctg

	ggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttc ttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaagg tgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggctt ccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtgg cagggcggggcgtaaggatcctaactcgagctcggtaccaaattccaga aaagaggcctcccgaaaggggggccttttttcgttttggtccgctgcca aaccagatgtcaacacagctacaacg J23115 promoter-[Insert 20 nucleotide gRNA protospacer]-Single gRNA body-TrrnB terminator-TetR, pTet, BBa_B0034 RBS-[Insert Cas9 variant]-23 amino acid linker-[Insert DNA Polymerase variant]-L3S2P55 terminator-Kanamycin resistance cassette-pBR322 origin of replication
pEvolvR-Phi29	tttatagctagctcagcccttggtacaatgctagc [insert 20 nucleotide gRNA protospacer] gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaa cttgaaaaagtggcaccgagtcggtgcttttttt gaagcttgggcccga acaaaaactcatctcagaagaggatctgaatagcgccgtcgaccatcat catcatcatcattgagtttaaacggactccagcttggctgttttggcgg atgagagaagattttcagcctgatacagattaaatcagaacgcagaagc ggtctgataaaacagaatttgcctggcggcagtagcgcggtggtcccac ctgaccccatgccgaactcagaagtgaaacgccgtagcgccgatggtag tgtggggactccccatgcgagagtagggaactgccaggcatcaaataaa acgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttg tcggtgaacgacagg gctgccaaaccagatgtcaacacagctacaacgt taagacccactttcacatttaagttgtttttctaatccgcatatgatca attcaaggccgaataagaaggctggctctgcaccttggtgatcaaataa ttcgatagcttgtcgtaataatggcggcatactatcagtagtaggtgtt tccctttcttctttagcgacttgatgctcttgatcttccaatacgcaac ctaaagtaaaatgccccacagcgctgagtgcatataatgcattctctag tgaaaaaccttgttggcataaaaaggctaattgattttcgagagtttca tactgtttttctgtaggccgtgtacctaaatgtacttttgctccatcgc gatgacttagtaaagcacatctaaaacttttagcgttattacgtaaaaa atcttgccagctttccccttctaaagggcaaaagtgagtatggtgccta tctaacatctcaatggctaaggcgtcgagcaaagcccgcttatttttta catgccaatacaatgtaggctgctctacacctagcttctgggcgagttt acgggttgttaaaccttcgattccgacctcattaagcagctctaatgcg ctgttaatcactttacttttatctaatctagacatcattaattcctaat ttttgttgacactctatcgttgatagagttatttaccactccctatca gtgatagagaaaagaattcaaaagatctaaagaggagaaaagatctatg [Insert Phi29 variant] ggttctagtgaaaccccgggaacaagtgagtcggccacccctgaaggtg gatcaggggtagcggatccgttcagatcccgcagaacccgctgattct ggttgacggatctagttacctgtaccgtgcttaccatgctttcccgcct ttgaccaattctgctggtgaacctacgggagctatgtacggagttctga atatgttgcgttctttaattatgcagtacaagcctacccacgctgctgt tgttttcgatgctaaaggtaagacgttccgcgacgagttattcgagcac tataagtctcaccgtcctccgatgcctgatgacttacgcgctcagattg agccgctgcatgctatggtgaaggctatgggtttacctcttttggctgt

	cagcggtgttgaggctgatgatgtcattggcaccttagctcgtgaggct gagaaggctggtcgccctgttttgatttctaccggtgacaaggacatgg ctcaattggttaccccgaacatcaccctgatcaacaccatgaccaacac gattctgggtcctgaggaagttgttaacaaatatggtgttcctccggag ttgattattgactttcttgctcttatgggcgattcttcagacaatatcc cgggtgttccaggtgttggagagaagactgctcaagctctgcttcaggg tctgggtggtttggacacccttacgctgaaccggagaagatcgccggt ctgtcttttcgcggtgctaagaccatggctgctaaactggaacagaata aggaggtcgcatacctgtcttatcaattggctaccatcaagacggatgt ggagttagaacttacgtgcgagcagcttgaggttcaacagcctgctgct gaggaactgctgggtcttttaagaaatacgaatttaagcgttggaccg ccgacgttgaggctggtaagtggctgcaagctaagggtgctaagccggc tgctaaaccgcaagaaacgagtgtcgctgatgaggctccggaggttacc gctaccgttatcGGTTCTagtgaaaccccgggaacaagtgagtcggcca cccctgaaggtggatcaggggtagcGGATCCgacaagaagtattctat cggactggctatcgggactaatagcgtcgggtgggccgtgatcactgac gagtacaaggtgccctctaagaagttcaaggtgctcgggaacaccgacc ggcattccatcaagaaaaatctgatcggagctctcctctttgattcagg ggaaaccgctgaagcaacccgcctcaagcggactgctagacggcggtac accaggaggaagaaccggatttgttaccttcaagagatattctccaacg aaatggcaaaggtcgacgacagcttcttccataggctggaagaatcatt cctcgtggaagaggataagaagcatgaacggcatcccatcttcggtaat atcgtcgacgaggtggcctatcacgagaaatacccaaccatctaccatc ttcgcaaaaagctggtggactcaaccgacaaggcagacctccggcttat ctacctggccctggcccacatgatcaagttcagaggccacttcctgatc gagggcgacctcaatcctgacaatagcgatgtggataaactgttcatcc agctggtgcagacttacaaccagctctttgaagagaaccccatcaatgc aagcggagtcgatgccaaggccattctgtcagcccggctgtcaaagagc cgcagacttgagaatcttatcgctcagctgccgggtgaaaagaaaaatg gactgttcgggaacctgattgctctttcacttgggctgactcccaattt caagtctaatttcgacctggcagaggatgccaagctgcaactgtccaag gacacctatgatgacgatctcgacaacctcctggcccagatcggtgacc aatacgccgacctttttccttgctgctaagaatctttctgacgccatcct gctgtctgacattctccgcgtgaacactgaaatcaccaaggcccctctt tcagcttcaatgattaagcggtatgatgagcaccaccaggacctgaccc tgcttaaggcactcgtccggcagcagcttccggagaagtacaaggaaat cttctttgaccagtcaaagaatggatacgccggctacatcgacggaggt gcctcccaagaggaattttataagtttatcaaacctatccttgagaaga tggacggcaccgaagagctcctcgtgaaactgaatcgggaggatctgct gcggaagcagcgcactttcgacaatgggagcattccccaccagatccat cttggggagcttcacgccatccttcggcgccaagaggacttctacccct ttcttaaggacaacagggagaagattgagaaaattctcactttccgcat cccctactacgtgggacccctcgccagaggaaatagccggtttgcttgg atgaccagaaagtcagaagaaactatcactccctggaacttcgaagagg tggtggacaagggagccagcgctcagtcattcatcgaacggatgactaa cttcgataagaacctccccaatgagaaggtcctgccgaaacattccctg ctctacgagtactttaccgtgtacaacgagctgaccaaggtgaaatatg tcaccgaagggatgaggaagcccgcattcctgtcaggcgaacaaaagaa

