DiversiPhi29: an orthogonal system for the continuous directed evolution of genes *in vivo*

UOregon 2021 iDEC Team
Nora Kearns, Amanda Kreppel, Calin Plesa
Phil and Penny Knight Campus for Accelerating Scientific Impact, University of Oregon

ABSTRACT

Continuous directed evolution techniques leverage error-prone replication machinery for rapid *in vivo* mutagenesis of target genes. However, promiscuous activity by error-prone machinery can lead to damage of the host genome.

Here, we propose an orthogonal replication system, DiversiPhi29, for the continuous directed evolution of genes *in vivo* in a bacterial chassis. This system leverages the well-studied bacteriophage ϕ29 and *E. coli*. In this project, we first attempt to modify four ϕ29 proteins to enable the heterologous replication of a linear construct consisting of a single gene-of-interest and origin sequences. Once orthogonal replication is established, we will implement a system capable of tuning the mutation rate of the linear construct's replication by altering the ratio of two ϕ29 DNA polymerases, one of which contains fidelity-reducing mutations. This approach will enable high throughput molecular evolution in the best understood model organism.

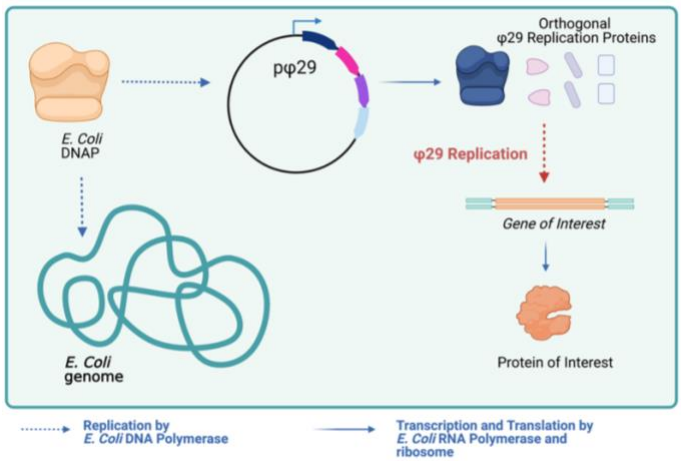

Figure 1. Overview of ϕ29 orthogonal replication system. pϕ29 plasmid carrying the four replication proteins is transcribed and translated by host machinery. A second linear construct (**pL**) carrying the gene of interest is inserted and only replicated by the ϕ29 replication proteins, which do not interact with the host genome.

INTRODUCTION

Proteins possess numerous biological functions which can be repurposed for applications in medicine and industry. Further, the ability to mutate DNA, and thereby proteins, presents a convenient mechanism by which to generate novel and useful functions.

Directed evolution allows scientists to guide and accelerate the mutational path of a sequence in a random or targeted manner towards a certain functional or phenotypic ideal. Current approaches to directed evolution consist of rounds of mutation *in vitro* to develop a library of differentially mutated genes, followed by expression of the mutant library *in vivo* to assess phenotypic outcomes. Variants are screened for a select function, and the most suitable variants undergo further rounds of mutation, expression, and screening to achieve a suitable phenotype.

This process is time consuming and labor intensive, necessitating multiple transformation steps during which genes of interest are inserted in cells for replication and expression. Each transformation step also results in a loss of library diversity, or a "bottleneck," as not all gene variants are successfully taken up by cells. As such, these approaches are limited in their ability to characterize many sequence-function relationships in parallel.

Continuous methods for directed evolution circumvent many of these limitations by allowing multiple cycles of mutation to take place entirely *in vivo* without manual interference. Mutagenesis in continuous evolution relies on an increase in the mutational rate of the host cells above typical genomic error rates. This can be done through chemical mutagenesis using alkylating agents, UV irradiation, or base analogues[1]. More recently, mutational frequency has been increased by deactivating host proofreading enzymes[2] or mutating the host DNA Polymerase to reduce copying fidelity[3]. However, both methods generate unwanted mutations in the host genome. Over time, mutations accumulate in essential genes and kill the host, limiting the number of evolutionary generations possible. Such short experimental durations prevent access to certain adaptations of interest that require long mutational paths.

Orthogonal replication architectures applied to directed evolution provide the advantages of continuous *in vivo* mutagenesis while maintaining stability of the host genome. Previously established systems in yeast have relied on an error-prone DNA polymerase which interacts specifically with an engineered DNA construct and cannot interact with the host genomic DNA[4].

The minimal replication system of bacteriophage ϕ29, a well-characterized virus which naturally infects *Bacillus subtilis*, represents an ideal candidate for orthogonal replication.

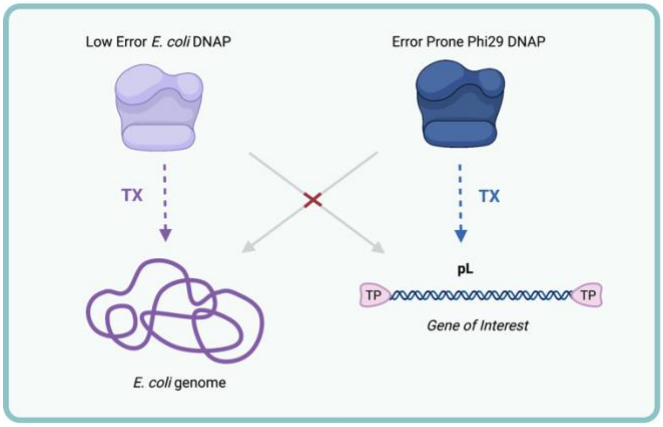

Figure 2. Orthogonal Replication. Error prone replication machinery interacts exclusively with and mutates the gene of interest, while the high-fidelity host machinery maintains integrity of the host genome. The error-prone ϕ29 DNAP recognizes a terminal protein to initiate replication, while the *E. coli* DNAP recognizes an RNA primer. Adapted from Ravikumar et al., 2014.

