ABSTRACT

ZHANG, ZHONGTIAN. Discovery, Functional Characterization, and Utility of a Novel Type III-D CRISPR-Cas System from *Saccharopolyspora erythraea* (Under the direction of Dr. Gavin J. Williams).

Clustered regularly interspaced short palindromic repeat (CRISPR)-Cas (CRISPR-associated protein) systems found in prokaryotes can provide protection against invading genetic elements. Using CRISPR RNAs (crRNAs) as a guide, CRISPR-Cas effector complexes can locate and degrade the target nucleic acids to avoid viral infection and block plasmid transfer. The mechanisms to silence invasive genetic elements in prokaryotes are diverse. Here we report the discovery of an active type III-D CRISPR-Cas system in the macrolide antibiotic producer *Saccharopolyspora erythraea* (hereafter referred to as SeryCRISPR). *In silico* analysis showed unique arrangement of *cas* genes. The endogenous CRISPR-Cas system was demonstrated to block plasmid transfer by recognizing the RNA transcripts of the plasmid. The crRNA processing and maturation was elucidated through small RNA-Seq. Furthermore, the SeryCRISPR was reconstituted in a heterologous host and six core genes were demonstrated to be necessary for CRISPR interference. A high-throughput, protospacer adjacent motif (PAM) screening platform was developed which led to the observation that no PAM was required for SeryCRISPR interference. Protein pull-down assays revealed two subunits of the effector complex. Plasmid transformation assays revealed 26 to 36 bp of spacers are functional. The SeryCRISPR can be programmed to target designated sequences by providing a mini-CRISPR array. In the native host of *S. erythraea*, it was demonstrated that genome targeting using endogenous CRISPR-Cas, supplemented with editing template results in the desired gene integration. In *Escherichia coli*, SeryCRISPR can be used to regulate gene expression through RNA interference. Furthermore, it was observed that the RNA targeting of SeryCRISPR induces

deleterious mutation in *E. coli*. The results establish SeryCRISPR as a novel RNA targeting system that does not require PAM recognition and can be utilized for gene editing as well as transcriptional perturbation.

© Copyright 2021 by Zhongtian Zhang

All Rights Reserved

Discovery, Functional Characterization and Utility of a Novel Type III-D CRISPR-Cas System from *Saccharopolyspora erythraea*

by
Zhongtian Zhang

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy

Microbiology

Raleigh, North Carolina
2021

APPROVED BY:

Dr. Gavin J. Williams Committee Chair	Dr. Rodolphe Barrangou
Dr. Eric S. Miller	Dr. José M. Bruno-Bárcena

DEDICATION

To my loving wife, Siyao Li, without whom this would not have been possible.

BIOGRAPHY

Zhongtian Zhang was born in Shaoxing, China and raised by his loving parents and grandparents. He grew up with an aptitude for science and engineering. He was tutored by his great grandfather in English and had the idea of studying abroad when he met Charlotte Doyle, who later became his godparent. He attended the University of Illinois at Urbana-Champaign where he studied Chemistry. Then he did his master's study in Biosystems Engineering at Auburn University. In 2017, he joined the Williams Lab at North Carolina State University in pursuit of a Ph.D. in microbiology.

ACKNOWLEDGMENTS

I would first like to acknowledge my advisor Dr. Gavin Williams. His mentorship, support and encouragement have made me an independent researcher. I would like to thank my committee Dr. Rodolphe Barrangou, Dr. Eric Miller, Dr. José Bruno-Bárcena on teaching me CRISPR-Cas fundamentals, microbial genetics, and physiology. I would also like to thank Dr. Chase Beisel for helpful discussions.

I would like to acknowledge my coworkers. Their hands-on training, advice, and support enabled my benchwork as well as scientific thinking. They are my mentors and my friends.

Finally, I thank my family who have supported me and made me the person I am today. My parents, my grandparents, my wife, and my godparent who always encouraged my curiosity while instilling the values of integrity, kindness, and hard work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

CRISPR-Cas technologies and their applications in natural product biosynthesis, engineering, and synthetic biology

1.1 CRISPR-Cas adaptive immunity machinery and CRISPR-Cas based genetic engineering

Natural products (NPs) derived from secondary metabolites of actinomycetes have important biological activities, including antibacterial, antifungal, immunosuppressive, anticancer, and antivirus activities.[1,2] Within the phylum of Actinobacteria, the genus of *Streptomyces* are most abundant in biosynthetic gene clusters (BGCs) for producing NPs.[3] They are significant sources for clinically essential medicines and inspire new drug development, such as the anticancer compound daunorubicin from *Streptomyces peucetius* and the antibacterial daptomycin from *Streptomyces roseosporus*.[4,5] Genetic manipulation of *Streptomyces* has been challenging due to the lack of established molecular biology toolboxes and traditionally relies on double-crossover integration and selectable markers.[6] However, with recent advances in genetic engineering technologies, many previously hard-to-manipulate *Streptomyces* strains have been successfully edited with unprecedented efficiency.

Bacteria and bacteriophages are constantly co-evolving.[7] During this process, bacteria (and archaea) have developed intricate immune systems for preventing infections of phages and other mobile genetic elements (MGEs), among which the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) adaptive immune system remains to be a hot topic in microbiology and molecular biology to date.[8–10]

In 1987, Ishino and colleagues reported a unique repetitive sequence found in *Escherichia coli*.[11] Independently, Mojica and colleagues found curious sequences in *Haloferax*

mediterranei that are multiple copies of palindromic sequences separated by different spacers.[12,13] It was not until 2002 that those sequences were first named "CRISPR".[14] With the development of sequencing technologies and bioinformatic tools, some of the sequences in CRISPR arrays were found to match fragments of phage genomes or plasmids, raising the speculation that CRISPR may function for inhibiting phages.[15–17] In 2007, Barrangou and colleagues observed acquired immunity against phage infections by acquiring phage-derived sequence at CRISPR loci in *Streptococcus thermophilus.*[18] They also proved that CRISPR adaptation and interference are linked to CRISPR-associated (Cas) proteins. The current understanding of CRISPR-Cas bacterial adaptive immune systems is that they consist of CRISPR arrays and the *cas* genes. Their main features can be described by three distinct stages, namely protospacer acquisition, crRNA maturation, and interferences (Fig. 1.1).[19–21]

Ideally, the adaptive immune system is robust. However, phages are constantly evolving to bypass or interrupt it. As seen in most identified CRISPR-Cas systems and their gene arrangements, the genetic region is frequently attacked by viral elements such as transposons.[22] Such evolutionary pressure and mutations have created huge diversity in Cas protein sequences, gene compositions, and architectures of the genomic loci.[23] Classification of the vastly different CRISPR-Cas systems relies on signature Cas proteins and locus architecture.[22,24] Firstly, they are divided into two general classes based on whether the system utilizes a multi-subunit protein complex as the effector in the interference step (Class 1, type I, III, and IV) or only one single protein (Class 2, type II, V and VI). Various types within each class are distinguished according to the nuclease protein. For example, type I systems use Cas3, type II systems use Cas9; type III systems use Cas10, type V systems use Cas12 (previously named Cpf1), and type VI systems use Cas13 while less is known about type IV.[25] The CRISPR ribonucleotide protein effector complex

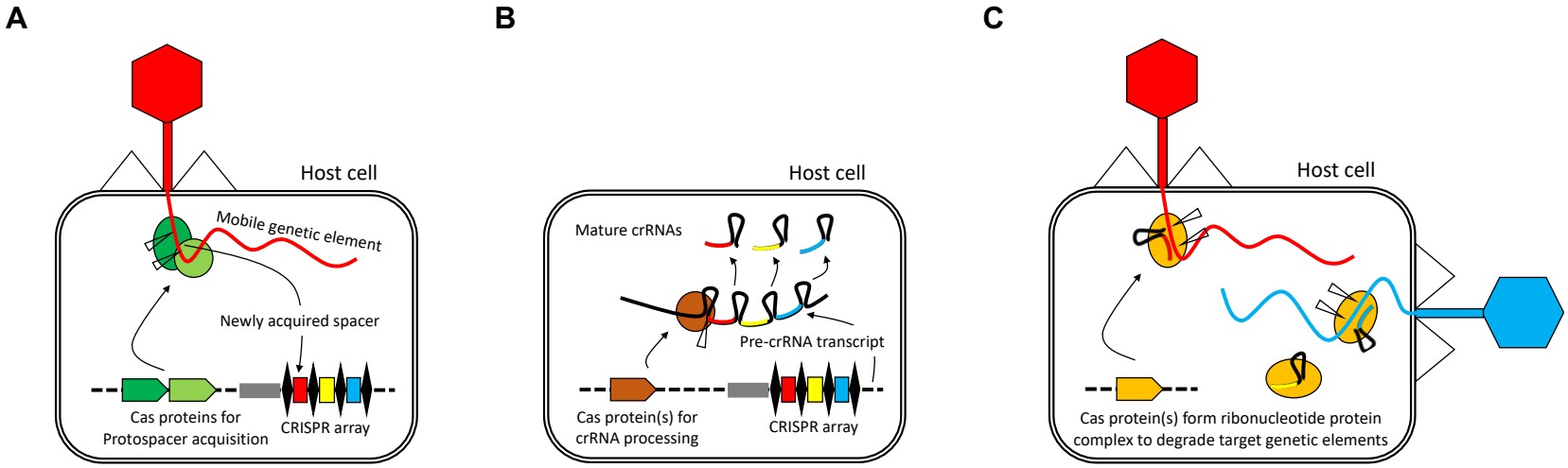


Figure 1.1. Cartoon illustration of the CRISPR-Cas system providing immunity against mobile genetic elements such as phages. (A) Acquisition. When cells are infected by MGEs such as phage, a protospacer on the invading DNA species (indicated as a red bar) is recognized by a protospacer acquisition set of Cas proteins. The protospacer is inserted to the leader end (grey rectangle) of the CRISPR array as a newly acquired spacer (red rectangle) with repeats (black diamonds). **(B)** CRISPR RNA (crRNA) processing. The CRISPR array is transcribed as pre-crRNA and processed by a crRNA maturation set of Cas proteins into mature crRNAs. **(C)** Interference. Mature crRNAs are loaded into an interference set of Cas proteins to form ribonucleotide protein complexes that find and cleave invading DNA species based on sequence complementarity.

recognizes and cleaves specific DNA/RNA sequences based on sequence complementarity between the CRISPR RNA (crRNA) and the target. To differentiate the invading protospacers from the spacers of host CRISPR loci, most CRISPR-Cas effectors recognize protospacer adjacent motif (PAM) or the equivalent to trigger binding and cleavage.[26,27]

The sequence-specific DNA cutting capability of CRISPR-Cas immediately found application in genetic engineering. Notably, though, DNA double-strand breakage (DSB) in the chromosome is lethal unless the cell responds to repair it through different pathways, and this is leveraged to enable genetic engineering. There are two main repair pathways for DSB: non-homologous end joining (NHEJ) and homology-directed repair (HDR). Since NHEJ is a random process that stitches the cleaved ends of the cut site, it can insert or delete DNA fragments that introduce mutations. HDR can introduce desired DNA sequences embedded in the homologous template that bridges the cleavage site while disabling the target recognition site but requires an additional editing template to be supplemented (Fig. 1.2). The Class 2 CRISPR-Cas systems have been most widely studied and applied for genetic engineering due to their robustness and portability. They require only two components to perform a sequence-specific DNA cleavage: a single Cas nuclease protein and a crRNA. The best-known example would be the *Streptococcus pyogenes* Cas9 system (spCas9). The spCas9 in complex with engineered single guide (sg)RNA can efficiently cleave target double-stranded DNA when the target DNA is flanked by 3'-'NGG' PAM.[28] Within ten years of their discovery, CRISPR-Cas9 based genome editing platforms have been developed for a wide range of prokaryotes and eukaryotes.[29,30]

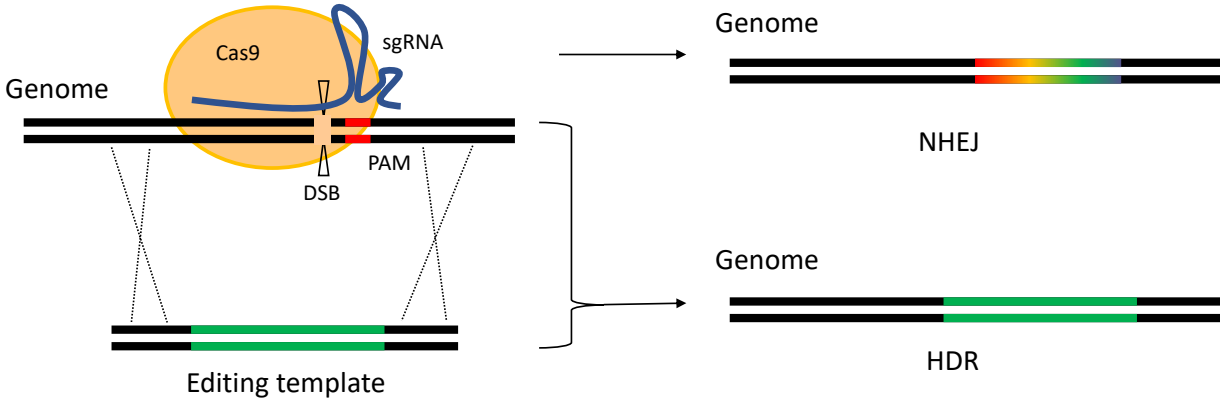


Figure 1.2. CRISPR-Cas mediated genome editing, Cas9 as an example. The Cas9 in complex with engineered single guide (sg)RNA recognizes the target DNA based on sequence complementarity. Cas9 can efficiently cleave target double-stranded DNA when the target DNA is flanked by a protospacer adjacent motif (PAM) sequence. DNA double-strand breakage (DSB) is repaired through non-homologous end joining (NHEJ) or homology-directed repair (HDR) when an editing template with flanking homology sequences is available.

Dissecting the essential nuclease further, using spCas9 as an example, the nuclease has two catalytic domains, each responsible for cutting one of the two strands of the target DNA. Silencing either of the two catalytically active residues (D10A or H840A) results in nickase Cas9 (nCas9), which alleviates lethality induced by DSB.[31,32] Silencing both of the residues turns spCas9 into dCas9, which can still perform sequence-specific DNA binding, but loses the endonuclease activities. dCas9 has been developed into the CRISPRi method for down regulation of transcription by blocking elongation (Fig. 1.3A).[33] In contrast, when fused with transcription activator modules, dCas9 fusion proteins can promote transcription of the targets (Fig. 1.3B).[34,35] Both CRISPRi and CRISPR activation (CRISPRa) are utilized for transcriptional perturbation without changing the target DNA sequence. The programmable, sequence-specific DNA binding characteristics of dCas9 have been further exploited to develop Base Editor toolbox by fusing the dCas9 with cytidine deaminase or adenosine deaminase (Fig. 1.3C).[36,37]

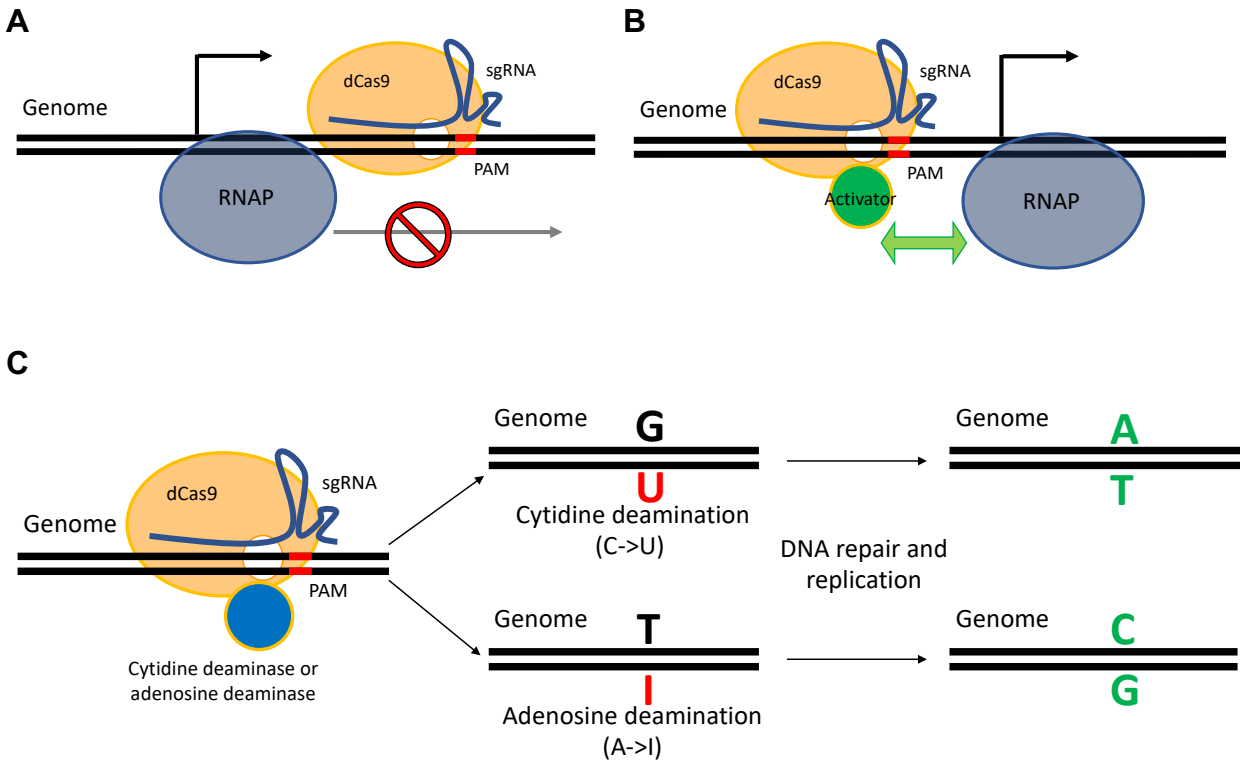


Figure 1.3. The dCas9 based CRISPR interference (CRISPRi), CRISPR activation (CRISPRa), and Base Editor systems. (A) CRISPRi system consists of a catalytically dead Cas9 (D10A and H840A mutations indicated by white circle), denoted as dCas9. The dCas9-sgRNA complex binds to the DNA target, blocking the elongation of RNA polymerase (RNAP), consequently inhibiting the expression level. **(B)** CRISPRa is applied to gene activation by the fusion of dCas9 and transcription activators to increase transcription level, upregulating the target gene expression. **(C)** Cytosine Base Editor relies on cytidine deaminase protein fused to a Cas protein to covert Cytosine to Uridine within target DNA. Adenine Base Editor converts Adenosine to Inosine within target DNA. Inosine is recognized as Guanosine (G) by DNA replication machinery. Therefore, Cytosine Base Editor can achieve C-to-T change after DNA repair and replication, while Adenine Base Editor can achieve A-to-G change.

