PROGENIE

Programmable Gene Disruption in Pathogenic Bacteria

2021 UCSC iGEM

SUBMITTED TO THE UNIVERSITY OF CALIFORNIA, SANTA CRUZ FOR CONSIDERATION OF THE DEANS' AND CHANCELLOR'S AWARD

MARCH 24TH, 2022

Submitted By

Alfonso GAMINO[2], David KELAITA[1], Denise CALDERON[1], Emily HALLAMASEK[2], Franklin ZHENG[1], Julia HOWARD[2], Nabil MOHAMMED[1], Rhea KAMATH[1], Rose DELVILLAR[1], Stephen HWANG[1], Tanya IVANOV[1], TaraBryn GRISMER[3], Tobin BERGER-CAHN[2], Torrey BROWNELL[2], Wen LIU[1], & Yi Chi CHU[2]

Thesis Advisor

Dr. David BERNICK
Professor of Biomolecular Engineering

UNIVERSITY OF CALIFORNIA, SANTA CRUZ

1156 HIGH STREET, SANTA CRUZ, CA 95064

[1]Department of Biomolecular Engineering
[2]Department of Molecular, Cell, and Developmental Biology
[3]Department of Chemistry and Biochemistry

Abstract

Virulence factors and antibiotic resistance genes in certain bacteria aid in pathogenicity and host infection. These undesirable genes are characteristic of pathogenic strains of bacteria, such as Shiga toxin-producing *Escherichia coli* (STEC), one of the leading causes of food recalls in the United States. Individuals who consume STEC-contaminated foods suffer stomach cramps diarrhea, vomiting, fever, and in severe cases, kidney failure and death. Here we develop a method to target the spread of STEC at its source, preventing outbreaks and recalls through the sequence-specific elimination of the Shiga toxin gene, *stx2*, in pathogenic *E. coli* without disturbing commensal bacteria. We use chromosomally-integrated *mcherry* in *E. coli* as a model for elimination of *stx2* in a mixed-culture. Our sequence-specific gene elimination approach, Progenie, involves three aspects: delivery, elimination, and propagation. We use an M13 phagemid that disrupts gene expression through the insertion of DNA encoding an autonomous CRISPR-guided DNA integration system into the target gene. The delivered plasmid can propagate through a mixed-culture by conjugation, removing the target gene from a bacterial population without disrupting or killing microbial communities. As a proof of concept to help local farmers and distributors avoid recalls and reduce food waste, our sequence-specific elimination of *stx2* demonstrates the generalizability to eliminate pathogenicity and undesirable genes in other microbes.

Acknowledgements

Progenie thanks Dr. David Bernick, our principal investigator, and James Hahn, our graduate student mentor, for assistance and guidance on Progenie. We thank Maria Diaz Angel (Hartnell College) for supporting Progenie as an intern. We thank IISER-K's Namooste iGEM team for their help on our mathematical population modeling. We acknowledge the following individuals for informing our human practices research: Dr. Arie Havelaar (University of Florida) for discussing the global impacts of Shiga toxin-producing *E. coli*, Dr. Joji Maramuto (University of California, Santa Cruz) for discussing Shiga-toxin producing *E. coli* pathology, Drew McDonald (Taylor Farms) for explaining the preventative measures against foodborne illness in leaf vegetables, Diego Vasques (Pacific International Marketing) for explaining outbreak prevention, Eric Wilhelmsen and Chris McGinnis (SmartWash Solutions) for discussing the food sanitation, packaging, and testing industry, Elis Owens (Birko Corporation) for discussing potential applications of Progenie. We acknowledge the following individuals for informing our experimental design research: Dr. Josh Kittleson (Bolt Threads) for bacteriophage vector advice, Dr. Manel Camps (University of California, Santa Cruz) for conjugation and flow cytometry advice, Dr. Chad Saltikov (University of California, Santa Cruz) for advice on conjugation experiments, Mays Salih (University of California, Santa Cruz) for flow cytometry advice, and Leo Vo (Columbia University) for advice on utilizing the INTEGRATE system. We acknowledge the Haussler Lab and Carpenter Lab (University of California, Santa Cruz) for use of the SpectraMax ID3 and Attune NxT instruments, respectively. Progenie is funded by Aether Biomachines, the College Ten Senate, the College Ten Student Project Fund, the Cowell Senate, the Cynthia Matthews Scholarship, the Foley-Mendelssohn Travel Scholarship, Kevin Karplus, the Merrill Senate, the Porter College Creative and Innovative Grant, the Student Fee Advisory Committee, the UC Santa Cruz Alumni Special Project Fund, and the Undergraduate Research in Science and Technology Award. In addition, we acknowledge the iGEM Competition sponsors Benchling, Geneious Prime, Integrated DNA Technologies, and NEB for providing the tools and materials that makes Progenie possible. We would like to thank BioLink Depot for providing laboratory equipment and supplies. We thank the Sternberg Lab (Columbia University) for providing pSPAIN and pSPIN plasmids. We would like to thank the University of California, Santa Cruz and the Jack Baskin School of Engineering for supporting Progenie, and the next generation of synthetic biologists.

Contents

List of Figures

Chapter 1

Introduction

Generalizing a strategy to both target and edit the genome of a microorganism within a mixed community in a loci-specific manner opens up the possibility of controlling a microbial population at the most fundamental level [1]. Editing microbes in their natural community may also provide insight on how individual species influence others. Recently, researchers have optimized ways for integrating large genetic payloads into bacterial chromosomes in a highly sequence-specific manner [2, 3]. If the correct genetic payload is integrated into the proper locus, gene expression in that region can be extinguished [3]. We aim to apply these gene-editing technologies to pathogenic bacteria and increase the efficiency of gene elimination by making the system propagable in a bacterial population. Additionally, applications of highly specific targeting technology include microbiome gene therapies that could be extended to soil, plants, or even humans.

For example, Shiga toxin-producing *Escherichia coli* (STEC) is one such microbe that could be targeted with this gene-editing technology. STEC is one of the top three leading causes of food recalls in the United States due to its ability to infect humans through contaminated meat and produce [4]. Although the data on STEC outbreaks are limited, it is estimated that over 2.8 million people globally are infected with STEC every year [4]. Symptoms of STEC infection include bloody diarrhea, hemolytic uremic syndrome (HUS), and even death. HUS—a type of kidney failure—is the most common disease caused by STEC infection, and predominantly affects infants and children [5]. When STEC-contaminated food is ingested, Shiga toxin enters the bloodstream from the intestines and damages glomeruli in the kidney. Damaged blood cells and platelets clog the glomeruli, disrupting normal function [6].

The Shiga toxin-producing gene (*stx2*) originates from a prophage that integrated into the chromosome of some serotypes of *Escherichia coli* [7]. In our study, we focus on the most toxic form of Shiga toxin, Stx2A, produced by *E. coli* serotype O157:H7, as it is frequently responsible for severe STEC infections [7]. *E. coli* O157:H7 colonizes the gut of several different types of animals, such as sheep, goats, deer, and even some birds, but cattle have been identified as the source for three out of every four STEC infections in humans [8]. Transmission occurs predominantly through the shedding of infectious cells in feces, due to *E. coli* O157:H7 residing within the bovine rectum [9].

In this study, we aim to apply a programmable, gene-editing technology to the Shiga toxin-producing gene in STEC as a relevant proof-of-concept for our programmable system, hereinafter referred to as "Progenie." Additionally, we aim to engineer Progenie to be an autonomous system that can migrate through a population in order to ensure the persistence of the editing technology [3]. Thus, we investigated ways for Progenie to propagate throughout a heterogeneous bacterial population using a combination of bacteriophage vectors and conjugation. We test our design on a model *E. coli* strain with a chromosomally-integrated *mcherry* gene. Because Progenie has programmable RNA guides (gRNA), we hypothesize that a working *mcherry* elimination system can be expanded to *stx2a* using different guide sequences [3].

The *mcherry*-integrated *E. coli* model is described in detail in the following chapters, along with the designs of our propagable, programmable gene-elimination mechanism. We also describe a mathematical model that outlines the population dynamics of our system as it propagates through a mixed culture of bacteria. The model was created in collaboration with the IISER-K iGEM team, Namooste. (Appendix A.2 describes the details of our collaboration.) *Finally, we plan to explore field*

detection of Shiga toxin with DNA aptamers, which have the potential to help farmers and ranchers detect the toxins before shipping food.

1.1 *E. coli* O157:H7 transmission pathways

In order to design an effective technology to prevent STEC outbreaks, we investigated the source of food contamination as well as current food safety practices. We contacted various farms and cattle handlers in order to better understand how we can engineer our propagable gene-elimination system to be of practical use to those in need of it. The following sections give an overview of the transmission pathways associated with STEC infections and the legalities behind food safety in the United States and internationally.

1.1.1 *E. coli* O157:H7 transmission pathways

When we initially tried to identify the source of STEC in leaf vegetables, multiple professionals and professors with expertise in agricultural farming confirmed that contamination from cattle ranches directly affects leafy green farms. A soilborne pathogen specialist at UC Santa Cruz, Dr. Joji Muramoto, revealed that STEC is only known to contaminate leaf vegetables through tainted water or compost run-off containing manure. This would imply that contamination is specifically related to the presence of cattle in proximity to a leafy green farm. However, according to an interview with Drew McDonald, chairman of food safety at Taylor Farms, STEC outbreaks occur despite having no near-by cattle ranches. Theories about how STEC outbreaks occur without near-by cattle ranches include: *E. coli* O157:H7 that are indigenous to the land; the bacteria remain dormant until warmer environmental conditions allow it to bloom, making it difficult to detect early on [10]; or certain soil amendments farmers use for their produce may cause contamination; there may be common practices that farms use that are not yet known to cause contamination [11].

Due to the unidentifiable sources of contamination, we decided to investigate easy-to-use detection methods. Section 4.2 elaborates the methods of detection we explored. Despite leafy green farmers experiencing non-cattle related outbreaks, we continued to pursue a gene therapy for cattle due to beef being identified as the most STEC-contaminated food product in the Americas and in Europe [4]. A bovine-STEC vaccine developer [12], Dr. Matias Fingerman, later revealed that any cattle-targeting technology has to essentially be free for farmers because farmers have no economic or legal incentive to treat their cattle; they suffer no economic loss over meat recalls and there is no legislation requiring them to utilize new technology. Given that farmers are not likely to employ our elimination method, we identified other avenues in which a programmable and propagable gene-editing technology may be applied (further discussed in Chapter 5). However, we still pursue a *stx2*-targeting mechanism as a proof-of-concept to demonstrate the programmability of Progenie in a real-world application.

1.1.2 Current methods of addressing STEC outbreaks

We investigated where our technology could be applied as well as who would use it. As the result of a STEC outbreak that infected 140 people from romaine lettuce from Taylor Farms in 2018 [13], Taylor Farms increased their preventative measures to maintain the highest level of food safety. The chairman of food safety at Taylor Farms, Drew McDonald, revealed that leafy green sampling happens before harvest; rapid tests are used to minimize the time of transportation between the farm and the packaging facilities. Testing for contamination requires over 8,000 samples to identify a singular positive test in water. Additionally, well water is the only water source used to grow leafy greens because open-water sources get contaminated by wildlife or cattle farm run-off. They also treat their water with a chlorine solution to prevent contamination. Despite these preventative measures, there is no specific treatment to eliminate STEC.

After learning about Taylor Farms' in-depth testing and treatment plan, we spoke with other companies in the food-sanitation industry, such as Pacific International Marketing (PIM) and Birko Corporation, about their food-recall prevention measures. PIM regulates all water irrigation systems with a chlorine treatment to reduce potential *E. coli* contamination. Birko conducts multiple rounds of washes to eliminate *E. coli* from the surface of beef carcasses. Initial washes are done with water and are followed with lactic and citric acid washes, which have been shown to be effective at removing contaminates, according to Director of Technical Services Elis Owens at Birko. In terms of testing, PIM outsources testing to BioAg services, who collect and analyze food samples for coliform bacteria.

Even with all these preventative measures, the amount of food borne illnesses is expected to rise if additional safety measures are not taken [4]. Thus, as we proceed with our project, we will continue to work closely with food safety experts to develop the best technology that they would use in order to prevent STEC—and possibly other pathogenic bacteria—outbreaks in the future.

Chapter 2

Engineering a Programmable and Propagable Gene-Editing Mechanism

The objective of this study is to develop a method for eliminating or disrupting a gene of interest, while simultaneously propagating the mechanism throughout a bacterial metapopulation. We hypothesize that integrating an autonomous, CRISPR-guided DNA integration system [3] will disrupt the expression of a target gene and propagate through a mixed bacterial culture, ensuring the persistence of our editing mechanism amongst target cells. To test this hypothesis, we first engineered a host *E. coli* strain containing a chromosomally-integrated *mcherry* gene to model gene disruption. The following sections discuss our *mcherry* model and our gene-editing design along with any results we have collected so far. Construction of our gene-elimination system is currently underway, and we are still working on making our system propagable. This chapter will be updated as we build and test our plasmids.

2.1 Chromosomally integrated *mcherry* in *E. coli*

Fluorescent proteins such as mCherry have proven to be convenient tools for detecting the presence or absence of gene expression following genetic manipulation; the genetic sequences encoding these markers are often tagged onto proteins or transformed using plasmids. Because the toxin genes of many foodborne pathogens—including *stx2*—are chromosomal [7], we chose to integrate a fluorescent marker into the chromosome of *E. coli* to use as a model for a targeted gene disruption mechanism. Integrating *mcherry* into the *E. coli* genome, rather than using the *stx2* gene, avoids the production of the toxic Stx2A protein, which requires biosafety level (BSL) 2 containment. Upon verifying the efficacy of our gene elimination mechanism, we may begin working on eliminating the *stx* gene in *E. coli* O157:H7 in BSL-2 conditions. To generate the model *E. coli* organism, we used the optimized One-Step Integration Plasmid (pOSIP) to integrate the *mcherry* sequence into the chromosome of DH5α E. coli [14].

Cells possessing the chromosomally integrated *mcherry* gene will express the fluorescent protein; excitation with 590 nm wavelength light results in red fluorescence. If the gene elimination mechanism successfully disrupts the *mcherry* gene, cells will lose fluorescence. Because Progenie is programmable, we hypothesize that we can target the *stx2* gene for disruption by substituting the appropriate guide RNA (gRNA) sequence. This section discusses how we engineered our *mcherry*-integrated host *E. coli* strain.

2.1.1 Clonetegration

The pOSIP-KL-mcherry (Addgene) plasmid encodes for the machinery needed to integrate the *mcherry* gene into the chromosome of *E. coli* using "clonetegration", a method for inserting a gene of interest into a bacterial chromosome with phage integrase proteins [14]. Upon transformation of the pOSIP-KL-mcherry into *E. coli*, the plasmid's phage attachment site (*attP*) associates with the bacterial attachment site (*attB*) in the cell's genome. An integrase from λ phage catalyzes the integration of the complete pOSIP plasmid into the chromosome at the *attB* site. The pOSIP

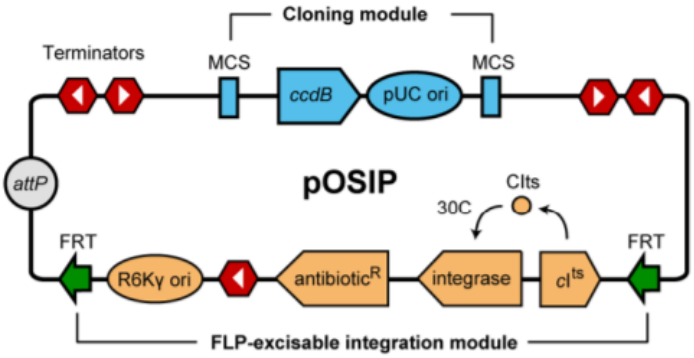

FIGURE 2.1: Figure of the pOSIP plasmid map with an intact cloning module. The *mcherry* gene replaces the cloning module in the pOSIP-KL-*mcherry* plasmid used to generate our *mcherry* model. This plasmid includes the payload for integration, *attP* site, and the integration module. The integration module consists of a phage integrase under a temperature sensitive λ repressor, which is active at 30°C. Integrase expression occurs within a 1 hour recovery at 37°C during transformation. The FRTs are used by Flippase to remove the integration module from the chromosome after the initial integration. We used the pE-FLP plasmid to express Flippase for excision of the integration module. Reproduced from St-Pierre, F.et al. *One-Step Cloning and Chromosomal Integration of DNA*. ACS Synthetic Biology 2,537–541. ISSN: 2161-5063, 2161-5063.(Sept. 20, 2013).

plasmid also encodes flippase recognition targets (FRTs), which are required for the excision of the integration module (λ integrase, λ repressor, kanamycin resistance, and origin of replication) from the chromosome upon transformation of a flippase-expressing plasmid, pE-FLP [14]. The final result is the integration of *mcherry* into the *attB* site in the bacterial genome. Figure 2.1 shows a schematic of the pOSIP plasmid.

2.1.2 Results: Engineering DH5α-*mcherry* model

Clonetegration is incompatible with Overexpress C41 pLysS *E. coli*. We initially attempted the clonetegration of pOSIP-KL-mcherry into Overexpress C41 pLysS *E. coli* cells (Sigma-Aldrich) [15]. Upon transforming competent cells with pOSIP-KL-mcherry, we observed growth on kanamycin plates. After restreaking on a fresh kanamycin plate, growth indicated that integration of the complete pOSIP cassette was successful because C41 cells lack the necessary machinery to replicate pOSIP-KL-mcherry [14]. However, we were unable to excise the integration module from these C41 using pE-FLP. After transformation with pE-FLP yielded no colonies on agar plates containing ampicillin, we repeated the transformation with known-competent C41 cells as a positive control. We were unable to transform pE-FLP into native Overexpress C41 pLysS *E. coli*, which revealed that the C41 strain's inability transform pE-FLP is likely due to plasmid incompatibility between ori sites on pE-FLP (a pSC101 ori) and the native pLysS plasmid (a p15A ori [16]).

Clonetegration into F′ WM3064 may be incompatible due to their ability to replicate the pOSIP plasmids. We integrated pOSIP-KL-mcherry into WM3064 for future phage infection and conjugation experiments. After the initial transformation of pOSIP-KL-mcherry, the resulting colonies had higher fluorescence than the other pOSIP-KL-mcherry transformants. Further investigation in the literature revealed that the WM3064 strain is *pir+* [17], which means it expressed the Pir protein needed to replicate pOSIP after transformation [14]. The cells' ability to replicate this plasmid makes it difficult to verify if transformed colonies are chromosomally integrating pOSIP or simply replicating the plasmid (or both). Additionally, after successful transformation with pOSIP-KL-mcherry, we were unable to transform WM3064 cells with pE-FLP to excise the integration module. The inability of WM3064 to take up pE-FLP after pOSIP-KL-mcherry transformation is likely due to plasmid incompatibility between pE-FLP's pSC101 ori and pOSIP's R6Kγ ori [18].

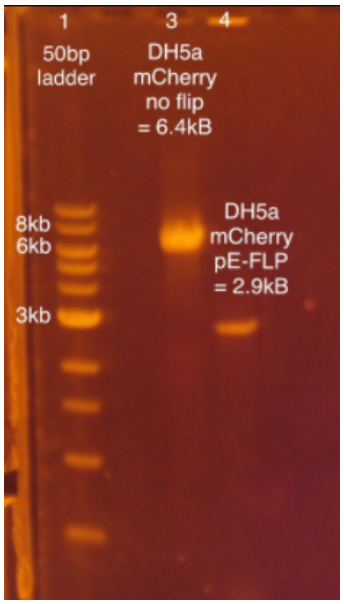

FIGURE 2.2: An image of an agarose gel of the colony PCR products from clonetegrated cells. We designed PCR primers for the *E. coli* genome to amplify the λ *attB* site. Before transformation of pOSIP-KL-mcherry, the expected PCR product is 182 bp. After transformation with the OSIP plasmid, the expected product is 6.4 kb, and the flipped integrant product is 2.9 kb. The untransformed colony yielded no result in lane 2.