ϕ29 replicates its small linear genome using just four proteins: DNA Polymerases (DNAP), terminal proteins (TP), single stranded binding proteins (SSBP), and double stranded binding proteins (DSBP)[5]. Here, we repurpose the replication machinery of bacteriophage ϕ29 to develop an orthogonal replication system, DiversiPhi29, for the continuous directed evolution of large gene libraries *in vivo.*

RESULTS

Restriction digest was used to validate successful assembly of a plasmid (pϕ29) carrying the ϕ29 DNAP, TP, SSBP, and DSBP. Complete sequencing of our initial assembly revealed a 70 base-pair deletion.

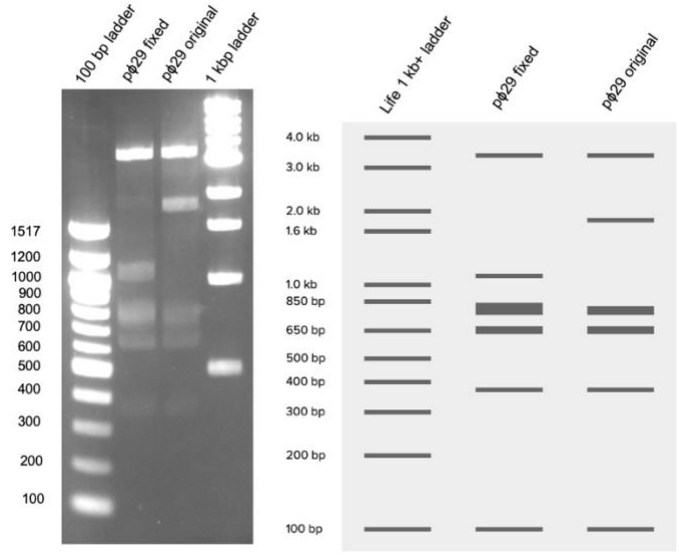

Figure 3. Restriction Digest Validation of pϕ29
The final plasmid was validated by restriction digest with MlyI. **Left image:** Lane 1: 100 bp ladder, Lane 2: MlyI digest of pϕ29 plasmid, corrected with primers carrying missing sequence. Lane 3: MlyI digest of uncorrected pϕ29 plasmid. Lane 4: 1 kb ladder. **Right image:** Anticipated gel results from virtual digest. Lane 1: Life 1 kb+ ladder. Lane 2: virtual MlyI digest of pϕ29 plasmid, corrected with primers carrying missing sequence. Lane 3: virtual MlyI digest of uncorrected pϕ29 plasmid.

To fix the deletion, the plasmid was PCR amplified with restriction sites, and the 70 base-pair segment was inserted by digest and ligation. Correct assembly was validated by gel electrophoresis of a MlyI digest on the original and fixed plasmid (Fig. 3).

pϕ29, which expresses the ϕ29 DNAP, TP, SSBP, and DSBP under control of a lac-inducible promoter, was transformed into *E. coli* strain DH5α. Induction of these cells with variable concentration of the lactose analog Isopropylthio-β-galactoside (IPTG) showed the relative burden posed to host cells by the ϕ29 proteins.

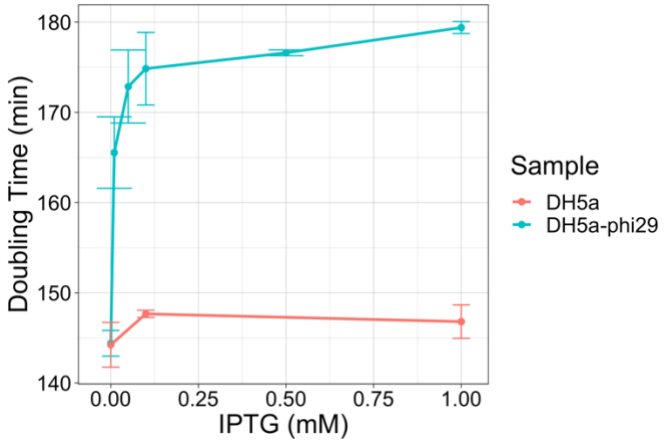

Figure 3. Doubling Time of Cells Expressing pϕ29. Growth of pϕ29 cells (blue) when induced at 0, 0.05, 0.1, 0.5, and 1.0 mM IPTG compared to growth of DH5α cells (orange) at 0, 0.1, and 1.0 mM IPTG.

As IPTG induction level increased from 0 mM to 0.01 mM, the doubling time slowed by 14.6%, from 144.4 ± 1.96 min to 165.6 ± 6.53 min. Increasing IPTG concentration to 0.05, 0.1, 0.5, and 1.0 mM further slowed the doubling time to 172.85 ± 3.11 min (19.7% slower), 174.84 ± 0.61 min (21.1% slower), 176.60 ± 6.55 min (22.3% slower), and 179.39 ± 0.94 min (24.2% slower). Successfully expressing ϕ29 proteins may be complicated by their potential for toxicity to the host. Depending on a protein's characteristics and function, overexpressing recombinant proteins in *E. coli* can interfere with cell survival and proliferation[6]. The increased doubling time indicates that ϕ29 cells divert metabolic resources away from normal replication to make ϕ29 proteins. The minor increase in doubling time from 0.1 to 1.0 mM suggests that pϕ29 expression reached saturation at 0.1 mM IPTG.