1.2 CRISPR-Cas based genetic engineering in actinomycetes for natural product discovery and diversification

Traditionally single-crossover integrations facilitated by suicide plasmids are employed for gene disruption and editing in *Streptomyces*.[6] The spontaneous exchange of genetic contents

between *Streptomyces* genomes and the plasmids is selected via resistance markers. However, the number of available selectable markers is limiting. Moreover, the selection cassette may 'pop out', resulting in restoration of the wild-type allele under insufficient selection pressure. Alternatively, utilizing flanking recombinase target sites and expression of the corresponding recombinase can enable the recycling of markers, but this requires additional counter selection steps and is often labor-intensive and time-consuming.

During 2015, four research groups successively developed CRISPR-Cas9-based genome editing tools for *Streptomyces* species, achieving gene deletions and point mutations.[38–41] Utilizing codon-optimized spCas9 and HDR machinery from plasmid-borne repair templates (pCRISPomyces-1 and pCRISPomyces-2), Cobb and colleagues achieved the precise deletion of genomic DNA ranging from 20 bp to 31.4 kb in sizes with an efficiency of 21–100% by homology-directed repair (HDR) in three different *Streptomyces* species.[38] The platform was later applied to perform promoter knock-in's to activate silent BGCs of other classes, discovering novel type II polyketides.[42] Huang and colleagues developed pKCcas9dO for *S. coelicolor*, which achieved single deletion at high efficiency (60–100%) and double deletion at about 50% efficiency.[39] Tong and colleagues also performed gene deletion in *S. coelicolor*.[40] In addition, they experimented with NHEJ repair machinery by coupling the *ligD* gene from *Streptomyces carneus.* All these genome editing plasmids were constructed with temperature-sensitive replicon pSG5 for curing the plasmid. However, Zeng and colleagues developed a segregationally unstable pIJ101-derived shuttle vector. They found it better suited for carrying spCas9 and performing HDR because pIJ101 has a higher copy number to facilitate recombination yet can be easily cured when coupled with CodA counter-selectable marker.[41]

Both pSG5- and pIJ101 replicon-based spCas9 genome editing platforms have been developed for *Saccharopolyspora erythraea* NRRL2338 and industrial strains, namely the pKECas9 system by Liu and colleagues,[43,44] and the pMWCas9 system by Mo and colleagues.[45] In the pMWCas9 editing system, a thiostrepton-inducible promoter *tipA*p was employed to replace constitutive promoters to delay the expression of Cas9, which mitigated the Cas9 toxicity issue and enhanced transformation efficiency.[45] More importantly, pMWCas9 was successfully used to delete the *eryAIII* gene, which is highly repetitive within its coding sequence, from the polyketide synthase (PKS) involved in erythromycin biosynthesis in *S. erythraea.*[45]

Development of CRISPR-Cas systems using different Cas proteins, including *Streptococcus thermophilus* Cas9 (Sth1Cas9), *Staphylococcus aureus* Cas9 (SaCas9), and *Francisella novicida* Cpf1 (FnCpf1), for genome editing and transcriptional regulation, has offered more versatile applications in manipulating *Streptomyces* and other actinomycetes genome.[46–48] Notably, Cpf1 can process pre-crRNA into mature crRNAs enabling simultaneous expression of multiple crRNAs to guide multiplex genome editing, which is more advantageous than installing tandem sgRNA constructs to achieve dual targeting.[47] To date, several CRISPRi based gene silencing tools have been developed for *Streptomyces* based on dCas9 and ddCpf1.[47,49] In *Mycobacterium tuberculosis* and *Rhodococcus opacus*, CRISPRi gene knockdown was achieved with codon-optimized inactive Sth1Cas9.[50,51] It is worth mentioning that temporal transcriptional perturbation studies usually involve simultaneous repression of multiple targets; thus the multiplex ability of ddCpf1 based CRISPRi is slightly more advantageous than dCas9 based platform, though these two proteins have very different efficiencies in different *Streptomyces* strains.

dCas9 or nCas9 guided base editors enable single-nucleotide-resolution DNA mutagenesis in the genome without causing lethal DSB. Additionally, base editors do not rely on NHEJ or HDR, thus eliminating random mutations and oversized donor templates. Tong and colleagues developed two base editing systems, CRISPR-cBEST and CRISPR-aBEST, by fusing rAPOBEC1 cytidine deaminase and the ecTadA adenosine deaminase to the codon-optimized nCas9, respectively.[52] In *S. coelicolor*, CRISPR-cBEST converted C to T within −11 to −17 bp window upstream of the PAM sequence while CRISPR-aBEST converted A to G within −12 to −17 bp window. Furthermore, the authors used CRISPR-cBEST to successfully introduce STOP codons into the target locations in *Streptomyces griseofuscus* and mutagenized *kirN* in *Streptomyces collinus Tü365* with close to 100% editing efficiency.[52] Zhao and colleagues also achieved CRISPR-cBEST with different cytidine deaminase in *S. coelicolor* and *Streptomyces rapamycinicus*.[53] Base Editor was also applied to other Actinobacteria species such as *Corynebacterium glutamicum*, with IPTG-inducible expression of nCas9 and activation-induced cytidine deaminase fusion protein.[54,55] In the *Corynebacterium* systems, temperature-sensitive sgRNA plasmids are employed for ease of plasmid curing.

Another aspect of using CRISPR-Cas technology in NP discovery and diversification is in heterologous hosts and *in vitro*. Successful heterologous BGC expression requires cloning and refactoring the required biosynthetic pathways and provision of substrates and post-translational modification enzymes.[56] Traditional cloning methods rely heavily on the availability of restriction enzyme digestion sites. The *in vitro* application of CRISPR-Cas9 as a programmable DNA cutter circumvents this limitation. Jiang and Zhu developed Cas9-assisted targeting of chromosome segments (CATCH) and successfully excised up to 100 kb of BGC from bacterial chromosomes *in vitro* and subsequently capturing the DNA fragment via Gibson assembly.[57]

CRISPR-Cas9 and TAR cloning have enabled multiplex promoter refactoring for BGCs in yeast.[58,59] *In vitro* Cas9 digestion coupled with Gibson assembly has also enabled accurate protein engineering of type I modular PKSs with highly repetitive sequences.[60]

1.3 Challenges and Future perspectives

The most critical feature of CRISPR-Cas systems that limit their scope and utility is the toxicity of the Cas proteins, which hampers transformation and editing efficiencies. Cas9 toxicity has been widely reported in bacterial genome engineering.[61–63] These toxic effects can be overcome by using inducible promoters to delay the expression of nucleases. Using paired nCas9 may also help circumvent the toxicity and off-targeting. Another major bottleneck is that many actinomycetes lack an efficient delivery method for recombinant DNA. Large constructs can only be introduced into cells by conjugation, which significantly limits editing throughput. A third challenge is that most non-model strains lack a basic molecular genetics toolbox. For example, for several *Streptomyces* species, only two plasmid replicons are being utilized, pSG5 and pIJ101, and only one inducible promoter *tipA*p. Finally, there are no suitable bioinformatics tools for selecting target sites and currently rely on trial and error.

Future CRISPR-Cas-based genome editing opportunities in actinomycetes for NP discovery and engineering should focus on two realms. The first is enzyme-level engineering. Currently, CRISPR-Cas tools are powerful, yet the NP engineering efforts are limited to promoter insertion to turn on silent BGCs, gene deletion of competing pathways. With single-amino-acid level editing precision enabled by Base Editors, more structural diversification of NP in native hosts can be expected. Secondly, endogenous CRISPR-Cas systems are superior for genome engineering in some archaeal and bacterial strains but are yet to be adopted for

Streptomyces. For example, repurposing endogenous type I CRISPR-Cas has been successful in *Clostridium pasteurianum*, *Clostridium tyrobutyricum*, *E. coli*, *Streptococcus mutans*, *Lactobacillus crispatus*, and *Haloferax volcanii*.[64–69] Notably, endogenous CRISPR-Cas systems appear widespread in *Streptomyces* species.[70] Comparative bioinformatic analysis on various *Streptomyces* species revealed that 46 of them contained CRISPR loci, among which only the type I-E system of *Streptomyces avermitilis* was characterized.[71] By harnessing endogenous CRISPR-Cas systems, the intrinsic toxicity issue of heterologous nucleases could be mitigated. Furthermore, the genome-editing shuttle vector can be significantly reduced in size by eliminating the plasmid-encoded nuclease gene, thus promoting transformation efficiencies. Additionally, the endogenous machinery is capable of processing multiple crRNA, which will make multiplex editing feasible.

1.4 Scope of this Dissertation

This dissertation explores the novel endogenous CRISPR-Cas system found in the erythromycin-producing actinomycete *S. erythraea* (SeryCRISPR). Chapter 2 describes the *in silico* discovery of the CRISPR loci. The direct repeat sequences were extracted from the CRISPR arrays and different spacers were studied. Chapter 2 also describes the characterization of SeryCRISPR system in the native host as well as in *E. coli* to elucidate the functionality and targeting rules. The work discussed within Chapter 3 focuses on the utility of SeryCRISPR as a potential genetic engineering tool for both *S. erythraea* and heterologous host. In Chapter 4, *in vitro* Cas9 editing was explored as an advantageous tool for engineering polyketide synthases. Additionally, it was demonstrated that *Saccharomyces cerevisiae* is a promising alternative microbial chassis for natural product biosynthesis. This dissertation concludes with a vision for

how SeryCRISPR and other CRISPR technologies can be leveraged in the future to genetically engineer natural product synthetic pathways, leading to vertical advances in natural product diversification and in synthetic biology.

CHAPTER 2

In silico discovery and functional characterization of SeryCRISPR

2.1 Introduction

Clustered regularly interspaced short palindromic repeats (CRISPR) and associated proteins (Cas) provide adaptive immunity in prokaryotes against mobile genetic elements (MGE) such as phages and plasmids.[72–74] DNA or RNA sequences acquired from different MGE snippets are incorporated into the host genome, separated by short, direct repeats, forming the CRISPR arrays.[75] The CRISPR arrays are transcribed into long RNA transcripts and then diced before further processing to generate mature CRISPR RNAs (crRNA).[76–78] The mature crRNAs are assembled into a single Cas effector protein or protein complexes, directing them to detect and interfere with the cognate targets.

CRISPR-Cas systems are widespread in bacteria (46%) and archaea (90%), though the distribution of the types and subtypes vary greatly within genus and species.[79] Currently, two major CRISPR-Cas system classes have been described, encompassing six types and 34 subtypes.[80,81] Class 1 includes types I, III, and IV, defined by the presence of a multiprotein effector complex to carry out the interference. In contrast, class 2 systems are composed of types II, V, and VI, which rely on single effector nucleases to do DNA or RNA cutting.[76,82,83]

Class 1 effectors are multi-subunit protein complex systems that use a backbone built around Cas7-like proteins that bind to the crRNAs. This backbone subunit is present in multiple copies and is conserved across the Class 1 effectors. In the *Escherichia coli* type I-E Cascade system, the Cas protein effector complex binds target DNA and recruits the Cas3 enzyme to degrade DNA.[81,84] In comparison, type III effector complexes bind to target RNA transcripts,

activating several different enzymatic activities, including target RNA degradation, adjacent DNA cutting, and non-specific RNA degradation.[85] In some type III systems, the Cas7-mediated RNA cleavage activities show characteristic 6-nucleotide spacing degradation.[86–88] The large subunit Cas10 contains an HD-nuclease domain responsible for target DNA cleavage.[85,89–92] This provides a tolerance mechanism for transcriptionally dormant targets such as lysogenic phages or potentially beneficial moderate MGEs.[93] Cas10 also has the capability of synthesizing cyclic oligoadenylate (cOA) species from the GGDD-motif of the Palm domain and gives a signal to the CRISPR-associated Rossman fold (CARF) domain-containing proteins upon recognition of targets.[94,95] The CARF domain-containing proteins, Csm6 (in type III-A) or Csx1 (in type III-B), can degrade both invading RNA and host RNAs, leading to cell dormancy or cell death when activated by cOA.[96,97] Such non-discriminatory RNA degradation can be triggered to serve as an abortive immunity (Abi) mechanism if MGE breaches the CRISPR-Cas line of defense.[98] Type III CRISPR-Cas systems are observed to be more tolerant to mismatches within nucleic acid target sequences than other CRISPR types, making viral escape difficult.[99]

Actinomycetes are well known for producing many bioactive secondary metabolites. However, the endogenous CRISPR-Cas discovered in Actinomycetes are limited to type I systems within the genus of *Streptomyces*.[71] Here, we report the discovery of a novel type III-D CRISPR-Cas system found in the genome of *Saccharopolyspora erythraea* NRRL2338, dubbed SeryCRISPR hereafter. It was demonstrated that the endogenous CRISPR-Cas machinery can defend the native host against MGE by targeting the RNA transcripts. The core effector complex is also effective in a heterologous host. Its target recognition rules were characterized by a high-throughput positive readout screen. Finally, the crRNA processing pattern was elucidated through small RNA-Seq and the crRNA processing gene was identified.

2.2 Results and Discussion

2.2.1 *In silico* analysis of the CRISPR-Cas system in *S. erythraea*

The CRISPR arrays of *S. erythraea* NRRL2338 complete genome sequences (NC_009142.1)[100] were first identified by the CRISPR Recognition Tool implemented in the Geneious software (Fig. 2.1A).[101,102] The adjacent genetic region was poorly annotated with hypothetical protein CDS and transposases. We then used CRISPRdisco and CRISPRCasTyper bioinformatic pipelines to search for and annotate *cas* genes based on protein homology to the reference sets.[79,81,103–105] The analysis revealed a third CRISPR array that originally belonged to Array 2 but was interrupted by two transposase-encoding genes. The CRISPR array orientation is determined by the AT% of the flanking regions and the degree of conservation of individual repeats compared to the consensus sequence.[14,106]

The first CRISPR array (Array 1) contains 19 spacers (length: 36–42 nt) flanked by direct repeat sequences of 36 nt with a consensus sequence of (5′-CCGTCACCGACGCGCAGCGTCGGTCCTCATTGCGGC-3′) created by WebLogo (Fig. 2.1B)[107]. The second CRISPR array (Array 2) was comprised of 15 spacers (length: 35–59 nt) flanked by direct repeat sequences of 36 nt (5′-CTGTCACCGGCGCCAAGCGTCGGTCCTCATTGCGGC-3′), which is only four underlined nucleotides different from that of Array1 (Fig. 2.1C). The RNAfold predicted secondary structures of the repeats form similar hairpin loops, while the repeat from Array 1 has a chance of forming an additional stem and internal loop (Fig. 2.1D and 2.1E).[108] The nucleotide sequences search of all spacers using nucleotide BLAST revealed that all hits except one (17/18) are found to align with the reverse complementary of coding sequences from various bacteria genomes and MGEs (Table 2.1).[109] The genome-targeting spacers are hypothesized to participate

in gene regulation and bacterial genome evolution.[110,111] The significant bias in template strand targeting suggest that SeryCRISPR may recognize RNA transcripts rather than DNA. In addition, the adjacent motif sequences of all potential protospacers on both 5' and 3' ends of alignment were extracted. If the query spacer only had hits on one end and gaps on the other, only the adjacent sequence of the aligning end was extracted. The consensus sequences of protospacer adjacent motifs (PAM) of 5' and 3' are examined separately. For the 5' end extracted PAM, there is no conserved nucleotide sequence from -4 to -1 region, while the -7 to -5 region is showing biased preference toward G and C, which is from multiple hits of a single spacer, Array 2 spacer 12 (A2.12) (Fig. 2.1F, Table 2.1). No consensus sequences can be found in either 3' or 5' flanking regions (Fig. 2.1F and 2.1G).

Table 2.1 *Saccharopolyspora erythraea* CRISPR spacers are found to match with the reverse complementary of coding sequences from various bacteria genomes using nucleotide BLAST.