	ggcaattgtggaccttctgttcaagaccaatagaaaggtgaccgtgaag cagctgaaggaggactatttcaagaaaattgaatgcttcgactctgtgg agattagcggggtcgaagatcggttcaacgcaagcctgggtacttacca tgatctgcttaagatcatcaaggacaaggattttctggacaatgaggag aacgaggacatccttgaggacattgtcctgactctcactctgttcgagg accgggaaatgatcgaggagaggcttaagacctacgcccatctgttcga cgataaagtgatgaagcaacttaaacggagaagatataccggatgggga cgccttagccgcaaactcatcaacggaatccgggacaaacagagcggaa agaccattcttgatttccttaagagcgacggattcgctaatcgcaactt catgcaacttatccatgatgattccctgacctttaaggaggacatccag aaggcccaagtgtctggacaaggtgactcactgcacgagcatatcgcaa atctggctggttcacccgctattaagaagggtattctccagaccgtgaa agtcgtggacgagctggtcaaggtgatgggtcgccataaaccagagaac attgtcatcgagatggccagggaaaaccagactacccagaagggacaga agaacagcagggagcggatgaaaagaattgaggaagggattaaggagct cgggtcacagatccttaaagagcacccggtggaaaacacccagcttcag aatgagaagctctatctgtactaccttcaaaatggacgcgatatgtatg tggaccaagagcttgatatcaacaggctctcagactacgacgtggacca tatcgtccctcagagcttcctcgcagacgactcaattgacaataaggtg ctgactcgctcagacaagaaccggggaaagtcagataacgtgccctcag aggaagtcgtgaaaaagatgaagaactattggcgccagcttctgaacgc aaagctgatcactcagcggaagttcgacaatctcactaaggctgagagg ggcggactgagcgaactggacaaagcaggattcattaaacggcaacttg tggagactcggcagattactaaacatgtcgcccaaatccttgactcacg catgaataccaagtacgacgaaaacgacaaacttatccgcgaggtgaag gtgattaccctgaagtccaagctggtcagcgatttcagaaaggactttc aattctacaaagtgcgggagatcaataactatcatcatgctcatgacgc atatctgaatgccgtggtgggaaccgccctgatcaagaagtacccagca ctggaaagcgagttcgtgtacggagactacaaggtctacgacgtgcgca agatgattgccaaatctgagcaggagatcggaaaggccaccgcaaagta cttcttctacagcaacatcatgaatttcttcaagaccgaaatcaccctt gcaaacggtgagatccggaaggcgccgctcatcgagactaatggggaga ctggcgaaatcgtgtgggacaagggcagagatttcgctaccgtgcgcaa agtgctttctatgcctcaagtgaacatcgtgaagaaaaccgaggtgcaa accggaggcttttctaaggaatcaatcctccccaagcgcaactccgaca agctcattgcaaggaagaaggattgggaccctaagaagtacggcggatt cgattcaccaactgtggcttattctgtcctggtcgtggctaaggtggaa aaaggaaagtctaagaagctcaagagcgtgaaggaactgctgggtatca ccattatggagcgcagctccttcgagaagaacccaattgactttctcga agccaaaggttacaaggaagtcaagaaggaccttatcatcaagctccca aagtatagcctgttcgaactggagaatgggcggaagcggatgctcgcct ccgctggcgaacttcagaagggtaatgagctggctctcccctccaagta cgtgaatttcctctaccttgcaagccattacgagaagctgaaggggagc cccgaggacaacgagcaaaagcaactgtttgtggagcagcataagcatt atctggacgagatcattgagcagatttccgagttttctaaacgcgtcat tctcgctgatgccaacctcgataaagtccttagcgcatacaataagcac agagacaaaccaattcgggagcaggctgagaatatcatccacctgttca ccctcaccaatcttggtgcccctgccgcattcaagtacttcgacaccac

	catcgaccggaaacgctatacctccaccaaagaagtgctggacgccacc ctcatccaccagagcatcaccggactttacgaaactcggattgacctct cacagctcggaggggattaatggcctcggtaccaaagacgaacaataag acgctgaaaagcgtctttttcgttttggtccgctgagcagttacagag atgttacgaaccactagtgcactgcagtacagtgttacaaccaattaac caattctgattagaaaaactcatcgagcatcaaatgaaactgcaattta ttcatatcaggattatcaataccatattttgaaaaagccgtttctgta atgaaggagaaaactcaccgaggcagttccataggatggcaagatcctg gtatcggtctgcgattccgactcgtccaacatcaatacaacctattaat ttcccctcgtcaaaaataaggttatcaagtgagaaatcaccatgagtga cgactgaatccggtgagaatggcaaaagcttatgcatttctttccagac ttgttcaacaggccagccattacgctcgtcatcaaaatcactcgcatca accaaaccgttattcattcgtgattgcgcctgagcgaggcgaaatacgc gatcgctgttaaaaggacaattacaaacaggaatcgaatgcaaccggcg caggaacactgccagcgcatcaacaatattttcacctgaatcaggatat tcttctaatacctggaatgctgttttcccggggatcgcagtggtgagta accatgcatcatcaggagtacggataaaatgcttgatggtcggaagagg cataaattccgtcagccagtttagtctgaccatctcatctgtaacatca ttggcaacgctacctttgccatgtttcagaaacaactctggcgcatcgg gcttcccatacaatcgatagattgtcgcacctgattgcccgacattatc gcgagcccatttatacccatataaatcagcatccatgttggaatttaat cgcggcctggagcaagacgtttcccgttgaatatggctcataacacccc ttgtattactgtttatgtaagcagacagttttattgttcatgatgatat atttttatcttgtgcaatgtaacatcagagattttgagacacaacgtgg ctttgttgaataaatcgaacttttgctgagttgaaggatcaggagttca tgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccc cgtagaaaagatcaaaggatcttcttgagatccttttttctgcgcgta atctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtt tgccggatcaagagctaccaactcttttccgaaggtaactggcttcag cagagcgcagataccaaatactgtccttctagtgtagccgtagttaggc caccacttcaagaactctgtagcaccgcctacatacctcgctctgctaa tcctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgg gttggactcaagacgatagttaccggataaggcgcagcggtcgggctga acgggggttcgtgcacacagcccagcttggagcgaacgacctacaccg aactgagatacctacagcgtgagctatgagaaagcgccacgcttcccga agggagaaaggcggacaggtatccggtaagcggcagggtcggaacagga gagcgcacgagggagcttccagggggaaacgcctggtatctttatagtc ctgtcgggtttcgccacctctgacttgagcgtcgattttgtgatgctc gtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttta cggttcctggccttttgctggccttttgctcacatgttctttcctgcgt tatcccctgattctgtggataaccgtattaccgcctttgagtgagctga taccgctcgccgcagccgaacgaccgagcgcagcgagtcagtgagcgag gaagcggaagagcgcctgatgcggtattttctccttacgcatctgtgcg gtatttcacaccgcaatggtgcactctcagtacaatctgctctgatgcc gcatagttaagccagtatacactccgctatcgctacgtgactgggtcat ggctgcgccccgacacccgccaacacccgctgacgcgccctgacgggct tgtctgctcccggcatccgcttacagacaagctgtgaccttctccggga gctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgaggca

	gctgcggtaaagctcatcagcgtggtcgtgaagcgattcacagatgtct gcctgttcatccgcgtccagctcgttgagtttctccagaagcgttaatg tctggcttctgataaagcgggccatgttaagggcggtttttcctgttt ggtcactgatgcctccgtgtaagggggatttctgttcatccctgaattc gcatctagactgaactggccgataattgcagacgaacggatttacttta cagctagctcagtcctaggtattatgctagcgctagcaaagaggagaaa agatctatggagaaaaaaatcactggatataccaccgttgatatatccc aatggcatcgtaaagaacattttgaggcatttcagtcagttgctcaatg tacctataaccagaccgttcagctggatattacggccttttaaagacc gtaaagaaaaataagcacaagtttaaccggcctttattcacattcttg cccgcctgatgaatgctcatccggaatttcgtatggcaatgaaagacgg tgagctggtgatatgggatagtgttcacccttgttacaccgttttccat gagcaaactgaaacgttttcatcgctctggagtgaataccacgacgatt tccggcagtttctacacatatattcgcaagatgtggcgtgttacggtga aaacctggcctatttccctaaagggtttattgagaatatgttttcgtt tcagccaatccctgggtgagtttcaccagttttgatttaaacgtggcca atatggacaacttcttcgcccccgttttcaccatgggcaaatattatac gcaaggcgacaaggtgctgatgccgctggcgattcaggttcatcatgcc gtttgtgatggcttccatgtcggcagaatgcttaatgaattacaacagt actgcgatgagtggcagggcggggcgtaaggatcctaactcgagctcgg taccaaattccagaaaagaggcctcccgaaaggggggccttttttcgtt ttggtccgctgccaaaccagatgtcaacacagctacaacg J23115 promoter-[Insert 20 nucleotide gRNA protospacer]-Single gRNA body-TrrnB terminator-TetR, pTet, BBa_B0034 RBS-[Insert Phi29 variant]-PolI(1-324)-enCas9-L3S2P55 terminator-Kanamycin resistance cassette-pBR322 origin of replication
pEvolvR expressing two gRNAs	tttatagctagctcagcccttggtacaatgctagc [insert 20 nucleotide gRNA protospacer] gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaa cttgaaaaagtggcaccgagtcggtgctttttttctcggtaccaaattc cagaaaagaggcctcccgaaaggggggccttttttcgttttggtccgct gccaaaccagatgtcaacacagctacaacgtttatagctagctcagccc ttggtacaatgctagc [insert 20 nucleotide gRNA protospacer] gttttagagctagaaatagcaagttaaaataaggctagtccgttatcaa cttgaaaaagtggcaccgagtcggtgctttttttt gaagcttgggcccgaacaaaaactcatctcagaagaggatctgaatagc gccgtcgaccatcatcatcatcatcattgagtttaaacggactccagct tggctgttttggcggatgagagaagattttcagcctgatacagattaaa tcagaacgcagaagcggtctgataaaacagaatttgcctggcggcagta gcgcggtggtcccacctgaccccatgccgaactcagaagtgaaacgccg tagcgccgatggtagtgtggggactccccatgcgagagtagggaactgc caggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgt tttatctgttgtttgtcggtgaacgacaggctgccaaaccagatgtca acacagctacaacgttaagacccactttcacatttaagttgttttctta atccgcatatgatcaattcaaggccgaataagaaggctggctctgcacc ttggtgatcaaataattcgatagcttgtcgtaataatggcggcatacta

	tcagtagtaggtgtttccctttcttctttagcgacttgatgctcttgat cttccaatacgcaacctaaagtaaaatgccccacagcgctgagtgcata taatgcattctctagtgaaaaaccttgttggcataaaaaggctaattga ttttcgagagtttcatactgtttttctgtaggccgtgtacctaaatgta cttttgctccatcgcgatgacttagtaaagcacatctaaaacttttagc gttattacgtaaaaaatcttgccagctttccccttctaaagggcaaaag tgagtatggtgcctatctaacatctcaatggctaaggcgtcgagcaaag cccgcttatttttacatgccaatacaatgtaggctgctctacacctag cttctgggcgagtttacgggttgttaaaccttcgattccgacctcatta agcagctctaatgcgctgttaatcactttacttttatctaatctagaca tcattaattcctaatttttgttgacactctatcgttgatagagttattt taccactccctatcagtgatagagaaaagaattcaaaagatctaaagag gagaaaagatctatg [Insert Cas9 variant] tggttctagtgaaaccccgggaacaagtgagtcggccaccctgaaggt ggatcaggggtagcggatcc [Insert DNA Polymerase variant] taataatggcctcggtaccaaagacgaacaataagacgctgaaaagcgt ctttttttcgttttggtccgctgagcagttacagagatgttacgaaccac tagtgcactgcagtacagtgttacaaccaattaaccaattctgattaga aaaactcatcgagcatcaaatgaaactgcaatttattcatatcaggatt atcaataccatattttgaaaaagccgtttctgtaatgaaggagaaaac tcaccgaggcagttccataggatggcaagatcctggtatcggtctgcga ttccgactcgtccaacatcaatacaacctattaatttcccctcgtcaaa aataaggttatcaagtgagaaatcaccatgagtgacgactgaatccggt gagaatggcaaaagcttatgcatttctttccagacttgttcaacaggcc agccattacgctcgtcatcaaaatcactcgcatcaaccaaaccgttatt cattcgtgattgcgcctgagcgaggcgaaatacgcgatcgctgttaaaa ggacaattacaaacaggaatcgaatgcaaccggcgcaggaacactgcca gcgcatcaacaatattttcacctgaatcaggatattcttctaatacctg gaatgctgttttcccggggatcgcagtggtgagtaaccatgcatcatca ggagtacggataaaatgcttgatggtcggaagaggcataaattccgtca gccagtttagtctgaccatctcatctgtaacatcattggcaacgctacc tttgccatgtttcagaaacaactctggcgcatcgggcttcccatacaat cgatagattgtcgcacctgattgcccgacattatcgcgagcccatttat acccatataaatcagcatccatgttggaatttaatcgcggcctggagca agacgtttcccgttgaatatggctcataacacccttgtattactgttt atgtaagcagacagttttattgttcatgatgatatattttatcttgtg caatgtaacatcagagatttgagacacaacgtggctttgttgaataaa tcgaactttgctgagttgaaggatcaggagttcatgaccaaaatccct taacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatca aaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgca aacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagag ctaccaactcttttccgaaggtaactggcttcagcagagcgcagatac caaatactgtccttctagtgtagccgtagttaggccaccacttcaagaa ctctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtg gctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagac gatagttaccggataaggcgcagcggtcgggctgaacgggggttcgtg cacacagcccagcttggagcgaacgacctacaccgaactgagataccta

	cagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcgg acaggtatccggtaagcggcagggtcggaacaggagagcgcacgaggga gcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgc cacctctgacttgagcgtcgattttgtgatgctcgtcaggggggcgga gcctatggaaaaacgccagcaacgcggccttttacggttcctggcctt ttgctggccttttgctcacatgttctttcctgcgttatcccctgattct gtggataaccgtattaccgcctttgagtgagctgataccgctcgccgca gccgaacgaccgagcgcagcgagtcagtgagcgaggaagcggaagagcg cctgatgcggtattttctccttacgcatctgtgcggtatttcacaccgc aatggtgcactctcagtacaatctgctctgatgccgcatagttaagcca gtatacactccgctatcgctacgtgactgggtcatggctgcgccccgac acccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggc atccgcttacagacaagctgtgaccttctccgggagctgcatgtgtcag aggttttcaccgtcatcaccgaaacgcgcgaggcagctgcggtaaagct catcagcgtggtcgtgaagcgattcacagatgtctgcctgttcatccgc gtccagctcgttgagtttctccagaagcgttaatgtctggcttctgata aagcgggccatgttaagggcggtttttcctgtttggtcactgatgcct ccgtgtaaggggatttctgttcatccctgaattcgcatctagactgaa ctggccgataattgcagacgaacggatttactttacagctagctcagtc ctaggtattatgctagcgctagcaaagaggagaaaagatctatggagaa aaaaatcactggatataccaccgttgatatatcccaatggcatcgtaaa gaacattttgaggcatttcagtcagttgctcaatgtacctataaccaga ccgttcagctggatattacggccttttaaagaccgtaaagaaaaataa gcacaagtttaaccggcctttattcacattcttgcccgcctgatgaat gctcatccggaatttcgtatggcaatgaaagacggtgagctggtgatat gggatagtgttcacccttgttacaccgttttccatgagcaaactgaaac gttttcatcgctctggagtgaataccacgacgatttccggcagtttcta cacatatattcgcaagatgtggcgtgttacggtgaaaacctggcctatt tccctaaagggtttattgagaatatgttttcgtttcagccaatccctg ggtgagtttcaccagttttgatttaaacgtggccaatatggacaacttc ttcgcccccgttttcaccatgggcaaatattatacgcaaggcgacaagg tgctgatgccgctggcgattcaggttcatcatgccgtttgtgatggctt ccatgtcggcagaatgcttaatgaattacaacagtactgcgatgagtgg cagggcggggcgtaaggatcctaactcgagctcggtaccaaattccaga aaagaggcctcccgaaaggggggccttttttcgttttggtccgctgcca aaccagatgtcaacacagctacaacg J23115 promoter-[Insert 20 nucleotide gRNA protospacer]-Single gRNA body-TrrnB terminator-TetR, pTet, BBa_B0034 RBS-[Insert Cas9 variant]-23 amino acid linker-[Insert DNA Polymerase variant]-L3S2P55 terminator-Kanamycin resistance cassette-pBR322 origin of replication
Amino acid sequences of EvolvR parts used in this study	
nCas9	DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE

	ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLKDDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY HHAHDAYLNAVVGTALIKKYPKLESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKRPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
enCas9	DKKYSIGLAIGTNSVGWAVITDEYKVPSKKFKVLGNTDRHSIKKNLIGA LLFDSGETAEATRLKRTARRRYTRRKNRICYLQEIFSNEMAKVDDSFFH RLEESFLVEEDKKHERHPIFGNIVDEVAYHEKYPTIYHLRKKLVDSTDK ADLRLIYLALAHMIKFRGHFLIEGDLNPDNSDVDKLFIQLVQTYNQLFE ENPINASGVDAKAILSARLSKSRRLENLIAQLPGEKKNGLFGNLIALSL GLTPNFKSNFDLAEDAKLQLSKDTYDDDLDNLLAQIGDQYADLFLAAKN LSDAILLSDILRVNTEITKAPLSASMIKRYDEHHQDLTLLKALVRQQLP EKYKEIFFDQSKNGYAGYIDGGASQEEFYKFIKPILEKMDGTEELLVKL NREDLLRKQRTFDNGSIPHQIHLGELHAILRRQEDFYPFLKDNREKIEK ILTFRIPYYVGPLARGNSRFAWMTRKSEETITPWNFEEVVDKGASAQSF IERMTNFDKNLPNEKVLPKHSLLYEYFTVYNELTKVKYVTEGMRKPAFL SGEQKKAIVDLLFKTNRKVTVKQLKEDYFKKIECFDSVEISGVEDRFNA SLGTYHDLLKIIKDKDFLDNEENEDILEDIVLTLTLFEDREMIEERLKT YAHLFDDKVMKQLKRRRYTGWGRLSRKLINGIRDKQSGKTILDFLKSDG FANRNFMQLIHDDSLTFKEDIQKAQVSGQGDSLHEHIANLAGSPAIKKG ILQTVKVVDELVKVMGRHKPENIVIEMARENQTTQKGQKNSRERMKRIE EGIKELGSQILKEHPVENTQLQNEKLYLYYLQNGRDMYVDQELDINRLS DYDVDHIVPQSFLADDSIDNKVLTRSDKNRGKSDNVPSEEVVKKMKNYW RQLLNAKLITQRKFDNLTKAERGGLSELDKAGFIKRQLVETRQITKHVA QILDSRMNTKYDENDKLIREVKVITLKSKLVSDFRKDFQFYKVREINNY HHAHDAYLNAVVGTALIKKYPALESEFVYGDYKVYDVRKMIAKSEQEIG KATAKYFFYSNIMNFFKTEITLANGEIRKAPLIETNGETGEIVWDKGRD FATVRKVLSMPQVNIVKKTEVQTGGFSKESILPKRNSDKLIARKKDWDP

	KKYGGFDSPTVAYSVLVVAKVEKGKSKKLKSVKELLGITIMERSSFEKN PIDFLEAKGYKEVKKDLIIKLPKYSLFELENGRKRMLASAGELQKGNEL ALPSKYVNFLYLASHYEKLKGSPEDNEQKQLFVEQHKHYLDEIIEQISE FSKRVILADANLDKVLSAYNKHRDKPIREQAENIIHLFTLTNLGAPAAF KYFDTTIDRKRYTSTKEVLDATLIHQSITGLYETRIDLSQLGGD
PolI1M	VQIPQNPLILVDGSSYLYRAYHAFPPLTNSAGEPTGAMYGVLNMLRSLI MQYKPTHAAVVFDAKGKTFRDELFEHYKSHRPPMPDDLRAQIEPLHAMV KAMGLPLLAVSGVEADDVIGTLAREAEKAGRPVLISTGDKDMAQLVTPN ITLINTMTNTILGPEEVVNKYGVPPELIIDFLALMGDSSDNIPGVPGVG EKTAQALLQGLGGLDTLYAEPEKIAGLSFRGAKTMAAKLEQNKEVAYLS YQLATIKTDVELELTCEQLEVQQPAAEELLGLFKKYEFKRWTADVEAGK WLQAKGAKPAAKPQETSVADEAPEVTATVISYDNYVTILDEETLKAWIA KLEKAPVFAFDTETDSLDNISANLVGLSFAIEPGVAAYIPVAHDYLDAP DQISRERALELLKPLLEDEKALKVGQNLKYARGILANYGIELRGIAFDT MLESYILNSVAGRHDMDSLAERWLKHKTITFEEIAGKGKNQLTFNQIAL EEAGRYAAEDADVTLQLHLKMWPDLQKHKGPLNVFENIEMPLVPVLSRI ERNGVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTI LFEKQGIKPLKKTPGGAPSTSEEVLEELALDYPLPKVILEYRGLAKLKS TYTDKLPLMINPKTGRVHTSYHQAVTATGRLSSTDPNLQNIPVRNEEGR RIRQAFIAPEDYVIVSADYSQIELRIMAHLSRDKGLLTAFAEGKDIHRA TAAEVFGLPLETVTSEQRRSAKAINFGLIYGMSAFGLARQLNIPRKEAQ KYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLPDIKSSNGA RRAAAERAAINAPMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDE LVFEVHKDDVDAVAKQIHQLMENCTRLDVPLLVEVGSGENWDQAH
PolI2M	VQIPQNPLILVDGSSYLYRAYHAFPPLTNSAGEPTGAMYGVLNMLRSLI MQYKPTHAAVVFDAKGKTFRDELFEHYKSHRPPMPDDLRAQIEPLHAMV KAMGLPLLAVSGVEADDVIGTLAREAEKAGRPVLISTGDKDMAQLVTPN ITLINTMTNTILGPEEVVNKYGVPPELIIDFLALMGDSSDNIPGVPGVG EKTAQALLQGLGGLDTLYAEPEKIAGLSFRGAKTMAAKLEQNKEVAYLS YQLATIKTDVELELTCEQLEVQQPAAEELLGLFKKYEFKRWTADVEAGK WLQAKGAKPAAKPQETSVADEAPEVTATVISYDNYVTILDEETLKAWIA KLEKAPVFAFDTETDSLDNISANLVGLSFAIEPGVAAYIPVAHDYLDAP DQISRERALELLKPLLEDEKALKVGQNLKYARGILANYGIELRGIAFDT MLESYILNSVAGRHDMDSLAERWLKHKTITFEEIAGKGKNQLTFNQIAL EEAGRYAAEDADVTLQLHLKMWPDLQKHKGPLNVFENIEMPLVPVLSRI ERNGVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTI LFEKQGIKPLKKTPGGAPSTSEEVLEELALDYPLPKVILEYRGLAKLKS TYTDKLPLMINPKTGRVHTSYHQAVTATGRLSSTDPNLQNIPVRNEEGR RIRQAFIAPEDYVIVSADYSQNELRIMAHLSRDKGLLTAFAEGKDIHRA TAAEVFGLPLETVTSEQRRSAKAINFGLIYGMSAFGLARQLNIPRKEAQ KYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLPDIKSSNGA RRAAAERAAINAPMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDE LVFEVHKDDVDAVAKQIHQLMENCTRLDVPLLVEVGSGENWDQAH
PolI3M	VQIPQNPLILVDGSSYLYRAYHAFPPLTNSAGEPTGAMYGVLNMLRSLI MQYKPTHAAVVFDAKGKTFRDELFEHYKSHRPPMPDDLRAQIEPLHAMV