Cells induced at varying levels of IPTG were then transformed with a linear plasmid (pL) carrying DHFR R67, a variant of dihydrofolate reductase which confers high levels of resistance to the antibiotic trimethoprim[7]. Cells both with (pL+) and without pL (pL-) were plated on media with IPTG concentrations corresponding to induction level. A negative control not transformed with pL was also plated on each IPTG condition. As expected, pL+ and pL- cells plated on 10 ug/mL trimethoprim and without IPTG did not grow. Although growth was observed on one of four replica plates from the 10 ug/mL

trimethoprim/1.0mM IPTG/ pL+ condition, no growth was seen on the other 3 plates. Significant growth on all four

replica plates was only observed on the 5 ug/mL trimethoprim and 0.1 mM IPTG condition, both with and without pL (Table 1).

Table 1. Number of replica plates on which growth was observed with varying induction and selection levels

pL	0 IPTG / 10 ug/mL TRI	IPTG / 5 ug/mL TRI	IPTG / 10 ug/mL TRI
–	0	4	0
+	0	4	1

The 50% inhibitory concentration of DHFR R67 to trimethoprim is 5.8 g/L[7]. If pL replication by ϕ29 proteins was successful, the expressed DHFR R67 would easily confer resistance of 10 ug/mL (0.01 g/L). The growth observed at the lower trimethoprim resistance regardless of pL presence suggests a high frequency of escape mutations: the host naturally acquires resistance to the antibiotic through genomic mutations rather than the host acquiring resistance through the plasmid. This was confirmed by running a colony PCR to assess the presence of pL in the surviving colonies. A colony that observed growth from each of the conditions was tested. The inverted terminal repeats in the ϕ29 ORIs make PCR amplification of the entire pL impossible; so, a PCR of the middle sequence, DHFR R67, was done and the product was visualized on a gel (Fig. 4). Negative PCR results from all colonies verified that cells acquired resistance to trimethoprim through a separate mechanism independent of pL replication by ϕ29 proteins. Colony PCR also verified that pL did not integrate into the genome, as DFHR R67 would have been amplified off the 0.1 mM IPTG -pL colonies.

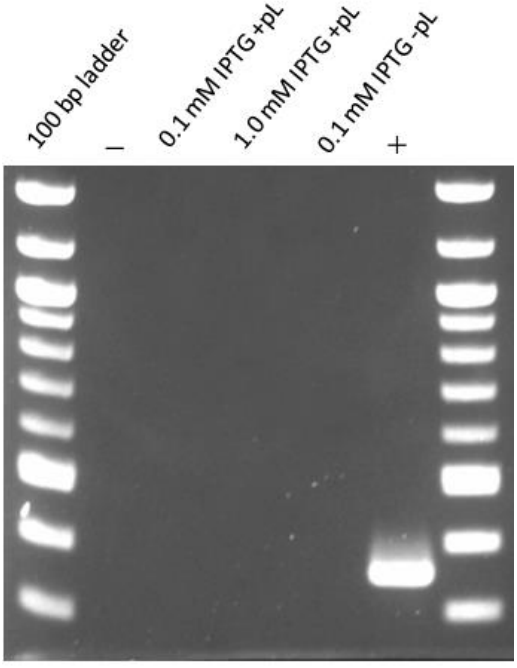

Figure 4. Colony PCR for DHFR R67
Lane 1: 100 bp ladder. Lane 2: no template negative control. Lane 3: 0.1 mM IPTG/5ug/mL Tri /pL+. Lane 4: 1.0 mM IPTG/10 ug/mL/pL+. Lane 5: 0.1 mM IPTG/5 ug/mL/pL–. Lane 6: DHFR R67 template positive control

The ϕ29 TP, DSBP, and SSBP were ligated into an alternative plasmid backbone, pGS21a, that carried the wildtype ϕ29 DNAP with 6XHis and GST tags to enable easier purification. In the resulting plasmid, pGSϕ29, expression of the ϕ29 proteins is under control of a T7 promoter. pGSϕ29 was validated by restriction digest, using the original pϕ29 as a control. pGSϕ29 was further validated by Oxford Nanopore sequencing.

pGSϕ29 was transformed into *E. coli* strain T7 express (C2566H), which expresses T7 polymerase in the lac operon. T7 express is also optimized for protein expression.

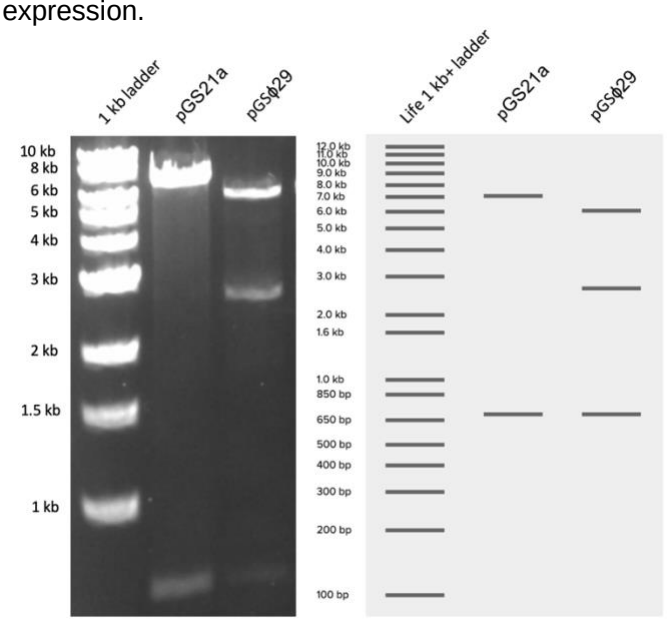

Figure 5. Restriction Digest Validation of pGSϕ29
The reconstructed pGSϕ29 plasmid was validated by restriction digest with KpnI and NdeI. **Left image:** Lane 1: 1 kb ladder, Lane 2: KpnI and NdeI digest of pϕ29 plasmid. Lane 3: KpnI/NdeI digest of pGSϕ29 plasmid. **Right image:** Anticipated gel results from virtual digest. Lane 1: Life 1 kb+ ladder. Lane 2: virtual KnpI/NdeI digest of pGSϕ29 plasmid, corrected with primers carrying missing sequence. Lane 3: virtual KpnI/NdeI digest of uncorrected pGSϕ29 plasmid.