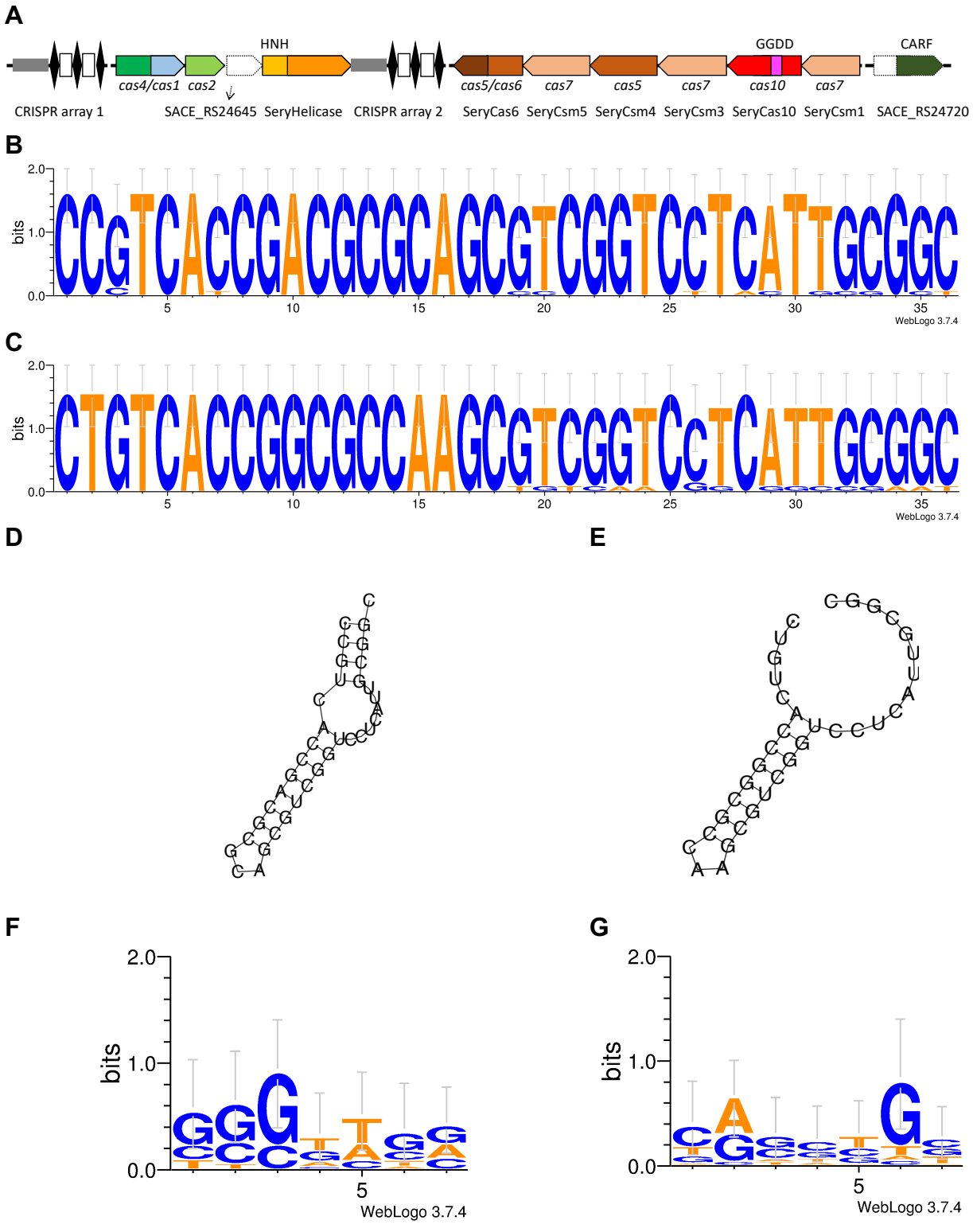
Spacer position	Targeting strand	Potential targeting species	Potential targeting gene
A1-1	template	*Corallococcus coralloides* strain B035 chromosome	cell devision protein ftsB
A1-5	template	*Cobetia sp.* AM6 chromosome	type I secretion protein
	template	*Enterobacter cloacae* strain NH77 chromosome	Ig-like domain-containing protein
	template	*Erythrobacter sp.* HKB08 chromosome	hypothetical
A1-6	template	*Actinoalloteichus sp*. AHMU CJ021 chromosome	purine NTP pyrophosphatase
A2-5	template	*Saccharopolyspora spinosa* strain CCTCC M206084 chromosome	hypothetical
	template	Self-targeting SACE_3649	hypothetical
A2-11	template	*Rhodococcus sp*. djl-6-2 chromosome	hypothetical
A2-12	template	*Streptomyces hygroscopicus subsp. jinggangensis* 5008 chromosome	DNA methylase
	template	*Azotobacter chroococcum* strain B3 plasmid pacX50fB3	helicase
	template	*Azotobacter salinestris* strain KACC 13899 plasmid unnamed1	hypothetical
	template	*Pseudomonas rhodesiae* strain BS2777 genome	helicase/REP
	template	*Anthracocystis flocculosa* PF-1	hypothetical
	template	*Streptomyces sp*. W1SF4 chromosome	helicase
	template	*Deinococcus ficus* strain CC-FR2-10 plasmid pDFI3	hypothetical
	template	*Siphoviridae sp*. isolate ctR2M13	terminase
A2-15	template	*Bradyrhizobium sp*. SK17 chromosome	hypothetical
	coding	*Nostoc sp*. C057 plasmid p057_B	hypothetical

A set of 4 putative genes was found downstream of Array 1. Two of them are identified to encode Cas4/Cas1 and Cas2 homologs responsible for protospacer acquisition as the first step of the adaptive immunity for the host.[112] A set of six genes, including the *serycas10,* are identified as the putative effector complex. The SeryCas10 is missing the canonical HD nuclease domain, but a GGDD-like minimal CRISPR polymerase (mCpol) domain was identified through conserved protein domain analysis. SeryCRISPR is considered a type III-D CRISPR-Cas system since Cas10 is the signature protein of type III CRISPR-Cas systems, and the subtype III-D is categorized by a Cas10 that lacks an HD nuclease domain but retains the GGDD-motif for cOA synthesis.[81]

Notably, SACE_RS24645 and SACE_RS24650, located in the genome directly downstream of the set protospacer acquisition genes *cas4/cas1* and *cas2,* are not identified to be relevant to the CRISPR-Cas machinery through the bioinformatics pipelines. Little can be found about the function of the hypothetical 249 amino acid polypeptide encoded by SACE_RS24645. Previously published transcriptome data of *S. erythraea* was re-analyzed in this study and we discovered that this gene is actively transcribed in a similar transcription profile to the other *cas* genes in the locus based on RNA-Seq coverage and reads, suggesting it might be an integral part of the protospacer acquisition machinery (Appendix A, Fig. A1).[113,114] On the other hand, upon examining the conserved domain of SACE_RS24650, an HNH nuclease domain and an ATP-dependent helicase domain were found. It was named SeryHelicase and is predicted to play a significant role in the CRISPR interference due to the similar domain architecture to Cas3 proteins found in type I CRISPR-Cas systems. However, no homology can be linked to any known Cas3 proteins by amino acid sequence or structural prediction (I-TASSER).[115] Additionally, SACE_RS24720 at 5kb distance from the effector complex gene cluster was

identified to encode a SMODS-associated and fused to various effector domains (SAVED) containing protein. SAVED domains are known to sense cyclic AMP/GMP, while some are also described as CRISPR Associated Rossmann Fold (CARF).[116–119]

Figure 2.1. *In silico* analysis of CRISPR-Cas system found in the genome of *Saccharopolyspora erythraea* NRRL2338. (A) Schematic representation of SeryCRISPR locus. Two CRISPR arrays are found. Homologous genes are color-coded and identified by a systematic name following the previous classification. Specialty names are given under systematic names. The subunits encoded by *serycsm1*, *serycsm3*, *serycsm4*, *serycsm5* all carry Repeat Associated Mysterious Protein (RAMP) domain. The SACE_RS24720 gene encodes a CRISPR Associated Rossmann Fold (CARF) domain. Gene region colored yellow represents the HNH nuclease domain, fused to a helicase. No HD nuclease domain can be found in *serycas10*, but a GGDD-like Minimum CRISPR Polymerase (mCpol) domain is identified. (B) The consensus sequence of the 36 nt direct repeat extracted from 20 entries of Array 1 using the frequency plot of WebLogo. (C) The consensus sequence of the 36 nt direct repeat extracted from 16 entries of Array 2 using the frequency plot of WebLogo. (D) Predicted secondary structure of the direct repeat from Array 1 using RNAfold. (E) Predicted secondary structure of the direct repeat from Array 2 using RNAfold. (F) Nucleotide frequency plot of the 5' flanking sequence of matching protospacers generated by WebLogo, the x-axis represents -7 to -1 position from left to right. (G) Nucleotide frequency plot of the 3' flanking sequence of matching protospacers generated by WebLogo, the x-axis represents +1 to +7 position from left to right.


A
HNH
GGDD
CARF
cas4/cas1
cas2
cas5/cas6
cas7
cas5
cas7
cas10
cas7
CRISPR array 1
SACE_RS24645
SeryHelicase
CRISPR array 2
SeryCas6
SeryCsm5
SeryCsm4
SeryCsm3
SeryCas10
SeryCsm1
SACE_RS24720
B
bits
WebLogo 3.7.4
C
bits
WebLogo 3.7.4
D
E
F
bits
WebLogo 3.7.4
G
bits
WebLogo 3.7.4

2.2.2 SeryCRISPR is active in defending against MGE

In silico analysis suggests the SeryCRISPR is a novel type III-D system with a nuclease-inactive Cas10. The model systems for studying type III CRISPR-Cas are Csm (type III-A) and Cmr (type III-B) complexes where Cas10 with HD domain acts as ssDNase.[89–91,120] The DNase activity is activated by RNA binding and recognition. Cas10 (Csm1/Cmr2) also generates cOA from ATP using the GGDD motif of the conserved Palm domain to signal trans-acting Csm6/Csx1 ribonuclease for non-specific ssRNA degradation.[94,95,121] To distinguish "non-self" from "self" RNA, type III systems rely on the base-pairing potential between the 5′-handle of crRNA and the 3′-flanking region of the target transcript, rather than PAM recognition.[122] The noncomplementary 3′-flanking region of the target RNA specifies target recognition and interference, while complementarity to the 5′-handle of crRNA indicates that the sequence derives from the CRISPR array. Based on the genetic arrangements, functional protein domain occurrences, and that most spacers are targeting onto the reverse complementary of the aligned genes, it was hypothesized that while SeryCas10 is a nuclease-deactivated Cas10 variant, the SeryCRISPR could target and degrade RNA transcripts, with the possibility of recruiting the SeryHelicase to activate DNA interference. In addition to that, it could be synthesizing cOA to signal SACE_24720 or other CARF domain ribonuclease to commence Abi.

To validate the functionality of SeryCRISPR in defending against MGE and to probe the self *vs*. non-self recognition machinery, plasmid interference assays were designed that tested the ability of the native system to prevent the uptake of plasmids carrying actively transcribed targets, where the repeat-spacer installed in reverse complementary orientation serves as the negative control as it is identical to array orientation in the native host (if the arrays were transcribed in the opposite direction). The sequences of the first, middle and last spacers from

both arrays (A1.01, A1.10, A1.19 and A2.01, A2.10, A2. 15) were selected and installed onto Cas9-deleted, pCRISPomyces-2. The spacers are constitutively transcribed under the control of a *gapdh* promoter in the reverse complementary orientation in reference to the native spacer sequences with and without flanking direct repeats. The corresponding series of plasmids are called pCRISPomyces-Target (Fig. 2.2A).[123–125]

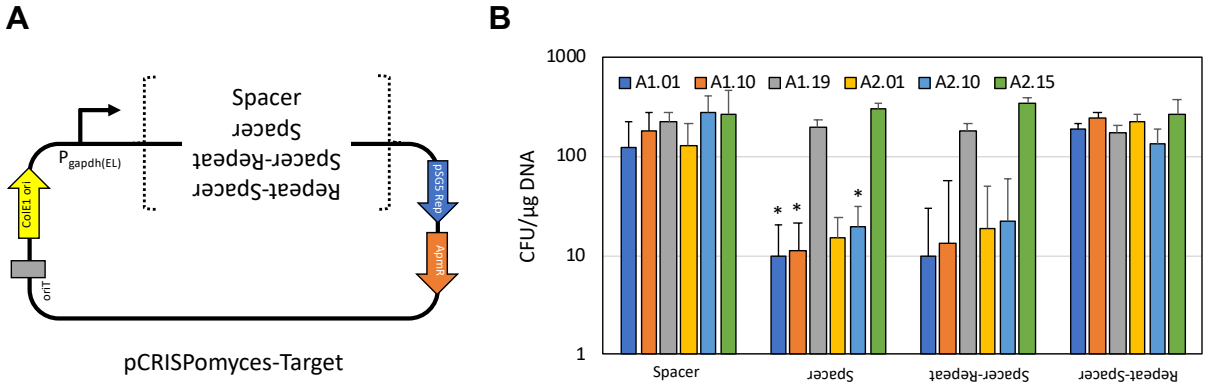


Figure 2.2. Plasmid interference assays in *Saccharopolyspora erythraea* reveals the native CRISPR-Cas is active in defending against mobile genetic elements. (A) Schematic of the target plasmid used in the interference assays. Configurations that allow the protospacer RNA to be targeted by the endogenous CRISPR-Cas system led to plasmid elimination and cell death on selective medium. Upside down text denotes the reverse-complement. (B) Transcription-dependent plasmid interference using six native CRISPR spacers from the two arrays. Upside down text indicates that the DNA sequence of the installed target is the reverse-complement to the CRISPR array and spacer sequence. The bar graphs represent the mean of 3 independent biological replicates, and the error bars represent the SD. *P < 0.05 after Welch's *t*-test to compare each sample with corresponding negative control ("Spacer" samples).

The interference assay results showed that the interference plasmids could successfully trigger a defense mechanism when spacer sequences are installed in reverse complementary under constitutive transcription (for spacer A1.01, A1.10, A2.01, and A2.10) (Fig. 2.2B). Specifically, the transformation efficiency was reduced by 12-fold, 16-fold, 9-fold, and 15-fold, respectively, reflecting that both arrays are active in the native host of *S. erythraea*. Furthermore, the plasmid interference is abolished when the 5'-flanking sequence of the protospacer RNA transcript is identical to the repeat but not when a direct repeat is present at 3'. These results

suggest that the target recognition happens between crRNA and RNA transcript of protospacers, while self *vs*. non-self discrimination is determined by base pairing of 5'-handle of the crRNA. No interference was observed for target plasmids carrying spacer A1.19 or A2.15 in either orientation. When examining the flanking repeats of those two spacers, the repeat sequences on the 5' end of the two spacers both have one bp deviation than the consensus that may interfere with the formation of the pre-crRNA secondary structure and crRNA maturation.

Overall, these results suggest the ribonucleotide protein complex recognizes RNA transcripts and triggers a defense mechanism when the 3' handle of the target RNA species lacks complementary to the 5' handle of the crRNA in the native host of *S. erythraea*. However, the negative functional readout from the plasmid interference assays (successful CRISPR-Cas defense results in cell death) makes it difficult to distinguish whether it is from direct DNA interference triggered by RNA recognition or Abi triggered by signaling of cOA species and indiscriminately RNases.

2.2.3 Csm complex genes are minimally needed for CRISPR interference

To further elucidate the mode of action and the genes involved, the entire effector complex gene operon (encoding SeryCsm1-SeryCas10-SeryCsm3-SeryCsm4-SeryCsm5-SeryCas6) was reconstructed in *E. coli*, along with the potential ancillary proteins for DNA interference (SeryHelicase), SAVED domain-containing protein (SACE_RS24720), and another hypothetically CRISPR related unknown proteins (SACE_RS24645) on the plasmid backbone of pCDFDuet-1 with lactose inducible control (Fig. 2.3A). In addition, a second plasmid pZZ04 harboring a mini CRISPR array targeting the essential gene *ftsZ* of the *E. coli* genome under

arabinose inducible promoter was also constructed. Upon L-arabinose and IPTG induction, the CRISPR interference machinery was successfully triggered, and heterologous host cell viability was reduced by more than 20-fold (Fig. 2.3B), as judged by colony-forming units.

The system was further reduced by deleting each one of the hypothetical genes (encoding SeryHelicase/SACE_RS24645/SACE_RS24720) and effector complex genes (encoding SeryCsm1/SeryCas10/SeryCsm3/SeryCsm4/SeryCsm5/SeryCas6) and testing for the interference activity via cell viability. The results showed that removing all three hypothetically related genes encoding SeryHelicase, SACE_RS24645, and SACE_RS24720 does not impair interference activities. On the other hand, none of the effector complex genes is dispensable (Fig. 2.3B). These results proved that the entire effector complex operon of the six genes is minimally

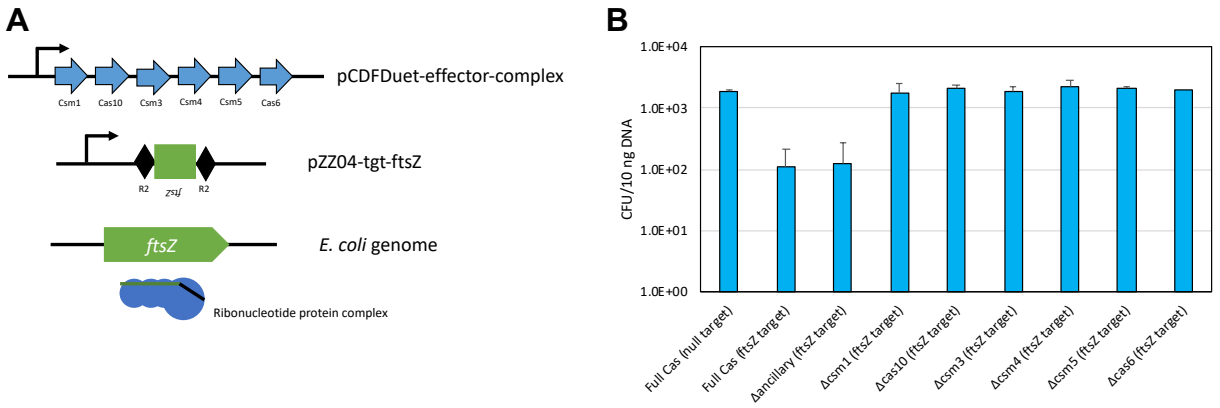


Figure 2.3. Interference assays in *Escherichia coli* reveal the six core effector complex genes are minimally required for CRISPR interference in a heterologous host. (A) Schematic of the Cas protein expression plasmid and crRNA plasmid used in the interference assays. The mini CRISPR array is transcribed, and crRNA is made to guide the protein complex to target the *E. coli* essential gene *ftsZ*. Upside down text denotes the spacer sequence of the installed target is reverse-complement to the coding sequence of *ftsZ*. (B) Transformation efficiency of the *ftsZ* targeting plasmid into *E. coli* cells harboring single gene deleted effector complex plasmid. "*ftsZ* target" is the positive control where the *ftsZ* targeting plasmid is transformed into *E. coli* cells harboring a complete set of effector complex genes. "null target" is the negative control where the null targeting plasmid is transformed into *E. coli* cells harboring a complete set of effector complex genes. The bar graphs represent the mean of 3 independent biological replicates, and the error bars represent the SD. *P < 0.05 after Welch's *t*-test to compare each sample with negative control ("null target" sample) where the six core effector complex genes and a null target crRNA are introduced.

required for the interference activities in the heterologous host, *E. coli*. Within the core effector complex, there is no predicted DNase functionality; thus, the mode of action is likely through target RNA binding and shredding. Targeting other non-essential loci such as *arsB* was attempted, but a significant reduction in cell viability was not observed. This result is contradictory to the previous plasmid clearance activity in the native host. Accordingly, this suggests either there are different DNase and RNase genes other than SeryHelicase and SACE_24720 in the genome of *S. erythraea* that performs such functionalities, or that DNase recruitment/signaling of the RNase requires additional conditions that are not replicated in the heterologous system.