Normally, plasmid incompatibility between the pOSIP and pE-FLP plasmids is not a concern, because pOSIP is not meant to be replicated by the host cell. Since WM3064 cells can replicate pOSIP and cannot be transformed with pE-FLP, we transitioned to using DH5α and BL21(DE3) in further experiments.

DH5α cells can successfully host the pOSIP cassette and can take up pE-FLP for removal of pOSIP's integration module. We successfully integrated pOSIP-KL-mcherry and excised the integration module in NEB DH5α *E. coli*. We verified successful integration into the λ *attB* site with colony PCR on native DH5α, DH5α integrants with no pE-FLP, and DH5α integrants after transformation with pE-FLP. An agarose gel of the resulting colony PCR can be found in Figure 2.2. We also streaked flipped integrants on kanamycin plates to test for the loss of the integration module. No growth on kanamycin LB/agar plates further verified that pE-FLP successfully removed the integration module from the DH5α-*mcherry* integrants.

Integrants expressing *mcherry* fluoresce more than native cells lines. We used a plate reader and flow cytometry to validate the levels of fluorescence in our *mcherry*-integrated *E. coli* model. Figure 2.3 shows plate reader data for non-fluorescent DH5α and DH5α-*mcherry* integrants. Figure 2.4 shows flow cytometry data collected for non-fluorescent DH5α, DH5α-*mcherry*, and WM3064-*mcherry* cells.

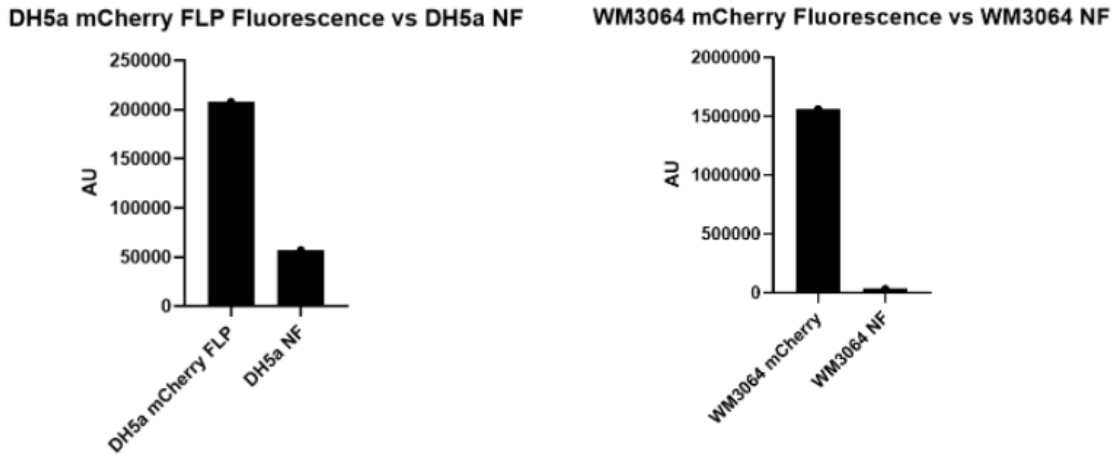

FIGURE 2.3: Plots showing plate reader data for non-fluorescing and fluorescing *E. coli* cells. Here, we see that our *mcherry*-containing cells appear to fluoresce more than untransformed cells. These results verified that our model *E. coli* strains are expressing fluorescent protein, and show a significant difference in fluorescence from an untransformed control. This data is not normalized, and the experiment will need to be repeated. For future experiments, we will normalize by diluting OD_{600} to 0.05 before analysis. After we build our gene-elimination system, we will have more data to show for different levels of decreased gene expression. Plots made with *Prism* [19].

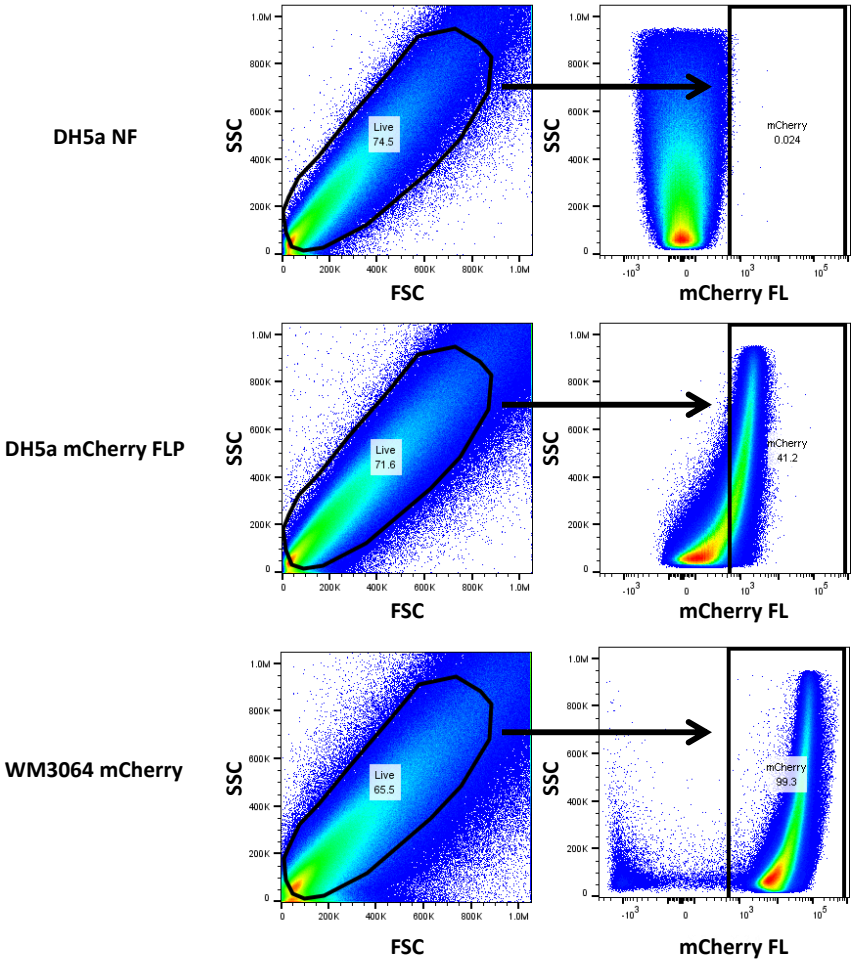

FIGURE 2.4: Flow cytometry analysis of DH5α-*mcherry* and WM3064-*mCherry* cells in comparison to a DH5α non fluorescent (NF) control. Red and orange indicate areas of high cell density; yellow indicates moderate cell density; and blue and green indicate low cell density. We excited mCherry with a 561 nm laser and detected mCherry fluorescence (mCherry FL) in the YL2 channel of an Attune NxT flow cytometer. Side scatter (SSC) and forward scatter (FSC) were detected and plotted on the left. SSC is correlated to cell granularity and FSC is correlated to cell size, allowing for identification of single live cells from a total population [20]. A subpopulation of live cells is gated from SSC and FSC dot plots and subsequently plotted against mCherry FL (right). Here, we observed approximately 41% of DH5α-*mcherry* cells to be *mcherry* expressing in the live single cell subpopulation, compared to 99% of WM3064-*mcherry* cells. Data analysis performed and figures made with *FlowJo* [21].

2.2 Gene elimination mechanism

We explored multiple methods of using the clustered regularly interspaced short palindromic repeats (CRISPR) system to eliminate a gene of interest in bacterial cells. CRISPRs are part of an acquired immune response of many bacteria in which chromosomally integrated bacteriophage DNA flanked by the palindromic repeats serve as a template for RNA transcripts to guide CRISPR-associated proteins (Cas) to a target sequence. Some Cas proteins, like Cas9, cleave DNA at the target site in order to render that DNA sequence unusable [22]. Recently, another CRISPR-Cas system has been discovered that performs integration of genetic payload into a target region, potentially eliminating gene function via insertional knockout [3]. The following sections discuss what gene-editing mechanism we chose to work with and how we constructed that mechanism into an expression vector.

2.2.1 Choosing a CRISPR system for propagable gene-elimination

Although many CRISPR-Cas systems exist for gene disruption, an optimal system should be encoded in the least amount of bases possible due decreasing plasmid transfer efficiency with increasing plasmid size using delivery vectors like conjugation and phagemid infection. Additionally, an M13 bacteriophage cargo capacity limits the plasmid size to approximately 12kb, explained further in Section 2.3.1. For the designed mechanism to be propagated through horizontal and vertical gene transfer, the cell must remain alive. DSB cleavage of genomic DNA in a bacterium lacking an efficient non-homologous end joining DNA repair machinery results in cell death, so the designed mechanism must also include alternative machinery [22]. We investigated two CRISPR systems for propagable gene-elimination: CRISPR-Cas9 and INTEGRATE.

CRISPR-Cas9, a class II CRISPR-Cas system, creates double stranded breaks (DSB) in the target DNA. For our system to propagate in a mixed community of bacteria, we need to reduce the toxicity of our gene-editing mechanism so that cells can pass on the necessary components to their offspring and to their neighbors (Section 2.4). To reduce toxicity of Cas9 DNA cleavage, Jian et. al and Pyne et. al included λ red phage components Beta, Gam, and Exo. These components generate and protect overhanging sequences with which homologous recombination can efficiently occur [23, 24]. We considered pursuing homologous recombination as a means of disrupting *stx2* gene function using Cas9 and λ red phage components, however disrupting *stx2* using homologous recombination and CRISPR-Cas9 alone requires a homologous template with which recombination can occur. This method requires the supplementation of the microbial community with recombinant oligonucleotides, which is a major obstacle in making the system propagable after many generations [23].

Another repair mechanism we explored involves adding non-homologous end joining (NHEJ) genes to a CRISPR-Cas9 system. Su et al. used a CRISPR-Cas9 system and NHEJ mechanism derived from *Mycobacterium tuberculosis* to substantially improve the ability of *E. coli* to survive and repair DSBs [25]. In the process of repair, researchers observed that DSBs are subject to damage within the cell and are imprecisely repaired, resulting in loss of function mutations, insertions, or deletions. Su et al. reported disruption of $64.5\% \pm 7.5\%$ of targeted *lacZ* gene function in transformed cells. We decided against this system due to the unreliability of random DNA mutations and due to the inability to insert a desired DNA sequence that would aid the system in propagation.

Chromosome cleavage may be avoided completely by utilizing dead Cas systems, such as dead Cas9 (dCas9), in which the catalytically active cleavage site is mutated, resulting in a loss of cleavage ability but retaining guide ability [26]. Utilizing dCas9, Banno et al. attached a cytidine deaminase, which allowed for specific deamination of cytosine to uracil [27]. Both the CRISPR-Cas9-NHEJ system employed by Su et al., and the dCas9-cytidine-deaminase system employed by Banno et al. are capable of crRNA directed mutagenesis of the Stx2A subunit. We chose not to use this system because, like using NHEJ genes with Cas9, this system does not allow the insertion of

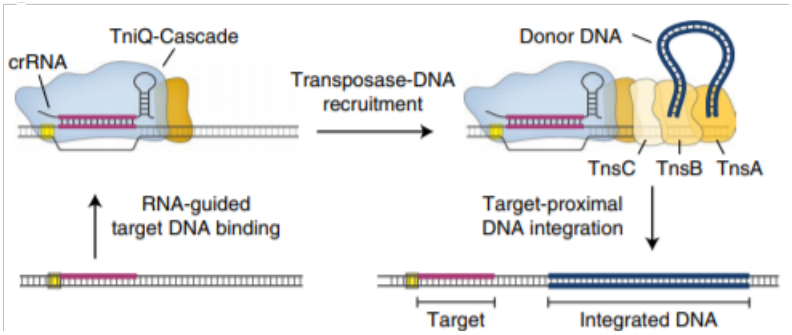

FIGURE 2.5: A schematic of the INTEGRATE system. The TniQ-Cascade complex is guided to the target locus by a programmable guide RNA. Once the complex reaches the target site, transposase proteins TnsA, TnsB, and TnsC associate with TniQ-Cascade, and catalyze the integration of donor DNA into a site roughly 50 base pairs downstream of the target. Reproduced from Vo, P. L. H., Ronda, C., Klompe, S. E., Chen, E. E., Acree, C., Wang, H. H., & Sternberg, S. H. (2021). CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. Nature Biotechnology, 39(4), 480–489. https://doi.org/10.1038/s41587-020-00745-y [3].

any cargo DNA, which prevents us from introducing a propagation mechanism. All three systems described so far utilize Cas9 as the guide effector, and retain similar on-target accuracy [24, 25, 27].

Alternatively, Klompe et al. demonstrated a three-plasmid system for integrating a genetic payload into sequence-specific loci in bacterial chromosomes [2]. The first plasmid contained a CRISPR array followed by a TniQ-Cascade operon composed of *tniQ*, *cas5*, *cas7*, and *cas6*. The second plasmid had three transposase genes: *tnsA*, *tnsB*, and *tnsC*. The final plasmid was equipped with a customizable genetic payload for integration. Klompe et al. showed that they could use specific CRISPR guide RNAs to integrate the payload into target-specif loci. In a subsequent study, Vo et al. increased the efficiency and specificity even further by combining the three plasmids' components into a single-plasmid integration system termed INTEGRATE [3]. A schematic of the INTEGRATE system can be seen in Figure 2.5.

The INTEGRATE system uses CRISPR-guided DNA transposases for DNA integration in which a programmable guide RNA guides a TniQ-cascade complex to the target site [2]. TniQ, a homologue of *E. coli* TnsD [2], recruits the transposase proteins TnsA, TnsB, and TnsC. The transposase proteins catalyze the integration of the DNA cargo roughly 50 bases downstream of the target sequence. This genetic payload is defined by specific left and right transposon sequences flanking the cargo DNA [3]. According to Vo et. al, the efficiency of integration is greater than 90%, and on-target specificity is greater than 95%. Additionally, Vo et al. developed a single-plasmid autonomous INTEGRATE system (pSPAIN) in which the transposon ends flank the entire INTEGRATE gene cassette. Figure 2.6 shows a schematic of the pSPAIN plasmid. In this system, the integrated DNA cargo is the entire INTEGRATE cassette spanning from R to L transposon ends, which will allow for propagation of the system via vertical gene transfer. We are pursuing insertional-gene knockout utilizing the autonomous INTEGRATE system, first in the *mcherry* gene of our model *E. coli*, then in the *stx2* gene of *E. coli* O157:H7 [7]. We hypothesize that inserting a span of DNA into the gene's promoter or the active site coding sequence will render the toxin protein harmless without killing the bacteria.

2.2.2 Construction of an *mcherry*-targeting CRISPR editing system

Engineering the *mcherry* integration plasmid (pMINT) requires alteration of the guide RNA sequence in pSPAIN. Figure 2.6 shows a schematic of the pMINT plasmid, and a detailed plasmid map can be found in Appendix B.4. Geneious Prime [29] was used to predict 34-nt target sequences with a 5′ CC PAM [3] in the *mcherry* gene, and chose targets that lead to the integration of our system into either the promoter of *mcherry* or into the beginning of the gene itself. From

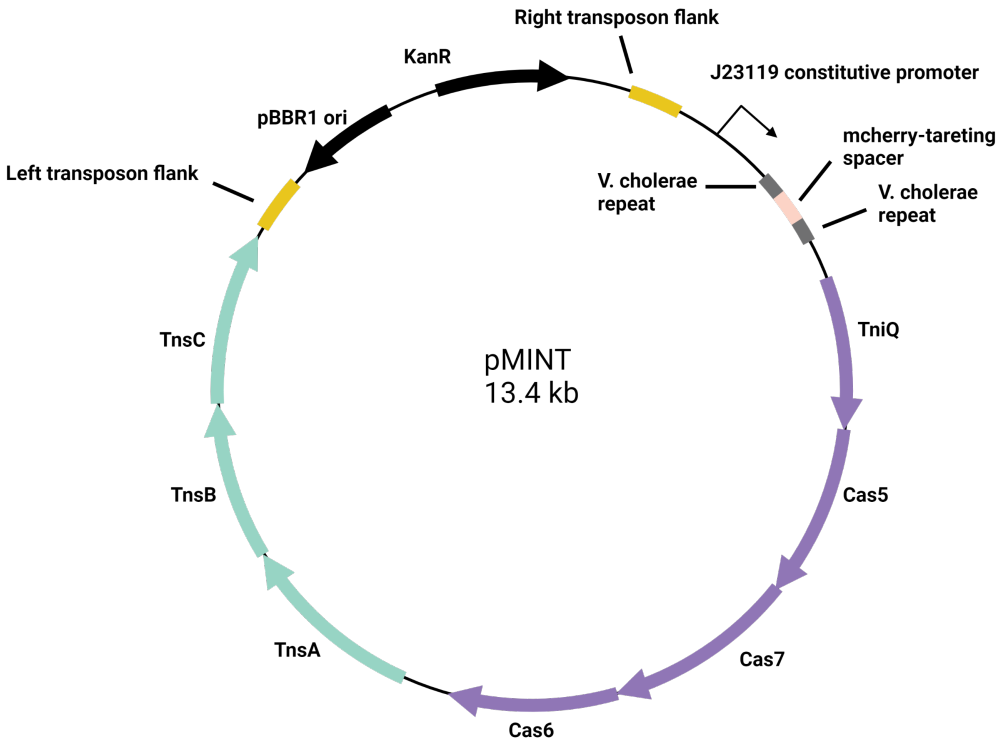

FIGURE 2.6: Plasmid map of the entire INTEGRATE gene cassette, which includes the CRISPR array, TniQ-Cascade complex (*tniQ, cas5, cas7, and cas6*) and transposase genes *tnsA, tnsB, and tnsC* all under control of the constitutive J23119 promoter. The CRISPR array includes Cas guide sequences that target the *mcherry* promoter in our fluorescent *E. coli* cells. The L and R transposon ends flank the entire INTEGRATE cassette, marking the transposon cargo region. Figure adapted from Vo et. al [3] with *Adobe Illustrator* [28].

the target sequences, we picked three guides for insertion into the promoter and two for insertion into the gene. The sequences for the targeting spacers can be seen in Appendix B.2.

We first attempted site-directed mutagenesis (SDM) to insert five different *mcherry*-targeting spacers into the pSPAIN plasmid. We ran the SDM PCR according to the methods listed in Section 6.3. Competent DH5α cells were transformed with the final SDM reaction mix, and grown on antibiotic media. We tested for the presence of the CRISPR array in the five potential pMINT constructs with colony PCR. Upon running an agarose gel on the PCR products, we observed no bands at the expected size for all pMINT samples. The pSPAIN positive control had a band at the expected 227 bp, indicating that the primers work and that SDM failed.

The SDM likely failed during the inverse PCR step of pSPAIN—a 13.4 kb plasmid. The SDM primers used in the inverse PCR were suboptimal with, with the melting temperatures (T_m) as high as 82°C— more than 10°C higher than the working temperature of the Q5 HiFi DNA polymerase [30]. Running the PCR with 10% DMSO reduced primer T_m by 5.5 to 6°C, but even with the added DMSO, the melting temperatures were too high. The SDM primer melting temperatures are high due to the complementary annealing bases that span the 28-nt direct repeats in the CRISPR array. Running SDM at a lower temperature results in non-specific binding of SDM primers and a failed SDM.

After site-directed mutagenesis failed, we developed a different method for cloning in the desired CRISPR arrays into pSPAIN using restriction enzymes and Golden Gate assembly. The CRISPR array in the original pSPIN contains BsaI restriction enzyme sites in its spacer sequence that allow for scarless Golden Gate assembly of any spacer sequence into the proper locus on the plasmid; thus, we wanted to put the pSPIN CRISPR array into the pSPAIN backbone in order to easily add any guide sequence. Using SalI and BamHI restriction enzymes to digest pSPAIN and pSPIN, we plan to remove the CRISPR array from pSPIN and clone it into pSPAIN. After restriction enzyme cloning pSPIN's CRISPR array into pSPAIN, we will perform Golden Gate assembly of *mcherry*-targeting spacers into pSPAIN to complete construction of our pMINT variants. We also decided to go forward with only the three *mcherry*-targeting spacers that cause integration of the entire INTEGRATE system into the promoter of the *mcherry* gene. Further verification of the successful cloning of pMINT constructs will include Sanger sequencing of the CRISPR array as well as sequencing the entire plasmid. Spacer sequences and Golden Gate gene blocks can be seen in Appendix B.2.