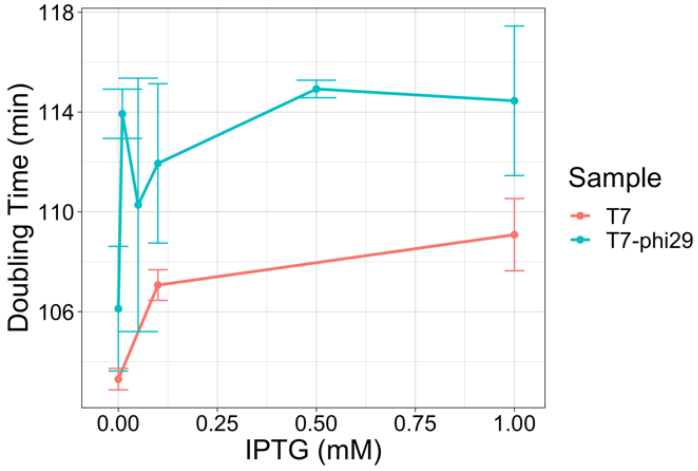

Figure 6. Doubling Time of T7 Cells Expressing pGSϕ29
Growth of pϕ29 cells (blue) when induced at 0, 0.01, 0.05, 0.1, 0.5, and 1.0 mM IPTG compared to growth of DH5α cells (orange) at 0, 0.1, and 1.0 mM IPTG.

T7 cells expressing pGSϕ29 were then induced at increasing concentrations of IPTG to express the ϕ29

proteins. As IPTG induction level increased from 0 mM to 0.01 mM, doubling time slowed by 7.36%, from 106.1 ± 2.49 min to 113.9 ± 0.98 min. Increasing IPTG concentration to 0.05, 0.1, 0.5, and 1.0 mM slowed doubling time to 110.28 ± 5.08 min (3.92% slower), 111.94 ± 3.19 min (5.49% slower), 114.92 ± 3.50 min (8.30% slower), and 114.45 ± 3.00 min (7.85% slower). Doubling time in T7 express was significantly faster than in DH5α, with T7 replicating an average of 64.9 minutes (39%) faster at the highest induction level. T7 cells induced at varying levels of IPTG were then transformed with pL. Cells both with (pL+) and without pL (pL-), were plated on media with IPTG concentrations corresponding to induction level (0 mM, 0.1 mM, or 1 mM).

It was hypothesized that host exonuclease activity may degrade pL and prevent its replication. To counteract exonuclease degradation, pL was constructed with both phosphorothioated (PT) and non-phosphorothioated (NPT) primers. A negative control not transformed with pL was also plated on each IPTG condition. To avoid escape mutations, trimethoprim concentration was increased to 0.5 mg/mL (a 50 fold increase from the first trial with DH5α). No growth was observed on any of the plates, indicating that replication of pL by the ϕ29 proteins was unsuccessful (Table 2).

Table 2. Number of replica plates on which growth was observed with varying induction and selection levels

pL	0 IPTG / 0.5 mg/mL TRI	IPTG / 0.5 mg/mL TRI	1 IPTG / 0.5 mg/mL TRI
–	0	0	0
+ NPT	0	0	0
+ PT	0	0	0

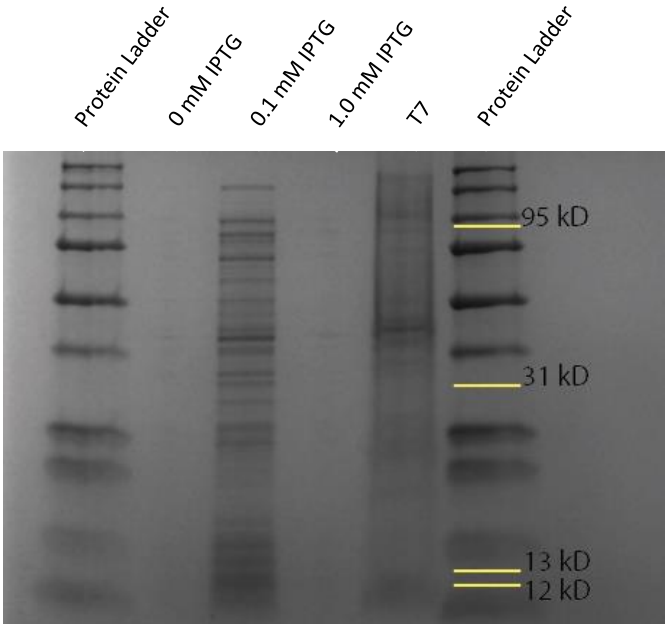

Figure 7. Protein Gel of T7 Cells Expressing pGSϕ29 induced at 0, 0.1, and 1.0 mM IPTG. T7 cells not expressing pGSϕ29 was run as a control. Far left and right lanes are the Precision Plus Protein Ladder, and bands at the expected molecular weights are marked: DSBP: 12 kD, SSBP: 13 kD, TP: 31 kD, GST-tagged DNAP: 95 kD.

To determine whether the ϕ29 proteins were actually expressed in T7, a protein gel was run on cells induced at 0, 0.1, and 1.0 mM IPTG. Proteins bands were only clear in the sample expressed at 0.1 mM IPTG. A distinct band at the molecular weight of the terminal protein (31 kD) was observed, but it was difficult to discern if bands at 95, 13, and 12 kD were distinct from bands in the T7 negative control.