2.2.4 PAM screening assay reveals no PAM requirements for target RNA interference

All CRISPR-Cas immunity requires robust self *vs*. non-self recognition machinery. Most well-studied DNA targeting systems rely on recognition of specific PAM sequences to commence interference. In contrast, type III CRISPR-Cas systems recognize target RNA transcripts that lack complementary sequence to the 5' handle of the crRNA. Studies have shown that in the type III-B system in *Pyrococcus furiosus*, the target RNAs containing an RNA PAM (rPAM) triggers DNA cleavage activities while the RNA cleavage is independent of PAM.[89]

Bioinformatics analysis of CRISPR spacers was initially used to identify matching target sequences and their adjacent motifs (Fig. 2.1F and 2.1G). Although evidence of a conserved PAM within the 7-nt region was not found, despite quality matches to sequences in the NCBI database, a possible limitation due to the availability of the matching sequences in the genomic databases cannot be ruled out. Thus, a high-throughput screen is needed to determine functional PAMs of SeryCRISPR. *In vitro* screens based on the depletion of a target plasmid or the

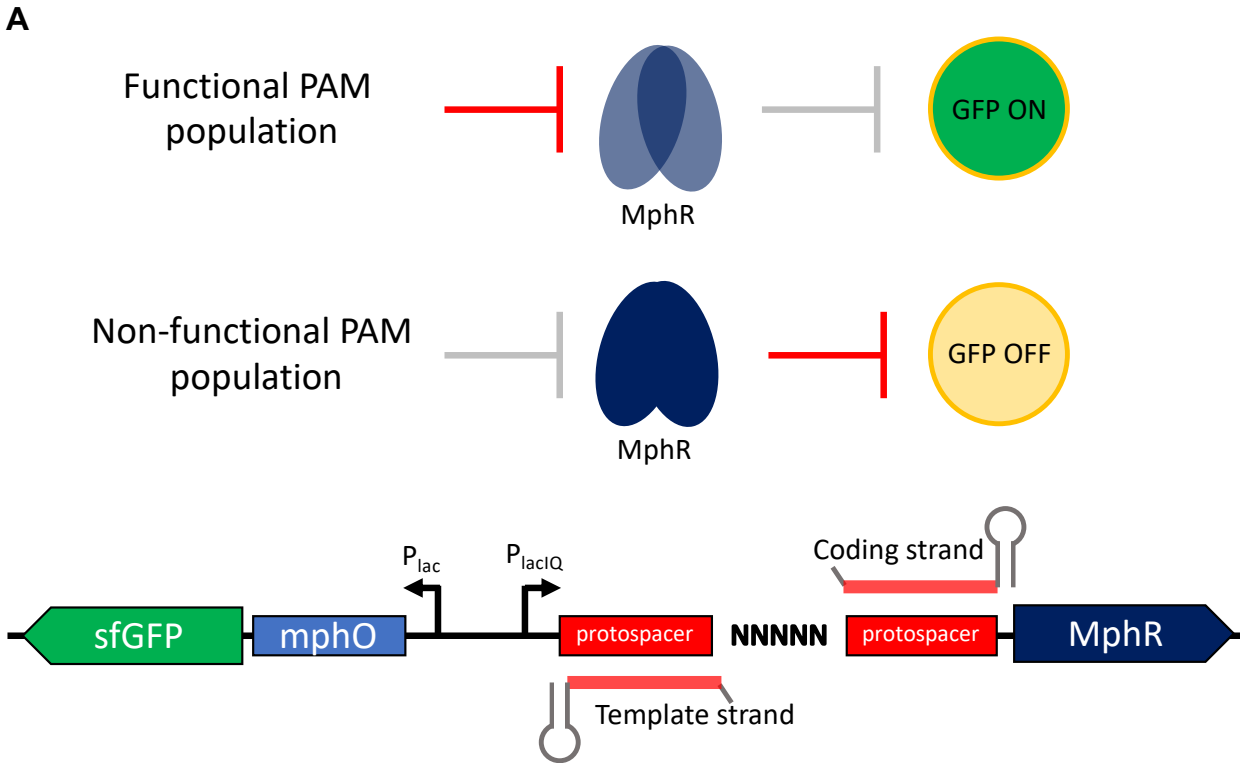
introduction of a double-stranded break have been developed to determine functional PAMs,[126–128] but they require high library coverage to identify depleted PAM sequences quantitatively. Besides, *in vitro* screens that rely on NGS are incompatible with type III systems that cleave and degrade DNA or RNA. An *in vivo*, positive screen was reported to elucidate functional PAMs by programming the nuclease deactivated CRISPR-Cas machinery to target a repressor protein in control of a reporter gene.[129] Upon successful PAM recognition, the transcription of the repressor protein gets downregulated, which would result in upregulation of the reporter gene.

The MphR-mphO repressor and operator system (Fig. 2.4A) was leveraged to construct a robust and sensitive NOT-gate repression. MphR is a repressor protein that belongs to the TetR family of regulators. Natively, MphR binds to *mphO* and tightly regulates the downstream genes responsible for resistance to macrolide antibiotics.[130] The screening platform consists of a five bp random library cloned downstream of the transcription initiation site of the *lac* operon promoter before the RBS of *MphR*. Immediately downstream of the LacIQ promoter is the *mphO* operator sequence controlling the expression of sfGFP. Since the effector complex of SeryCRISPR already has the DNA nuclease inactive SeryCas10, the entire effector complex cassette was introduced directly along with the crRNA targeting to protospacers adjacent to the random library. Theoretically, if SeryCRISPR targets DNA, the machinery will result in a portion of the library population expressing GFP in both coding strand targeting and template strand targeting. In contrast, RNA targeting will only include a GFP fluorescent population in the template strand targeting group (Fig. 2.4B). To eliminate the chance of any false positive hits caused by randomly mutated ON populations of the library, Fluorescence Activated Cell Sorting (FACS) was used to sort the library for the strict OFF population before introducing the library to *E. coli* cells already harboring *cas* gene cassette and mini CRISPR arrays.

Upon L-arabinose induction of the array transcription, no colonies with sfGFP turned on were seen from plates of 10^3 colonies in the group of coding strand targeting. In contrast, the template strand targeting resulted in most of the colonies having definitive green fluorescence (Fig. 2.4C). Next, 24 colonies were picked, and the plasmid DNA of the PAM library region was sequenced. Notably, the frequency chart of the nucleotide sequences shows significant variance at each position (Fig. 2.4D). Taken together, this data indicates that the SeryCRISPR targets RNA transcripts and the interference activity is independent of the PAM sequence.

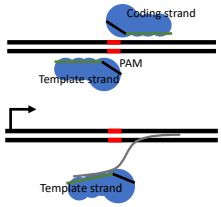
Figure 2.4. (A) The screening platform consists of a five-nucleotide random library of potential PAM sequences cloned downstream of the lacIQ promoter flanked by two identical protospacers controlling the transcription of the MphR repressor protein. Immediately downstream of the lac promoter is the MphR binding site (mphO) controlling the expression of sfGFP. The nuclease deactivated SeryCRISPR is targeted to a protospacer. sfGFP expression would indicate the PAM sequence successfully triggered interference. (B) Theoretical outcomes of different target recognition modes (DNA or RNA) by target strands. DNA targeting will result in a portion of the library population have GFP ON signal in both coding strand targeting and template strand targeting. In contrast, RNA targeting will only have a GFP ON population in the template strand targeting group. (C) Cells harboring a random PAM library were negatively screened by fluorescence-activated cell sorting (FACS) before transforming into the CRISPR-Cas machinery. Coding strand targeting yielded no ON population, while the template strand targeting resulted in ON population. (D) 24 Colonies of the ON populations were picked and sequenced. The PAM library region shows no conserved sequence as seen in the nucleotide frequency plot generated by WebLogo from representative colonies.

A



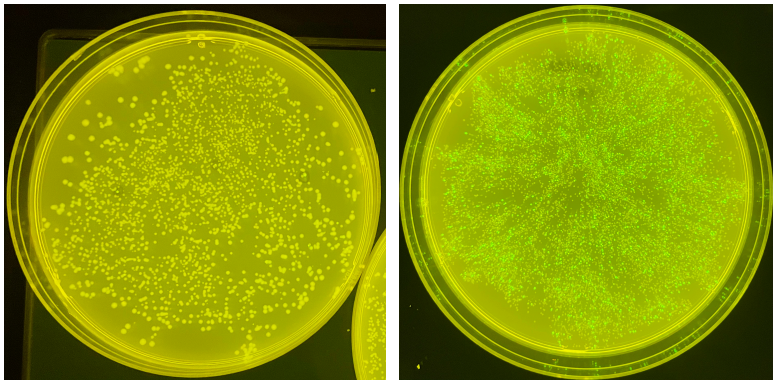
B

Will the library have GFP-ON population?



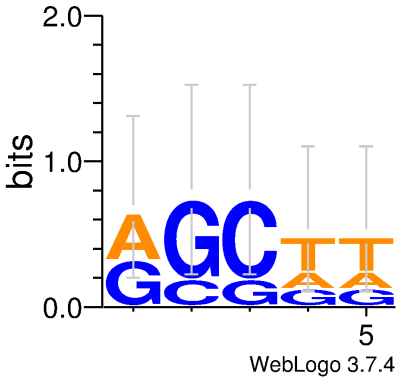
	Coding strand	Template strand
DNA Targeting	Yes	Yes
RNA Targeting	No	Yes

C



Coding strand targeting | Template strand targeting

D



2.2.5 Protein complex pull-down assays reveal the composition of effector complex

To gain insight into the ribonucleotide protein complex of the SeryCRISPR effector, heterologous expression and subsequent pull-down assay was performed. Briefly, the first Cas7-like protein (SeryCsm1) in the operon was appended with an N-terminal hexahistidine tag via PCR mutagenesis. The protein complex was expressed with a non-targeting crRNA in *E. coli* (Fig. 2.5A). The cells were lysed, and proteins were purified with nickel affinity chromatography. The His-tagged SeryCsm1 protein and any proteins in complex with it were separated and collected. The collected fractions were passed through 100 kDa molecular weight

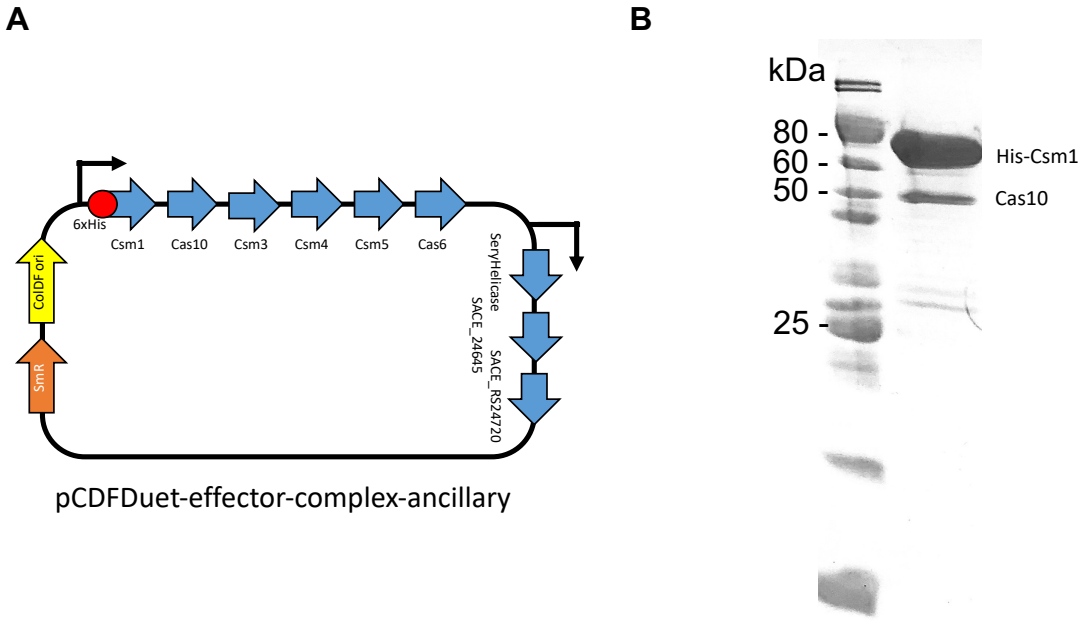


Figure 2.5. Protein pull-down assay shows at least Cas10 is forming a complex with Csm1. (A) Schematic representation of the Cas protein expression plasmid with the core effector complex genes and three related genes (SeryHelicase, SACE_24645, and SACE_24720). The Csm1 is His-tagged at N-terminus. (B) The Cas proteins were expressed with crRNA present. His-tagged protein and proteins in complex with is were purified by nickel affinity column chromatography. His-tagged SeryCsm1 protein that was not forming the complex was removed by size filtration. The protein complex was collected and analyzed by SDS-PAGE.

size filtration and washed thoroughly to remove the His-tagged SeryCsm1 protein that was not forming the complex. After size filtration, the protein complex was then analyzed by SDS-PAGE

to reveal the subunit composition (Fig. 2.5B). Two of the bands shown are apparent sizes matching the two subunits encoded by the expression system, namely the His-SeryCsm1 and SeryCas10. At the same time, a third protein of approximately 28 kDa was also observed. Based on the molecular weight, this unknown protein may be encoded by an in-frame, internal open reading frame within the CDS of SeryCas6.

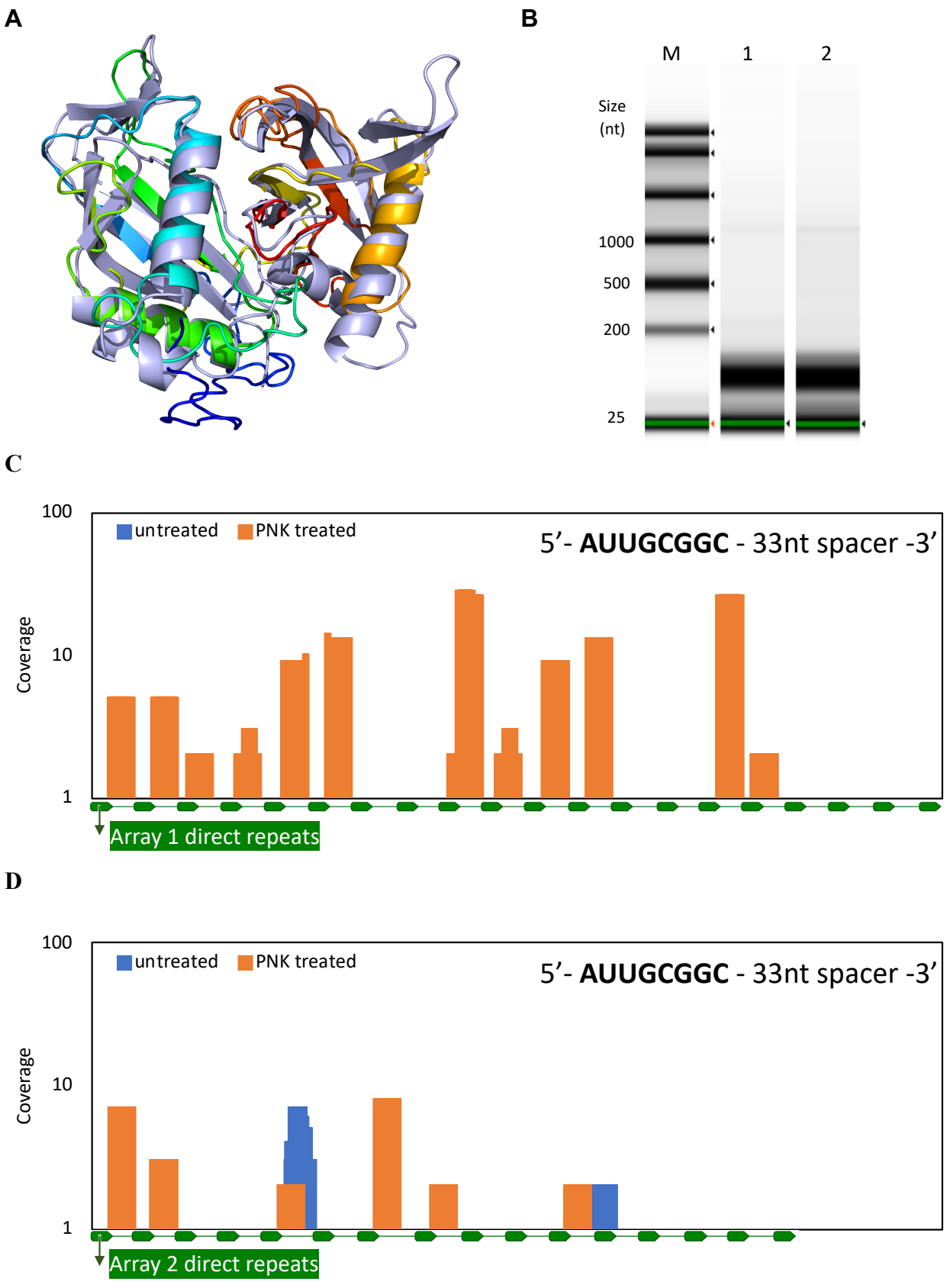
2.2.6 RNAseq elucidates crRNA processing pattern and Cas6 homolog identified

To identify the enzyme that generates crRNAs, the *cas* genes of the *S. erythraea* CRISPR loci were revisited. While a set of protospacer acquisition genes and six interference *cas* genes were identified, no confident Cas6 homolog was found. Genome annotation of originally assigned the 1173bp SACE_RS36875 as "*cas5u/6u*" CDS of uncharacterized Cas I-U subtype, derived by automated computational analysis using gene prediction method by protein homology. Upon inspection, a 258 amino acid polypeptide (28 kDa) can be translated from an in-frame GTG start codon within the *cas5u/6u*. This potential internal ORF-translated Cas protein shares 50% amino acid identity to other uncharacterized type I-U Cas5/Cas6 proteins and uncharacterized Csb2 proteins.[109] Since the amino acid sequence provides minimal insight, an attempt to elucidate the structure of the potential internal ORF protein by structural modeling with I-TASSER was carried out.[115] PfCas6 was identified as the closest structural homolog functionally verified and shared a similar overall architecture (Fig. 2.6A).