	KAMGLPLLAVSGVEADDVIGTLAREAEKAGRPVLISTGDKDMAQLVTPN ITLINTMTNTILGPEEVVNKYGVPPELIIDFLALMGDSSDNIPGVPGVG EKTAQALLQGLGGLDTLYAEPEKIAGLSFRGAKTMAAKLEQNKEVAYLS YQLATIKTDVELELTCEQLEVQQPAAEELLGLFKKYEFKRWTADVEAGK WLQAKGAKPAAKPQETSVADEAPEVTATVISYDNYVTILDEETLKAWIA KLEKAPVFAFDTETDSLDNISANLVGLSFAIEPGVAAYIPVAHDYLDAP DQISRERALELLKPLLEDEKALKVGQNLKYARGILANYGIELRGIAFDT MLESYILNSVAGRHDMDSLAERWLKHKTITFEEIAGKGKNQLTFNQIAL EEAGRYAAEDADVTLQLHLKMWPDLQKHKGPLNVFENIEMPLVPVLSRI ERNGVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTI LFEKQGIKPLKKTPGGAPSTSEEVLEELALDYPLPKVILEYRGLAKLKS TYTDKLPLMINPKTGRVHTSYHQAVTATGRLSSTDPNLQNIPVRNEEGR RIRQAFIAPEDYVIVSADYSQNELRIMAHLSRDKGLLTAFAEGKDIHRA TAAEVFGLPLETVTSEQRRSAKRINFGLIYGMSAFGLARQLNIPRKEAQ KYMDLYFERYPGVLEYMERTRAQAKEQGYVETLDGRRLYLPDIKSSNGA RRAAAERAAINAPMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDE LVFEVHKDDVDAVAKQIHQLMENCTRLDVPLLVEVGSGENWDQAH
PolI5M	VQIPQNPLILVDGSSYLYRAYHAFPPLTNSAGEPTGAMYGVLNMLRSLI MQYKPTHAAVVFDAKGKTFRDELFEHYKSHRPPMPDDLRAQIEPLHAMV KAMGLPLLAVSGVEADDVIGTLAREAEKAGRPVLISTGDKDMAQLVTPN ITLINTMTNTILGPEEVVNKYGVPPELIIDFLALMGDSSDNIPGVPGVG EKTAQALLQGLGGLDTLYAEPEKIAGLSFRGAKTMAAKLEQNKEVAYLS YQLATIKTDVELELTCEQLEVQQPAAEELLGLFKKYEFKRWTADVEAGK WLQAKGAKPAAKPQETSVADEAPEVTATVISYDNYVTILDEETLKAWIA KLEKAPVFAFDTETDSLDNISANLVGLSFAIEPGVAAYIPVAHDYLDAP DQISRERALELLKPLLEDEKALKVGQNLKYARGILANYGIELRGIAFDT MLESYILNSVAGRHDMDSLAERWLKHKTITFEEIAGKGKNQLTFNQIAL EEAGRYAAEDADVTLQLHLKMWPDLQKHKGPLNVFENIEMPLVPVLSRI ERNGVKIDPKVLHNHSEELTLRLAELEKKAHEIAGEEFNLSSTKQLQTI LFEKQGIKPLKKTPGGAPSTSEEVLEELALDYPLPKVILEYRGLAKLKS TYTDKLPLMINPKTGRVHTSYHQAVTATGRLSSTDPNLQNIPVRNEEGR RIRQAFIAPEDYVIVSADYSQNELRIMAHLSRDKGLLTAFAEGKDIHRA TAAEVYGLPLETVTSEQRRSAKRINFGLIYGMSAFGLARQLNIPRKEAQ KYMDLYFERYHGVLEYMERTRAQAKEQGYVETLDGRRLYLPDIKSSNGA RRAAAERAAINAPMQGTAADIIKRAMIAVDAWLQAEQPRVRMIMQVHDE LVFEVHKDDVDAVAKQIHQLMENCTRLDVPLLVEVGSGENWDQAH
PolI3M-TBD	VQIPQNPLILVDGSSYLYRAYHAFPPLTNSAGEPTGAMYGVLNMLRSLI MQYKPTHAAVVFDAKGKTFRDELFEHYKSHRPPMPDDLRAQIEPLHAMV KAMGLPLLAVSGVEADDVIGTLAREAEKAGRPVLISTGDKDMAQLVTPN ITLINTMTNTILGPEEVVNKYGVPPELIIDFLALMGDSSDNIPGVPGVG EKTAQALLQGLGGLDTLYAEPEKIAGLSFRGAKTMAAKLEQNKEVAYLS YQLATIKTDVELELTCEQLEVQQPAAEELLGLFKKYEFKRWTADVEAGK WLQAKGAKPAAKPQETSVADEAPEVTATVISYDNYVTILDEETLKAWIA KLEKAPVFAFDTETDSLDNISANLVGLSFAIEPGVAAYIPVAHDYLDAP DQISRERALELLKPLLEDEKALKVGQNLKYARGILANYGIELRGIAFDT MLESYILNSVAGRHDMDSLAERWLKHKTITFEEIAGKGKNQLTFNQIAL