Given the change in doubling time of cells expressing pGSϕ29 (Fig. 6), and the protein gel (Fig. 7) which indicated potential expression of the ϕ29 proteins, it was hypothesized that the ϕ29 likely were expressed but were not interacting successfully with pL, potentially as the result of deleterious interaction with heterologous host proteins. To study the protein-protein interactions between *E. coli* and ϕ29, and rationally design the ϕ29 proteins to avoid them would not have been feasible, so we proceeded with randomized mutagenesis.

The pGSϕ29 backbone was amplified normally with PCR, and different sets of the replication proteins were amplified with error-prone PCR (EP PCR) to introduce base substitutions at a rate of ~10^{-4} substitution per replicated base[1]. Successful amplification was validated with gel electrophoresis (Fig. 8). All amplicons were the expected length. In future work, mutations will be evaluated with Sanger sequencing.

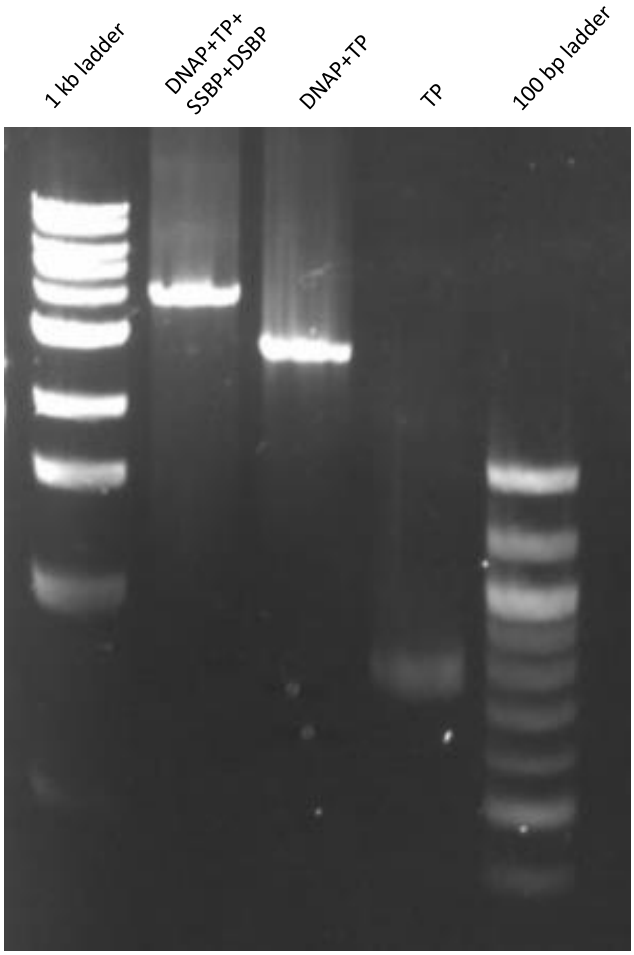

Figure 8. Validation of PCR mutagenesis
Lane 1: 1 kb ladder. Lane 2: EP PCR of ϕ29 DNAP, TP, SSBP, and DSBP. Lane 3: EP PCR of ϕ29 DNAP+TP. Lane 4: EP PCR of ϕ29 TP. Lane 5: 100 bp ladder.

DISCUSSION

Here, we propose an orthogonal replication system for the continuous directed evolution of genes *in vivo* which repurposes the minimal replication system of bacteriophage ϕ29. We have not yet established replication of pL by the ϕ29 proteins in *E. coli.* However, amplification of a linear plasmid by ϕ29 proteins has been demonstrated *in vitro*[8], providing insight into potential issues with expression *in vivo*.

Bacteriophage ϕ29 naturally infects *B. subtilis*. Over evolutionary time, its proteins have evolved to minimize deleterious interactions with *B. subtilis* host proteins which impede phage genome amplification. We attempt to reconstitute the ϕ29 system in *E. coli*, a non-native host for which phage ϕ29 has not evolved. Interactions between host and ϕ29 machinery may inhibit binding between the DNAP-TP complex and the ORIs.

Directed evolution is an ideal approach to molecular engineering when knowledge of how to rationally tune a protein or system is limited. Though our objective is to leverage this platform as a continuous directed evolution system, we are also implementing directed evolution to minimize potential host-ϕ29 protein-protein interactions and optimize the ϕ29 system itself.

Induction of the ϕ29 cells showed that the ϕ29 proteins burden to the host and slow replication. In current work to optimize our system we are implementing error-prone PCR to randomly mutate the ϕ29 proteins, with the objective of tuning expression of the ϕ29 proteins and minimizing toxicity to the host. We are also attempting to add a fifth protein, p1, a membrane associated proteins which interacts with the TP to assist in DNA replication.

Future Directions

Once orthogonal replication of pL has been successfully established, we will reduce the copying fidelity of the ϕ29 DNAP to mutate pL. The ϕ29 DNAP is a well characterized polymerase with several mutations known to impact copying fidelity. We will first use targeted mutagenesis to change residues involved in proofreading and nucleotide selection to optimize a low-fidelity polymerase. Then, we plan to create a system to tune the rate of mutagenesis on pL by alternating expressions of a low and high-fidelity polymerase (Fig. 9).

DiversiPhi29 will be a powerful tool for the continuous directed evolution of genes *in vivo*. Relative to traditional directed evolution models, DiversiPhi29 circumvents transformation bottlenecks and consequent loss of library diversity, as library diversity is generated over many generations.

While an orthogonal replication platform exists in yeast[4], no such platform exists in *E. coli*, a well-characterized host which achieves higher transformation efficiency and associated library diversity. Though DiversiPhi29 generates diversity *in vivo*, we ultimately plan to combine this system with *in vitro* mutagenesis to optimize evolutionary scale and depth. Developing a highly divergent starting library through *in vitro* mutagenesis maximizes scale, while using orthogonal continuous directed evolution optimizes depth, allowing access to complex variants which require long mutational paths.