To study the crRNA processing pattern in the native host, small RNA-Seq of *S. erythraea* was performed. The crRNA processing pattern of type III CRISPR-Cas may yield crRNA with a 5'-OH and a 2',3'-P, which can prohibit the ligation of the adapters.[131] To optimize the detection and the coverage of mature crRNA, the small RNA from the total RNA of *S. erythraea* was

extracted. The purified small RNA species are then treated with T4 PNK to remove the 2',3'-P and then further incubated with ATP to reinstall the 5'-P.[132] The library was then prepared from 20nt-80nt RNA species extracted from the gel (Fig. 2.6B). The small RNA libraries were sequenced through Illumina HiSeq2000 RNA-Seq methodology. Over 300 million individual sequence reads were trimmed and mapped to the reference genomes and elucidated the processing patterns of the two CRISPR arrays. Processing occurs within the repeat elements of both arrays despite the difference in ribonucleotide sequences and predicted secondary structure, generating uniform crRNAs with a 5'-terminal AUUGCGGC 8-nt tag and 33-nt spacers. Both CRISPR arrays are actively producing mature crRNAs. This result supports the previous observation that protospacer sequences from both arrays can trigger plasmids clearance (Fig. 2.2B). It is also consistent with previous studies that the type III mature crRNAs do not maintain 5'- P, thus can only be detected when treated with T4-PNK.[131]

Figure 2.6. (A) Prediction of an internal open reading frame encoded protein within s*eryCas6* may be responsible for processing crRNA. The structural model (I-TASSER) of SeryCas6 (internal ORF, rainbow) shows high similarity to PfCas6 (light blue). (B) Small RNA was extracted from total RNA of *S. erythraea*, followed by T4 PNK and ATP treatment. (C) Small RNA-Seq coverage of Array 1, no coverage is observed in untreated small RNA sample while PNK-treated sample reveals mature crRNA of 41-nt, with an 8-nt (AUUGCGGC) tag at 5'-end. (C) Small RNA-Seq coverage of Array 2, PNK-treated sample reveals mature crRNA of 41-nt, with an 8-nt (AUUGCGGC) tag at 5'-end.


A
B
M
1
2
Size
(nt)
1000
500
200
25
C
100
10
1
Coverage
untreated
PNK treated
5’- AUUGCGGC - 33nt spacer -3’
Array 1 direct repeats
D
100
10
1
Coverage
untreated
PNK treated
5’- AUUGCGGC - 33nt spacer -3’
Array 2 direct repeats

2.3 Conclusions

SeryCRISPR, the endogenous type III-D CRISPR-Cas system discovered in *S. erythraea,* has now been demonstrated to effectively defend the host against plasmids, despite nuclease-inactive Cas10. The plasmid interference assay also suggests that SeryCRISPR avoids self-targeting when 5'-flanking regions of the crRNAs are complementary to direct repeats. A novel Cas protein, SeryHelicase, has also been identified that has similar domain architecture to the canonical Cas3 proteins from type I systems, though the protein's function is yet to be determined.

Reconstitution of SeryCRISPR in *E. coli* revealed six effector complex genes are all functionally required. However, the protein pull-down assay could only verify two subunits of the effector complex, SeryCsm1, and SeryCas10. The nuclease-inactive SeryCas10 enabled a high-throughput functional PAM screen and proved that the SeryCRISPR targets RNA transcripts and does not rely on PAM to commence interference. The small RNA-Seq revealed that both arrays are active and the mature crRNAs are 41-nt in length, with 5'- AUUGCGGC tag followed by 33-nt spacers.

Overall, this study provides a framework to discover and characterize novel endogenous CRISPR-Cas systems based on *in silico* examination, small RNA-Seq, plasmid interference assays, high-throughput functional PAM screening, and protein expression and pull-down assays. Additionally, this research may provide new insight into the evolutionary diversification of type I and type III CRISPR-Cas systems.

2.4 Materials and Methods

2.4.1 Microorganisms and Culture Medium

The bacterial strains used in this study are listed in Appendix A, Table A1. *Escherichia coli* DH5α was used for the construction of recombinant plasmids. Cells were cultivated on Luria–Bertani medium (tryptone 10 g/L, yeast extract 5 g/L, and NaCl 10 g/L) at 30 or 37 °C, flasks with shaking at 220 rpm, supplemented with apramycin (100 μg/mL) during plasmid cloning. *E. coli* BL21 (DE3) was used for protein expression in the pull-down assay. *E. coli* BL21 (DE3) was also used to carry out interference assays and PAM screening. For the selection of apramycin-resistant *S. erythraea* after transformation, 50 μg/mL apramycin was used. The *S. erythraea* was grown on R2YE agar plates. For seed-stock preparation, strains were cultured in 250 mL flasks containing 30 mL of tryptic soy broth (TSB) for 48 h at 30 °C, with shaking at 220 rpm. Following this, 0.5 mL of seed cultures were inoculated into a 500 mL flask containing 50 mL TSB medium under the same culture conditions. The cell pellets were harvested at 48 h for total RNA extraction.

2.4.2 CRISPR-Cas System Detection and Characterization *in silico*

The *S. erythraea* NRRL2338 genome available in the National Center for Biotechnology Information (NCBI) GenBank in June 2020 was mined to determine the occurrence of CRISPR-Cas systems. The *in silico* analyses were performed using CRISPRdisco and CRISPRCasTyper.[103,133] First, BLAST was used to retrieve the SeryCRISPR Cas proteins using previously identified Cas proteins in other species as queries (identified Cas Proteins are listed in Appendix A, Table A2). Then, the putative CRISPR arrays were identified using the CRISPR Recognition Tool implemented in Geneious 10.0.6 software. After that, the CRISPR-Cas

systems were manually curated and annotated to specialty names. Finally, the CRISPR subtypes were designated based on signature Cas proteins (HD domain deactivated Cas10) and associated proteins.

2.4.3 CRISPR Array Analysis and PAM Prediction

Bioinformatic analysis was performed to predict the PAM sequence based on spacer–protospacer match, using the spacers of *S. erythraea* NRRL2338 strain and BLAST algorithm. The WebLogo server was used to represent the consensus sequences of the repeats from the two arrays and represent the PAM sequence based on a frequency chart. The RNAfold server was used to predict the secondary structure of the repeat sequences when transcribed to crRNAs.

2.4.4 RNA Extraction, PNK Treatment, and RNA Sequencing Analysis

Total RNA of *S. erythraea* NRRL2338 was isolated from a 10-mL TSB culture, with two independent biological replicates grown to an optical density at 600 nm (OD600nm) ~ 0.6. Cells were harvested by centrifugation (3,200 × g, 10 min, 4 °C), and the cell pellets were flash-frozen and stored at −80 °C until RNA extraction was performed. Total RNA was isolated using a Zymo DirectZol RNA Miniprep Kit (Zymo Research), following the manufacturer's protocol. Small RNA from total RNA was isolated by gel extraction followed by T4 PNK treatment to remove potential 2',3'-P and reinstall the 5'-P. The small RNA library preparation and sequencing were performed at GENEWIZ, and data analysis was performed by Geneious software.

2.4.5 DNA Manipulations

Plasmid DNA from *E. coli* was obtained using a Monarch miniprep kit (NEB) following the manufacturer's instructions. PCR primers, double-stranded synthetic DNA for plasmid interference assays, and single-stranded DNA for annealing oligonucleotides were synthesized by Integrated DNA Technologies (IDT; Morrisville, NC). PCR amplicons for colony screening were generated using standard PCR protocols and Phusion Hot Start High-Fidelity Polymerase (Invitrogen). DNA sequencing was performed by Genewiz to confirm sequence content. Restriction digestions were performed using high-fidelity restriction enzymes (NEB). Digested products for ligation were purified using a Monarch PCR & DNA Cleanup Kit (NEB) or Monarch DNA Gel Extraction Kit (NEB). Blunt end ligation reactions were performed at a 3:1 insert/vector ratio using 50 ng of vector in a final volume of 10 µL, using T4 ligase and PNK (NEB) based on the manufacturer's instructions. To anneal two single-stranded DNA oligonucleotides, each oligonucleotide was resuspended in water to a final concentration of 100 µM, each strand (A + B) was mixed, and the final volume was adjusted to 50 µL with IDT Duplex Buffer. Both strands were annealed by incubating at 95 °C for 2 min, followed by a gradient decrease of temperature at 0.1 °C/s to room temperature. Gibson Assembly was performed using NEBuider HiFi DNA Assembly kit (NEB) following the manufacturer's protocol.

2.4.6 Construction of Plasmids

pCRISPomyces-Target series of plasmids used in plasmid transformation assay in *S. erythreae* was constructed by first deleting *cas9* from pCRISPomyces-2 using primers ΔCas9_F/R, followed by BbsI digestion. pCRISPomyces-2 was a gift from Huimin Zhao

(Addgene plasmid # 61737 ; http://n2t.net/addgene:61737 ; RRID:Addgene_61737).[134] Insert constructs were made by annealing primer pairs listed in Appendix A, Table A3. Inserts were ligated to BbsI digested pCRISPomyces-ΔCas9 and transformed into *E. coli* DH5α, then selected by blue-white screening and confirmed by Sanger sequencing.

The pCDFDuet-effector-complex-ancillary plasmid was constructed by amplifying each individual gene insert from *S. erythraea* genomic DNA using primer pairs listed in Table A3, followed by Gibson Assembly. Then, single-gene deletion of pCDFDuet-effector-complex-ancillary was achieved by PCR amplification looping out the gene to be deleted, followed by blunt-end ligation, using the primers listed in Table A3 (Appendix A, Fig. A2).

pZZ04 series of plasmids were constructed by PCR amplification of (1). apramycin resistance gene from pCRISPomyces-2, (2). pSC101 ori from pKD46, (3) *araC*/pBAD promoter cassette from pKD46, followed by Gibson Assembly with two gBlocks encoding the repeat-lacZ-repeat cassette (Appendix A, Fig. A3). To install spacers targeting *ftsZ* or *mphR*, BbsI was used to digest pZZ04. Insert constructs were made by annealing primer pairs. Inserts were ligated to BbsI digested pZZ04 and transformed into *E. coli* DH5α, then selected by blue-white screening and confirmed by Sanger sequencing. Primers and gBlocks used are listed in Table A3.

The pSense-5N library was constructed by PCR amplification of pSense-mphA-mphR by degenerate primer pairs, followed by blunt-end ligation and transformation to *E. coli* DH5α (Fig. A4). The resulting library was negatively sorted using FACS before interference to ensure no false-positive ON population. Next, the plasmids from the sorted library were extracted and transformed into *E. coli* BL21 (DE3) harboring pZZ04 and pCDFDuet-effector-complex plasmids, followed by plating on LB agar plates with L-arabinose and IPTG.

2.4.7 Protein Expression and Pull-down Assay

pCDFDuet-effector-complex-ancillary and pZZ04-null were co-transformed into *E. coli* BL21 (DE3) cells. A single colony was picked and used to inoculate 3 mL LB media (50 μg/mL apramycin and streptomycin). The 3 mL culture was incubated overnight at 37 °C with shaking at 250 rpm. 1 mL of cell culture was used to inoculate a 200 mL LB expression culture (50 μg/mL apramycin and streptomycin, 10 mM L-arabinose). The culture was grown to an OD600 of approximately 0.6 at 37 °C and with shaking at 250 rpm. Effector com expression was induced by adding IPTG to a final concentration of 0.1 mM. After incubating the cell culture at 22 °C and 250 rpm overnight, cells were pelleted, resuspended in 3 mL lysis buffer (recipe below), and lysed for 8 minutes by sonication. The cells were centrifuged again for clearing the cell debris, and the supernatant was collected. His-tagged SeryCsm1 and proteins in complex with it were purified from the soluble fraction by fast protein liquid chromatography (FPLC) using a Ni-NTA column. The gradient protocol is: 5 mL 0% B; inject sample onto column, 10 mL; 5 mL 0% - 45% B; 10mL 45% - 50% B; 25mL 50% - 100% B; 10 mL 100% A. Flow rate was 1.5 mL/min. Fractions were collected and analyzed by SDS- PAGE analysis. All the fractions with His-SeryCsm1 were combined and filtered through 100 kDa Millipore centrifuge filters and washed three times with 20 volumes of lysis buffer. The final collected protein complex was analyzed by SDS- PAGE.

Buffer recipe: lysis buffer and wash buffer (A): 20 mM NaH_2PO_4, 0.5 M NaCl, and 20 mM imidazole, pH 7.5. Elution buffer (B): 20 mM NaH_2PO_4, 0.5 M NaCl, and 300 mM imidazole, pH 7.5.

2.4.8 Statistical Analyses

All bar graphs represent the mean of 3 independent biological replicates, and the error bars represent the SD. Data distribution was analyzed with Welch's *t*-test, used to compare 2 unpaired groups (sample vs. control) under the hypothesis that the 2 groups contain equal means. Comparisons with a P value <0.05 were considered statistically significant.

CHAPTER 3

Utility of a Novel Type III-D CRISPR System: (1) Genome Editing in the Native Host and (2) Transcriptional Regulation/Mutagenesis in a Heterologous Host

3.1 Introduction

CRISPR-Cas systems, especially those based on type II Cas9 and type V Cas12a, have revolutionized the genetic and metabolic engineering of bacteria. However, the intrinsic toxicity of Cas9 and Cas12a-mediated CRISPR-Cas tools can lead to cell death in some strains, which led to leveraging endogenous type I and III CRISPR-Cas systems.[135–138] The type-I systems use multiple Cas proteins in conjunction with mature crRNA to form a CRISPR-related antiviral defense complex (Cascade) to target and guide Cas3 protein to cleave foreign DNA fragments.[139–141] Similarly, type III effector complexes are comprised of several protein subunits.[142] A unique feature of characterized type III-A (Csm) and type III-B (Cmr) CRISPR-Cas systems is that they are capable of three types of nuclease activities.[143,144] Firstly, the large subunits Csm3 and Cmr4 sequence specifically cleaves targeted RNA.[145–147] The targeted RNA is recognized by crRNA in the ribonucleotide protein complex and cleaved at fixed, 6-nt long intervals, known as the 'ruler mechanism.'[86,148–150] It was shown that complementarity between crRNA and targeted RNA does not have to be exact, and the presence of mismatches does not abolish this nuclease activity.[99] Secondly, the HD nuclease domain of the Cas10 is capable of non-specific ssDNA cleavage.[151] It was observed that when RNA polymerase opens up DNA double helix during transcription, the ssDNA is exposed to adjacent Cas10 and cleaved.[152] Thirdly, type III CRISPR-Cas systems also possess non-specific RNA degradation activities enabled by signaling between the effector complex and an additional endoribonuclease protein

not incorporated in the complex.[94,97] Briefly, the palm domain of Cas10 synthesizes cyclic oligoadenylates (cOA) from ATP upon activation.[94,97] cOA is sensed by the CRISPR-associated Rossman fold (CARF) domain of the endoribonuclease and in higher eukaryotes and prokaryotes, allosterically activates the nucleotide-binding (HEPN) domain to degrade RNA.[153,154] The triggered non-discriminatory RNA degradation results in cell death or cell dormancy and serves as an abortive immunity response. Unlike the type I CRISPR-Cas systems or other widely applied systems (e.g., type II), the self versus non-self recognition of type III systems rely on the complementarity between the 5′crRNA handle and the 3′ protospacer region of targeted RNA to inhibit cleavage, rather than relying on the presence of a protospacer adjacent motif (PAM) to commence cleavage, and thus prevents the host's CRISPR array from being targeted.[152]

The complexity of type III effector complexes makes their utilization challenging. Nonetheless, type III-A CRISPR-Cas systems from *Lactococcus lactis*, *Staphylococcus epidermidis,* and *Streptococcus thermophilus* have been successfully reconstructed in *E. coli*, achieving plasmid interference and elimination by the crRNA.[136] In the native host of *Staphylococcus aureus,* the type III-A system was programmed to target the genome, and unpredictable fragment deletion was observed.[155] The non-discriminatory endoribonuclease Csm6 was used in the SHERLOCKv2 method of specific RNA detection.[155]

However, little is known about the type III-D CRISPR-Cas where the signature protein Cas10 lacks the HD nuclease domain.[156] In the previous chapter, the putative *cas* genes of the type III-D system found in *Saccharopolyspora erythraea* were identified (Chapter 2), the plasmid clearance activity was observed in the native host, and the mode of action was determined in a heterologous host as RNA targeting. The targeting rule of the system was

elucidated, revealing that no PAM sequence is required. To further understand the different nuclease activities of the type III-D CRISPR-Cas and probe the potential utility and application of the system, plasmids transcribing self-targeting CRISPR arrays were designed to harness it for precise genome editing. The interference activity with various spacer lengths were also determined in the heterologous host of *E. coli*. Although DNA interference in *E. coli* could not be reconstructed, downregulation through CRISPR-Cas machinery was observed. Genome editing in *E. coli* using the type III-D system was also attempted.

3.2 Results and Discussion

3.2.1 Determination of Optimal Spacer Length for Repurposing the Type III-D CRISPR-Cas in a Heterologous Host

crRNA maturation in type III systems consists of two steps. The first is the digestion of pre-crRNA transcripts by Cas6.[157] Subsequently, crRNAs in type III systems are further trimmed down.[158] The exact mechanism of secondary processing is not known. However, it was shown that this trimming generates different lengths of functional crRNA molecules, the most abundant being 39 nt and 45 nt crRNAs in type III-B system, which corresponds to 31 bp and 39 bp in the spacers.[159,160] A diverse range of spacer lengths were observed in the CRISPR arrays of *S. erythraea*, from 35 bp to 59 bp (Chapter 2), which led to the hypothesis that crRNA in the type III-D system also undergoes secondary trimming regardless of spacer length to produce mature crRNA. To probe the minimal required and optimal spacer length, the efficacy of spacers ranging from 21 bp to 36 bp was tested, using the same platform designed for elucidating necessary *cas* genes as described in Chapter 2. Briefly, the set of core effector complex genes were determined to be minimally required (Fig. 3.1A). Next, a second plasmid harboring different spacers in

length with the identical sequences from the 5'-end while using a non-functional 36 bp spacer as control was introduced. The cells were then plated on LB agar plates with L-arabinose to induce the transcription of the mini array. When a 36 bp spacer is tested, a 30-fold reduction in colony-forming units (CFU) was observed. A 10-fold CFU reduction was observed with the 26 bp and 31 bp spacer groups. The shortest spacer tested was 21 bp and showed no reduction in CFU when targeting the essential gene (Fig. 3.1B). However, it was also observed that expression of the Cas proteins alone, with non-targeting spacer, showed significant toxicity as seen in heterogeneous and tiny transformant colonies. One possible explanation could be that some of the subunits are capable of degrading RNA, which could be obstructing cellular metabolism. One way to mitigate such toxicity is to fine tune the expression level for each subunit as well as to optimize crRNA production.