2.2.3 Testing viability of gene disruption in *mcherry* model cells

We successfully clone three *mcherry* promoter-targeting spacers into pMINT, and tested each plasmids' ability to disrupt *mcherry* gene expression and reduce fluorescence in DH5α-*mcherry E. coli*. Upon transformation of pMINT into competent DH5α-*mcherry*, the INTEGRATE gene cassette will be inserted into the promoter of the *mcherry* gene, reducing the level of mCherry protein production. Dependent on mRNA and protein degradation, a reduction in the levels of gene expression of mCherry will translate to a reduction in fluorescence in the cell population. Figure 2.7 shows a plot of relative fluorescence of cell cultures transformed with the different pMINT plasmids.

After performing these experiments, we will chose the guide 2 as our guide for following experiments.

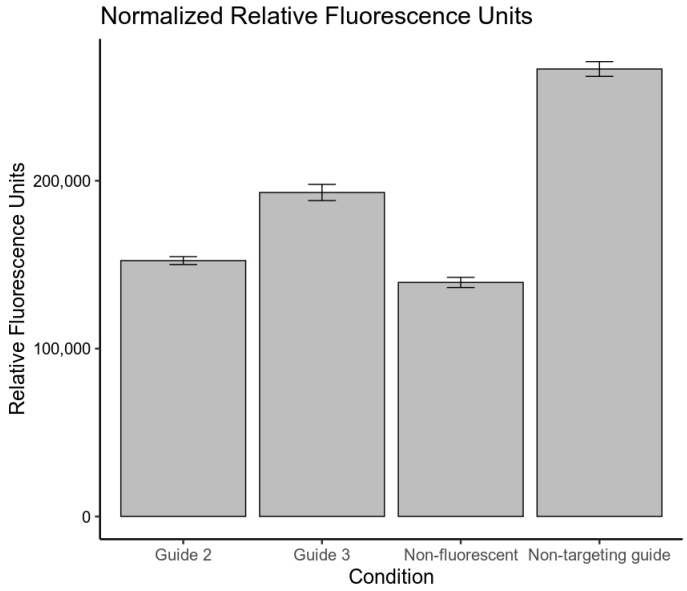

FIGURE 2.7: Plot of relative fluorescence units for plain DH5 transformed with control plasmid and DH5-mcherry transformed with pMADRID (non-targeting guide), pMINT Guide 2, and pMINT Guide 3 as measured using a plate reader. Data was normalized using OD600. Results indicate transformation of pMINT Guides 2 and 3 decrease fluorescence of DH5α-mcherry integrants. A Mann-Whitney U test performed between the DH5-mcherry non-targeting control and pMINT Guide 2 ($p = 2.20E\text{-}10 < 0.05$) and pMINT Guide 3 ($p = 1.48E\text{-}8 < 0.05$), indicate significantly decreased fluorescence.

2.3 Using bacteriophages for initial delivery

After we engineer a plasmid encoding a specific gene-targeting and elimination mechanism, we need a way to deliver it to our target microbial population. Thus, we needed a delivery vector that is large enough to carry our machinery, can target a specific limited host range, and is nonlytic.

Initially, we considered using a bacterium engineered to carry our plasmid like a probiotic. A bacterium has no size restriction for our plasmid and also does not kill hosts. However, transmission would rely solely on bacterial conjugation (explained in Section 2.4), meaning that our bacterial vector would have to be able to transport plasmids to recipient cells that don't have pili [31]. We plan to apply Progenie to Shiga toxin-producing *E. coli*—a bacterium that already has a pilus, meaning it could not be a recipient of our plasmid during conjugation [32]. Thus, we explored using bacteriophages to transport our engineered plasmid delivery method because they are able to delivery to cells with pili.

2.3.1 Use of bacteriophages as vectors

Bacteriophages (phages) are viruses that selectively infect bacteria. They consist of a protein-based capsid that transports the viral genome into a susceptible host cell upon infection. During the typical viral life cycle, the virus infects a host, hijacks the host's cellular machinery to replicate its own genome, packages the genome into capsids, and releases them into the environment. Bacteriophages are strong potential tools for inserting, modifying, or destroying bacterial genes of interest due to their specificity in host, simple construction, and ability to transport genetic material between hosts [33–35].

We will utilize particles derived from a bacteriophage to deliver our engineered genes to target cells in the population. These particles, known as pseudovirions [36], utilize the virus's capability to package and transport genes while avoiding real viral infection. Pseudovirions are constructed *in vitro* by providing a lab strain of bacteria with two plasmids: a helper phage and the phagemid.

The helper phage contains most of a bacteriophage's genome, encoding for the capsid structural and packaging proteins. However, the phage origin of replication (ori) is damaged, so the helper phage cannot be packaged into viral particles. The phagemid contains the desired synthetic cargo, a resistance marker, and two origins of replication: one bacterial ori and one phage ori [37]. The bacterial ori allows the phagemid to be replicated in the host bacterial cell, while phage ori is included so the helper phage will recognize the phagemid as DNA to be packaged into pseudovirions [36]. The pseudovirions carrying the phagemid will be isolated and applied to a target population of bacterial cells (see methods in Section 6.5). When a target cell is infected by a pseudovirion, the cell receives the phagemid DNA and begins expressing the synthetic genes as intended.

Since we aim to specifically target *E. coli* O157:H7, we investigated two *E. coli*-specific bacteriophages: M13 and P1. The M13 is a filamentous, single-stranded DNA virus that infects *E. coli* possessing pili, such as the strain WM3064 [38]. The M13 genome is 6.4 kb, coding for 11 structural and assembly proteins. The phage particle binds to the pilus of a bacterium before injecting its genome into the cell. The growth in a bacterium infected by M13 is attenuated, but hosts do not die because the phage is nonlytic [38]. The P1 phage, alternatively, is a head-tail, double-stranded DNA virus of *E. coli*. The P1 genome is 93.6 kb, coding for at least 117 genes [39]. After infection, P1 lies in a dormant lysogenic cycle, transferring its genome to daughter cells during cellular replication of the host. However, P1 is a temperate phage; when the inhibitory protein Coi is present, the phage's C1 repressor is blocked and the commences the lytic cycle, killing the host and releasing viral progeny [34].

Both of these phage options could be utilized in a similar manner as the delivery vector for the phagemid. However, there are two key differences: genome size and replication cycle. M13 and P1 have different capsid sizes, which influences the size of the cargo enclosed. M13 has a 12 kb capacity of ssDNA, whereas P1 has a capacity greater than 90 kb dsDNA. Depending on the final gene targeting and elimination mechanism, the size may exceed the 12 kb capacity of M13 phage. Despite the superior cargo capacity of P1, it must be accompanied by a lysogenic phage which ultimately leads to cell death [34]. Because the goal of Progenie is to edit the genome without killing the host, we ultimately decided to use the M13 phage as the initial delivery vector. M13 does not lyse its host, is extensively characterized, and is commonly used in other applications such as phage display [37, 40]. Finally, the INTEGRATE machinery is only 8.6 kb, which is small enough to use the M13 phage near its upper size limitation. Figure 2.8 summarizes pseudovirion formation.

2.3.2 Construction of a phagemid-deliverable INTEGRATE

We chose the M13-derived helper phage M13KO7 that contains 11 genes that encode for the various capsid and packaging proteins of M13, a Kanamycin resistance gene, and a p15A origin of replication. This helper phage recognizes plasmids that contain an F1 origin of replication, then packages it in single-stranded circular form.

For a plasmid to be packaged into the M13 bacteriophage capsid, the plasmid must include an F1 phage ori [37]. We designed a phagemid version of pMINT that contains an F1 ori for phage packaging (referred to henceforth as ϕMINT). Like pMINT, the M13K07 helper phage contains kanamycin resistance (KanR), so we designed ϕMINT to contain a chloramphenicol resistance gene (CmR) in place of the original kanamycin resistance gene in pMINT. Having two different antibiotic resistance markers will allow us to reliably select for bacteria that have taken up both the helper phage and ϕMINT. Figure 2.9 shows a plasmid map of the ϕMINT plasmid.

To construct ϕMINT, we will use Golden Gate assembly to add the F1 ori and chloramphenicol resistance gene into the pMINT plasmid. The added sequences were dsigned to assemble into the pMINT plasmid outside of the INTEGRATE transposon flanks, meaning that neither of the additional sequences will be part of the integrated payload. Adding the F1 ori makes the phagemid about 12.3 kb long, which is the upper size limitation for M13 packaging. We hypothesize that the

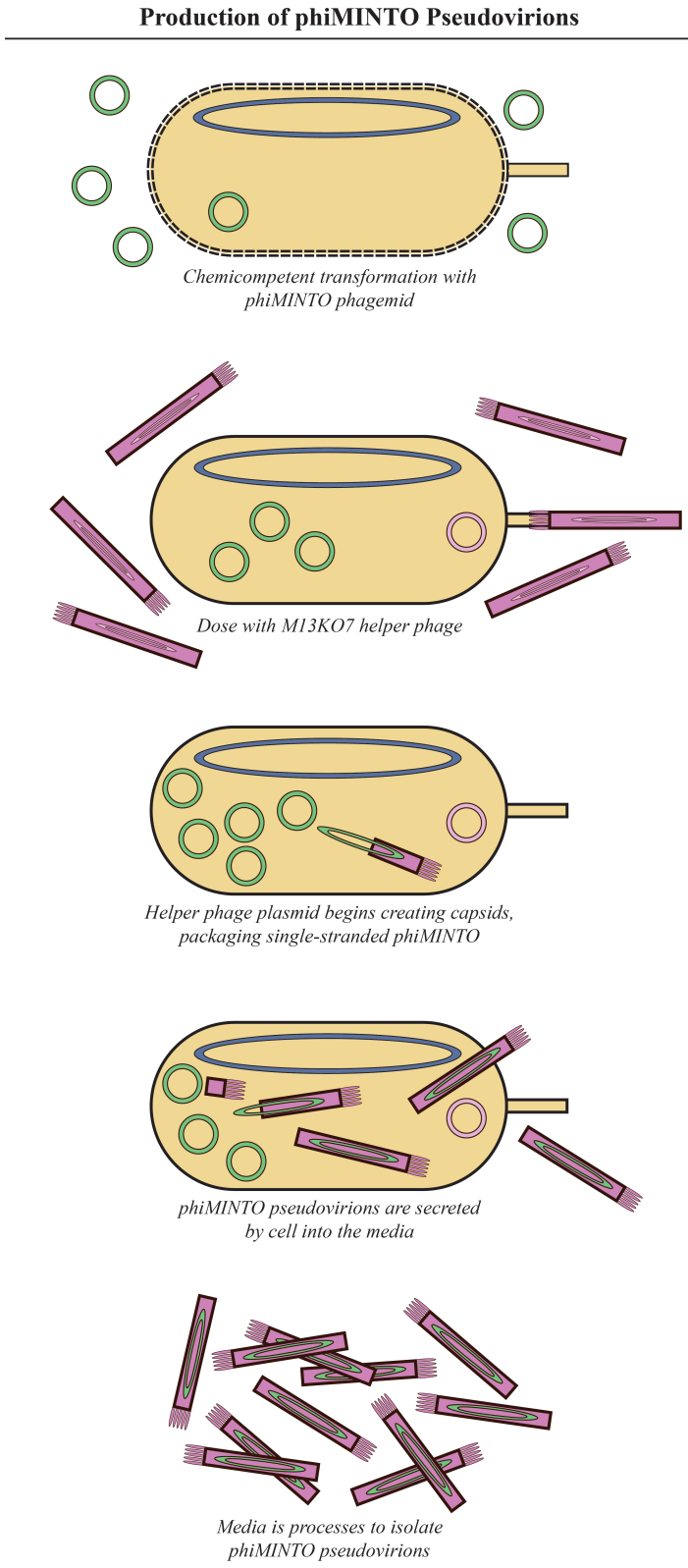

FIGURE 2.8: To produce pseudovirions, ϕMINT is transformed into a host cell while virions infect the host with the helper phage (we use M13KO7, refer to methods in Section 2.3.2). The helper phage (pink circle) forms M13 pseudovirions and packages single-stranded ϕMINT. The host secretes pseudovirions containing ϕMINT into the surrounding media, where they are isolated and purified for experimental use. Figure made in *Adobe illustrator* [28].

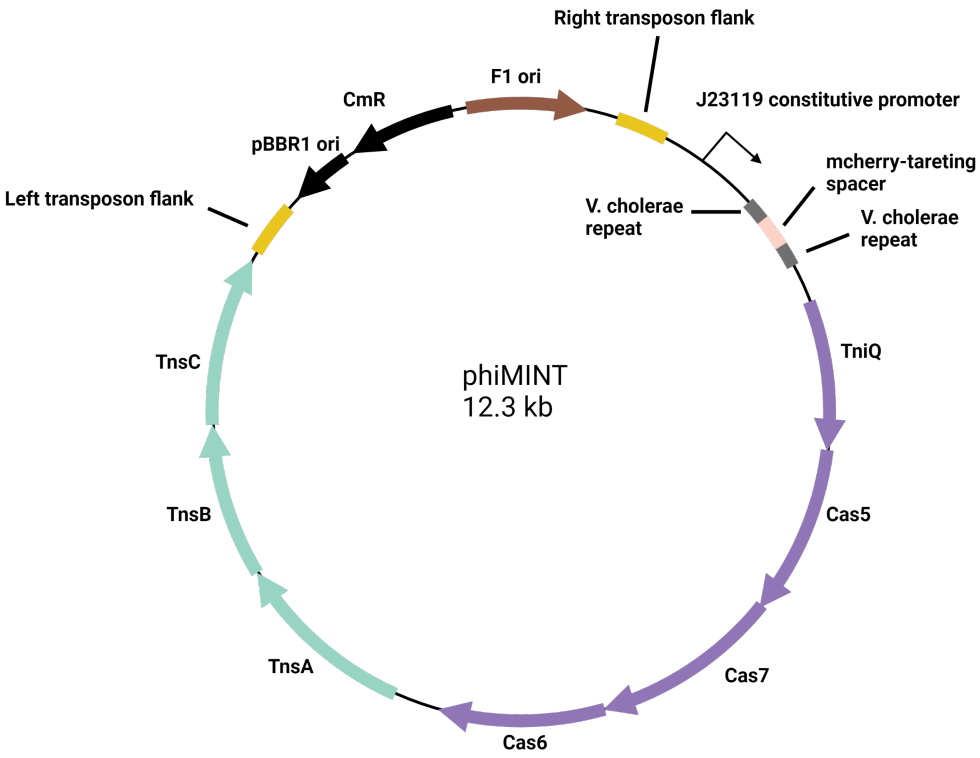

FIGURE 2.9: A plasmid map of our ϕMINT phagemid. The addition of an F1 ori will allow the M13 helper phage to package the phagemid for delivery. We will add the selective marker chloramphenicol (CmR) in order to use a different resistance than what the helper phage has (kanamycin) as we want to select for both plasmids separately. The F1 ori and CmR gene blocks will be assembled into the pMINT plasmid that shows the highest fidelity in disrupting *mcherry* gene expression. The ϕMINT phagemid design results in a length of 12.3 kb. Figure made with *Adobe Illustrator* [28].

addition of an F1 ori will still allow our system to be packaged into an M13 phage that can then infect *E. coli* cells with our gene-elimination mechanism. The host cell does not receive the helper phage during infection with ϕMINT and so cannot make more M13 phages. To allow Progenie to spread after this initial infection, we investigated the use of bacterial conjugation as a secondary method for horizontal gene transfer, described further in Section 2.4.

2.3.3 Phage infection experimental design

To test the delivery capabilities of ϕMINT, we plan to perform infection experiments using *E. coli* DH5α-F′. Because this strain has an F pilus, it is susceptible to infection by the M13 bacteriophage and thus our pseudovirions.

We first verified that the gene-elimination mechanism works by transforming the ϕMINT into mcherry-expressing DH5α-F′ cells. pMADRID, a plasmid that does not target mCherry was transformed into those mcherry-expressing DH5α-F′ as a negative control. Cells were grown on chloramphenicol agar plates to select for transformants. After an incubation period, plates were analyzed for the presence or absence of mCherry fluorescence. All fluorescence analysis was performed via microscopy, plate reader, or flow cytometry (see Section 6.2).

Once the phagemid's machinery was verified, we packaged and precipitated our pseudovirions. We transformed MINT into DH5α-F′ cells and selected for transformants with chloramphenicol. We then grew up the transformants and transduced that culture with M13KO7 phage (Thermo Fisher), the modified helper phage form of M13 phage. Cells that now had both helper phage and phagemid plasmids proceeded to package copies of the ϕMINT into M13 pseudovirions and

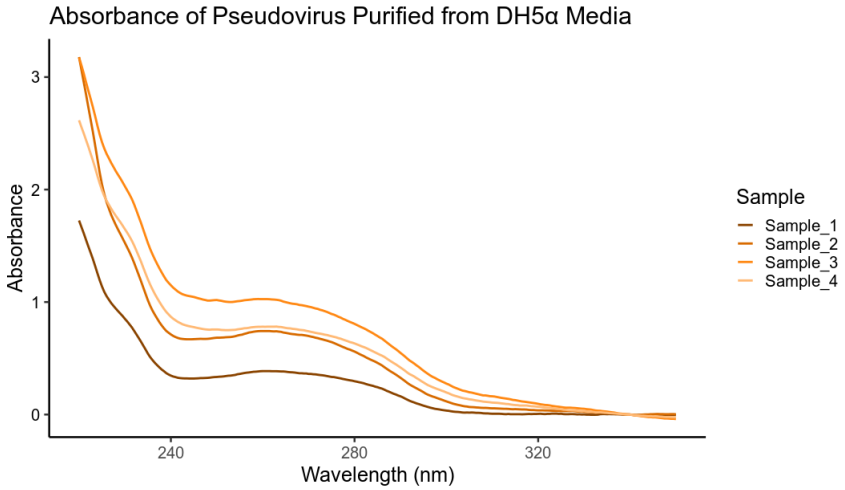

FIGURE 2.10: Graph of absorbance of media containing ϕMINT pseudovirions, showing the characteristic curve of filamentous phage. Sample contains high proportion of phage protein to DNA, resulting in a large arc between 260 nm to 280 nm, and may also contain contaminates, resulting in some absorbance around 320 nm. Absorbance results are used to calculate concentration of phage particles in media.

secrete them into the media. Concentration of purified virion samples was determined with spectrophotometry and the equation for quantifying filamentous phage (see Section 6.5).

Following quantification, we performed a DNA cleanup on the phage precipitate samples and ran a gel electrophoresis to visualize what sizes of plasmid were being packaged.

We then performed an initial phage-mediated fluorescence knockdown experiment. We grew up cultures of DH5α-mcherry in ampicillin media, then transduced with our ϕMINT-phage when the cell culture OD600 reached approximately 0.450. Transduction was performed with a flat 1uL of phage per 1mL of culture. Negative control culture received no phage.

Further experiments with the ϕMINT-phage will investigate the impact of varied multiplicities of infection (the MOI, the number of added per cell during infection), how transduction and gene elimination progress over time, and if ϕMINT-phage infection accelerates cell death. Infection rate and gene elimination success will be determined by analyzing the proportion of target cells with reduced fluorescence for each dosage of phage, at each time step, and by comparing relative amounts of dead cells. Then, the optimal MOI for this system that maximizes gene elimination with minimal damage will be determined.