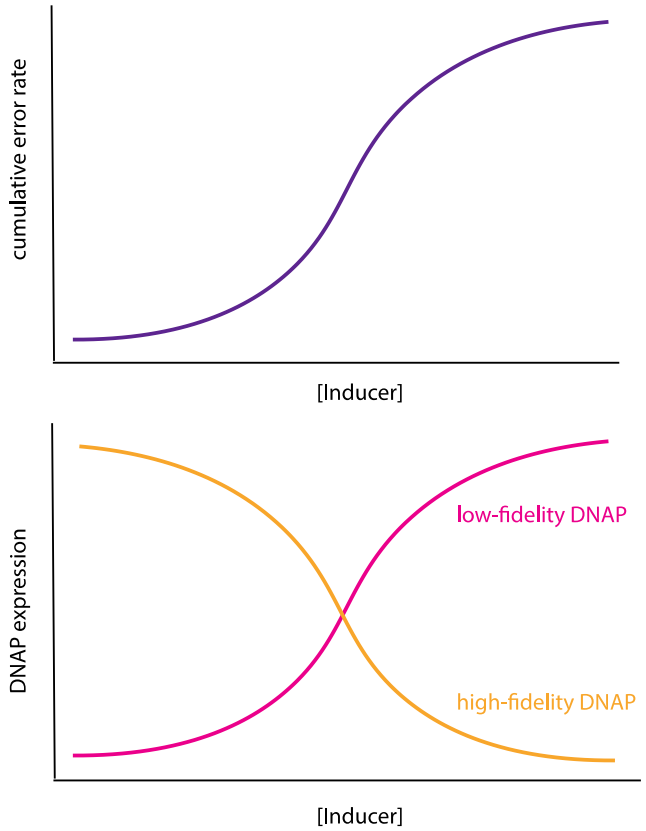

Figure 9. Controlled expression of low- and high-fidelity ϕ29 DNA Polymerases tunes mutagenesis rate on pL.

ACKNOWLEDGEMENTS

This project was funded by the University of Oregon Undergraduate Research Opportunities Program and the Peter O'Day Fellowship in the Biological Sciences. We would also like to thank the Phil and Penny Knight Campus for Accelerating Scientific Impact for providing additional support.

METHODS

DNA for the four ϕ29 replication proteins was ordered as two separate g-blocks (one encoding the ϕ29 DNA Polymerase, (DNAP), and one encoding the ϕ29 TP, SSBP, and DSBP (TPBP) from Integrated DNA technologies (IDT) and amplified by PCR with primers containing BsaI restriction sites at the 5' ends [SI A-B]. PCR products were validated using gel electrophoresis. Off-target amplification resulting in shorter and longer-than expected fragments were seen on the gel, so bands of the correct length were gel extracted and cleaned with ZymoGen gel DNA recovery kit, and further cleaned by drop dialysis on 0.025 µm Millipore DNA filter paper membranes. The pϕ29 backbone, pSEVA224, was ordered from Centro Nacional de Biotecnologia and amplified by PCR with primers carrying BsaI sites in their tails [SI C]. pSEVA224 was

then pre-digested with BsaI and rSAP to prevent backbone self-ligation [SI F].

The two inserts (DNAP and TPBP) were pre-digest with BsaI and ligated before transfer to the GGA [SI D-E], such that there was only one long insert as opposed to two shorter inserts. Equimolar quantities of the two BsaI-digested sequences were ligated using T7 ligase, which joins complementary cohesive ends, The ligated sequence was then joined into the rSAP digested SEVA224 backbone in another round of GGA [SI G].

To verify that the plasmid contained the complete insert (DNAP and TPBP, ligated), a PCR that amplified the insert region was performed off the assembled plasmid [SI H], and length was validated by gel electrophoresis. Bands of the correct length (3.5 kb) were observed, but due to incomplete ligation a band was also observed at 2 kb, the length of the insert region with only a single cargo sequence (TPBP). Extracting plasmids with the full insert from the assembly product would have been difficult, so the mixed assembly product was transformed, and cells containing correct plasmids were identified at a later step.

Transformation

The assembled plasmid was transformed into NEB DH5α electrocompetent cells. 25 ul of cells were mixed with 1.5 ul of DNA in a PCR tube on ice, transferred to a chilled cuvette, and electroporated at 1.8 mV. Cells were resuspended in 950 ul of SOC outgrowth medium and immediately transferred to a 37°C shaker for 1 hour. The same procedure was done using puc19 DNA, a control plasmid which carries a gene encoding resistance to the antibiotic carbenicillin. After 1 hour of shaking, 100 ul of ϕ29 cell mixture was plated on 4 kanamycin plates, and 100 ul of control cell mixture was plated on 1 carbenicillin plate. Plates were incubated at 37°C for 24 hours.

pϕ29 Validation

Due to the mixture of dual-insert and single-insert plasmids products from the GGA, a colony PCR was performed off 64 colonies to identify colonies which carried plasmids with the complete insert [SI I]. The PCR product from each colony was visualized on a gel, and bands at 3.5 kbp represented amplicons of the correct length and corresponded to colonies with the correct plasmid. Several of these colonies were selected and further validated by Sanger sequencing on the insert (DNAP-TPBP) and the 300 base pairs of pSEVA224 directly flanking the insert [SI J]. One colony with Sanger sequencing data most closely aligned to the original plasmid design was selected for downstream use.

pϕ29 Correction

A 70 bp deletion was discovered in the selected plasmid, which was the result of an initial error in the g-block. This error was corrected by PCR with primers carrying the missing sequence in their tails, as well as BsaI restriction site [SI K]. PCR of the entire plasmid was performed, followed by digest with BsaI and ligation with T7 [SI L-M]. Restriction digest with MlyI, an enzyme that cut 9 times within the final plasmid, was used to validate the corrected plasmid [SI N]. Importantly, MlyI cut within the missing sequence. The lengths between MlyI cut sites on the fixed plasmid differed from those on the original plasmid, which demonstrated whether the fix had been successful.