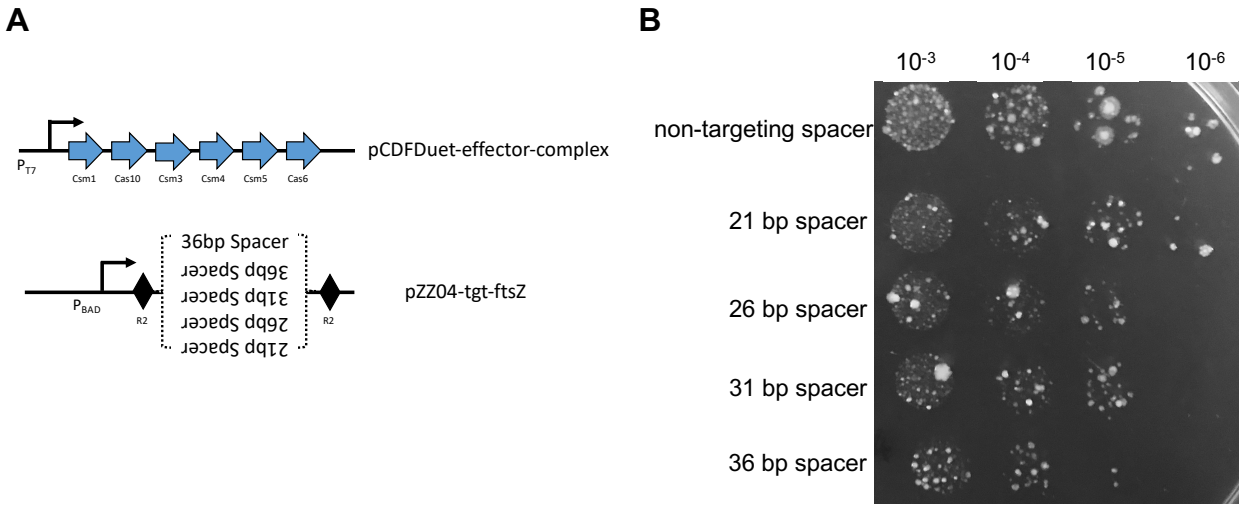


Figure 3.1. Interference assays in *Escherichia coli* revealed that a spacer length of 26 to 36 bp triggers CRISPR interference in a heterologous host. (A) Schematic of the Cas protein expression plasmid and crRNA plasmid used in the interference assays. The mini CRISPR array is transcribed, and crRNA is made to guide the protein complex to target *E. coli* essential gene *ftsZ*. Upside down text denotes the spacer sequence of the installed target is reverse-complement to the coding sequence of *ftsZ*. (B) Transformation efficiency of pZZ04 plasmids harboring different spacers into *E. coli* cells shown as colony-forming units in serial 10-fold dilutions.

3.2.2 Genome editing in *S. erythraea* NRRL2338 with native type III-D CRISPR

S. erythraea produces the clinically significant macrolide antibiotic erythromycin A. It is also a model system for the study and manipulation of polyketide biosynthesis.[161,162] CRISPR-Cas9 based methods have been developed to engineer the genome of *S. erythraea* and increase the yields of natural products by removing competing pathways, improving precursor and cofactor pools, and overexpressing transcriptional regulators.[44,163] However, the endogenous CRISPR-Cas system has not yet been harnessed for genome editing in *S. erythraea.* As a first step, the pSG5 temperature-sensitive replicon-based shuttle vector was constructed by first deleting the Cas9 of the pCRISPomyces-2 and swapping the sgRNA cassette with a gblock containing two direct repeat sequences from the CRISPR array 2, interspaced by *lacZ* for ease of selection in cloning steps. To install the designed spacers, the plasmid backbone is digested by the type IIS restriction enzyme BbsI and ligated to spacers generated by primer annealing. The loci SACE_RS08680 (erythromycin esterase, old locus tag SACE_1765) was chosen as a target to test for genome editing efficiency. A 36 bp spacer was selected from the beginning of the erythromycin esterase CDS, and 1 kb homology arms from upstream and downstream of the site were cloned to flank *mcherry* as the plasmid-borne editing template for homology-directed repair (Fig. 3.2A). After conjugal transformation, 16 colonies were observed with 10^6 donor cells. One was visibly red and was subsequently propagated for two rounds of selection on TSB agar plates containing both of the antibiotics nalidixic acid and apramycin. The plasmid was then cured by propagating the selected colonies at 37 °C. Colony PCR detection of the locus was performed with primers flanking the region outside of the homology arms (Fig. 3.2B). Red fluorescence was observed by fluorescence microscopy (Fig. 3.2C and D).

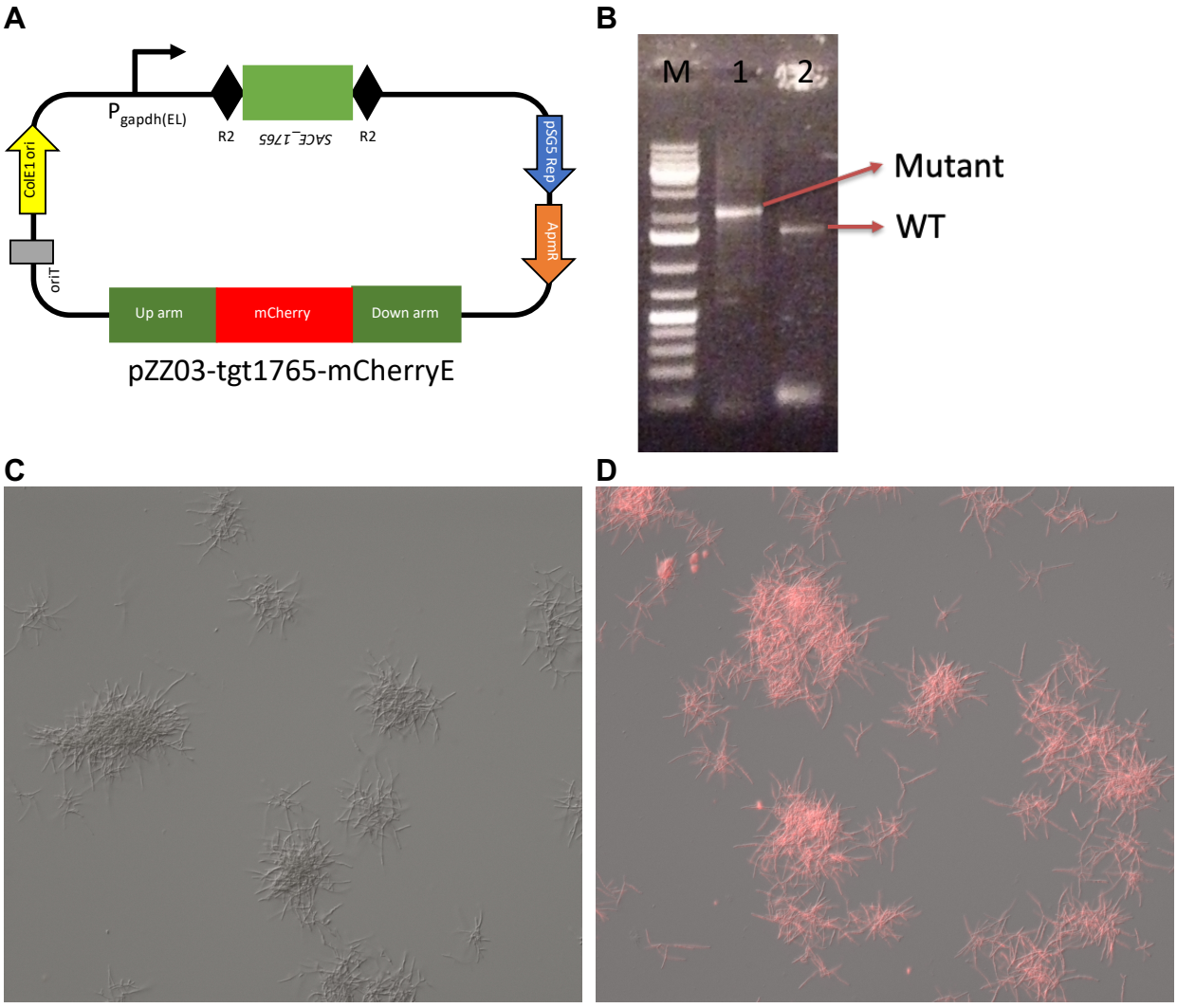


Figure 3.2. Genome integration of *mCherry* in *Saccharopolyspora erythraea* NRRL2338 at the SACE_RS08680 locus (old locus tag SACE_1765) using endogenous type III-D CRISPR-Cas and plasmid-borne editing template. (A) Design of the genome editing shuttle vector pZZ03-tgt1765-mCherryE. The genome-editing plasmids harbor pSG5 temperature-sensitive replicon, apramycin selection marker, a homology-directed repair fragment with a matching sequence to the flanking region of the target site with 1kb arms, and a mini CRISPR array with repeat-spacer-repeat construct targeting on the chromosome. (B) cPCR verification of the genome editing with primers flanking the region outside of the homology arms. Lane M is GeneRuler 1kb plus DNA ladder. Lane 1 is the mutant with genome integration of *mCherry,* and lane 2 is the wide-type negative control. (C) Fluorescence microscopy images of wild-type *S. erythraea* under white light and overlaid with red fluorescence. (D) Fluorescence microscopy images of mutant *S. erythraea* with *mCherry* integration at SACE_RS08680 locus under white light and overlaid with red fluorescence.

3.2.3 Type III-D CRISPR-Cas core effector complex can downregulate gene expression via RNA interference in a heterologous host

A type I-E system has been reconstructed in *E. coli* by deleting the Cas3 nuclease to enable transcriptional perturbation.[65] Without the nuclease, the type I-E CRISPR-associated complex resorts to a targeting DNA binding effect.[65] Using green fluorescent protein (GFP) as a reporter, it was demonstrated that the endogenous I-E CRISPR-Cas system could downregulate target gene expression by 82% in *E. coli*.[65] Since the type III-D system from *S. erythraea* contains a Cas10 protein without a canonical HD nuclease domain, it is hypothesized here that the core effector complex can be readily employed for transcriptional regulation through target RNA binding and degradation. To test this hypothesis, a plasmid was constructed that constitutively expresses sfGFP by deleting *mphR* from pSense and was introduced into *E. coli*.[164] Next, a second plasmid was introduced expressing the core SeryCRISPR effector complex (six genes) with or without three additional hypothetical ancillary *cas* genes. Empty pCDFDuet plasmid was used as a negative control. Finally, a third plasmid was introduced, transcribing *sfgfp* targeting crRNAs upon L-arabinose induction (Fig. 3.3A). It was observed that the relative fluorescence (fluorescence/cell density) significantly decreased in both groups expressing effector complex genes, with or without the additional hypothetical genes, compared to the control group transcribing the mini CRISPR array without Cas proteins (Fig. 3.3B). The impact of the hypothetical genes, SeryHelicase, SACE_RS24645, and SACE_RS24720, were indistinguishable, verifying the previous results that they are not required for CRISPR interference in a heterologous host (Chapter 2). This method provided a simple and powerful strategy for regulating metabolic fluxes.

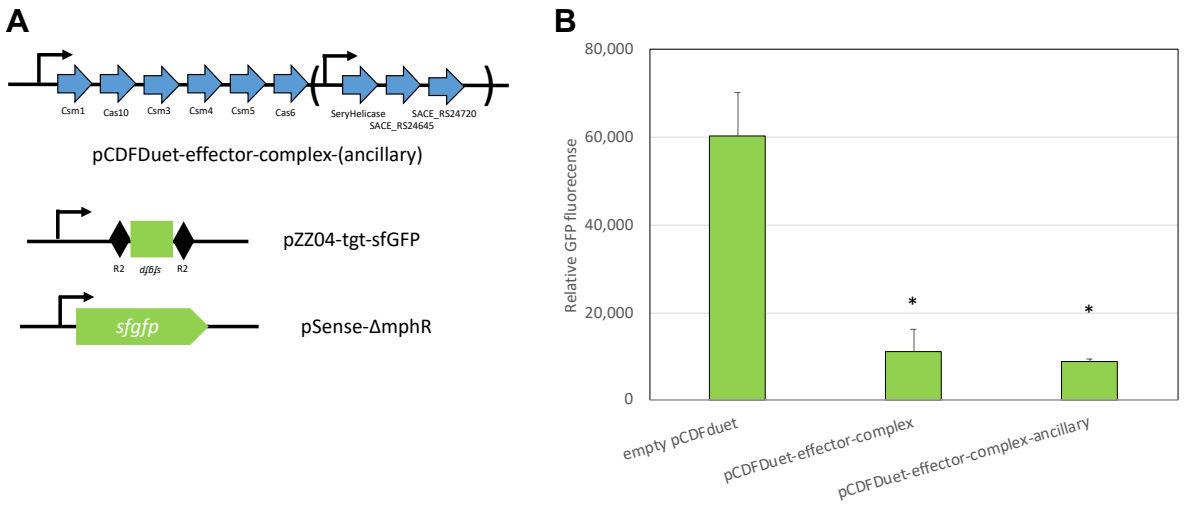


Figure 3.3. CRISPR Interference in *Escherichia coli* using type III-D CRISPR-Cas from *Saccharopolyspora erythraea*. (A) Schematic of the Cas protein expression plasmids and crRNA plasmid used in the interference assays. The pSense-ΔmphR is constitutively expressing GFP. The mini CRISPR array is transcribed, and crRNA is made to guide the protein complex with and without the presence of hypothetical ancillary genes to target on *sfgfp* from pSense-ΔmphR. Upside down text denotes the spacer sequence of the installed target is reverse-complement to the coding sequence of *sfgfp*. (B) The relative fluorescence (fluorescence intensity normalized by cell density) of *E. coli* harboring SeryCRISPR machinery targeting on *sfgfp*, with and without different hypothetical ancillary genes. The crRNA plasmid with empty pCDFDuet serves as a negative control. Relative fluorescence is determined by a plate reader on a 96-well plate. The bar graphs represent the mean of 3 independent biological replicates, and the error bars represent the SD. *P < 0.05 after Welch's *t*-test to compare each sample with the negative control.

3.2.4 Type III-D core effector complex induces mutagenesis at a target site in a heterologous host

The *S. aureus* type III-A system can achieve large-fragment genomic deletions when self-targeting crRNA is introduced.[155] The results suggest that the type III-A system regulates the stability of the bacterial genome and can be used as an efficient tool for gene knockout in bacteria. In the previous study, a PAM screen platform was developed to elucidate targeting rules of the type III-D CRISPR-Cas (Chapter 2.2.4). Briefly, the effector complex was introduced to target protospacers adjacent to a random PAM library. Successful interference would turn ON sfGFP expression. Before the assay, the random PAM library was sorted using fluorescence-

activated cell sorting (FACS) to select a strict OFF population (Appendix B, Fig. B1). The negative sort ensured no false-positive cells were carried into the CRISPR interference assay. During the CRISPR interference, a significant portion of *E. coli* colonies showed that sfGFP expression was turned ON by the SeryCRISPR machinery. Following the screen, the DNA of the ON population (twenty-two colonies) at the PAM library region was sequenced. Conserved PAM sequences were not identified. Unexpectedly, the DNA sequences of two clones had truncations of *mphR* at the targeted site. It was found that hundreds of base pairs were deleted in the genetic region downstream of the targeted region (Fig. 3.4). This sequencing information suggests that the CRISPR interference of SeryCRIPSR has likely introduced deleterious mutations. To investigate the frequency of such deleterious mutations, the ON population from the interference assay was collected and re-plated on agar plates without antibiotic selection for the crRNA plasmid. In addition, to ensure the removal of crRNA, the agar plate was incubated at 37 °C to cure the temperature-sensitive crRNA plasmid. By removing the crRNA producing plasmid, most of the library colonies reverted to no green fluorescence (sfGFP OFF), while 2% remained to be fluorescent (sfGFP ON), suggesting a high frequency of mutations caused by SeryCRISPR targeting.

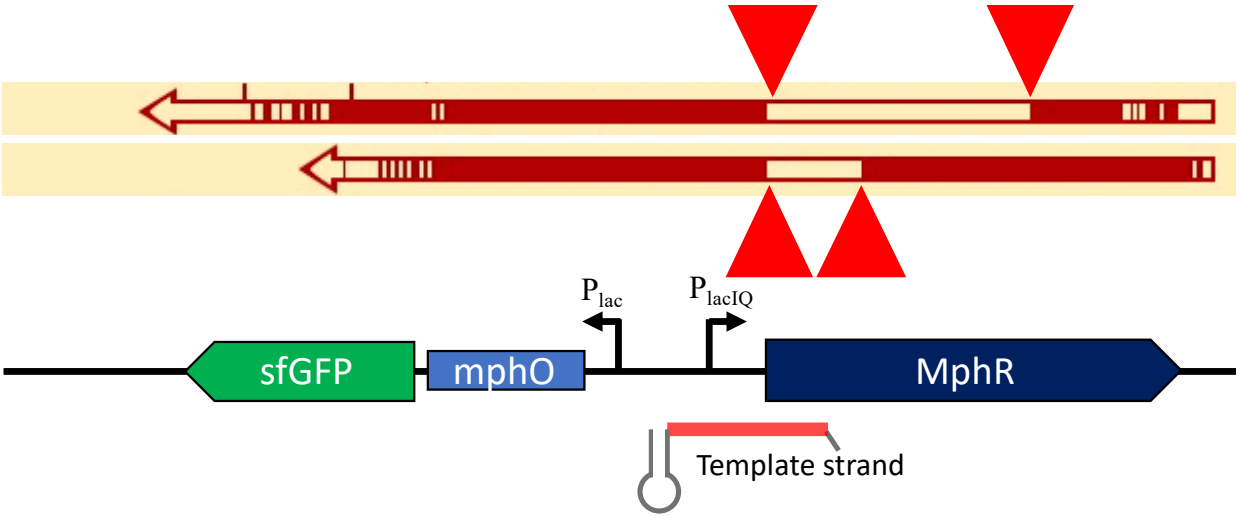


Figure 3.4. Type III-D effector complex with nuclease inactive Cas10 can induce deleterious mutation. The screening platform consists of MphR repressor protein and MphR binding site (mphO) controlling the expression of sfGFP. The core effector complex was targeted to a protospacer upstream of *mphR*. After interference, *mphR* truncated population emerged. The sequence alignment above the schematics is proportionate in size. The red triangles indicate missing sequences. The truncation happened at the targeted site shown by the representative alignments.