	EEAGRYAAEDADVTLQLHLKMWPDLQKHKGPLNVFENIEMPLVPVLSRI ERNGVKIDPKVLHNHSEELTLRLAELEKKATETFGSWYQPKGGTEMFCH PRTGKPLPKYPRIKTPKVGGIFKKPKNKAQREGREPCELDTREYVAGAP YTPVEHVVFNLSSTKQLQTILFEKQGIKPLKKTPGGAPSTSEEVLEELA LDYPLPKVILEYRGLAKLKSTYTDKLPLMINPKTGRVHTSYHQAVTATG RLSSTDPNLQNIPVRNEEGRRIRQAFIAPEDYVIVSADYSQNELRIMAH LSRDKGLLTAFAEGKDIHRATAAEVFGLPLETVTSEQRRSAKRINFGLI YGMSAFGLARQLNIPRKEAQKYMDLYFERYPGVLEYMERTRAQAKEQGY VETLDGRRLYLPDIKSSNGARRAAAERAAINAPMQGTAADIIKRAMIAV DAWLQAEQPRVRMIMQVHDELVFEVHKDDVDAVAKQIHQLMENCTRLDV PLLVEVGSGENWDQAH
Phi29 (N62D)	KHMPRKMYSCDFETTTKVEDCRVWAYGYMNIEDHSEYKIGNSLDEFMAW VLKVQADLYFHDLKFDGAFIINWLERNGFKWSADGLPNTYNTIISRMGQ WYMIDICLGYKGKRKIHTVIYDSLKKLPFPVKKIAKDFKLTVLKGDIDY HKERPVGYKITPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAGSDSLKG FKDIITTKKFKKVFPTLSLGLDKEVRYAYRGGFTWLNDRFKEKEIGEGM VFDVNSLYPAQMYSRLLPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELK EGYIPTIQIKRSRFYKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYN VEYISGLKFKATTGLFKDFIDKWTYIKTTSEGAIKQLAKLMLNSLYGKF ASNPDVTGKVPYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTI TAAQACYDRIIYCDTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKR AKYLRQKTYIQDIYMKEVDGKLVEGSPDDYTDIKFSVKCAGMTDKIKKE VTFENFKVGFSRKMKPKPVQVPGGVVLVDDTFTIK
Phi29 (N62D, L384R)	KHMPRKMYSCDFETTTKVEDCRVWAYGYMNIEDHSEYKIGNSLDEFMAW VLKVQADLYFHDLKFDGAFIINWLERNGFKWSADGLPNTYNTIISRMGQ WYMIDICLGYKGKRKIHTVIYDSLKKLPFPVKKIAKDFKLTVLKGDIDY HKERPVGYKITPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAGSDSLKG FKDIITTKKFKKVFPTLSLGLDKEVRYAYRGGFTWLNDRFKEKEIGEGM VFDVNSLYPAQMYSRLLPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELK EGYIPTIQIKRSRFYKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYN VEYISGLKFKATTGLFKDFIDKWTYIKTTSEGAIKQLAKRMLNSLYGKF ASNPDVTGKVPYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTI TAAQACYDRIIYCDTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKR AKYLRQKTYIQDIYMKEVDGKLVEGSPDDYTDIKFSVKCAGMTDKIKKE VTFENFKVGFSRKMKPKPVQVPGGVVLVDDTFTIK
iPhi29 (N62D)	KHMPRKRYSCDFETTTKVEDCRVWAYGYMNIEDHSEYKIGNSLDEFMAW ALKVQADLYFHDLKFDGAFIINWLERNGFKWSADGLPNTYNTIISRTGQ WYMIDICLGYKGKRKIHTVIYDSLKKLPFPVKKIAKDFKLTVLKGDIDY HKERPVGYKITPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAGSDSLKD FKDIITTKKFKKVFPTLSLGLDKKVRYAYRGGFTWLNDRFKEKEIGEGM VFDVNSLYPAQMYSRLLPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELK EGYIPTIQIKRSRFYKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYN VEYISGLKFKATTGLFKDFIDKWTYIKTTSEGAIKQLAKLMLNSLYGKF ASNPDVTGKVPYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTI TAAQACYDRIIYCDTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKR

	AKYLRPKTYIQDIYMKEVDGKLVEGSPDDYTDIKLSVKCAGMTDKIKKE VTFENFKVGFSRKMKPKPVQVPGGVVLVDDTFTIK
iPhi29 (N62D, L384R)	KHMPRKRYSCDFETTTKVEDCRVWAYGYMNIEDHSEYKIGNSLDEFMAW ALKVQADLYFHDLKFDGAFIINWLERNGFKWSADGLPNTYNTIISRTGQ WYMIDICLGYKGKRKIHTVIYDSLKKLPFPVKKIAKDFKLTVLKGDIDY HKERPVGYKITPEEYAYIKNDIQIIAEALLIQFKQGLDRMTAGSDSLKD FKDIITTKKFKKVFPTLSLGLDKKVRYAYRGGFTWLNDRFKEKEIGEGM VFDVNSLYPAQMYSRLLPYGEPIVFEGKYVWDEDYPLHIQHIRCEFELK EGYIPTIQIKRSRFYKGNEYLKSSGGEIADLWLSNVDLELMKEHYDLYN VEYISGLKFKATTGLFKDFIDKWTYIKTTSEGAIKQLAKRMLNSLYGKF ASNPDVTGKVPYLKENGALGFRLGEEETKDPVYTPMGVFITAWARYTTI TAAQACYDRIIYCDTDSIHLTGTEIPDVIKDIVDPKKLGYWAHESTFKR AKYLRPKTYIQDIYMKEVDGKLVEGSPDDYTDIKLSVKCAGMTDKIKKE VTFENFKVGFSRKMKPKPVQVPGGVVLVDDTFTIK
Phi29 SSB	ENTNIVKATFDTETLEGQIKIFNAQTGGGQSFKNLPDGTIIEANAIAQY KQVSDTYGDAKEETVTTIFAADGSLYSAISKTVAEAASDLIDLVTRHKL ETFKVKVVQGTSSKGNVFFSLQLSL

11. Bibliography

Aguirre, Andrew J., Robin M. Meyers, Barbara A. Weir, Francisca Vazquez, Cheng-Zhong Zhang, Uri Ben-David, April Cook, et al. 2016. "Genomic Copy Number Dictates a Gene-Independent Cell Response to CRISPR/Cas9 Targeting." *Cancer Discovery* 6 (8): 914–29.

Alper, Hal, Joel Moxley, Elke Nevoigt, Gerald R. Fink, and Gregory Stephanopoulos. 2006. "Engineering Yeast Transcription Machinery for Improved Ethanol Tolerance and Production." *Science* 314 (5805). http://science.sciencemag.org/content/314/5805/1565.

Badran, Ahmed H., and David R. Liu. 2015. "Development of Potent in Vivo Mutagenesis Plasmids with Broad Mutational Spectra." *Nature Communications* 6 (1): 8425.

Bambara, Robert A., and Theodore Choi. 1973. "On the Processive Mechanism of Escherichia Coli DNA Polymerase I Quantitative Assessment of Processivity." *The Journal of Biological Chemistry* 253 (2): 413–23.

Boer, J. G. de, and L. S. Ripley. 1988. "An in Vitro Assay for Frameshift Mutations: Hotspots for Deletions of 1 Bp by Klenow-Fragment Polymerase Share a Consensus DNA Sequence." *Genetics* 118 (2).

Brocklehurst, Peter, and Peter. 2002. "Antibiotics for Gonorrhoea in Pregnancy." In *Cochrane Database of Systematic Reviews*, edited by Frances J. Kellie. Chichester, UK: John Wiley & Sons, Ltd.

Camps, Manel, Jussi Naukkarinen, Ben P. Johnson, and Lawrence A. Loeb. 2003. "Targeted Gene Evolution in Escherichia Coli Using a Highly Error-Prone DNA Polymerase I." *Proceedings of the National Academy of Sciences of the United States of America* 100 (17): 9727–32.

Choi, Peter S., and Matthew Meyerson. 2014. "Targeted Genomic Rearrangements Using CRISPR/Cas Technology." *Nature Communications* 5 (April): 4728.

Cobb, Ryan E., Ning Sun, and Huimin Zhao. 2013. "Directed Evolution as a Powerful Synthetic Biology Tool." *Methods* 60 (1): 81–90.

Costantino, Nina, and Donald L. Court. 2003. "Enhanced Levels of Lambda Red-Mediated Recombinants in Mismatch Repair Mutants." *Proceedings of the National Academy of Sciences of the United States of America* 100 (26): 15748–53.

Cremer, Marion, Florian Grasser, Christian Lanctôt, Stefan Müller, Michaela Neusser, Roman Zinner, Irina Solovei, and Thomas Cremer. 2012. "Multicolor 3D Fluorescence In Situ Hybridization for Imaging Interphase Chromosomes." In , 205–39. Humana Press, Totowa, NJ.