Construction of pL

A linear plasmid, pL, was constructed carrying the gene for DHFR R67, a variant highly resistant to the antibiotic trimethoprim, flanked by the ϕ29 ORIs. Due to the presence of similar inverted terminal repeats (ITRs) at the ends of the sequences, it was not possible to PCR amplify the entire sequence as one. During PCR, ITRs anneal to each other rather than the primers and interfere with amplification. pL was instead ordered as three separate sequences: ORI Left (ORI_L), ORI Right (ORI_R), and DHFR_R67. Each sequence was amplified by PCR with primers containing BspQI sites in their 5' ends [SI Q-S]. Following amplification, each sequence was independently digested with BspQI to create 3 base pair complementary overhangs, and the three sequences were joined together using T7 ligase [SI T].

Induction and pL transformation

DH5α cells previously transformed with pϕ29 were then prepared for a second transformation with the linear plasmid, pL. The ϕ29 proteins were under control of lac promoter variant P_{trc} which activates transcription of downstream coding sequences in the presence of IPTG. Prior to transformation with pL, cells carrying pϕ29 were induced with IPTG, such that transcription and translation of the ϕ29 proteins would occur and the proteins would be present and ready to replicate pL once transformed. Induction with IPTG showed the relative burden posed to the cells by the ϕ29 proteins.

Cells were first inoculated in 10 mL Luria-Bertani Broth (LB) and 10 ul Kanamycin for 16 hours. The culture was then diluted to a starting Optical Density (OD600) of 0.05 in 50 ml of LB. 50 mL of culture was induced at either 0 mM IPTG, 0.1 mM IPTG, or 1.0 mM IPTG. OD600 readings were taken every 30 minutes until each culture reached an OD of 0.6-0.7, at which point 2 5 ml tubes of each culture (0 mM IPTG, 0.1 mM IPTG, and 1.0 mM IPTG) were centrifuged at 3000 rcf for 4 minutes.

Making cells electrocompetent

Supernatant from each tube was extracted and discarded, and the pellet was resuspended in 700 ul of water by pipetting up and down gently. The mixture was centrifuged again, supernatant was extracted, and the pellet was again resuspended in 700 ul of water. A final centrifugation step was performed, supernatant was extracted, and each pellet was resuspended for the final time in 40 ul of water. For each IPTG condition, one group of cells was combined with 1.5 ul of pL DNA (30 ng/ul) and placed on ice. The other group was left without pL.

Each mixture was electroporated at 1.8 mV. Cells were immediately resuspended in 950 ul SOC outgrowth medium and transferred to a 37 °C shaker for 1 hour. 100 ul of each mixture was plated on 4 10ug/mL trimethoprim plates with their respective IPTG induction condition (0 mM, 0.1 mM, or 1.0 mM) and allowed to grow overnight at 37 °C. Plates were then transferred to 30 °C for one more night.

Assessing Presence of pL

Colony PCR using primers flanking the *DHFR 67* gene was run on one colony from each of the plates on which growth was observed [SI U]. Length and presence of the PCR product was verified by gel electrophoresis.

Transfer to pGS21a/T7 Express

Due to issues with expression in DH5α, the ϕ29 proteins were switched to an expression strain. Cloning strains such as DH5α are optimized for DNA replication and stability, but often cannot tolerate high recombinant protein expression. Expression strains, such as BL21, are optimized for high recombinant protein expression and stability. Mutations found in most expression strains include ompT (outer membrane protease) deficiency, and lon protease deficiency, which reduce the rate of protein degradation[11].

In addition to switching strains, the ϕ29 TP, DSBP, and SSBP were transferred to a pGS21a plasmid. pGS21a carries the wildtype ϕ29 DNA Polymerase with 6X His and GST tags which enable purification, as well as a gene encoding resistance to carbenicillin. pGS21a and the TPBP sequence in pϕ29 were PCR amplified with Golden Gate primers containing PaqCI restriction sites in their 5' tails [SI V-W]. The two sequences were joined together using the NEB PaqCI GGA standard protocol [SI X].[12] In the resulting plasmid, pGSϕ29, all four ϕ29 proteins are under control of a T7 promoter, which ensures strong expression of the ϕ29 proteins upon IPTG induction. A PCR of the TPBP insert and flanking regions of the pGS21a backbone was done to verify correct assembly [SI Y].

pGSϕ29 was then cloned into NEB DH5α cells by electroporation methods previously described, and the transformed cells were plated on carbenicillin. Colony PCR with primers flanking the ϕ29 TPBP insert region was done to identify candidates with the correct assembly [SI Z]. This transformation step into DH5a, a cloning strain, was done before cloning into an expression strain to generate many copies of the pGSϕ29 plasmid, as cloning strains achieve higher transformation efficiency. Further, DNA recovered from cloning strains by miniprep is supercoiled, which increases transformation efficiency into the expression strain as cells are more likely to take up smaller, tightly coiled pieces of DNA.

Three candidates were miniprepped using the NEB Monarch Miniprep Kit to recover their plasmid. They were validated with whole-plasmid sequencing by SNPsaurus. A correct plasmid was identified and transformed into T7 Express Competent *E. coli* (C2566H), a chemically competent strain which expresses T7 polymerase when induced with IPTG as previously described. After transformation with pGSϕ29, the cells were be made electrocompetent again as previously described, transformed with pL, and screened on high concentrations of trimethoprim (0.5 mg/mL) with either no IPTG or 0.1 mM IPTG.