3.3 Conclusions

In conclusion, a type III-D CRISPR-Cas system discovered in *S. erythraea* was successfully harnessed for genome engineering to integrate *mCherry* into the chromosome. SeryCRISPR has significant potential to be developed into valuable genome editing tools. Firstly, the size of the genome-editing shuttle vector can be reduced by not having to carry the exogenous Cas protein, thus handling and transformation of the vectors can be more efficient. Additionally, the toxicity of exogenous Cas nucleases can be avoided, further improving the portability of the genome editing plasmids. Thirdly, SeryCRISPR does not require a PAM to recognize targets, potentially expanding the selection of targetable regions. Finally, the endogenous CRISPR-Cas can perform multiplex editing if an artificial CRISPR array is supplied. However, there are intrinsic limitations to the type III-D system in genome engineering

applications. The effector complex recognizes target RNA; thus, the editing efficiency theoretically correlates to the transcription rate of the target region. Moreover, in other reported type III CRISPR-Cas systems, the target recognition machinery tolerates mismatches between protospacers and crRNA. In genome editing applications, this could potentially cause off-target effects.

In a heterologous host, transcription interference in *E. coli* was also achieved by SeryCRISPR. The core effector complex can repress the expression of GFP five-fold. However, the functions of the hypothetical proteins, namely SeryHelicase (HNH domain-containing helicase) and the CARF domain-containing protein, remain unknown. Furthermore, a deleterious mutation was also observed at the target site after CRISPR interference assays despite the core effector complex lacking the nuclease domain in Cas10.

3.4 Materials and Methods

3.4.1 Microorganisms and Culture Medium

The bacterial strains used in this study are listed in Appendix B, Table B1. *Escherichia coli* DH5α was used for the construction of recombinant plasmids. Cells were cultivated on Luria–Bertani medium (tryptone 10 g/L, yeast extract 5 g/L, and NaCl 10 g/L) at 30 or 37 °C, flasks with shaking at 220 rpm, supplemented with apramycin (100 μg/mL) during plasmid cloning. *E. coli* BL21 (DE3) was used to carry out interference assays. For the selection of apramycin-resistant *S. erythraea* after transformation, 50 μg/mL apramycin was used. The *S. erythraea* was grown on R2YE agar plates. For seed-stock preparation, strains were cultured in 250 mL flasks containing 30 mL of tryptic soy broth (TSB) for 48 h at 30 °C, with shaking at 220 rpm. Following this, 0.5 mL of seed cultures were inoculated into a 500 mL flask containing

50 mL TSB medium under the same culture conditions. The cell pellets were harvested at 48 h for total RNA extraction.

3.4.2 Conjugal Transformation of *S. erythraea*

Intergeneric conjugal transfer of plasmid DNA from *E. coli* to Streptomyces is adopted.[165,166] Briefly, *S. erythraea* spores were washed twice with 2x YT medium followed by heat shock at 50 °C for 10 min. *E. coli* ET12567 (pUZ8002) was grown in LB to an OD600 of 0.6 at 37 °C and washed twice to remove antibiotics. 500 µL culture of *E. coli* was added to the spores and mixed by inversion. The cells were briefly centrifuged and spread on mannitol/soya agar plates supplemented with 10 mM $MgCl_2$ and incubated for 20 h at 30°C. One mL of nalidixic acid (0.5 mg/ml) and apramycin (1 mg/ml) was added to the agar surface. Plates were further incubated for 5 days at 30°C.

3.4.3 Relative Fluorescence Measurement

Individual colonies were picked and used to inoculate 3 mL cultures in LB media supplemented with ampicillin (100 µg/mL), apramycin (100 µg/mL), and streptomycin (50 µg/mL). The cultures were grown at 30 °C with agitation at 250 rpm to OD 0.6, at which point 10 mM of L-arabinose and IPTG were added. The cells were cultured for ~16 h, then centrifuged at 3,000 rpm for 10 min, and the supernatant was discarded. The cell pellet was then resuspended in 1 mL of phosphate buffered saline (PBS). Then, 100 µL of the cell suspension was transferred to clear flat-bottom and black flat-bottom 96-well plates for analysis of the optical density (OD600) and fluorescence (ex 485 nm/ em 510 nm). The fluorescence intensity was divided by the OD600 to yield a relative sfGFP fluorescence value.

3.4.4 DNA Manipulations

Plasmid DNA from *E. coli* was obtained using a Monarch miniprep kit (NEB) following the manufacturer's instructions. PCR primers, double-stranded synthetic DNA for plasmid interference assays, and single-stranded DNA for annealing oligonucleotides were synthesized by Integrated DNA Technologies (IDT; Morrisville, NC). PCR amplicons for colony screening were generated using standard PCR protocols and Phusion Hot Start High-Fidelity Polymerase (Invitrogen). DNA sequencing was performed by Genewiz to confirm sequence content. Restriction digestions were performed using high-fidelity restriction enzymes (NEB). Digested products for ligation were purified using a Monarch PCR & DNA Cleanup Kit (NEB) or Monarch DNA Gel Extraction Kit (NEB). Ligation reactions were performed at a 3:1 insert/vector ratio using 50 ng of vector in a final volume of 10 µL, using T4 ligase and PNK (NEB) based on the manufacturer's instructions. To anneal two single-stranded DNA oligonucleotides, each oligonucleotide was resuspended in water to a final concentration of 100 µM, each strand (A + B) was mixed, and the final volume was adjusted to 50 µL with IDT Duplex Buffer. Both strands were annealed by incubating at 95 °C for 2 min, followed by a gradient decrease of temperature at 0.1 °C/s to room temperature. Gibson Assembly was performed using NEBuider HiFi DNA Assembly kit (NEB) following the manufacturer's protocol.

3.4.5 Construction of Plasmids

The pCDFDuet-effector-complex-ancillary plasmid was constructed by amplifying each gene insert from *S. erythraea* genomic DNA as described in Chapter 2. pCDFDuet-effector-

complex was achieved by PCR amplification looping out the ancillary genes using primer pairs listed in Appendix B, Table B2.

pZZ04 was constructed as described in Chapter 2. To test for functional spacer lengths, a series of pZZ04 plasmids harboring different lengths of spacers targeting on *ftsZ* were constructed. BbsI was used to digest pZZ04. Insert constructs were made by annealing primer pairs. Inserts were ligated to BbsI digested pZZ04 and transformed into *E. coli* DH5α, then selected by blue-white screening and confirmed by Sanger sequencing. pZZ04-sfGFP was constructed by annealing sfGFP targeting primers. Primers used are listed in Table B2.

S. erythreae genome editing plasmid pZZ03-tgt1765-mCherryE was constructed by first deleting *cas9* from pCRISPomyces-2 using primers ΔCas9_F/R, followed by Gibson Assembly with two gBlocks encoding the repeat-lacZ-repeat cassette to make pZZ03. To install spacers targeting on SACE_1765, BbsI was used to digest pZZ03. Spacer constructs were made by annealing primer pairs listed in Table B2. Inserts were ligated to BbsI digested pZZ03 to make pZZ03-tgt1765. To install the editing template, 1kb up homology arm, mCherry, and 1kb down homology arm were amplified individually and amplified by overlap extension PCR followed by restriction cloning to XbaI site of pZZ03-tgt1765 to make pZZ03-tgt1765-mCherryE.

3.4.6 Fluorescence Microscopy

5 μL of the bacterial culture was smeared on a glass slide, and mCherry fluorescence was observed using a fluorescence microscope (Zeiss Axioimager M.2) with excitation at 587 nm and emission at 610 nm.

CHAPTER 4

Future directions for SeryCRISPR studies and approaches to genetic manipulation of polyketide biosynthesis

4.1 Future directions for SeryCRISPR studies

The original inspiration for this dissertation was based on an interest in developing native CRISPR-based genome editing tools for *Saccharopolyspora erythraea* and probing its scope of utility, aiming to complement traditional genetic manipulation tools. Before this study, an automated bioinformatic pipeline was used to identify the putative CRISPR arrays, yet no *cas* genes were accurately annotated or functionally characterized.[100,133] Herein, the SeryCRISPR was identified as a type III-D system through computational analysis (Chapter 2), based on the signature protein Cas10 that lacks an HD domain. The activity of the endogenous CRISPR system in defending against actively transcribed mobile genetic elements (MGEs) was also verified. Furthermore, the CRISPR interference machinery was reconstructed in *E. coli,* and the targeting mode was elucidated as PAM-independent RNA targeting (Chapter 2). Finally, such activities were harnessed for genome editing in the native host and transcriptional regulation in *E. coli* (Chapter 3). However, these novel advances notwithstanding, a few aspects of SeryCRISPR are worthy of more in-depth future studies.

Firstly, the plasmid clearance activity demonstrated in *S. erythraea* indicates DNA interference or abortive immunity via RNA degradation by trans-acting endoribonucleases. Yet, such plasmid clearance activity was not observed in our interference assays in *E. coli*. One possibility is that the nuclease domain-containing SeryHelicase is involved in the DNA interference in the native host but was not functionally reconstituted in *E. coli*. One way to test

out the potential nuclease activity of SeryHelicase is by expression, purification, and subsequent *in vitro* cleavage assay of its activity by feeding radiolabeled nucleotide species. To directly investigate the involvement of the SeryHelicase in the native host, *S. erythraea*, a SeryHelicase deletion strain needs to be created first, followed by plasmid interference assays to test if the Δ*seryhelicase* strain retains the capability of defending against MGEs. Such knock-out can be achieved by traditional recombination-based methods or Cas9-based genome editing tools.[39,163] Another possibility is that such 'plasmid clearance' results from abortive immunity, i.e., the host cells underwent suicidal response when transformed with the plasmid. The destruction of both invader RNA and host RNA was observed by *in vitro* assays in type III-A CRISPR-Cas systems (Csm6) and type III-B-Cas CRISPR systems (Csx1).[167,168] These trans-acting RNase proteins have higher eukaryotes and prokaryotes nucleotide-binding (HEPN) and CRISPR-associated Rossmann fold (CARF) domains, each responsible for non-specific RNA degradation and sensing cyclic oligoadenylate (cOA) species.[169] In our studies, the 10kb regions upstream and downstream of the SeryCRISPR loci were inspected for proteins potentially capable of picking up cOA signals and displaying endoribonuclease activities. The SACE_RS24720 was identified as the most plausible candidate because it contains a CARF-like domain and a domain of unknown function (Chapter 2). However, the analyzed genomic region may not be sufficient in identifying such trans-acting endoribonuclease involved in the CRISPR-Cas machinery. Thus, a more thorough bioinformatic analysis of the entire genome may be helpful.

Secondly, it was demonstrated that the six core effector genes are functionally required (Chapter 2), but the exact composition of the effector protein complex has yet to be unequivocally determined. A protein pull-down assay was conducted in *E. coli* and revealed two subunits, SeryCsm1 and SeryCas10. One way to better resolve the composition is to carry out the

pull-down assay in the native host. One of the subunits must be provided as a His-tagged fusion protein, either directly via the genome or by deletion of the genomic copy and complementation with a tagged version from a plasmid. The ideal position to insert the His-tag without interrupting the transcription or translation machinery is at the N-terminus of the operon's first gene, the SeryCsm1.

The protein complex could be purified through nickel affinity column chromatography and subsequent analysis by SDS-PAGE and native PAGE. To identify each subunit and study the stoichiometry of the subunits, high-resolution protein mass spectrometry (Protein MS) is required. The in-gel digestion and protein MS of the *E. coli* expressed SeryCRISPR effector complex was carried out and signature peptide signals were detected from SeryCsm1, SeryHelicase, and SeryCas6, but the overall coverage wasn't enough to quantitatively elucidate stoichiometry. Thus, this approach requires further optimization.

A third aspect is the evolutionary significance of SeryCRISPR. The type III CRISPR-Cas systems are believed to be distantly related to type I Cascade systems.[170,171] Although SeryHelicase is not similar to any reported Cas3 proteins in terms of amino acid sequences or protein homology, its functional domain architecture comprising an HNH nuclease fused to a helicase is similar to canonical Cas3 proteins.[172] The large subunit Cas8 of type I systems are structurally similar to Cas10, with inactivated RNA recognition motifs.[173] Based on structural similarities of the effector protein complexes, it was proposed that type I CRISPR systems evolved from type III systems.[174,175] Briefly, it was hypothesized that the abortive immunize (Abi) system could give rise to the type III-like CRISPR-Cas effector module. The ancestral class 1 CRISPR-Cas system is inferred to have emerged from the casposon derived CRISPR array and adaptation module and the type III-like effector module derived from the Abi

system.[174] The subsequent acquisition of the HD nuclease domain by the effector module provided RNA-guided DNA cleavage. It was proposed that inactivation of the oligoA polymerase domain in the effector complex and acquisition of the Cas3 helicase led to the emergence of type I systems, which lack the cyclic oligoA-dependent signaling pathway and exclusively cleave double-stranded DNA (Fig. 4.1).[174] If SeryHelicase is indeed involved in the interference, and no ancillary trans-acting ribonuclease can be found, the SeryCRISPR could fill in the evolutionary gap between type III and type I systems.

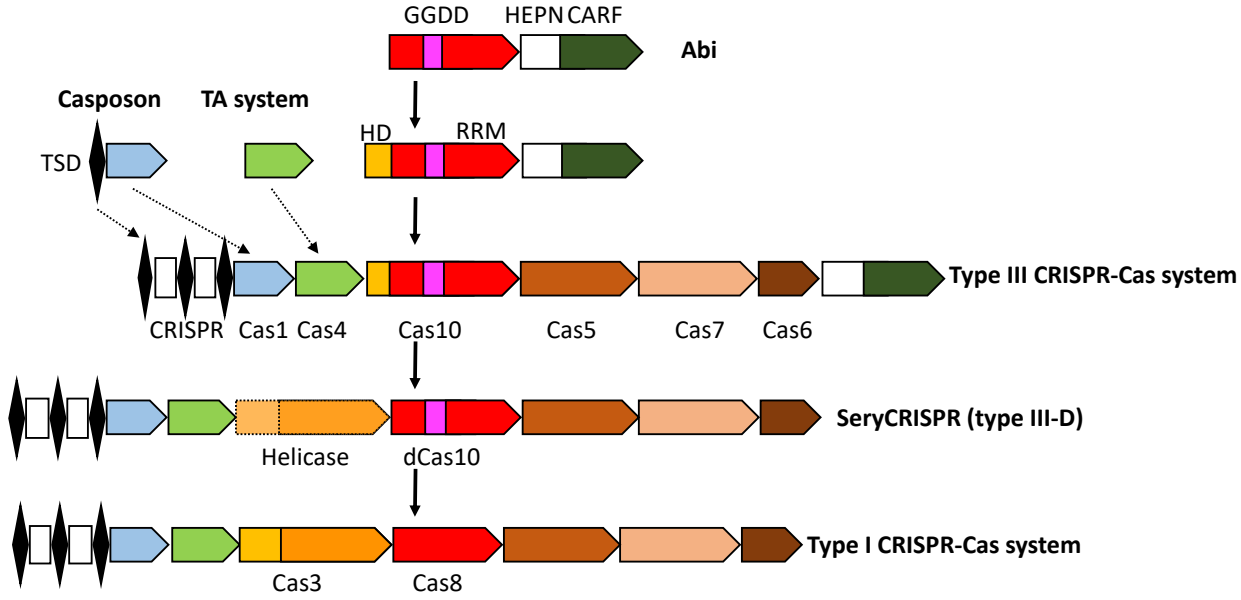


Figure 4.1 SeryCRISPR provides a hypothetical evolution scenario from type III CRISPR-Cas to type I systems. The figure depicts a hypothetical scenario of the evolution of Class I type III CRISPR-Cas systems to obtain the nuclease domain-containing helicase gene (SeryHelicase as an example) while inactivating the HD-nuclease domain of Cas10 (SeryCas10 as an example). In addition, SeryCRISPR may have lost the ancillary trans-acting RNase module. Further inactivation of the polymerase domain of dCas10 would result in Cas8-like proteins and evolve to a type I-like system. The typical CRISPR-Cas operon organization is shown for each CRISPR-Cas type. Homologous genes are color-coded and identified by a systematic name. GGDD, a key catalytic motif of the cyclase or polymerase domain of Cas10 that is involved in the synthesis of cyclic oligoA signaling molecules; HEPN, higher eukaryotes, and prokaryotes nucleotide-binding (HEPN) domain that is an RNase; CARF, CRISPR-associated Rossmann fold domain; HD, HD nuclease domain; TSD, target site duplication, the potential origin of ancestral CRISPR repeats.

Finally, the mystery of genome targeting spacers remains. CRISPR-Cas is the adaptive immune system in prokaryotes, and the spacers are known to be derived from bacteriophages or conjugative plasmids. However, ~20% of the spacers (7/34) from SeryCRISPR arrays are homologous to chromosomal sequences in the NCBI database, which constitutes all the identifiable spacers in *S. erythraea* (Chapter 2, Table 2.1). Notably, spacer 5 of array 2 shares high sequence similarity to the host genome targeting a hypothetical gene. Self-targeting spacers and spacers matching chromosomal sequences of other prokaryotes are widespread in CRISPR-Cas systems.[17,176] The role of self-targeting spacers maybe that they participate in gene regulation and bacterial genome evolution.[110,111] However, the phenomenon of targeting non-MGE derived chromosomal sequences of other species is yet to be explored. To investigate this, bioinformatics tools could be utilized to identify whether those targeted chromosomal genes were of MGE origins. Future studies could also experimentally verify the regulatory functions of self-targeting spacers by quantitative reverse transcription PCR.

4.2 Tangible erythromycin PKS engineering approaches in *E. coli* and *Saccharomyces cerevisiae*

4.2.1 *In vitro* Cas9 digestion coupled with Gibson assembly in editing modular polyketide synthases

Among natural products, erythromycin biosynthesis is studied as a model for modular polyketide synthases (PKSs). Yet, the genetic modification of the erythromycin PKS gene cluster is challenging due to the high sequence homology among modules and large size. Traditional genetic strategies to modify modular PKSs and produce new analogs rely on restriction enzyme digestion or recombination-based cloning.[177] However, the availability of suitable restriction

sites is limiting. To circumvent this bottleneck, *in vitro* Cas9 digestion coupled with Gibson assembly was utilized, as demonstrated in rapamycin PKS engineering by Kudo and colleagues.[60] Briefly, BAC vectors encoding the biosynthetic gene clusters are precisely digested by Cas9 nuclease guided by synthesized sgRNA at desired sites *in vitro*, followed by Gibson assembly to repair the gap through sequence homology in the flanking sequence of the inserts (Fig. 3.2A). Conveniently, Fang and colleagues have successfully constructed BAC vectors encoding the entire biosynthetic pathway achieving erythromycin production in *E. coli*.[178] It can be expected that the BAC vector encoding DEBS1, DEBS2, and DEBS3 polyketide synthases can be cleaved with high precision and efficiency by *in vitro* Cas9 digestion. By supplementing insert fragments as PCR products, desired construct or libraries of polyketide synthase pathway can be made in *E. coli*.