Crook, Nathan, Joseph Abatemarco, Jie Sun, James M. Wagner, Alexander Schmitz, and Hal S. Alper. 2016. "In Vivo Continuous Evolution of Genes and Pathways in Yeast." *Nature Communications* 7 (October): 13051.

Drake, J. W. 1991. "A Constant Rate of Spontaneous Mutation in DNA-Based Microbes." *Proceedings of the National Academy of Sciences of the United States of America* 88 (16): 7160–64.

Driscoll, Carlos A., David W. Macdonald, and Stephen J. O'Brien. 2009. "From Wild Animals to Domestic Pets, an Evolutionary View of Domestication." *Proceedings of the National Academy of Sciences of the United States of America* 106 Suppl (Supplement 1): 9971–78.

Ducani, Cosimo, Giulio Bernardinelli, and Björn Högberg. 2014. "Rolling Circle Replication Requires Single-Stranded DNA Binding Protein to Avoid Termination and Production of Double-Stranded DNA." *Nucleic Acids Research* 42 (16): 10596–604.

Esvelt, Kevin M., Jacob C. Carlson, and David R. Liu. 2011. "A System for the Continuous Directed Evolution of Biomolecules." *Nature* 472 (7344): 499–503.

Finney-Manchester, Shawn P., and Narendra Maheshri. 2013. "Harnessing Mutagenic Homologous Recombination for Targeted Mutagenesis in Vivo by TaGTEAM." *Nucleic Acids Research* 41 (9):

e99.
Frock, Richard L., Jiazhi Hu, Robin M. Meyers, Yu-Jui Ho, Erina Kii, and Frederick W. Alt. 2014. "Genome-Wide Detection of DNA Double-Stranded Breaks Induced by Engineered Nucleases." *Nature Biotechnology* 33 (2): 179–86.
Funatsu, G., E. Schiltz, and H. G. Wittmann. 1972. "Ribosomal Proteins. XXVII. Localization of the Amino Acid Exchanges in Protein S5 from Two Escherichia Coli Mutants Resistant to Spectinomycin." *Molecular & General Genetics: MGG* 114 (2): 106–11.
Glaser, Astrid, Bradley McColl, and Jim Vadolas. 2016. "GFP to BFP Conversion: A Versatile Assay for the Quantification of CRISPR/Cas9-Mediated Genome Editing." *Molecular Therapy. Nucleic Acids* 5 (7): e334.
Greener, Alan, Marie Callahan, and Bruce Jerpseth. 1997. "An Efficient Random Mutagenesis Technique Using anE.coli Mutator Strain." *Molecular Biotechnology* 7 (2): 189–95.
Hall, Brandon M., Chang-Xing Ma, Ping Liang, and Keshav K. Singh. 2009. "Fluctuation Analysis CalculatOR: A Web Tool for the Determination of Mutation Rate Using Luria-Delbruck Fluctuation Analysis." *Bioinformatics* 25 (12): 1564–65.
Halperin, Shakked O., Connor J. Tou, Eric B. Wong, Cyrus Modavi, David V. Schaffer, and John E. Dueber. 2018. "CRISPR-Guided DNA Polymerases Enable Diversification of All Nucleotides in a Tunable Window." *Nature*. https://doi.org/10.1038/s41586-018-0384-8.
Hess, Gaelen T., Laure Frésard, Kyuho Han, Cameron H. Lee, Amy Li, Karlene A. Cimprich, Stephen B. Montgomery, and Michael C. Bassik. 2016. "Directed Evolution Using dCas9-Targeted Somatic Hypermutation in Mammalian Cells." *Nature Methods* 13 (12): 1036–42.
Jacob, Pierre, Adi Avni, and Abdelhafid Bendahmane. 2018. "Translational Research: Exploring and Creating Genetic Diversity." *Trends in Plant Science* 23 (1): 42–52.
Jagessar, Kevin L., and Chaitanya Jain. 2010. "Functional and Molecular Analysis of Escherichia Coli Strains Lacking Multiple DEAD-Box Helicases." *RNA* 16 (7): 1386–92.
Jinek, Martin, Krzysztof Chylinski, Ines Fonfara, Michael Hauer, Jennifer A. Doudna, and Emmanuelle Charpentier. 2012. "A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity." *Science* 337 (6096): 816–21.
Koboldt, Daniel C., Qunyuan Zhang, David E. Larson, Dong Shen, Michael D. McLellan, Ling Lin, Christopher A. Miller, Elaine R. Mardis, Li Ding, and Richard K. Wilson. 2012. "VarScan 2: Somatic Mutation and Copy Number Alteration Discovery in Cancer by Exome Sequencing." *Genome Research* 22 (3): 568–76.
Kulczyk, Arkadiusz W., Barak Akabayov, Seung-Joo Lee, Mihnea Bostina, Steven A. Berkowitz, and Charles C. Richardson. 2012. "An Interaction between DNA Polymerase and Helicase Is Essential for the High Processivity of the Bacteriophage" 287 (46): 39050–60.
Lin, Myat T., Alessandro Occhialini, P. John Andralojc, Martin A. J. Parry, and Maureen R. Hanson. 2014. "A Faster Rubisco with Potential to Increase Photosynthesis in Crops." *Nature* 513 (7519): 547–50.
Loh, Ern, Jesse J. Salk, and Lawrence A. Loeb. 2010. "Optimization of DNA Polymerase Mutation Rates during Bacterial Evolution." *Proceedings of the National Academy of Sciences of the United States of America* 107 (3): 1154–59.
Mao, Zhiyong, Michael Bozzella, Andrei Seluanov, and Vera Gorbunova. 2008a. "DNA Repair by Nonhomologous End Joining and Homologous Recombination during Cell Cycle in Human Cells." *Cell Cycle* 7 (18): 2902–6.
———. 2008b. "Comparison of Nonhomologous End Joining and Homologous Recombination in Human Cells." *DNA Repair* 7 (10): 1765–71.
Ma, Yunqing, Jiayuan Zhang, Weijie Yin, Zhenchao Zhang, Yan Song, and Xing Chang. 2016. "Targeted AID-Mediated Mutagenesis (TAM) Enables Efficient Genomic Diversification in Mammalian Cells." *Nature Methods* 13 (12): 1029–35.
Minnick, D. T., K. Bebenek, W. P. Osheroff, R. M. Turner, M. Astatke, L. Liu, T. A. Kunkel, and C. M. Joyce. 1999. "Side Chains That Influence Fidelity at the Polymerase Active Site of Escherichia Coli