Validating correct concentrations of Trimethoprim

To confirm that Trimethoprim concentration was appropriate to enable survival of cells successfully expressing DHFR R67, the DHFR R67 gene was inserted into a known SEVA224 plasmid with a Kanamycin antibiotic resistance marker and plated on LB media containing 50 ug/mL Kanamycin and variable concentrations of Trimethoprim. SEVA224 and DHFR R67 were PCR amplified using primers that contained BsaI cut sites [SI AA-AB]. SEVA224 was digested with DpnI [SI AC], and both fragments were digested BsaI [SI AD-AF]. 30 minutes into SEVA224 digestion, 1 uL of rSAP was added to prevent self-ligation. The two fragments were ligated together with T4 ligase and underwent a chemically competent transformation into T7 express. The transformed cells were then plated on LB media containing 50 ug/mL Kanamycin. After growth was observed, 4 colonies were re-streaked onto different plates containing 50 ug/mL Kan and 0.5, 0.25 or 0.1 mg/mL Trimethoprim. Growth was observed on all plates.

Inserting p1

p1 is a membrane associated protein from bacteriophage ϕ29 that interacts with the TP to assist in DNA replication. p1 was tested to be placed in two places on the pGSϕ29 backbone: between the stop codon of DSBP and the RBS of the SSBP and between the stop codon of the TP and the RBS of the DSBP. p1 and the pGSϕ29 plasmid backbone were amplified using primers flanked with PaqCI sites for Golden Gate Assembly [SI BA-BB]. Both fragments will be digested with PaqCI and ligated with T4 ligase.

Mutagenesis of ϕ29 proteins

Three sets of primers were made for each insert mutagenesis: pGS21a backbone and entire ϕ29 insert (including DNAP, TP, SSBP, DSBP), pGS21a and ϕ29 DNAP+TP, and pGS21a and ϕ29 TP. The PCR amplification was off of the full pGSϕ29 plasmid and only the inserts were mutagenized. Taq polymerase, two different concentrations of $MnCl_2$ and 55 mM $MgCl_2$ were used to mutagenize the whole insert and TP+DNAP [CA-CB]. Half of the tubes contained 0.01 mM $MnCl_2$, and the other half contained 0.15 mM $MnCl_2$ to increase mutation diversity. After PCR amplification and cleanup, the two concentrations were combined in an equimolar ratio. The Agilent Genemorph Mutagenesis Kit, using Mutazyme as the polymerase, was used to mutagenize the TP fragment [SI CC]. The fragments were first digested with DpnI to eliminate any leftover template and then PaqCI to create overhangs for ligation. The fragments will be joined to their respective backbones with T4 ligase.

References

(1) Packer, M. S.; Liu, D. R. Methods for the Directed Evolution of Proteins. *Nat. Rev. Genet.* **2015**, *16* (7), 379–394. https://doi.org/10.1038/nrg3927.

(2) Cox, E. C. Bacterial Mutator Genes and the Control of Spontaneous Mutation. *Annu. Rev. Genet.* **1976**, *10*, 135–156. https://doi.org/10.1146/annurev.ge.10.120176.001031.

(3) Greener, A.; Callahan, M.; Jerpseth, B. An Efficient Random Mutagenesis Technique Using AnE.Coli Mutator Strain. *Mol. Biotechnol.* **1997**, *7* (2), 189–195. https://doi.org/10.1007/BF02761755.

(4) Ravikumar, A.; Arrieta, A.; Liu, C. C. An Orthogonal DNA Replication System in Yeast. *Nat. Chem. Biol.* **2014**, *10* (3), 175–177. https://doi.org/10.1038/nchembio.1439.

(5) Blanco, L.; Lázaro, J. M.; Vega, M. de; Bonnin, A.; Salas, M. Terminal Protein-Primed DNA Amplification. *Proc. Natl. Acad. Sci.* **1994**, *91* (25), 12198–12202. https://doi.org/10.1073/pnas.91.25.12198.

(6) Rosano, G. L.; Ceccarelli, E. A. Recombinant Protein Expression in Escherichia Coli: Advances and Challenges. *Front. Microbiol.* **2014**, *5*. https://doi.org/10.3389/fmicb.2014.00172.

(7) Huovinen, P. Trimethoprim Resistance. *Antimicrob. Agents Chemother.* **1987**, *31* (10), 1451–1456.

(8) Self-replication of DNA by its encoded proteins in liposome-based synthetic cells | Nature Communications https://www.nature.com/articles/s41467-018-03926-1 (accessed 2021 -05 -20).

(9) Holguera, I.; Redrejo-Rodríguez, M.; Salas, M.; Muñoz-Espín, D. New Insights in the Φ29 Terminal Protein DNA-Binding and Host Nucleoid Localization Functions. *Mol. Microbiol.* **2014**, *91* (2), 232–241. https://doi.org/10.1111/mmi.12456.

(10) Muñoz-Espín, D.; Holguera, I.; Ballesteros-Plaza, D.; Carballido-López, R.; Salas, M. Viral Terminal Protein Directs Early Organization of Phage DNA Replication at the Bacterial Nucleoid. *Proc. Natl. Acad. Sci.* **2010**, *107* (38), 16548–16553. https://doi.org/10.1073/pnas.1010530107.

(11) Taylor-Parker, J. Plasmids 101: E. coli Strains for Protein Expression https://blog.addgene.org/plasmids-101-e-coli-strains-for-protein-expression (accessed 2021 -05 -15).

(12) Golden Gate Assembly Protocol using PaqCI (NEB #R0745) and T4 DNA Ligase (NEB #M0202) | NEB https://www.neb.com/protocols/2021/01/11/golden-gate-assembly-protocol-using-paqci-neb-r0745-and-t4-dna-ligase-neb-m0202 (accessed 2021 -05 -05).