2.3.4 M13 phage delivery of ϕMINT

***mcherry*-expressing cells have reduced fluorescence after transformation with ϕMINT.** We successfully reduced expression of mcherry through targeted gene knockout. When fluorescence output was measured with flow cytometry and plate reader, normalized with cell density, previously fluorescent cultures transformed with ϕMINT were measured to have fluorescence levels on par to the non-fluorescing control cells. As depicted in Figure 2.11, this significant reduction in fluorescence compared to their fluorescent counterparts is strong evidence that the ϕMINT gene-elimination mechanism was successful at stopping expression of the mcherry gene. The fluorescence level of ϕMINT-transformed cells was also on par with the pMINT-transformed cells, showing that the gene-targeting system is not affected by the phage delivery system.

Cells containing both phagemid and helper phage plasmids produce and secrete ϕMINT pseudovirions. To produce ϕMINT pseudovirions, ϕMINT DH5α-F' transformants were transduced with M13KO7 phage and incubated overnight. Media was processed to precipitate phage, allowing for verification of phage presence in each sample with the naked eye.

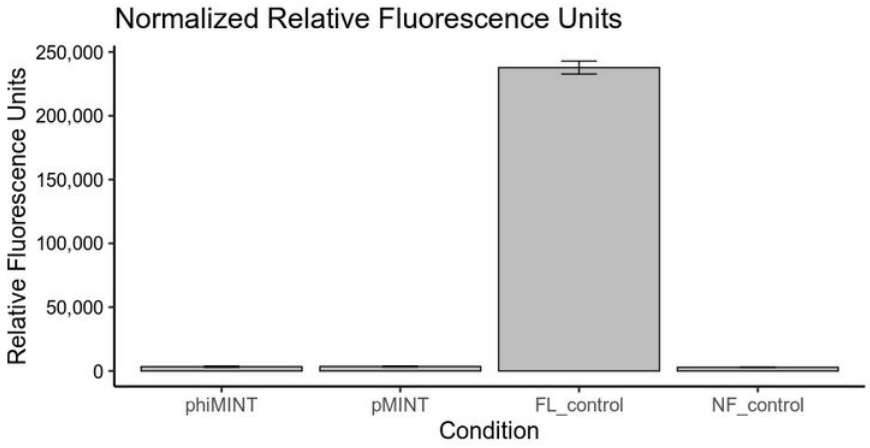

FIGURE 2.11: Plate reader data depicting the relative mCherry fluorescence of treated and control E. coli cultures, normalized with cell density. In order from left to right, bars depict DH5α-mcherry transformed with ϕMINT, DH5α-mcherry transformed with pMINT, DH5α-mcherry transformed with the non-targeting plasmid pMADRID, and plain DH5α as a non-fluorescent control.

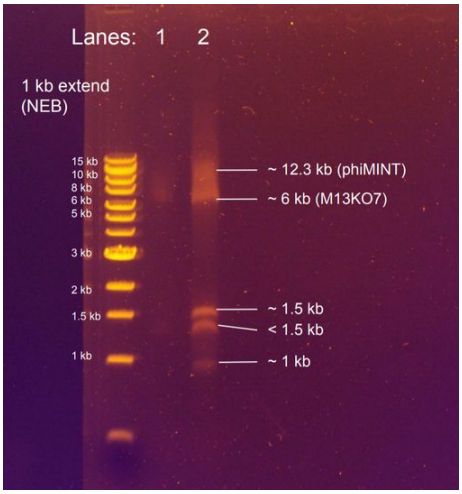

FIGURE 2.12: The DNA cleanup of media containing secreted packaged M13 pseudovirions run on gel electrophoresis. Bands and smear indicate DNA at the target length of the ϕMINT plasmid (12.3 kB), but also undesired lengths that are not full ϕMINT.

After precipitation, we calculated the concentration of phage in the media by measuring the absorbance of purified media using a spectrophotometer. We saw high absorbance between 260 nm to 280 nm with a relative maximum at 269 nm due to the high proportion of protein to DNA in filamentous phage [41]. We did also see a small amount of contaminants contaminants that have high absorbance around a wavelength of 320 nm [41]. From these data, we calculated that our sample contains 1.75×10^{12} pfu/mL.

One phage sample was processed using a DNA cleanup kit, then analyzed with gel electrophoresis. The gel indicates a presence of DNA at the proper length of ϕMINT (12.3kB), indicating that there were indeed full ϕMINT-phage in the sample (Figure 2.12). However, that DNA is not a defined band, but instead a smear that continues down to around 6kB. Thus, it is clear that our ϕMINT-phage production method gives us a mixture of full-length ϕMINT-phage and stunted pseudovirions.

***mcherry*-expressing cells have reduced fluorescence after infection with ϕMINT-carrying pseudovirions.** To confirm the infection ability of the ϕMINT pseudovirions into our target F'

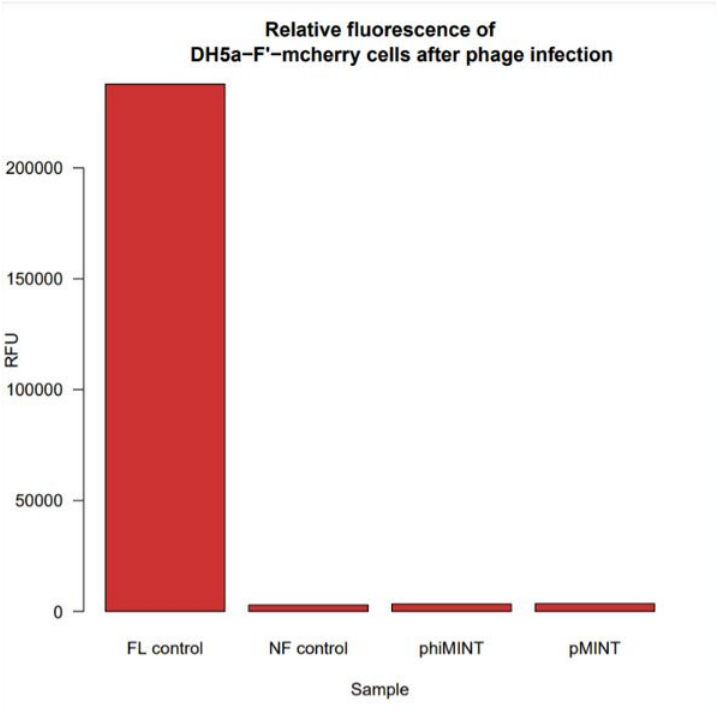

FIGURE 2.13: DH5α-F′-mcherry cultures were transduced with ϕMINT-phage, incubated with chloramphenicol, and grown at 30 C°. Relative fluorescence was measured on a plate reader and normalized with optical density. Samples transduced with ϕMINT-phage resulted in significantly lower levels of RFUs than the fluorescent control samples. These ϕMINT fluorescent levels were approximately even with the non-fluorescent DH5α cells and the pMINT-transformed cells.

cells, DH5α-F′-mcherry cultures were transduced with ϕMINT-phage and incubated with chloramphenicol. Fluorescence was measured with a plate reader and normalized with optical density. Once again, transduction with ϕMINT-phage resulted in significantly lower levels of fluorescence than the fluorescent control, with ϕMINT fluorescent levels approximately even with the non-fluorescent DH5α cells and the pMINT-transformed cells (Figure 2.13). In sum, ϕMINT-phage are highly effective at targeting and reducing expression of a desired gene.

2.4 Making the gene elimination mechanism propagable

Bacterial conjugation—a form of horizontal gene transfer—propagates genetic material through a population via plasmid-mediated transfer. It is essential for bacterial evolution because it often results in increased genetic diversity and fitness [42]. Antibiotic resistance, for example, spreads through a mixed population of bacteria through conjugation [43]. In addition to using bacteriophages to deliver our gene-editing machinery, we are exploring the use of conjugation to propagate our genetic payloads through a bacterial population.

2.4.1 Hijacking bacterial conjugation systems

Bacteriophage delivery is ideal for the initial infection of target cells with Progenie. The edited chromosome containing the non-functional target gene will be passed on through means of vertical gene transfer. However, it has no mechanism to propagate through the population horizontally, and is thus inefficient at affecting the entire target bacterial population. To surpass these limitations, we will employ bacterial conjugation in tandem with phage infection.

Bacterial conjugation, a form of horizontal gene transfer (HGT), propagates genetic material through a population via plasmid-mediated transfer [42], as seen in Figure 2.14. This is essential

in bacterial evolution as it often leads to genetic diversity and increased fitness. Antibiotic resistance, for example, spreads through a mixed population of bacteria through conjugation [44]. To deliver our gene-editing machinery, in this section we explore using conjugation in conjunction with bacteriophages to propagate our genetic payloads through a bacterial population.

Conjugation relies on the expression of *tra* (transfer) genes that code for the essential components such as a relaxosome and a conjugative pilus that contains type 4 secretion system (T4SS) proteins, also known as the mating pair formation (mpf) apparatus [45, 46]. The relaxosome, composed of a relaxase (formed by a group of proteins known as tra1) and three accessory proteins (TraY, TraM, IHF), nicks the dsDNA at a nic site located in the origin of replication for transfer (oriT). A type 4 coupling protein (T4CP) connects to the relaxosome, which is still attached to the 5′ end of the nicked sequence, in order to begin transfer of the nicked strand (the transfer strand) through the T4SS on the pilus into a recipient cell [46, 47]. The T4SS found in gram-negative bacteria have multiple roles: pilus biogenesis, DNA transportation, and protein transportation [46]—all of which are essential for bacterial conjugation. After transportation, the remaining ssDNA in the donor gets replicated to restore the complementary strand while the recipient cell replicates the transferred ssDNA, resulting a dsDNA in both the donor and recipient after conjugation [47].

All of the *tra* genes reside on a conjugative plasmid, such as the F plasmid found in *E. coli* by Lederberg et al. [48]. Entry exclusion genes present on such conjugative plasmids prevent conjugation of the same plasmid twice by expressing surface proteins that prevent the T-strand from entering into the recipient cell [49]. All conjugative plasmids have at least one entry exclusion gene, and the genes are hypothesized to be a safety mechanism to help avoid competition between identical plasmids within the host [50]. Additionally, there are two main types of conjugative plasmids: self-transmissible plasmids, those that have the complete T4SS genes to form pili and transfer over their genetic material on their own [51], and mobilizable plasmids, which only encode parts of the relaxosome and the oriT [52]. Mobilizable plasmids rely on self-transmissible plasmids to produce the rest of the machinery required for conjugation, thus self-transmissible plasmids are referred to as "helper" plasmids [50]. One such example of these two plasmids working in tandem is the IncQ RSF1010 mobilizable plasmid and IncPα RP4 conjugative plasmid [53]. On its own RSF1010 is able to nick its own oriT, but requires nine *trb* genes from RP4 for constructing the mpf apparatus and DNA transport [54]. We use a similar concept for propagation of our gene elimination vector by designing our conjugative plasmid to contain an oriT site (Figure 2.15).

For preliminary experiments on testing the conjugation ability of Progenie, we *will* use a known F′ cell line, *E. coli* WM3064. F′ strains contain the genes for conjugation on a plasmid that was once integrated into the chromosome of the cell. In the case of WM3064, it contains the IncPα RP4 conjugative system [55]. To hijack the encoded RP4 relaxosome replication machinery and mpf, we will add the RP4 oriT to our phagemid ϕMINT. The new phagemid with the oriT will hereby be called ϕMINTO. The presence of the oriT on the phagemid will allow it to be recognized and replicated by the WM3064 native RP4 relaxosome, and transferred to recipient cells via the RP4 mpf transfer system. In doing so, we have transformed our phagemid into a "mobilizable" unit, similar to the previously described IncQ plasmids, that can act as a passenger in conjugation encoded by a "helper" plasmid [56].

For experiments involving *E. coli* WM3064, the oriT is the only additional conjugative machinery needed for our phagemid [57]. However, in future applications involving other types of bacterial strains, ϕMINTO may need to contain its own relaxosome, oriT, and/or mpf, depending on how many functional conjugative genes the host bacteria contains. For example, pO157, the native F-like plasmid of *E. coli* O157:H7, is only able to make a type II secretion system and thus unable to mobilize on its own [58].

When constructing ϕMINTO for these experiments, plasmid incompatibility is the primary concern. Plasmid incompatibility occurs when certain plasmids are incapable of coexisting in a cell. Compatibility is determined by the similarity of the plasmids' vegetative replicon, which encodes and recruits replicative machinery for plasmid replication during cell growth and division [18]. If two plasmids within the same incompatibility group (Inc) are together in a cell, they

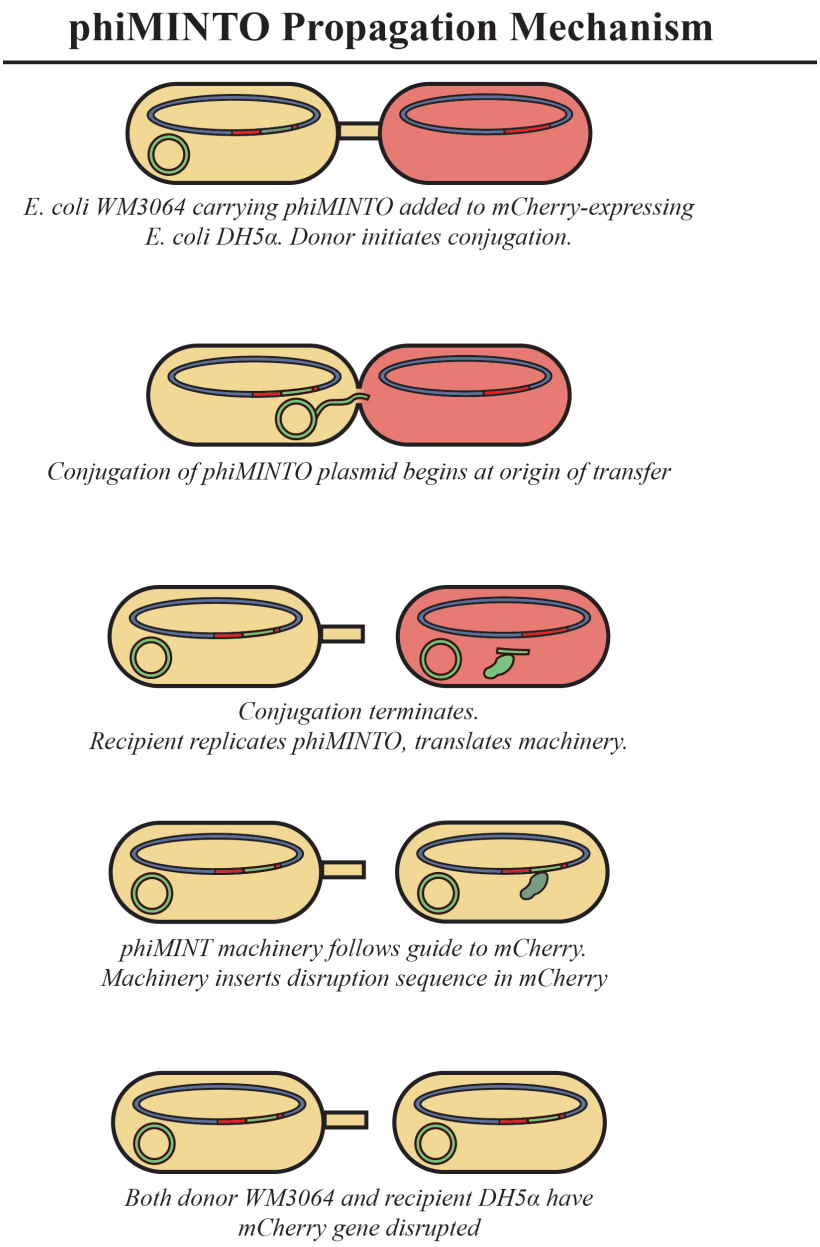

FIGURE 2.14: A schematic of ϕMINTO propagation via conjugation. Red cells express the and the yellow cells contains the ϕMINTO plasmid that disrupts *mcherry* fluorescence. Donor initiates the conjugation process at an oriT site, replicates the ϕMINTO plasmid, and uses its pilus to get the ssDNA plasmid into the recipient cell. The recipient replicates the transferred strand and expresses the INTEGRATE system, which mobilizes to disrupt *mcherry* and eliminate fluorescence. Figure made in *Adobe illustrator* [28].

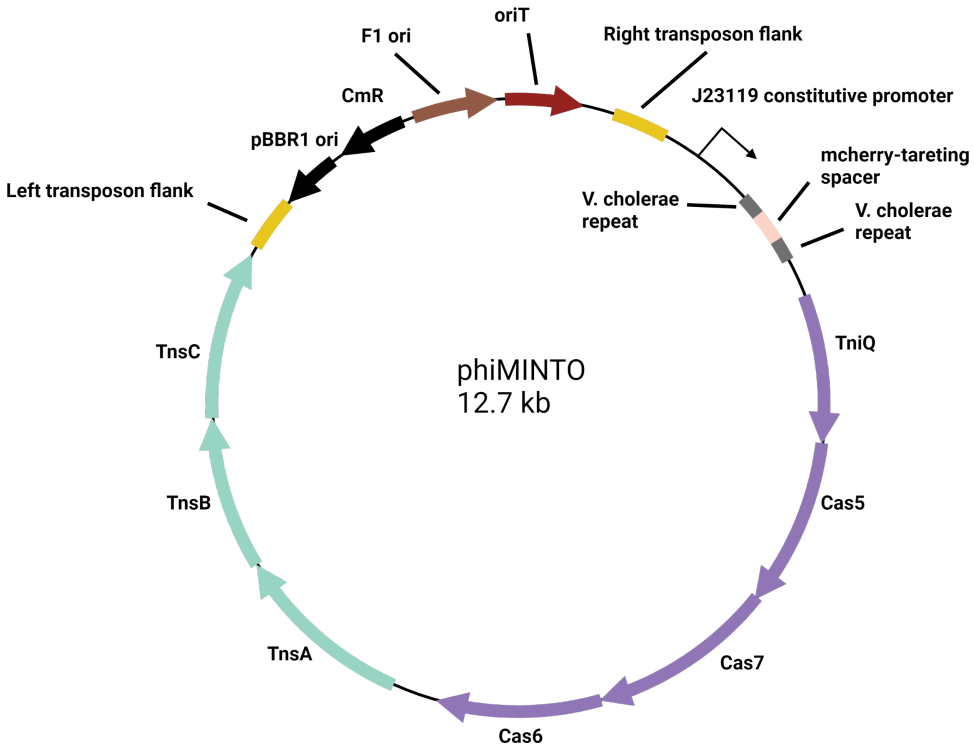

FIGURE 2.15: A plasmid map of the ϕMINTO phagemid. This phagemid contains an RP4 oriT, which will cause the phagemid to undergo conjugation via the host cell's own machinery. The F1 ori, RP4 oriT, and CmR gene blocks will all be assembled into the pMINT variant with the highest gene-elimination efficiency. The ϕMINTO phagemid is 12.7 kb. Figure made with *Adobe Illustrator* [28].

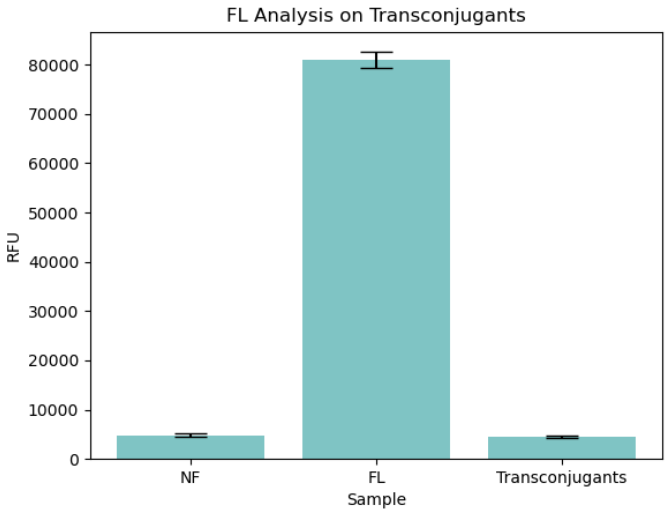

FIGURE 2.16: Plot of relative fluorescence units for DH5α-*mcherry* conjugative recipients as measured using a plate reader. Data was normalized using OD600. Error bars depict standard error. Results indicate succesful conjugation of pMINTO decreasing fluorescence of DH5α-*mcherry*.

compete for the same replicative machinery, which destabilizes the host cell and/or results in the expulsion of one of the plasmids [59]. However, it is crucial to recognize that incompatible plasmids have been experimentally observed to coexist in the same cell for days at a time before undergoing expulsion [60]. Despite the possibility of coexistence, we will design ϕMINTO to have a compatible replicon with any plasmids in *E. coli* WM3064 in order to ensure successful propagation during our conjugation experiments.