DNA Polymerase I (Klenow Fragment)." *The Journal of Biological Chemistry* 274 (5): 3067–75.
Povilaitis, Tadas, Gediminas Alzbutas, Rasa Sukackaite, Juozas Siurkus, and Remigijus Skirgaila. 2016. "In Vitro Evolution of phi29 DNA Polymerase Using Isothermal Compartmentalized Self Replication Technique." *Protein Engineering, Design & Selection: PEDS* 29 (12): 617–28.
Ravikumar, Arjun, Adrian Arrieta, and Chang C. Liu. 2014. "An Orthogonal DNA Replication System in Yeast." *Nature Chemical Biology* 10 (3): 175–77.
Richardson, Christopher D., Graham J. Ray, Mark A. DeWitt, Gemma L. Curie, and Jacob E. Corn. 2016. "Enhancing Homology-Directed Genome Editing by Catalytically Active and Inactive CRISPR-Cas9 Using Asymmetric Donor DNA." *Nature Biotechnology* 34 (3): 339–44.
Rosche, William A., and Patricia L. Foster. 2000. "Determining Mutation Rates in Bacterial Populations." *Methods* 20 (1): 4–17.
Rutledge, Peter J., and Gregory L. Challis. 2015. "Discovery of Microbial Natural Products by Activation of Silent Biosynthetic Gene Clusters." *Nature Reviews. Microbiology* 13 (8): 509–23.
Salis, Howard M. 2011. "The Ribosome Binding Site Calculator." In *Methods in Enzymology*, 498:19–42.
Sarkar, S., W. T. Ma, and G. v. H. Sandri. 1992. "On Fluctuation Analysis: A New, Simple and Efficient Method for Computing the Expected Number of Mutants." *Genetica* 85 (2): 173–79.
Schaefer, Kellie A., Wen-Hsuan Wu, Diana F. Colgan, Stephen H. Tsang, Alexander G. Bassuk, and Vinit B. Mahajan. 2017. "Unexpected Mutations after CRISPR–Cas9 Editing in Vivo." *Nature Publishing Group* 14 (6): 547–48.
Shimatani, Zenpei, Sachiko Kashojiya, Mariko Takayama, Rie Terada, Takayuki Arazoe, Hisaki Ishii, Hiroshi Teramura, et al. 2017. "Targeted Base Editing in Rice and Tomato Using a CRISPR-Cas9 Cytidine Deaminase Fusion." *Nature Biotechnology* 35 (5): 441–43.
Slaymaker, I. M., L. Gao, B. Zetsche, D. A. Scott, W. X. Yan, and F. Zhang. 2015. "Rationally Engineered Cas9 Nucleases with Improved Specificity." *Science* 351 (6268): 84–88.
Sohail, Anjum, Joanna Klapacz, Mala Samaranayake, Asad Ullah, and Ashok S. Bhagwat. 2003. "Human Activation-Induced Cytidine Deaminase Causes Transcription-Dependent, Strand-Biased C to U Deaminations." *Nucleic Acids Research* 31 (12): 2990–94.
Svitashev, Sergei, Joshua K. Young, Christine Schwartz, Huirong Gao, S. Carl Falco, and A. Mark Cigan. 2015. "Targeted Mutagenesis, Precise Gene Editing, and Site-Specific Gene Insertion in Maize Using Cas9 and Guide RNA." *Plant Physiology* 169 (2): 931–45.
Tai, T. H., D. Dahlbeck, E. T. Clark, P. Gajiwala, R. Pasion, M. C. Whalen, R. E. Stall, and B. J. Staskawicz. 1999. "Expression of the Bs2 Pepper Gene Confers Resistance to Bacterial Spot Disease in Tomato." *Proceedings of the National Academy of Sciences of the United States of America* 96 (24): 14153–58.
Timms, A. R., H. Steingrimsdottir, A. R. Lehmann, and B. A. Bridges. 1992. "Mutant Sequences in the rpsL Gene of Escherichia Coli B/r: Mechanistic Implications for Spontaneous and Ultraviolet Light Mutagenesis." *Molecular & General Genetics: MGG* 232 (1): 89–96.
Troll, Chris, David Alexander, Jennifer Allen, Jacob Marquette, and Manel Camps. 2011. "Mutagenesis and Functional Selection Protocols for Directed Evolution of Proteins in E. Coli." *Journal of Visualized Experiments: JoVE*, no. 49 (March). https://doi.org/10.3791/2505.
Truniger, Verónica, José M. Lázaro, Miguel de Vega, Luis Blanco, and Margarita Salas. 2003. "Phi 29 DNA Polymerase Residue Leu384, Highly Conserved in Motif B of Eukaryotic Type DNA Replicases, Is Involved in Nucleotide Insertion Fidelity." *The Journal of Biological Chemistry* 278 (35): 33482–91.
Tsai, Helen, Tyson Howell, Rebecca Nitcher, Victor Missirian, Brian Watson, Kathie J. Ngo, Meric Lieberman, et al. 2011. "Discovery of Rare Mutations in Populations :" 156 (July): 1257–68.
Vega, M. de, J. M. Lazaro, M. Salas, and L. Blanco. 1996. "Primer-Terminus Stabilization at the 3'-5' Exonuclease Active Site of phi29 DNA Polymerase. Involvement of Two Amino Acid Residues Highly Conserved in Proofreading DNA Polymerases." *The EMBO Journal* 15 (5): 1182–92.
Wang, Harris H., Farren J. Isaacs, Peter A. Carr, Zachary Z. Sun, George Xu, Craig R. Forest, and George M. Church. 2009. "Programming Cells by Multiplex Genome Engineering and Accelerated

Evolution." *Nature* 460 (7257): 894–98.
Wang, Lei, W. Coyt Jackson, Paul A. Steinbach, and Roger Y. Tsien. 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): 16745–49.
Wang, Y., Dennis E. Prosen, Li Mei, John C. Sullivan, Michael Finney, and Peter B. Vander Horn. 2004. "A Novel Strategy to Engineer DNA Polymerases for Enhanced Processivity and Improved Performance in Vitro." *Nucleic Acids Research* 32 (3): 1197–1207.
Whitney, Spencer M., Robert L. Houtz, and Hernan Alonso. 2011. "Advancing Our Understanding and Capacity to Engineer Nature's CO2-Sequestering Enzyme, Rubisco." *Plant Physiology* 155 (1). http://www.plantphysiol.org/content/155/1/27.short.
Wilson, Robert H., Hernan Alonso, and Spencer M. Whitney. 2016. "Evolving Methanococcoides Burtonii Archaeal Rubisco for Improved Photosynthesis and Plant Growth." *Scientific Reports* 6 (February): 22284.
Wright, Stephen I., Irie Vroh Bi, Steve G. Schroeder, Masanori Yamasaki, John F. Doebley, Michael D. McMullen, and Brandon S. Gaut. 2005. "The Effects of Artificial Selection on the Maize Genome." *Science* 308 (5726): 1310–14.
Xie, Kabin, Bastian Minkenberg, and Yinong Yang. 2015. "Boosting CRISPR/Cas9 Multiplex Editing Capability with the Endogenous tRNA-Processing System." *Proceedings of the National Academy of Sciences of the United States of America* 112 (11): 3570–75.
Zhang, Mingzi M., Fong Tian Wong, Yajie Wang, Shangwen Luo, Yee Hwee Lim, Elena Heng, Wan Lin Yeo, et al. 2017. "CRISPR–Cas9 Strategy for Activation of Silent Streptomyces Biosynthetic Gene Clusters." *Nature Chemical Biology* 13 (6): 607–9.
Zhao, Yongping, Congsheng Zhang, Wenwen Liu, Wei Gao, Changlin Liu, Gaoyuan Song, Wen-Xue Li, et al. 2016. "An Alternative Strategy for Targeted Gene Replacement in Plants Using a Dual-sgRNA/Cas9 Design." *Scientific Reports* 6 (April): 23890.
Zheng, Nai-Ying, Kenneth Wilson, Matthew Jared, and Patrick C. Wilson. 2005. "Intricate Targeting of Immunoglobulin Somatic Hypermutation Maximizes the Efficiency of Affinity Maturation." *The Journal of Experimental Medicine* 201 (9). http://jem.rupress.org/content/201/9/1467.short.
Zheng, Xiang, Xin-Hui Xing, and Chong Zhang. 2017. "Targeted Mutagenesis: A Sniper-like Diversity Generator in Microbial Engineering." https://doi.org/10.1016/j.synbio.2017.07.001.
Zhu, Xin-Guang, Stephen P. Long, and Donald R. Ort. 2010. "Improving Photosynthetic Efficiency for Greater Yield." *Annual Review of Plant Biology* 61 (1): 235–61.