2.4.2 Conjugation from *E. coli* WM3064 to target DH5α

ϕMINTO can be mobilized from WM3064 to DH5α. The optimal ratio of donors to recipients is to be determined, however conjugation is known to vary in efficiency between 1:10 and 1:10^{-7} donor to conjugated recipient cells (according to M. Camps). Recipients are much more likely to participate in conjugation when surrounded by a large number of donors, thus we expect higher donor to recipient ratios to result in more gene knock out. Additionally, spatial distribution of donors and recipients affects conjugation efficiency in that conjugation requires cell-to-cell contact and type IV pili have limited extension range. While the exact conjugation efficiency is unknown, we validated through plate reader that transconjugate DH5α cells had their $mcherry$ gene succesfully knocked out. (Figure 2.16)

mcherry-integrated DH5α cells that receive ϕMINTO through conjugation have reduced fluorescence. The fluorescence of the DH5α-*mcherry* recipient cells to decrease upon exposure to the ϕMINTO-infected WM3064 cells when compared to non-conjugated *mcherry*-DH5α cells. The rate of decrease, however, is more difficult to hypothesize. We expect the rate of attenuation to correlate with the conjugation rate between non-*mcherry* WM3064 and DH5α.

Chapter 3

Mathematical Modeling

Detection and elimination of cattle-related pathogens are two of the main aspects of our project, Progenie, and that of another iGEM team, Namooste. The Indian Institute of Science Education and Research Kolkata (IISER-K), located in West Bengal, India, hosts an iGEM team of 13 undergraduate students currently engineering a method to detect and vaccinate bovine mastitis, a disease caused by *Staphylococcus aureus*. With the SARS-CoV-2 pandemic limiting Namooste's access to a wet lab, Progenie is supporting the Namooste project by conducting wet lab experiments to develop their system to detect bovine mastitis (Appendix A.2 details our wet lab colaboration). In return, members of IISER-K are sharing their experience in mathematical modeling in both Python and Matlab's [61] SimBiology [62] in order to develop a computational model of how our gene-editing system will behave in the environment. The following sections give more detail on the mathematical model we made to simulate Progenie.

3.1 Mathematical modeling of the Progenie system

The goal of our chromosomally integrated *mcherry* model *in vitro* is to serve as evidence that our gene elimination machinery can work on the molecular and genetic level. However, for real-world applications, the various scenes in which our system will be used differ greatly. We are developing a mathematical model that will allow us to simulate the population dynamics in various circumstances, helping to guide our dosing in future applications. This section discusses the mathematical model built in collaboration with IISER-K and how we propose to apply it to our propagable, gene-editing system.

3.1.1 Biological foundation, equations, and parameters

Although our propagable gene-elimination system is not a pathogen, nor does it cause disease, it does spread through a population of hosts in a similar manner to an infectious diseases. To define variables that model the infectious-like behavior of our system, we use a visualized adaptation of an SIR compartment model [63], a model commonly used in epidemiology to simulate disease **S**usceptibility, **I**nfection, and **R**ecovery (Figure 3.1). In our model, we represent the proportion of individuals susceptible to the initial phage infection as S_V, and those susceptible to conjugation as S_C. The proportion of infected individuals that can conjugate, I, are regarded as infectious, and the proportion of those that cannot initiate conjugation, E, are considered non-infectious. The proportion of phagemid-carrying pseudovirion particles to cells in the population is represented by P, with a virion-half life of ϕ.

Before the phages are introduced, the population is made up entirely of susceptible cells, meaning $S_V + S_C = 1$, because none have yet received the phagemid. This model does not consider any cell types that are resistant to conjugation and pseudovirion infection to be a part of the population, as they are irrelevant in transmission dynamics. After dosing the total population with phage, the number of cells in S_V immediately decreases as they receive the phagemid and move into the infectious group at a rate of V_I, or the non-infectious group at a rate of V_E. The rate at which S_V decreases depends on the amount of phage available to bind, and the rate of binding, B.

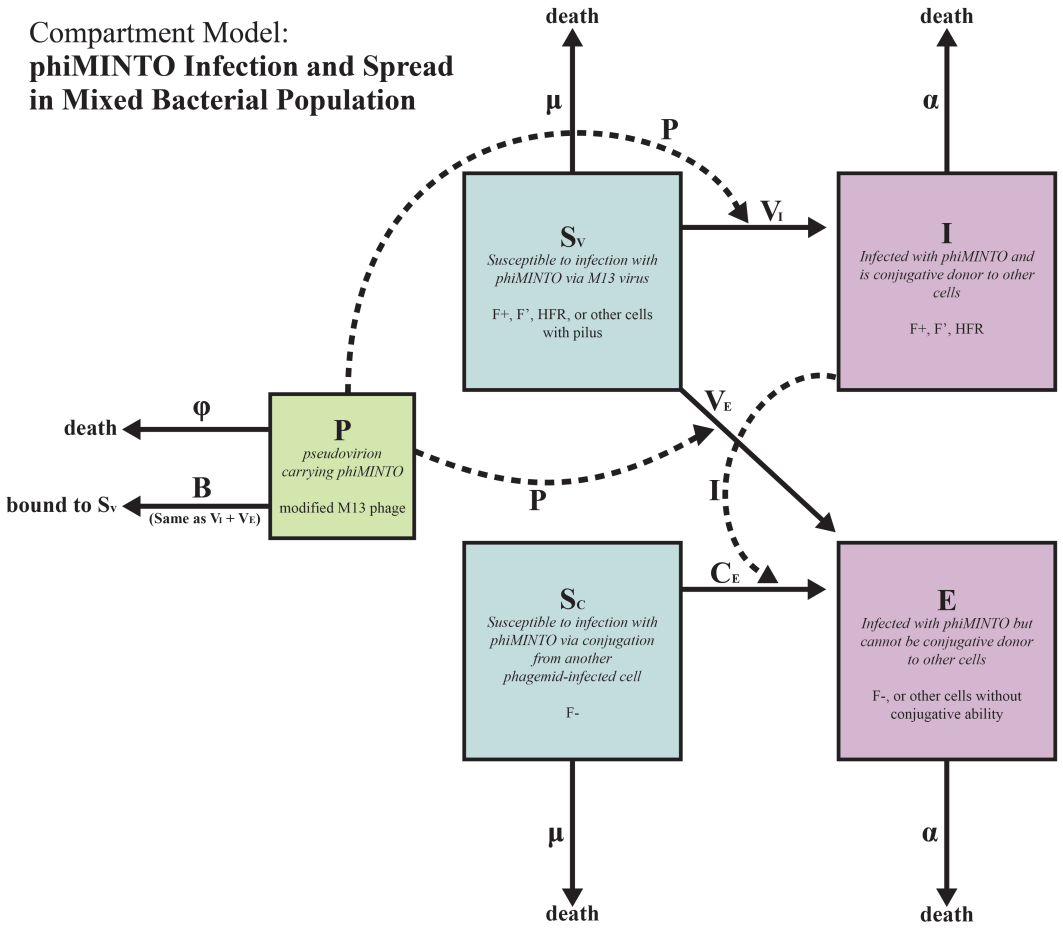

FIGURE 3.1: Compartment model depicting the transmission dynamics of the ϕMINTO plasmid in a bacterial population. Each compartment represents that cell group's proportion of the total population. Blue compartments (S_V and S_C) are cell types that do not yet carry ϕMINTO. Purple compartments (I and E) are cell types that have received ϕMINTO. The green compartment represents the multiplicity of infection (MOI) of initial ϕMINTO phage added. Solid arrows indicate the transition of individuals from one compartment to another, with the variable attached to the arrow indicating the rate of that movement. Dashed arrows indicate the influence of one compartment over the rate of another, such as P influencing the rates of arrows V_1 and V_2. Arrows denoted with α represent the rate at which ϕMINTO infected cell types die and arrows denoted with μ represent the rate at which uninfected cell types die. Figure made with *Adobe Illustrator* [28].

We determined that $B = V_I + V_E$, because the rates at which individuals move into the infectious or non-infectious groups directly relates to the rate of phage binding. When individuals in the S_C compartment are infected, they move into compartment E at a standard rate of conjugation C_E. Since our model tracks the population over time, we also included the rate of natural cell death, μ. Although we designed the system to have minimal cytotoxicity, it is possible that our machinery causes accelerated cell death, so cells that have been infected with our phagemid may die at an accelerated rate α.

From these compartments and rates, we developed five relevant equations on the rates at which the various compartments change in size. The first rate equation,

$$\frac{dP}{dt} = -\phi P - V_E P - V_I P, \tag{3.1}$$

describes the rate at which the amount of phage particles decreases, either through natural degradation or by infecting susceptible individuals.

We use

$$\frac{dS_V}{dt} = -V_E S_V P - V_I S_V P - \mu S_V \tag{3.2}$$

to model the rates at which individuals leave the susceptible group S_V and change into the infections or non-infections compartments. The equation also factors in the rate of natural cell death. For cells in S_C, we developed

$$\frac{dS_C}{dt} = -C_E S_C I - \mu S_C \tag{3.3}$$

to describe the rate at which they go into the infectious compartment I or die.

Since Equations 3.2 and 3.3 allow us to estimate the rate at which susceptible populations die or become infectious, we developed two rate equations to model the rate at which newly-infectious cells will start infecting other cells. The first,

$$\frac{dE}{dt} = V_E S_V P + C_E S_C I - \alpha E, \tag{3.4}$$

outlines the rate of change for the infected non-infectious population, and the second equation,

$$\frac{dI}{dt} = V_I S_V P - \alpha I, \tag{3.5}$$

describes the rate of change for the infected infectious population. By implementing the previous equations in SimBiology [62], we will be able provide a graphical representation of population dynamics with pre-established conditions. Relevant starting values for each compartment and rate will be found from either relevant literature or experimental data. Most initial values will be held constant, except for the initial dosage of pseudovirions (compartment P). We will alter the initial phage dosage in different model runs using the successful spread of the phagemid in a population to determine the optimal dosage of phage for a given situation. The following subsection further explains how we will implement this model.

3.1.2 Using population models

To adapt our mathematical model to the ϕMINTO system, we will first search through literature to create a set of relevant assumptions to set the proper parameters for our runs. We will be looking for variables for our compartments S_V, S_C, E, I, and P as well as the rates of change within the model: V_E V_I, C_E, μ, α, and ϕ. For example, because we are using the M13 phage for our experiments, we will determine appropriate infection rates for V_E and V_I from research performed with M13 phage. Similarly, we will be basing the rate of conjugation, C_E, on data measured from

F′ cells and plasmids with the same origin of transfer. We will continue this process for each parameter that we cannot measure directly.

After finding values for each variable, we will plug them into our set of equations in Section 3.1.1 and run simulations in SimBiology. A link to the modeling code we have developed so far can be found in Appendix B.3. Each run uses set parameters, except for the proportion of phage particles, P, which we will alter each time. The optimal phage dosage will consist of enough particles to move a significant portion of the susceptible population to the infected groups without being unreasonable to produce in a laboratory and administer in the field. We can obtain an estimate for the optimal dosage using the basic reproduction number R_0, which is defined as, "the expected number of secondary cases produced by a single (typical) infection in a completely susceptible population" [64]. In our phagemid-propagation model, R_0 is analogous to the number of recipients that the initial infected donor conjugates the phagemid into. Because the M13 phage only infects *E. coli* containing a pilus, the total amount of phage required to propagate the phagemid in a population is inversely proportional to the amount of cells that each infected donor can conjugate into. If $R_0 < 1$, our phagemid will be wiped out of the population. If R_0 is small—only around 2 or 3—than we will need to dose lots of phage to ensure propagation. If R_0 is very large, we can use less phage particles to begin the initial infection process. By comparing the outcomes of each model run and R_0 at various dosages of phage, we will be able to determine the optimal dosage of phage—one that we can use in the real world.

In addition to running the computational simulations, we will run a series of experiments measuring the fluorescence of our cells after the introduction of ϕMINTO via phage infection and conjugation. These experiments will test the fluorescence of *E. coli* WM3064 at 15 minute intervals for various dosages of phage and various proportions of conjugative donor and recipient. The WM3064 cell line is represented by compartments S_V and I, while DH5α is represented by S_C and E. Plate reading and flow cytometry will be used to obtain fluorescence data, to find an optimal *in vitro* phage dosage that can be compared to the optimal phage dosage determined *in silico*.

Chapter 4

Aptamer-Based Shiga Toxin Detection

In addition to a gene-elimination mechanism, we sought to develop a tool for detecting Shiga toxin in the field. The presence of Shiga toxin contamination in manure or water is typically detected by PCR, which takes time and requires access to a laboratory [65]. DNA aptamers that selectively bind against Shiga toxin are a potential cost-effective and efficient detection method for leafy green producers and cattle farmers due to the low cost of synthesizing synthetic oligonucleotides. To produce aptamers that bind and detect Shiga toxin, we will initially need to produce a non-toxic, recombinant form of the Shiga toxin protein, referred to here as Dead Shiga. After producing the recombinant protein, we will use the Systematic Evolution of Ligands by Exponential Enrichment (SELEX) protocol to generate aptamers that specifically target Dead Shiga. We predict that the recombinant Dead Shiga toxin model is structurally similar enough to Shiga toxin that aptamers with high affinity for Dead Shiga will also have high binding affinity for the full Shiga toxin protein. The following sections discuss the design of the Dead Shiga protein and the SELEX protocol for producing DNA aptamers against it.

4.1 Recombinant protein production

In order to execute the SELEX protocol, we need to first acquire antigen, which in our case is the Shiga toxin protein. Unfortunately the Shiga toxin protein is BSL2, and producing it in our lab would require permission from EH&S and the iGEM committee. Instead of expressing the entire Shiga toxin, we will design and express only the non-toxic subunits.

The Shiga toxin A and B genes encode for an A chain and five identical B chains that form a pentamer [7, 66]. The A subunit consists of an A1 subunit that contains the enzyme's active site, and an A2 subunit that is surrounded by the B pentamer. The Shiga toxin harms eukaryotic cells by cleaving ribosomal RNA, which interferes with protein translation. The B subunits bind to Gb3 receptors expressed on the surface of epithelial cells and intestinal linings in humans. When The B pentamer binds to Gb3, the A1 subunit is cleaved from the rest of the protein, and enters the cell to disrupt protein translation [7, 66]. The complete Shiga toxin protein is BSL2, so we designed a recombinant form of Shiga toxin that omits the toxic A1 subunit, rendering the recombinant protein non-pathogenic. A graphic of our recombinant Shiga toxin can be seen in Figure 4.1.

We obtained the protein sequence for Shiga toxin subunits A and B from the UCSC microbial genome browser [39], identified the sequence of the A2 subunit [7], and designed two synthetic DNA sequences: one for the A2 peptide, and another for the full B subunit. We added start and stop codons to both genes and optimized their sequences for expression in *E. coli*. Once we have identified an expression vector, we plan on cloning in these gene fragments separately from one another, then producing recombinant protein *in vitro*. The Dead Shiga protein will be used as the antigen for SELEX. We hope that our aptamer-based biosensors will be able to help farmers identify the presence of Shiga toxin in the field.

4.2 Aptamer-based detection

Aptamers are short, single stranded DNA or RNA molecules whose specific three dimensional folds lead to high binding affinity for specific ligands. DNA aptamers are cheaper and easier to synthesize than antibodies, and they are less likely to cause an immune response if accidentally ingested [67]. We plan to generate DNA aptamers that bind specifically to Shiga toxin using the Systematic Evolution of Ligands by Exponential enrichment (SELEX) protocol as shown in Figure 4.2.

In the SELEX protocol, a library of oligonucleotides is prepared with random DNA sequences in the middle, and fixed sequences on either end. The unique DNA sequence of each aptamer dictates the specific three-dimensional fold, which in turn dictates binding affinity to certain ligands. The library of potential aptamers is then tested against the target antigen to identify candidates with a high affinity for the target. Non-specific aptamers will not bind to the target, allowing them to be easily washed away, leaving behind the aptamers that bind to the target. The remaining aptamer molecules then get PCR amplified, and the procedure is repeated until the library of aptamers is reduced to a manageable size. The remaining aptamers are then tested against a non-target molecule to filter out non-specific binders. The remaining molecules in the aptamer library at the end of SELEX are only those that bind to our target and nothing else [67].

The benefits that come from aptamer-based detection are the ease, efficiency, and cost effectiveness of their production. Unlike antibodies which take months to culture and extract from animals, aptamers are generated in vitro and take weeks to produce [68]. Up to 1,015 unique sequences can be synthetically generated and sorted through the SELEX process, which reduces the library to a sizable pool of high affinity oligonucleotides. With the advent of high throughput sequencing we are now able to sequence entire pools of oligonucleotides between rounds of selection. This recent development has allowed us to find and predict which oligonucleotides will have higher binding affinities and specificities in a shorter time frame than previous SELEX methods [68]. Aptamers also have the advantage of greater stability than antibodies which is important for conditions in which there is a continuous need for detection of pathogens such as Shiga toxin.

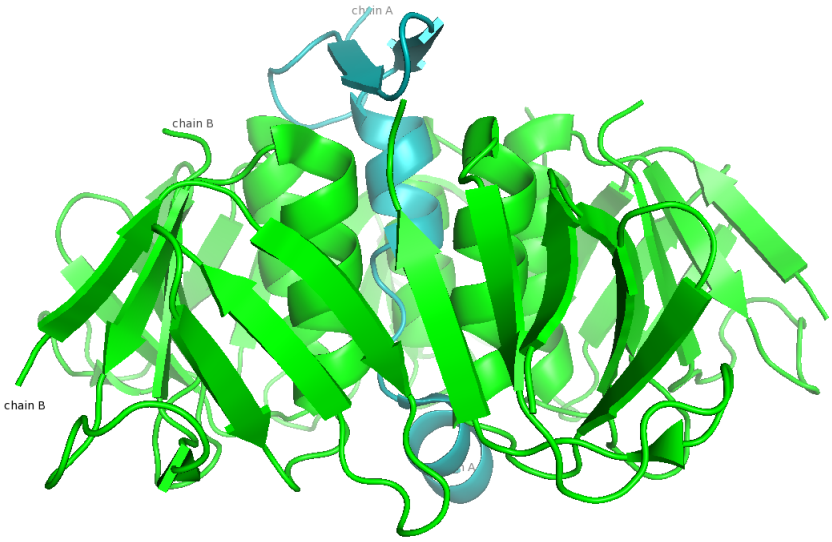

FIGURE 4.1: An image of the recombinant $A2B_5$ Shiga toxin protein we designed for aptamer development. In this figure, the A2 subunit is shown in blue and the surrounding B subunits are shown in green. This recombinant protein lacks the A1 subunit, which is solely responsible for ribosomal RNA cleavage and cell death. We hypothesize that this recombinant protein is structurally similar enough to the Shiga toxin that aptamers developed against this molecule will translate to a full Shiga toxin protein. Figure made with Pymol [69].

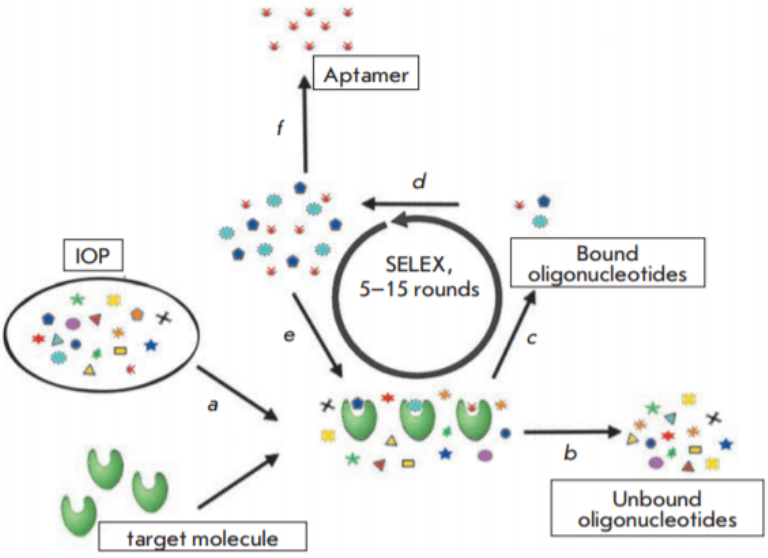

FIGURE 4.2: (a) The initial oligonucleotide pool (IOP) is tested to detect binding activity to the target molecule. In this case the target molecule is recombinant Shiga toxin protein, depicted in green. (b) Oligonucleotides that were not bound are washed away. (c) Bound oligonucleotides are eluted from the protein and recovered. (d) Once recovered, the leftover oligonucleotides are PCR amplified to create another testing pool. (e) This secondary pool is tested and selected for higher binding affinity and specificity against the target protein. This process is repeated for multiple rounds in order to select for aptamers with optimal binding affinity and specificity. (f) After multiple rounds, the remaining aptamers are sequenced and can be modified further with fluorescent tags. Reproduced from Lakhin, A. V., Tarantul, V. Z., & Gening, L. V. (2013). Aptamers: Problems, Solutions and Prospects. Acta Naturae, 5(4), 34–43.

Chapter 5

Discussion

Fundamentally, Progenie is a programmable and propagable gene-editing system. We use the CRISPR-guided INTEGRATE system first developed by Vo et al. [3] to knock out gene expression via insertion of the entire INTEGRATE module to a specific target sequence. That module is delivered to a mixed microbial community via phage infection, and propagates through it via bacterial conjugation. The INTEGRATE machinery is initially be delivered via an engineered M13 bacteriophage that selectively binds to the pilus of the target *E. coli*. Since all infected *E. coli* have a pilus, all possess the ability to conjugate into F- cells. The addition of an origin of transfer (oriT) to the plasmid carrying the INTEGRATE module (ϕMINTO) means that it can be mobilized by the infected cell's native conjugation machinery. Any cell containing the target gene will lose the ability to express that gene upon receiving ϕMINTO; cells without the target gene will be unaffected after receiving ϕMINTO. Thus, Progenie knocks out expression of a target gene from a population after phage-mediated delivery and propagation via bacterial conjugation and vertical gene transfer.

As a proof of concept in gene elimination, we designed the integration machinery to target the *mcherry* gene, whose expression can be easily measured qualitatively. To create our working model, we successfully used clonetegration to integrate *mcherry* into the chromosome of *E. coli* DH5α. We tested *mcherry* CRISPR guide sequences in different pMINT constructs in order to evaluate the efficacy of targeted DNA integration as mechanism for gene elimination. After successful disruption of *mcherry* with pMINT, proceeded to testing phage delivery of ϕMINT. Then we validated the success of MINTO conjugation. To support our experimental results, we developed a statistical model of the spread of ϕMINTO through a population. We plan to iteratively improve this model with our experimental data, then use it to further inform experimental design and simulate real-world application.

Progenie is first demonstrated on a *mcherry* model, as it is directly analogous to the *stx2* gene in *E. coli* O157:H7 we hope to target next. Editing the *stx* gene using Progenie will prevent production of Shiga toxin and ensure that the toxin cannot continue contaminating foods consumed by humans and affecting public health. One obstacle we face is the need for government to incentivize farms and ranches to implement these gene-eliminating and detecting technologies. Currently, the lack of legal and economic incentive permits cattle farmers to ignore food safety protocols, such as vaccination. According to Robert Russ, a veterinarian with Vaxxinova, cattle ranchers are reluctant to willingly implement these protocols. As a result, leafy green growers have even gone as far as to purchase vaccines themselves to supply to cattle ranches in order to protect their own crops from potentially dangerous runoff. However, our aptamer-based detection method will help target the source of outbreaks and prevent transportation of infected cattle to packaging facilities, where most cross contamination takes place [70].

Beyond STEC, we envision applications of our gene-editing tool in targeting other pathogenic, gram-negative bacteria. Gram-negative bacteria such as *Pseudomonas aeruginosa*, *Campylobacter*, *Cronobacter sakazakii*, or *Vibrio cholerae* are a few future targets. Progenie may be able to target and reduce pathogenicity of these microbes or remove antibiotic resistance. For example, *P. aeruginosa* is a known super mutator that confers multi-drug resistance in hospitals around the world [71]. If our system could effectively knock out the β-lactamase gene (bla) gene in *P. aeruginosa* and other

clinically significant bacteria, we could reduce the need for antibiotics and the risk for repeat infection [72]. Targeting of another pathogenic bacteria, *Campylobacter*, is elaborated within Appendix A.1.

The beauty of Progenie is its programmability and selectivity. Unlike antibiotics, the Progenie system has a broad host range, but will only impact the bacteria that contain the target gene of interest. The overall bacterial population remains healthy and genetically unaltered, instead of eradicated indiscriminately as with traditional antibiotics. Additionally, Progenie includes a means by which the edited cells further seek out and neutralize surrounding neighbors that still harbor the target gene. Editing individual genes should not greatly decrease the fitness of the bacterial hosts, and thus bacteria should not develop any sort of resistance to it. As the prevalence of antibiotic resistance and superbugs increases around the globe, the development of an alternative to traditional antibiotics is critical to combating disease and bacterial infection. While still in an early phase of testing, we envision Progenie as a major step towards a sequence-specific antibiotic, able to remove antibiotic resistance and reduce pathogenicity during infection while maintaining an intact healthy host microbiome.

Chapter 6

Experimental Methods

6.1 Preparing and transforming competent cells

Escherichia coli DH5α, C41, WM3064, and BL21 were made competent using the Inoue transformation protocol [73] with the following modifications. All cell types were grown in SOB, with 60mM DAP added for WM3064 (5μL per mL of media), at 37°C with rotary shaking at 250rpm. When an OD_{600} reached 0.55, cells were harvested and spun down at 2500×g for 20min. Cells were then resuspended in Inoue transformation buffer, and spun again at 2500×g for 20min. DMSO was then added before snap freezing the cells. When making competent cells containing pOSIP-KL-mcherry or pE-FLP, cells were grown at 30°C.

To transform the cells, we added 100ng of plasmid and heat shocked the cells at 42°C for 30-40sec. The cells were then incubated in SOC at 37°C with shaking at 200rpm for 45-60min. For transformations with pE-FLP, cells were recovered in SOC at 30°C for the 45-60min incubation period. The transformed cells were then plated on Lysogeny Broth plates (per liter: 10g Tryptone, 5g Yeast extract, 10g NaCl, 15 g Agar) supplemented with kanamycin (50μg/mL) or ampicillin (100μg/mL). Plates were incubated overnight at 30°C for pOSIP-KL-mcherry and pE-FLP, and 37°C for pSPIN and pSPAIN. For transformations with pOSIP-KL-mcherry, the kanamycin concentration was reduced to 15μg/mL. Cells were transformed with pOSIP-KL-mcherry (Addgene plasmid #116947), pE-FLP (Addgene plasmid #45978), M13K07 (Catalog no. 18311-019, Thermo fisher), pSPAIN, pMINT, and respective phagemids ϕMINT and ϕMINTO. Section 6.3 describes the methods used for plasmid and phagemid assembly.

6.2 Fluorescence analysis

To validate our *mcherry*-integrated *E. coli* model and assess gene knockout using our elimination mechanism, we used fluorescent microscopy, plate reader, and flow cytometry. A ZEISS Fluorescence Microscope with a Nikon camera was used to visualize the *E. coli* expressing *mcherry* by applying a dark field to locate the cells and excite mCherry using yellow light. mCherry excites at 587 nm and emits at 610 nm [74]. A SpectraMax iD3 plate reader was also used for quantitative analysis of mCherry fluorescence in *E. coli* NEB DH5α and WM3064. DH5α-*mcherry* integrant cultures were grown overnight for 16 hours at 30°C in liquid media, and transferred to a flat bottom Corning 96 well plate. All other samples were grown at 37°C. The connected Softmax Pro software was faulty, so settings were inputted directly into SpectraMax iD3's monitor. Samples were read on the fluorescence (FL) setting using the endpoint scheme that measures a single point from each well sample. Measurements were read from the top of the lidded 96 well plate at a plate height of 17.6mm with lid. The cells were excited at 587 nm and recorded for light at 627 nm, 17 nm over the ideal emission to satisfy the machine's recommended minimum difference of 40 nm. Samples were read in duplicates of 100μL and 200μL. Current samples are not normalized, future results will be normalized for cell count by OD_{600} and staining cells with SYBR Gold and DAPI. Additionally, an Attune NXT Flow Cytometer was used to assess *mcherry* gene expression in *E. coli*. Cell cultures were grown overnight for 16 hours at 37°C in liquid media and diluted to OD_{600} = 0.05

in 1× PBS 1mM EDTA. Fluorescence of mCherry was excited with a 561 nm laser and detected on the YL2 channel.

6.3 Plasmid purification and cloning

Plasmids were purified using the NEB Monarch Miniprep kit [75] with the following modifications. Overnight cultures were grown in 15 mL LB, supplemented with the proper antibiotics or DAP for WM3064 strain. After an overnight growth period, cells were spun down in a 15 mL centrifuge, supernatant was poured off, and the cells were resuspended in 600 μL of ddiH_2O. The resuspended cell mixture was then transferred to a clean 1.5 mL Eppendorf tube, and the protocol was carried out with no further changes.

Plasmid constructs were made using the site directed mutagenesis (SDM), restriction enzyme cloning, and Golden Gate assembly. SDM was performed using the protocol described by NEB [76] with the following modifications. The thermocycler conditions were set to 98°C for 30 seconds for the initial denaturation, 30 cycles of 98°C for 10 seconds, 72°C for 25 seconds, and 72°C for 400 seconds (30 sec/kb, for 13.4 kb). Then the final extension was held at 72°C for 2 minutes. SDM products were transformed into chemically competent DH5α cells and colony PCR and gel electrophoresis were used to verify success. Colony PCR using OneTaq was performed as described by the NEB Taq protocol [77], while using the NEB OneTaq protocol to set parameters for OneTaq [78].

Restriction cloning on pSPAIN *will* be done to construct an easily Golden Gate-able pSPAIN construct using pSPIN's BsaI-spacer sequence. This cloning method involves processing the pSPIN CRISPR region and the pSPAIN backbone separately before ligating the desired pSPAIN vector and pSPIN CRISPR array. Using the CRISPR array sequencing primers in Appendix B.1, we will PCR amplify the 227 bp CRISPR region, treat the mixture with DpnI, digest with SalI and BamHI, spin remove < 75 bp digestion fragments, and DNA cleanup the pSPIN CRISPR array. We will digest pSPAIN with SalI and BamHI, dephosphorylize the 5'- DNA ends with Thermosensitive Shrimp Alkaline Phosphatase (TSAP), and perform a DNA cleanup to remove enzymes. We will then ligate the processed pSPIN CRISPR array with pSPAIN vector backbone, transform the ligation product into DH5α, and select colonies with the properly constructed plasmi through colony PCR and gel electrophoresis. We will then perform Golden Gate Assembly on this intermediate plasmid (termed pMADRID) to insert the desired gene targeting CRISPR guide, finalizing the construction of pMINT.

6.4 Pseudovirion production and purification

DH5a-F' cells were be transformed with ϕMINT, transduced with the helper phage M13KO7 (ThermoFisher Scientific), and grown on LB plates with chloramphenicol to select for the transformants. Gel electrophoresis was used to verify presence of ϕMINT and M13KO7 in transformed/infected cells with expected bands around 12 kb and 9 kb, respectively. Precipitation and purification was be performed as described by the NEB procedure [79]. Purified virion samples will be measured in spectrophotometer, paying attention to the absorbance at A_{269} and A_{320} to be used to calculate the concentration of virions in the sample using the following equation for quantifying filamentous phage (George Smith)

$$\text{virions/mL} = \frac{(A_{269} - A_{320}) \cdot 6 \times 10^{16}}{\text{number of bases/virion}}.$$

6.5 Phage infection assays

All transduction experiments were performed in *E. coli* DH5a-F′ mcherry integrants. The ϕMINT plasmid was transformed into F′ cells using the same heat-shock transformation protocol described above. Cultures were grown with in chloramphenicol media (25 µg/mL) to select for transformants that took up ϕMINT. After an incubation period, plates were analyzed for the presence or absence of mCherry fluorescence. All fluorescence analysis was measured via microscopy, plate reader, or flow cytometry according to the methods described previously.

In the initial phage infection experiment, three cultures were grown: two DH5a-F′-mCherry and one DH5a-F′. When the OD600 reached between 0.400-0.500, ϕMINT-phage were transduced into one culture of DH5a-F′-mcherry at dosage of 1uL of phage per 1mL cell culture; the other two cultures received no additional treatments to serve as fluorescent and nonfluorescent controls. All three cultures were incubated for 1hr at 37C, then overnight at 30C in selective media. Relative fluorescence data of the experimental samples were compared to the data of both controls to determine Progenie's success.

In future experiments, we will grow six liquid cultures of mcherry-expressing *E. coli* DH5a-F′-mcherry. During the log phase of their growth, five cultures will be inoculated with a different concentration of ϕMINT virions, simulating multiplicities of infection of 0 (negative control), 0.01, 0.1, 1, 10, and 100. Fluorescence of each culture will be measured at 15-minute intervals after inoculating with pseudovirions. Infection rate and gene elimination success will be analyzed by comparing the proportion of target cells with reduced fluorescence between each dosage of phage. Then, the optimal multiplicity of infection for this system will be determined.

References

1. Rubin, B. E. *et al. Targeted Genome Editing of Bacteria Within Microbial Communities* preprint (Microbiology, July 17, 2020). http://biorxiv.org/lookup/doi/10.1101/2020.07.17.209189 (2021).
2. Klompe, S. E., Vo, P. L. H., Halpin-Healy, T. S. & Sternberg, S. H. Transposon-encoded CRISPR–Cas systems direct RNA-guided DNA integration. *Nature* **571,** 219–225. ISSN: 0028-0836, 1476-4687. http://www.nature.com/articles/s41586-019-1323-z (2021) (July 2019).
3. Vo, P. L. H. *et al.* CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nature Biotechnology* **39,** 480–489. ISSN: 1087-0156, 1546-1696. http://www.nature.com/articles/s41587-020-00745-y (2021) (Apr. 2021).
4. Pires, S. M., Majowicz, S., Gill, A. & Devleesschauwer, B. Global and regional source attribution of Shiga toxin-producing Escherichia coli infections using analysis of outbreak surveillance data. *Epidemiology and Infection* **147,** e236. ISSN: 0950-2688. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6625198/ (2021) (July 8, 2019).
5. Foundation, N. K. *Hemolytic Uremic Syndrome (HUS)* June 5, 2020. https://www.kidney.org/atoz/content/hemolytic (2021).
6. Clinic, M. *Hemolytic Uremic Syndrome (HUS)* July 22, 2021. https://www.mayoclinic.org/diseases-conditions/hemolytic-uremic-syndrome/symptoms-causes/syc-20352399 (2021).
7. Melton-Celsa, A. R. Shiga Toxin (Stx) Classification, Structure, and Function. *Microbiology Spectrum* **2** (eds Sperandio, V. & Hovde, C. J.) ISSN: 2165-0497. https://journals.asm.org/doi/10.1128/microbiolspec.EHEC-0024-2013 (2021) (Aug. 15, 2014).
8. Callaway, T. R., Carr, M. A., Edrington, T. S., Anderson, R. C. & Nisbet, D. J. Diet, Escherichia coli O157:H7, and Cattle: A Review After 10 Years, 14.
9. Stein, R. A. & Katz, D. E. Escherichia coli, cattle and the propagation of disease. *FEMS Microbiology Letters* **364.** ISSN: 0378-1097. https://doi.org/10.1093/femsle/fnx050 (2021) (Mar. 1, 2017).
10. Vogeleer, P., Tremblay, Y. D. N., Mafu, A. A., Jacques, M. & Harel, J. Life on the outside: role of biofilms in environmental persistence of Shiga-toxin producing Escherichia coli. *Frontiers in Microbiology* **0.** Publisher: Frontiers. ISSN: 1664-302X. https://www.frontiersin.org/articles/10.3389/fmicb.2014.00317/full (2021) (2014).
11. Agreement, C. L. G. H. M. *Food Safety Program* 2021. https://lgma.ca.gov/food-safety-program (2021).
12. Fingermann, M. *et al.* OMV-based vaccine formulations against Shiga toxin producing Escherichia coli strains are both protective in mice and immunogenic in calves. *Human Vaccines & Immunotherapeutics* **14,** 2208–2213. ISSN: 2164-554X (2018).
13. For Disease Control, C. & Prevention. *Outbreak of E. coli Infections Linked to Romaine Lettuce* Jan. 15, 2020. https://www.cdc.gov/ecoli/2019/o157h7-11-19/index.html (2021).
14. St-Pierre, F. *et al.* One-Step Cloning and Chromosomal Integration of DNA. *ACS Synthetic Biology* **2,** 537–541. ISSN: 2161-5063, 2161-5063. https://pubs.acs.org/doi/10.1021/sb400021j (2021) (Sept. 20, 2013).

15. Aldrich, S. *OverExpress C41(DE3) pLysS Chemically Competent Cells* https://www.sigmaaldrich.com/US/en/product/SIGMA/CMC0018.
16. Gene, S. *pLysS* https://www.snapgene.com/resources/plasmid-files/?set=pet_and_duet_vectors_(novagen)&plasmid=pLysS (2021).
17. OpenWetWare. *E. coli genotypes — OpenWetWare* [Online; accessed 31-August-2021]. 2021. https://openwetware.org/mediawiki/index.php?title=E._coli_genotypes&oldid=1098511.
18. Schwiesow, L. *Plasmids 101: Plasmid Incompatibility* https://blog.addgene.org/plasmids-101-plasmid-incompatibility (2021).
19. GraphPad. *Prism* version 9.1.0. https://www.graphpad.com/scientific-software/prism/ (2021).
20. Ormerod, M. G. *Home | Flow Cytometry - A Basic Introduction* Flow cytometry: a basic introduction. https://flowbook.denovosoftware.com/ (2021).
21. Biosciences, B. *FlowJo Software* https://www.flowjo.com/ (2021).
22. Cui, L. & Bikard, D. Consequences of Cas9 cleavage in the chromosome of Escherichia coli. *Nucleic Acids Research* **44,** 4243–4251. ISSN: 1362-4962 (May 19, 2016).
23. Jiang, W., Bikard, D., Cox, D., Zhang, F. & Marraffini, L. A. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology* **31,** 233–239. ISSN: 1087-0156, 1546-1696. http://www.nature.com/articles/nbt.2508 (2021) (Mar. 2013).
24. Pyne, M. E., Moo-Young, M., Chung, D. A. & Chou, C. P. Coupling the CRISPR/Cas9 System with Lambda Red Recombineering Enables Simplified Chromosomal Gene Replacement in Escherichia coli. *Applied and Environmental Microbiology* **81** (ed Kivisaar, M.) 5103–5114. ISSN: 0099-2240, 1098-5336. http://aem.asm.org/lookup/doi/10.1128/AEM.01248-15 (2021) (Aug. 1, 2015).
25. Su, T. *et al.* A CRISPR-Cas9 Assisted Non-Homologous End-Joining Strategy for One-step Engineering of Bacterial Genome. *Scientific Reports* **6,** 37895. ISSN: 2045-2322. http://www.nature.com/articles/srep37895 (2021) (Dec. 2016).
26. Bikard, D. *et al.* Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas system. *Nucleic Acids Research* **41,** 7429–7437. ISSN: 1362-4962, 0305-1048. https://academic.oup.com/nar/article-lookup/doi/10.1093/nar/gkt520 (2021) (Aug. 2013).
27. Banno, S., Nishida, K., Arazoe, T., Mitsunobu, H. & Kondo, A. Deaminase-mediated multiplex genome editing in Escherichia coli. *Nature Microbiology* **3,** 423–429. ISSN: 2058-5276. http://www.nature.com/articles/s41564-017-0102-6 (2021) (Apr. 2018).
28. Adobe Inc. *Adobe Illustrator* version CC 2020 (24.3.0). Aug. 26, 2021. https://adobe.com/products/illustrator.
29. Ltd., B. *Geneious Prime 2021.2.1* July 12, 2021. https://www.geneious.com/ (2021).
30. BioLabs, N. E. *NEB Tm Calculator* version 1.13.0. https://tmcalculator.neb.com/#!/main (2021).
31. Griffiths, A. J., Gelbart, W. M., Miller, J. H. & Lewontin, R. C. Bacterial Conjugation. *Modern Genetic Analysis.* Publisher: W. H. Freeman. https://www.ncbi.nlm.nih.gov/books/NBK21351/ (2021) (1999).
32. Lim, J. Y., Yoon, J. W. & Hovde, C. J. A Brief Overview of Escherichia coli O157:H7 and Its Plasmid O157. *Journal of microbiology and biotechnology* **20,** 5–14. ISSN: 1017-7825 (Jan. 2010).

33. Akhverdyan, V. Z. *et al.* Application of the bacteriophage Mu-driven system for the integration/amplification of target genes in the chromosomes of engineered Gram-negative bacteria—mini review. *Applied Microbiology and Biotechnology* **91,** 857–871. ISSN: 0175-7598. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3145075/ (2021) (2011).

34. Kittleson, J. T., DeLoache, W., Cheng, H.-Y. & Anderson, J. C. Scalable Plasmid Transfer using Engineered P1-based Phagemids. *ACS Synthetic Biology* **1,** 583–589. ISSN: 2161-5063, 2161-5063. https://pubs.acs.org/doi/10.1021/sb300054p (2021) (Dec. 21, 2012).

35. Lenneman, B. R., Fernbach, J., Loessner, M. J., Lu, T. K. & Kilcher, S. Enhancing phage therapy through synthetic biology and genome engineering. *Current Opinion in Biotechnology* **68,** 151–159. ISSN: 09581669. https://linkinghub.elsevier.com/retrieve/pii/S0958166920301646 (2021) (Apr. 2021).

36. Lund, P. E., Hunt, R. C., Gottesman, M. M. & Kimchi-Sarfaty, C. Pseudovirions as vehicles for the delivery of siRNA. *Pharmaceutical research* **27,** 400–420. ISSN: 0724-8741. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2831147/ (2021) (Mar. 2010).

37. Kehoe, J. W. & Kay, B. K. Filamentous Phage Display in the New Millennium. *Chemical Reviews* **105.** Publisher: American Chemical Society, 4056–4072. ISSN: 0009-2665. https://doi.org/10.1021/cr000261r (2021) (Nov. 1, 2005).

38. Smeal, S. W., Schmitt, M. A., Pereira, R. R., Prasad, A. & Fisk, J. D. Simulation of the M13 life cycle I: Assembly of a genetically-structured deterministic chemical kinetic simulation. *Virology* **500,** 259–274. ISSN: 0042-6822. https://www.sciencedirect.com/science/article/pii/S0042682216302239 (2021) (Jan. 1, 2017).

39. Łobocka, M. B. *et al.* Genome of Bacteriophage P1. *Journal of Bacteriology* **186,** 7032–7068. ISSN: 0021-9193. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC523184/ (2021) (Nov. 2004).

40. Marvin, D. A. Filamentous phage structure, infection and assembly. *Current Opinion in Structural Biology* **8,** 150–158. ISSN: 0959-440X (Apr. 1998).

41. Smith, G. P. *Phage-Display Vectors and Libraries Based on Filamentous Phage Strain fd-tet* Apr. 2006. http://www.biosci.missouri.edu/smithGp/PhageDisplayWebsite/PhageDisplayWebsiteIndex.html.

42. Choudhuri, S. in *Bioinformatics for Beginners* (ed Choudhuri, S.) 27–53 (Academic Press, Oxford, Jan. 1, 2014). ISBN: 978-0-12-410471-6. https://www.sciencedirect.com/science/article/pii/B9780124104716000025 (2021).

43. Koraimann, G. Spread and Persistence of Virulence and Antibiotic Resistance Genes: A Ride on the F Plasmid Conjugation Module. *EcoSal Plus* **8.** Publisher: American Society for Microbiology. https://journals.asm.org/doi/full/10.1128/ecosalplus.ESP-0003-2018 (2021) (July 17, 2018).

44. Koraimann, G. Spread and persistence of virulence and antibiotic resistance genes: a ride on the F plasmid conjugation module. *EcoSal Plus* **8** (2018).

45. Lawley, T., Klimke, W., Gubbins, M. & Frost, L. F factor conjugation is a true type IV secretion system. *FEMS Microbiology Letters* **224,** 1–15. ISSN: 0378-1097. https://doi.org/10.1016/S0378-1097(03)00430-0 (2021) (July 1, 2003).

46. Waksman, G. From conjugation to T4S systems in Gram-negative bacteria: a mechanistic biology perspective. *EMBO Reports* **20,** e47012. ISSN: 1469-221X. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6362355/ (2021) (Feb. 2019).

47. Virolle, C., Goldlust, K., Djermoun, S., Bigot, S. & Lesterlin, C. Plasmid Transfer by Conjugation in Gram-Negative Bacteria: From the Cellular to the Community Level. *Genes* **11.** Number: 11 Publisher: Multidisciplinary Digital Publishing Institute, 1239. https://www.mdpi.com/2073-4425/11/11/1239 (2021) (Nov. 2020).

48. Lederberg, J., Cavalli, L. L. & Lederberg, E. M. Sex Compatibility in Escherichia Coli. *Genetics* **37,** 720–730. ISSN: 0016-6731. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1209583/ (2021) (Nov. 1952).

49. Haase, J., Kalkum, M. & Lanka, E. TrbK, a small cytoplasmic membrane lipoprotein, functions in entry exclusion of the IncP alpha plasmid RP4. *Journal of Bacteriology* **178.** Publisher: American Society for Microbiology, 6720–6729. https://journals.asm.org/doi/10.1128/jb.178.23.6720-6729.1996 (2021) (Dec. 1, 1996).

50. Garcillán-Barcia, M. P. & de la Cruz, F. Why is entry exclusion an essential feature of conjugative plasmids? *Plasmid* **60,** 1–18. ISSN: 0147-619X. https://www.sciencedirect.com/science/article/pii/S0147619X08000231 (2021) (July 1, 2008).

51. Moran, R. A. & Hall, R. M. pBuzz: A cryptic rolling-circle plasmid from a commensal Escherichia coli has two inversely oriented oriTs and is mobilised by a B/O plasmid. *Plasmid* **101,** 10–19. ISSN: 0147-619X. https://www.sciencedirect.com/science/article/pii/S0147619X18300969 (2021) (Jan. 1, 2019).

52. Francia, M. V. *et al.* A classification scheme for mobilization regions of bacterial plasmids. *FEMS Microbiology Reviews* **28,** 79–100. ISSN: 0168-6445. https://doi.org/10.1016/j.femsre.2003.09.001 (2021) (Feb. 1, 2004).

53. Rawlings, D. E. & Tietze, E. Comparative Biology of IncQ and IncQ-Like Plasmids. *Microbiology and Molecular Biology Reviews* **65,** 481–496. ISSN: 1092-2172. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC99038/ (2021) (Dec. 2001).

54. Haase, J., Lurz, R., Grahn, A. M., Bamford, D. H. & Lanka, E. Bacterial conjugation mediated by plasmid RP4: RSF1010 mobilization, donor-specific phage propagation, and pilus production require the same Tra2 core components of a proposed DNA transport complex. *Journal of Bacteriology* **177,** 4779–4791. ISSN: 0021-9193. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC177245/ (2021) (Aug. 1995).

55. Wang, P. *et al.* Development of an efficient conjugation-based genetic manipulation system for Pseudoalteromonas. *Microbial Cell Factories* **14,** 11. ISSN: 1475-2859. http://www.microbialcellfact
com/content/14/1/11 (2021) (2015).

56. Meyer, R. Replication and conjugative mobilization of broad host-range IncQ plasmids. *Plasmid* **62,** 57–70. ISSN: 0147619X. https://linkinghub.elsevier.com/retrieve/pii/S0147619X09000699 (2021) (Sept. 2009).

57. Smillie, C., Garcillán-Barcia, M. P., Francia, M. V., Rocha, E. P. C. & de la Cruz, F. Mobility of Plasmids. *Microbiology and Molecular Biology Reviews* **74,** 434–452. ISSN: 1092-2172, 1098-5557. https://journals.asm.org/doi/10.1128/MMBR.00020-10 (2021) (Sept. 2010).

58. Remus-Emsermann, M. N. P., Aicher, D., Pelludat, C., Gisler, P. & Drissner, D. Conjugation Dynamics of Self-Transmissible and Mobilisable Plasmids into E. coli O157:H7 on Arabidopsis thaliana Rosettes. *Antibiotics* **10.** Number: 8 Publisher: Multidisciplinary Digital Publishing Institute, 928. https://www.mdpi.com/2079-6382/10/8/928 (2021) (Aug. 2021).

59. Neil, K., Allard, N. & Rodrigue, S. Molecular Mechanisms Influencing Bacterial Conjugation in the Intestinal Microbiota. *Frontiers in Microbiology* **12,** 673260. ISSN: 1664-302X. https://www.frontiersin.org/articles/10.3389/fmicb.2021.673260/full (2021) (June 4, 2021).

60. Velappan, N., Sblattero, D., Chasteen, L., Pavlik, P. & Bradbury, A. R. Plasmid incompatibility: more compatible than previously thought? *Protein Engineering, Design and Selection* **20,** 309–313. ISSN: 1741-0126 (July 2007).

61. The Mathworks, I. *MATLAB version 9.10.0.1613233 (R2021a)* (Natick, Massachusetts, 2021).

62. MATLAB. *SimBiology* https://www.mathworks.com/products/simbiology.html (The MathWorks Inc., Natick, Massachusetts, 2010).

63. Wikipedia contributors. *Compartmental models in epidemiology — Wikipedia, The Free Encyclopedia* [Online; accessed 29-August-2021]. 2021. https://en.wikipedia.org/w/index.php?title=Compartmental_models_in_epidemiology&oldid=1040638557.

64. Jones, J. H. *Notes on R_0* https://web.stanford.edu/~jhj1/teachingdocs/Jones-on-R0.pdf (2021).

65. Feng, P., Weagant, S. & Jinneman, K. *BAM Chapter 4A: Diarrheagenic Escherichia coli* https://www.fda.gov/food/laboratory-methods-food/bam-chapter-4a-diarrheagenic-escherichia-coli (2021).

66. Chan, Y. S. & Ng, T. B. Shiga toxins: from structure and mechanism to applications. *Applied Microbiology and Biotechnology* **100,** 1597–1610. ISSN: 0175-7598, 1432-0614. http://link.springer.com/10.1007/s00253-015-7236-3 (2021) (Feb. 2016).

67. Wang, T. *et al.* A Detailed Protein-SELEX Protocol Allowing Visual Assessments of Individual Steps for a High Success Rate. *Human Gene Therapy Methods* **30,** 1–16. ISSN: 1946-6536, 1946-6544. https://www.liebertpub.com/doi/10.1089/hgtb.2018.237 (2021) (Feb. 2019).

68. Zhuo, Z. *et al.* Recent Advances in SELEX Technology and Aptamer Applications in Biomedicine. *International Journal of Molecular Sciences* **18,** 2142. ISSN: 1422-0067. http://www.mdpi.com/1422-0067/18/10/2142 (2021) (Oct. 14, 2017).

69. Schrödinger, LLC. *The PyMOL Molecular Graphics System, Version 1.7.4.5 Edu* Nov. 2015.

70. Browne, A. S. *et al.* Transmission Dynamics of Shiga Toxin-Producing Escherichia coli in New Zealand Cattle from Farm to Slaughter. *Applied and Environmental Microbiology* **87.** Publisher: American Society for Microbiology, e02907–20. https://journals.asm.org/doi/10.1128/AEM.02907-20 (2021).

71. Al-Orphaly, M. *et al.* Epidemiology of Multidrug-Resistant Pseudomonas aeruginosa in the Middle East and North Africa Region. *mSphere* **6.** Publisher: American Society for Microbiology, e00202–21. https://journals.asm.org/doi/10.1128/mSphere.00202-21 (2021).

72. Daniel, J. W. & Philip, D. L. Mechanisms of -lactam Resistance Among Pseudomonas aeruginosa. *Current Pharmaceutical Design* **19,** 209–222. https://www.eurekaselect.com/104921/article (2021) (Dec. 31, 2012).

73. Im, H., Sambrook, J. & Russell, D. W. *The Inoue Method for Preparation and Transformation of Competent E. coli: "Ultra Competent" Cells* Aug. 20, 2011. https://bio-protocol.org/pdf/Bio-protocol143.pdf (2021).

74. Shaner, N. C. *et al.* Improved monomeric red, orange and yellow fluorescent proteins derived from Discosoma sp. red fluorescent protein. *Nature Biotechnology* **22.** Bandiera_abtest: a Cg_type: Nature Research Journals Number: 12 Primary_atype: Research Publisher: Nature Publishing Group, 1567–1572. ISSN: 1546-1696. https://www.nature.com/articles/nbt1037 (2021) (Dec. 2004).

75. BioLabs, N. E. *Monarch Plasmid DNA Miniprep Kit Protocol (NEB T1010)* https://www.neb.com/protocols/2015/11/20/monarch-plasmid-dna-miniprep-kit-protocol-t1010 (2021).

76. Inc., N. E. B. *Q5 Site-Directed Mutagenesis Kit Protocol (E0554)* https://www.neb.com/protocols/2013/01/26/q5-site-directed-mutagenesis-kit-protocol-e0554 (2021).

77. BioLabs, N. E. *PCR Protocol for Taq DNA Polymerase with Standard Taq Buffer (M0273)* https://www.neb.com/protocols/0001/01/01/taq-dna-polymerase-with-standard-taq-buffer-m0273 (2021).

78. BioLabs, N. E. *PCR Protocol for OneTaq® DNA Polymerase (M0480)* https://www.neb.com/protocols/2012/10/11/onetaqdnapolymerasem0480 (2021).

79. BioLabs, N. E. *Use of M13KO7 Helper Phage for isolation of single-stranded phagemid DNA* https://www.neb.com/protocols/0001/01/01/use-of-m13ko7-helper-phage-for-isolation-of-single-stranded-phagemid-dna (2021).

80. Igwaran, A. & Okoh, A. I. Human campylobacteriosis: A public health concern of global importance. *Heliyon* **5,** e02814. ISSN: 2405-8440. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6861584/ (2021) (Nov. 14, 2019).

81. Dasti, J. I., Tareen, A. M., Lugert, R., Zautner, A. E. & Groß, U. Campylobacter jejuni: A brief overview on pathogenicity-associated factors and disease-mediating mechanisms. *International Journal of Medical Microbiology* **300,** 205–211. ISSN: 1438-4221. https://www.sciencedirect.com/science/article/pii/S1438422109000575 (2021) (Apr. 1, 2010).

82. De Carvalho, A. F. *et al.* Detection of CDT toxin genes in Campylobacter spp. strains isolated from broiler carcasses and vegetables in São Paulo, Brazil. *Brazilian Journal of Microbiology* **44,** 693–699. ISSN: 1517-8382. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3910176/ (2021) (Jan. 15, 2014).

83. For Disease Control, C. & Prevention. *Cronobacter Infection and Infants* Feb. 8, 2021. https://www.cdc.gov/cronobacter/infection-and-infants.html (2021).

84. Kalyantanda, G., Shumyak, L. & Archibald, L. K. Cronobacter Species Contamination of Powdered Infant Formula and the Implications for Neonatal Health. *Frontiers in Pediatrics* **3,** 56. ISSN: 2296-2360. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4489094/ (2021) (July 2, 2015).

85. Singh, N., Goel, G. & Raghav, M. Insights into virulence factors determining the pathogenicity of Cronobacter sakazakii. *Virulence* **6,** 433–440. ISSN: 2150-5594. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4601314/ (2021) (May 7, 2015).

86. Faruque, S. M., Albert, M. J. & Mekalanos, J. J. Epidemiology, Genetics, and Ecology of Toxigenic *Vibrio cholerae*. *Microbiology and Molecular Biology Reviews* **62,** 1301–1314. ISSN: 1092-2172, 1098-5557. https://journals.asm.org/doi/10.1128/MMBR.62.4.1301-1314.1998 (2021) (Dec. 1998).

87. For Disease Control, C. & Prevention. *Cholera - Vibrio cholerae infection* Oct. 30, 2020. https://www.cdc.gov/cholera/index.html (2021).

88. Ezzat Alnakip, M. *et al.* The Immunology of Mammary Gland of Dairy Ruminants between Healthy and Inflammatory Conditions. *Journal of Veterinary Medicine* **2014,** 659801. ISSN: 2356-7708 (2014).

89. Shahbandeh, M. *Annual consumption of fluid cow milk worldwide in 2020, by country* https://www.statista.com/statistics/272003/global-annual-consumption-of-milk-by-region/ (2021).

90. Tsagkaris, A. *et al.* A microfluidic paper-based analytical device (PAD) with smartphone readout for chlorpyrifos-oxon screening in human serum. *Talanta* **222,** 121535. ISSN: 00399140 (Jan. 2021).

91. Abudayyeh, O. O. *et al.* C2c2 is a single-component programmable RNA-guided RNA-targeting CRISPR effector. *Science* **353.** Publisher: American Association for the Advancement of Science Section: Research Articles. ISSN: 0036-8075, 1095-9203. https://science.sciencemag.org/content/353/6299/aaf5573 (2021) (Aug. 5, 2016).

92. Gootenberg, J. S. *et al.* Nucleic acid detection with CRISPR-Cas13a/C2c2. *Science* **356.** Publisher: American Association for the Advancement of Science Section: Reports, 438–442. ISSN: 0036-8075, 1095-9203. https://science.sciencemag.org/content/356/6336/438 (2021) (Apr. 28, 2017).

93. Niewoehner, O. & Jinek, M. Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA* **22,** 318–329. ISSN: 1355-8382, 1469-9001. http://rnajournal.cshlp.org/lookup/doi/10.1261/rna.054098.115 (2021) (Mar. 2016).

94. Foster, K., Kalter, J., Woodside, W., Terns, R. M. & Terns, M. P. The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA Biology* **16,** 449–460. ISSN: 1547-6286, 1555-8584. https://www.tandfonline.com/doi/full/10.1080/15476286.2018.1493334 (2021) (Apr. 3, 2019).

95. Kazlauskiene, M., Kostiuk, G., Venclovas, Č., Tamulaitis, G. & Siksnys, V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science* **357,** 605–609. ISSN: 0036-8075, 1095-9203. https://www.sciencemag.org/lookup/doi/10.1126/science.aao0100 (2021) (Aug. 11, 2017).

96. Gootenberg, J. S. *et al.* Multiplexed and portable nucleic acid detection platform with Cas13, Cas12a, and Csm6. *Science* **360,** 439–444. ISSN: 0036-8075, 1095-9203. https://www.sciencemag.org/lookup/doi/10.1126/science.aaq0179 (2021) (Apr. 27, 2018).

Appendix A

Supplemental Information

A.1 Future directions and potential applications

Each year, bacteria in the *Campylobacter* genus cause over 400 million gastrointestinal infections worldwide [80]. The most common pathogenic *Campylobacter* species, *C. jejuni*, causes about 40,000-100,000 cases of diarrhea in developing countries as well as possesses antibiotic resistance, contributing to its threat to public health [80]. In order to apply Progenie to *C. jejuni*, we would target the cytolethal distending toxin (CDT) pathway, as it controls cell cycle and host cell apoptosis [81]. The CDT toxin consists of three subunits, *cdtA*, *cdtB*, and *cdtC*, but the protein produced by *cdtB* is mainly responsible for eukaryotic cell death [82]. Thus, a primary target for our Progenie system would be the *cdtB* gene.

Another relevant pathogenic bacteria, Cronobacter sakazakii, infects people of all age groups [83], but about 90% of cases are associated with contaminated powdered baby formula [84]. Although cases are rare, death occurs 40-80% of the time, and infection can cause neurological dis orders later in life [85]. The gene *ompA* produces a protein that crosses the blood-brain barrier and disrupts the central nervous system [85]. Thus, to disrupt the pathogenicity we would target *ompA*.

Cholera toxin (CT), encoded by the *ctxAB* gene in Vibrio cholerae serogroup O1 or O139 [86], causes about 2.9 million cases of profuse diarrhea and 95,000 deaths each year around the world [87]. In order to eliminate CT, we could target the *ctxAB*. All the bacteria mentioned, *C. Jejuni, Cronobacter sakazakii*, and *V. cholerae* are all alternative potential targets for Progenie. The programmability of our technology demonstrates the potential for genetic engineering to serve as a solution to many problems in public health and food safety.

A.2 IISER-K Collaboration: Wet lab

Bovine mastisis causes inflammation of the cows' udders, however the inflammation not only harms cattle, but also reduces the quality and quantity of milk produced, threatening the livelihood of the small-scale dairy farmers that make up 60% of the dairy economy [88]. With India being the largest dairy-dependent economy in the world, with approximately 200 million metric tons of milk produced in 2019, the IISER-K iGEM team aims to reduce the burden of bovine mastitis and to improve cattle health by developing a cost-effective detection method as well as an antibiotic-free solution for bovine mastitis in dairy farms [89].

As mentioned in Section 3, Progenie will conduct certain synthetic biology experiments for Namooste to further their progress on developing an easy-to-use device for detection of bovine mastitis. Namooste's device design has two main aspects: nucleic acid detection using an RNA-guided CRISPR system, known as SHERLOCKv2, and a colorimetric μPAD detection device [90]. Progenie will only be working with SHERLOCKv2, and will not be optimizing the μPAD. The following sections give more details on the components of SHERLOCKv2 we will be working on.

A.2.1 Detection using RNA-guided RNA-targeting CRISPR effector

The SHERLOCKv2 platform allows for rapid nucleic acid detection and elimination using Cas13, a class 2 type IV CRISPR-Cas effector [91, 92]. Nucleic acids can be detected by Cas13 upon recognition of a target RNA sequence, which activates non-specific RNase activity that degrades fluorescently-labeled RNA reporters; cleavage of the reporters results in fluorescence [92]. The signal is amplified by the inclusion of CRISPR type III effector nuclease Csm6.

In nature, the Csm6 protein is a ssRNA-specific endoribonuclease that acts in conjunction with the type III-A CRISPR effector complex (Csm complex), which includes the CRISPR-Cas protein Cas10 [93, 94]. Upon recognition of a nucleic acid sequence complementary to the guide RNA, Cas10 initiates cleavage using the Csm complex [93]. The RNA-cleavage products of the Csm complex consist of cyclic oligoadenylates that act as secondary signaling molecules to activate Csm6 [95]. Additionally, Kazlauskiene et al. 2017 show that linear adenine homopolymers terminated with a 2′,3′-cyclic phosphate are also capable of activating Csm6 from *Streptococcus thermophilus* (StCsm6) [95]. The cleavage activity of Cas13 results in hydroxylated 5′ ends and 2′,3′-cyclic phosphate ends [96], meaning that Cas13 can activate Csm6. Thus, when Cas13 cleaves the fluorescent reporters in SHERLOCKv2, the products activate Csm6 and result in additional non-specific RNase activity that cleaves more reporters.

The Csm6 protein has not been highly characterized in the context of the SHERLOCKv2 system or in its activation by linear adenine homopolymers terminated with a 2′,3′-cyclic phosphate. Thus in order to develop a highly-specific detection method for bovine mastitis using Cas13 from SHERLOCKv2, our team will characterize the Csm6 protein using different activators with a 2′,3′-cyclic phosphate end at varying concentrations as well as Csm6 protein sequences from different species. In Gootenberg et al. 2018, researchers were able to activate Csm6 from *Enterococcus italicus* (EiCsm6), *Thermus Thermophilus* (TtCsm6), and *Lactobacillus salivarius* (LsCsm6) using RNA adenylate molecules of different lengths and 2′,3′-cyclic phosphate ends. All three species' Csm6 protein preferentially cleaved A- and C-rich reporters of lengths 6-nt and 5-nt, respectively, with TtCsm6 being activated the least. To couple the activity of Cas13 to Csm6, Gootenberg et al. 2018 designed RNA activators containing a polyadenylate [poly(A)] stretch followed by a poly(U) stretch that could be cleaved by the uracil-preferring Cas13 enzyme.

Given that only four Csm6 proteins have been confirmed to be efficiently activated by linear adenine homopolymers terminated with 2′,3′-cyclic phosphate ends (StCsm6, EiCsm6, TtCsm6, and LsCsm6), we are interested in characterizing Csm6 from additional species. Furthermore, since reporters with a poly(A) stretch followed by a poly(U) stretch have been shown to be cleaved by Csm6 as well as Cas13, our characterization should include targets with these characteristics. Our current plan is to perform a nuclease assay with activators of varying lengths and 2′,3′-cyclic phosphate ends as well as 3′,5′-cyclic phosphate ends, as characteristic of cyclic oligoadenylates, along with reporters with known sequences. Once Csm6 gets activated and cuts the reporters, the reporters will serve as amplification templates in quantitative PCR (qPCR). If Csm6 was sufficiently activated, it will have cut the reporters (i.e. the templates for PCR) such that primers will not amplify the whole target and/or will not anneal properly, resulting in little to no amplicon. If Csm6 was not activated, the template will be intact and result in the presence of an amplicon. Results of the qPCR experiments will reveal which activators and reporter sequences work the best with Csm6, which can then be used in a TaqMan assay with digital PCR. The TaqMan assay involves a fluorescent probe that binds to a target sequence; in this case, the target sequence for the probe will be the same sequence as the Csm6 reporter. If Csm6 cuts the reporter, there is no target sequence for the probe and thus it will keep fluorescing. If Csm6 does not cut the reporter, then the probe can anneal to the template and get cut by Taq polymerase.

Appendix B

DNA Sequences and Modeling Code

B.1 PCR primers

Name	Forward primer	Reverse primer
SDM pMINT guide 1	GAATTGGGGATCGGCCGTGAACTGCCGAGTAGGTAGCTGATAACGGATCCGAATTCGAGCG	CTGGCAGTTTATGGCGGTTATCAGCTACCTACTCGGCAGTTCACCCATGGTATATCTCCACG
SDM pMINT guide 2	CCATCTAGATTTCCCCGTGAACTGCCGAGTAGGTAGCTGATAACGGATCCGAATTCGAGCG	TTTCCAAGGGCGAGGAGTTATCAGCTACCTACTCGGCAGTTCACCCATGGTATATCTCCACG
SDM pMINT guide 3	CCTTGGAAACCATCTAGTGAACTGCCGAGTAGGTAGCTGATAACGGATCCGAATTCGAGCG	GCGAGGAGGATAACATGTTATCAGCTACCTACTCGGCAGTTCACCCATGGTATATCTCCACG
SDM pMINT guide 4	ATGTGAACTTTGAAGCGTGAACTGCCGAGTAGGTAGCTGATAACGGATCCGAATTCGAGCG	GGAGGGTTCTGTTAACGTTATCAGCTACCTACTCGGCAGTTCACCCATGGTATATCTCCACG
SDM pMINT guide 5	TGACCGTTAACAGAACGTGAACTGCCGAGTAGGTAGCTGATAACGGATCCGAATTCGAGCG	CGAGTTCGAGATCGAAGTTATCAGCTACCTACTCGGCAGTTCACCCATGGTATATCTCCACG
NEB5α lambda attB site	CAAAAGCCAATGCCAGCG	GCGCAATGCCATCTGGTATC
pMINT CRISPR array Sanger sequencing	GTGAGCGAGGACACAAGC	CGCTGTAAGGTTTAGGTCTTTGC
pSPIN BsaI spacer cassette with AarI RE sites	CAAGACCACCTGCAGCGTCGACGTGGAGATATACCATG	CAAGACCACCTGCAGCGGATCCGTTATCAGCTACCTAC

B.2 Gene blocks: CRISPR guides and Golden Gate Assembly gene blocks

Name	Spacer sequence	Full gene block for Golden Gate assembly
pMINT CRISPR guide 1	CGCCATAAACTGCCAGGAATTGGGGATCGGCC	CAAGACGGTCTCCATAACCGCCATAAACTGCCAGGAATTGGGGATCGGCCGTGAACGAGACCCAAGAC
pMINT CRISPR guide 2	TCCTCGCCCTTGGAAACCATCTAGATTTCCCC	CAAGACGGTCTCCATAACTCCTCGCCCTTGGAAACCATCTAGATTTCCCCGTGAACGAGACCCAAGAC
pMINT CRISPR guide 3	ATGTTATCCTCCTCGCCCTTGGAAACCATCTA	CAAGACGGTCTCCATAACATGTTATCCTCCTCGCCCTTGGAAACCATCTAGTGAACGAGACCCAAGAC
pMINT CRISPR guide 4	GTTAACAGAACCCTCCATGTGAACTTTGAAGC	CAAGACGGTCTCCATAACGTTAACAGAACCCTCCATGTGAACTTTGAAGCGTGAACGAGACCCAAGAC
pMINT CRISPR guide 5	TTCGATCTCGAACTCGTGACCGTTAACAGAAC	CAAGACGGTCTCCATAACTTCGATCTCGAACTCGTGACCGTTAACAGAACGTGAACGAGACCCAAGAC

B.3 SimBiology modeling code

The SimBiology modeling code we are building can be found on GitHub at https://github.com/banilmohammed/UCSC-iGEM-2021-Model

B.4 Plasmid maps

B.4.1 pMINT

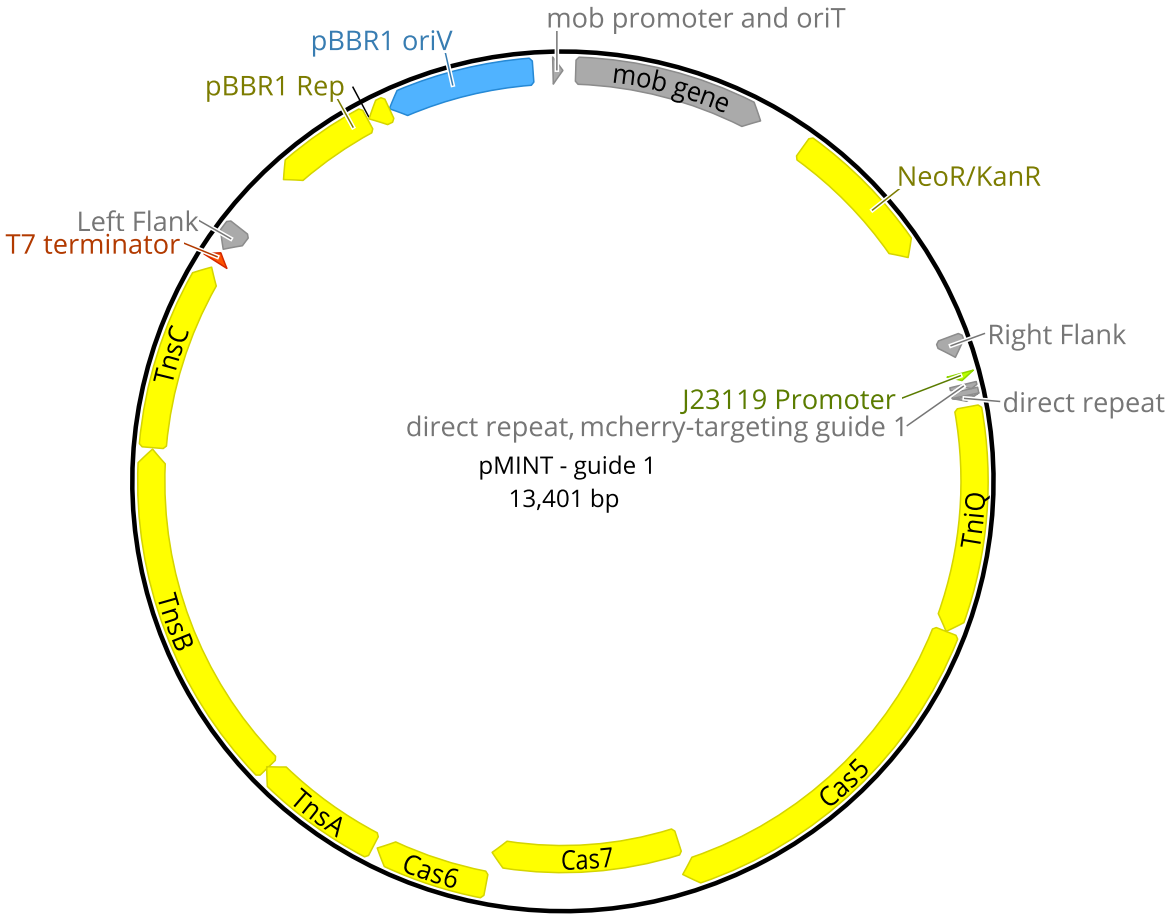

FIGURE B.1: This plasmid map is an edited version of pSPAIN, initially engineered by Vo et. al [3]. This pMINT plasmid contains *mcherry*-targeting spacer sequence. The Genbank file for this plasmid was shared by Vo et. al. Figure made with Geneious Prime [29].

B.4.2 ϕMINT

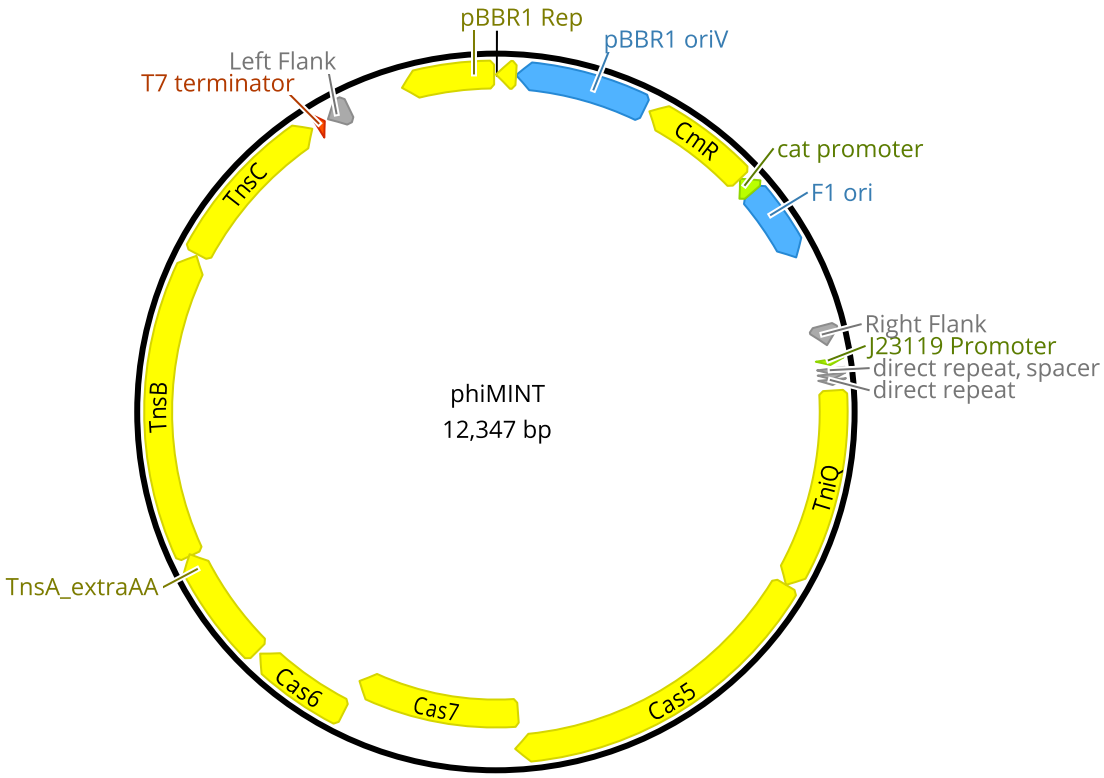

FIGURE B.2: This plasmid map is an edited version of pSPAIN, initially engineered by Vo et. al [3]. This ϕMINT plasmid contains *mcherry*-targeting spacer sequence, an F1 ori for phage packaging, and a chloramphenicol resistance marker in the place of pSPAIN's kanamycin resistance marker. The Genbank file for the original pSPAIN plasmid was shared by Vo et. al. Figure made with Geneious Prime [29].

B.4.3 ϕMINTO

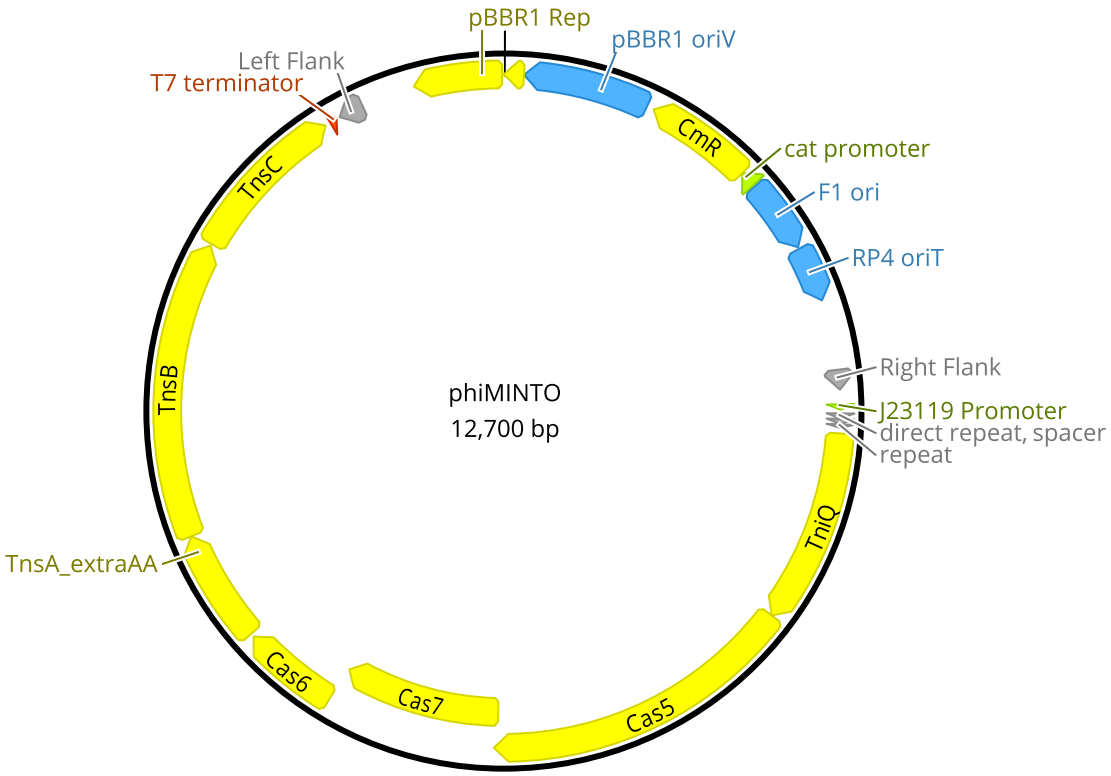

FIGURE B.3: This plasmid map is an edited version of pSPAIN, initially engineered by Vo et. al [3]. This ϕMINTO plasmid contains *mcherry*-targeting spacer sequence, an F1 ori for phage packaging, an RP4 oriT for mobilization, and a chloramphenicol resistance marker in the place of pSPAIN's kanamycin resistance marker. The Genbank file for the original pSPAIN plasmid was shared by Vo et. al. Figure made with Geneious Prime [29].