Our preliminary studies used rapid subculturing to cure one of the two BACs carried by the erythromycin-producing LF01 strain and isolated the DEBS encoding BAC DNA (49.6 kb).[178] Previous studies have identified that changing the tyrosine to arginine of the YASH motif within the acyltransferase domain of Module 6 shifts the substrate specificity and enable incorporation of non-natural extender units to potentially diversify the macrolactone ring of erythromycin.[179] As proof of concept, the YASH motif was selected to test the in vitro editing approach. First, a synthetic dsDNA fragment was used as a template to produce sgRNA *in vitro*, with the 20 nt targeting the YASH motif. Next, the purified BAC DNA was precisely and efficiently cleaved by the sgRNA-guided Cas9 protein. Then a pair of annealed DNA oligos were used as a repair template to 'stitch' the cleaved site while introducing the desired editing by Gibson assembly, changing the amino acid sequence from YASH to RASH (Fig. 4.2A. Finally, the edited BAC was transformed into *E.coli,* and sequencing the BAC DNA from single colonies

confirmed the *in vitro* Cas9 approach could achieve the single-amino-acid resolution editing at 25% efficiency (1/4) (Fig. 4.2B).

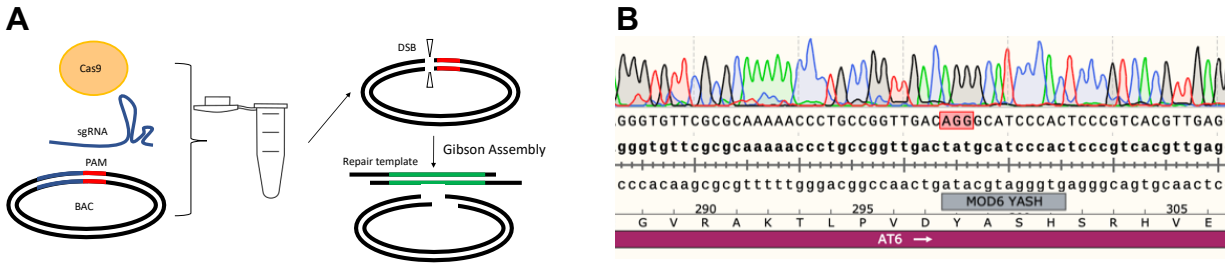


Figure 4.2 *in vitro* Cas9 digestion coupled with Gibson assembly in editing modular polyketide synthase, tyrosine to arginine as proof of concept. (A) Schematic workflow of *in vitro* Cas9 digestion of BAC followed by Gibson assembly. sgRNA can be produced by *in vitro* transcription from synthesized dsDNA. Purified Cas9 protein, sgRNA, and the DNA substrate are mixed and incubated. The cleaved DNA substrate and repair template are processed by 5' exonuclease, polymerase, and ligase to make the desired edits. (B) Tyrosine (ATA) from pDEBS was precisely cleaved, and Y to R (AGG) edit was made.

4.2.2 Genome integration of DEBS1-TE polyketide synthase in *Saccharomyces cerevisiae*

The chemical synthesis of structurally complex polyketides is challenging. Thus, large-scale production of pharmaceutically valuable polyketides relies on fermentation. Most natural polyketide-producing organisms are not optimal hosts for high-throughput PKS engineering or to produce drug analogs, as they tend to grow slowly and are often not amenable to genetic manipulation. On the other hand, a model organism such as *E. coli* can hardly produce the necessary precursor or extender units or host post-translational modification, nor can they tolerate the potential toxicity of the products. Additionally, the transformation efficiency of large BAC vectors encoding biosynthetic gene clusters is low in *E. coli*, impeding the engineering throughput. One tangible solution to PKS engineering and polyketide overproduction would be to develop a generic host in which PKS genes and other essential genes could be easily

transferred and functionally expressed. *Saccharomyces cerevisiae* is an ideal candidate due to its relatively rapid growth, highly developed genetic tools, and unmatched recombination capability to facilitate genetic manipulation. It has been demonstrated that precursors required for erythromycin biosynthesis, propionyl-CoA and methylmalonyl-CoA can be made in *S. cerevisiae* with exogenous pathways.[180] It has also been demonstrated that Sfp, a phosphopantetheinyl transferase from *Bacillus subtilis* can perform the required acyl carrier protein post-translational modifications.[181] Moreover, a precise and efficient Cas9-based genome editing platform was designed for *S. cerevisiae.*[182] In our preliminary studies, the entire recombinant DEBS1-TE polyketide synthase gene has been successfully integrated into the genome of *S. cerevisiae* BY4742 (Fig. 3.3A).[183] Briefly, Cas9 was expressed from a plasmid in *S. cerevisiae* before the sgRNA plasmid along with the purified PCR products as editing templates were co-transformed. The genetic region downstream of GRX5 was selected as an integration site. The GAL1 promoter and CYC1 terminator were chosen to control the gene expression. Since the entire integration cassette is more than 12 kb in size, the inserts were amplified as five fragments (GRX5_arm_up and GAL1 promoter as one fragment, first half of the DEBS1 as one fragment, the second half of DEBS1 as one fragment, TE from DEBS3 as one fragment, CYC1 terminator and GRX5_arm_down as one fragment) and utilized 200 bp of homology arms to facilitate the homology-directed repair. Two colonies were sequenced at the edited region and the observed genome editing efficiency was 100%. It can be predicted that with additional genome integration of the phosphopantetheinyl transferase and precursor pathways, a triketide lactone can be produced as proof-of-principle, followed by further genetic modification to afford drug-like macrolactones and macrolides (Fig. 3.3B).[180]

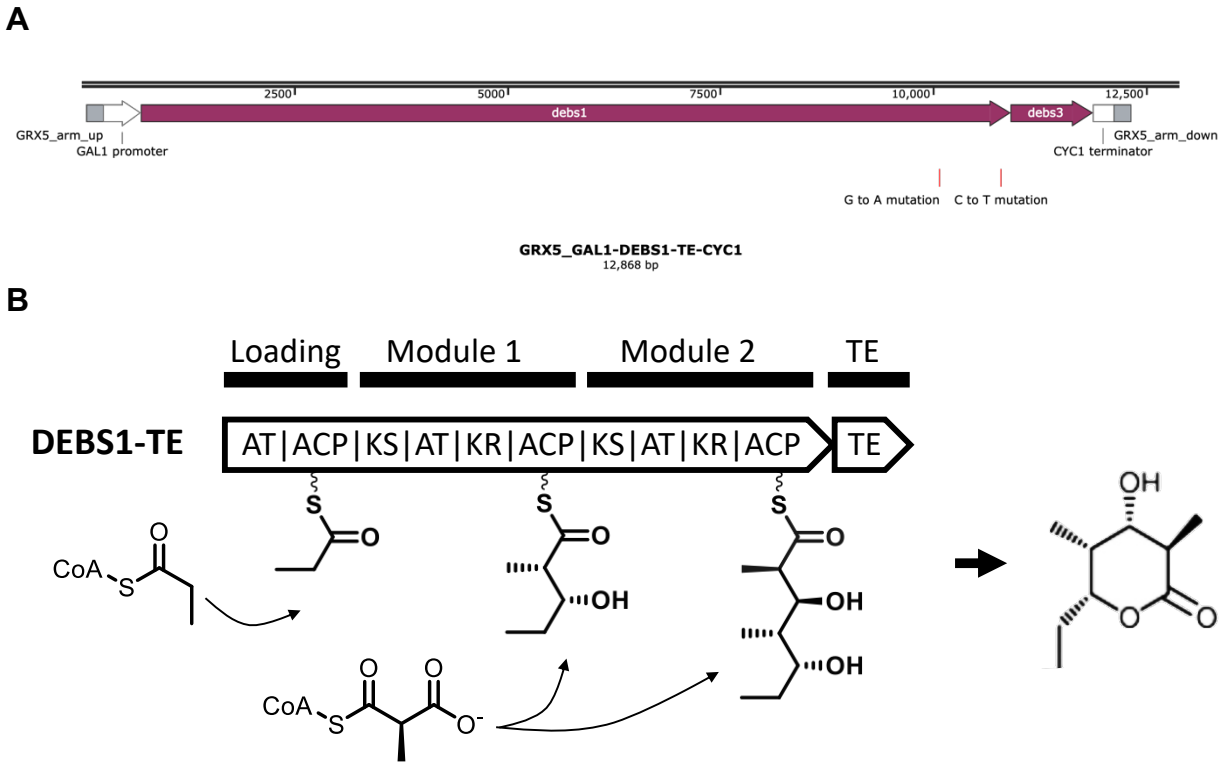


Figure 4.3 Constructing the biosynthetic pathway of a triketide lactone by deoxyerythronolide B synthase (DEBS1-TE) in *Saccharomyces cerevisiae* using CRISPR-Cas9. (A) Schematics of the genomic editing. GRX5 locus was selected as the integration site, and sgRNA was designed. A 200 bp homology arm was installed upstream and downstream of the inserted cassette to facilitate homology-directed repair. The entire insert cassette consisted of linear dsDNA fragments and was co-transformed with the sgRNA plasmid. (B) DEBS1-TE can make a triketide lactone product. The DEBS1 consists of a loading module (AT, acyl-transferase domain; ACP, acyl carrier protein) and two extender modules (KS, ketosynthase domains; AT; ACP; and KR, ketoreductase domain). The AT domain of the loading module recognizes and loads propionate from propionyl-CoA to the thiol of the loading domain ACP. Similarly, ATs of extender modules load methylmalonate from methylmalonyl-CoA to the corresponding ACPs . DEBS1-TE is a minimal, modular PKS comprising DEBS1 fused to the thioesterase (TE) domain of DEBS3 to form a triketide lactone product.

4.3 Materials and Methods

4.3.1 Microorganisms and Culture Medium

The bacterial strains used in this study are listed in Appendix C, Table C1. *Escherichia coli* DH5α was used for the construction of recombinant plasmids. LF01Cells were cultivated on Luria–Bertani medium (tryptone 10 g/L, yeast extract 5 g/L, and NaCl 10 g/L) at 30 or 37 °C,

flasks with shaking at 220 rpm, supplemented with kanamycin (25 μg/mL) during plasmid curing. The *Saccharomyces cerevisiae* was grown on YPD agar plates (1% yeast extract, 2% peptone, 2% glucose, w/v).

4.3.2 Transformation of *Saccharomyces cerevisiae*

Yeast transformation was carried out by a standard lithium acetate transformation method.[184] Briefly, 5mL of cells were harvested per transformation and washed with 0.2M LiAc. Next, the cells were resuspended in 100μL 0.2M LiAc supplemented with 50% PEG-4000 (v/v). Plasmid and DNA inserts were added and incubated at room temperature for 30 min. The mixture was then heat-shocked at 42°C for 30 min followed by cool down on ice for 5 min. The cells were then spun down to remove PEG-4000 and recovered in YPD for 4 h before plating on agar plates supplemented with appropriate antibiotic resistance markers.[182]

4.3.3 *In vitro* Cas9 digestion

In vitro digestion of DNA with Cas9 nuclease was performed following the manufacturer's protocol for *S. pyogenes* Cas9 Nuclease (NEB #M0386). *In vitro* RNA synthesis of sgRNA was performed following the manufacturer's protocol for HiScribe™ T7 High Yield RNA Synthesis Kit. A gBlock encoding T7 promoter and sgRNA was used as dsDNA template (Appendix C, Table. C2).

4.3.4 DNA Manipulations

Plasmid DNA from *E. coli* was obtained using a Monarch miniprep kit (NEB) following the manufacturer's instructions. PCR primers, double-stranded synthetic DNA for plasmid

interference assays, and single-stranded DNA for annealing oligonucleotides were synthesized by Integrated DNA Technologies (IDT; Morrisville, NC). PCR amplicons for colony screening were generated using standard PCR protocols and Phusion Hot Start High-Fidelity Polymerase (Invitrogen). DNA sequencing was performed by Genewiz to confirm sequence content. Restriction digestions were performed using high-fidelity restriction enzymes (NEB). Purification of digested products for ligation was performed using a Monarch PCR & DNA Cleanup Kit (NEB) or Monarch DNA Gel Extraction Kit (NEB). Ligation reactions were performed at a 3:1 insert/vector ratio using 50 ng of vector in a final volume of 10 μL, using T4 ligase and PNK (NEB) based on the manufacturer's instructions. To anneal two single-stranded DNA oligonucleotides, each oligonucleotide was resuspended in water to a final concentration of 100 μM, each strand (A + B) was mixed, and the final volume was adjusted to 50 μL with IDT Duplex Buffer. Both strands were annealed by incubating at 95 °C for 2 min, followed by a gradient decrease of temperature at 0.1 °C/s to room temperature. Gibson Assembly was performed using NEBuider HiFi DNA Assembly kit (NEB) following the manufacturer's protocol.

4.3.5 Construction of the *Saccharomyces cerevisiae* Genome editing Plasmid

sgRNA plasmid was obtained from Zhang et al.[182] A pair of primers were used to amplify the sgRNA looping out the existing 20 nt target sequence while introducing new GRX5 targeting 20 nt, followed by DpnI digestion and transformation to *E. coli* DH5α. pDEBS was used as a template to PCR amplify DEBS1 and TE. *S. cerevisiae* genomic DNA was used to amplify homology arms and GAL1 promoter and CYC1 terminator. The primers used are listed in Appendix C, Table C2.

REFERENCES

1. Barka EA, Vatsa P, Sanchez L, et al. Taxonomy, Physiology, and Natural Products of Actinobacteria. *Microbiol Mol Biol Rev*. 2016;80(1):1-43. doi:10.1128/MMBR.00019-15
2. Actinomycetes: Still a source of novel antibiotics. *Nat Prod Rep*. 2017;34(10):1203-1232. doi:10.1039/C7NP00026J
3. Challis GL. Genome mining for novel natural product discovery. *J Med Chem*. 2008;51(9):2618-2628. doi:10.1021/JM700948Z
4. Otten SL, Stutzman-Engwall KJ, Hutchinson CR. Cloning and expression of daunorubicin biosynthesis genes from Streptomyces peucetius and S. peucetius subsp. caesius. *J Bacteriol*. 1990;172(6):3427-3434. doi:10.1128/jb.172.6.3427-3434.1990
5. Baltz RH, Brian P, Miao V, Wrigley SK. Combinatorial biosynthesis of lipopeptide antibiotics in Streptomyces roseosporus. *J Ind Microbiol Biotechnol*. 2006;33(2):66-74. doi:10.1007/S10295-005-0030-Y
6. T. Kieser, M.J. Bibb, K.F. Chater, M.J. Butter, D.A. Hopwood, K.F. Chater, MJ Bittner, MJ Bibb, MJ Buttner, KF Chater DH. *Practical Streptomyces Genetics: A Laboratory Manual*. Vol 1.; 2000.
7. Chibani-Chennoufi S, Bruttin A, Dillmann ML, Brüssow H. Phage-host interaction: An ecological perspective. *J Bacteriol*. 2004;186(12):3677-3686. doi:10.1128/JB.186.12.3677-3686.2004
8. Mruk I, Kobayashi I. To be or not to be: Regulation of restriction-modification systems and other toxin-antitoxin systems. *Nucleic Acids Res*. 2014;42(1):70-86. doi:10.1093/NAR/GKT711
9. Gerdes K, Christensen SK, Løbner-Olesen A. Prokaryotic toxin-antitoxin stress response

loci. *Nat Rev Microbiol*. 2005;3(5):371-382. doi:10.1038/NRMICRO1147

10. Makarova KS, Grishin N V., Shabalina SA, Wolf YI, Koonin E V. A putative RNA-interference-based immune system in prokaryotes: Computational analysis of the predicted enzymatic machinery, functional analogies with eukaryotic RNAi, and hypothetical mechanisms of action. *Biol Direct*. 2006;1. doi:10.1186/1745-6150-1-7
11. Ishino Y, Shinagawa H, Makino K, Amemura M, Nakatura A. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isoenzyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol*. 1987;169(12):5429-5433. doi:10.1128/JB.169.12.5429-5433.1987
12. Mojica FJM, Juez G, Rodríguez-Valera F. Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites. *Mol Microbiol*. 1993;9:613-621.
13. Mojica FJM, Ferrer C, Juez G, Rodríguez-Valera F. Long stretches of short tandem repeats are present in the largest replicons of the Archaea Haloferax mediterranei and Haloferax volcanii and could be involved in replicon partitioning. *Mol Microbiol*. 1995;17:85-93.
14. Jansen R, Embden JDA van, Gaastra W, Schouls LM. Identification of genes that are associated with DNA repeats in prokaryotes. *Mol Microbiol*. 2002;43(6):1565-1575. doi:10.1046/J.1365-2958.2002.02839.X
15. Bolotin A, Quinquis B, Sorokin A, Dusko Ehrlich S. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology*. 2005;151(8):2551-2561. doi:10.1099/MIC.0.28048-0
16. Pourcel C, Salvignol G, Vergnaud G. CRISPR elements in Yersinia pestis acquire new

repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology*. 2005;151(3):653-663. doi:10.1099/MIC.0.27437-0

17. Mojica FJM, Díez-Villaseñor C, García-Martínez J, Soria E. Intervening Sequences of Regularly Spaced Prokaryotic Repeats Derive from Foreign Genetic Elements. *J Mol Evol 2005 602*. 2005;60(2):174-182. doi:10.1007/S00239-004-0046-3
18. Barrangou R, Fremaux C, Deveau H, et al. CRISPR provides acquired resistance against viruses in prokaryotes. *Science (80-)*. 2007;315(5819):1709-1712. doi:10.1126/SCIENCE.1138140
19. Deveau H, Garneau JE, Moineau S. CRISPR/Cas system and its role in phage-bacteria interactions. *Annu Rev Microbiol*. 2010;64:475-493. doi:10.1146/ANNUREV.MICRO.112408.134123
20. Van Der Oost J, Westra ER, Jackson RN, Wiedenheft B. Unravelling the structural and mechanistic basis of CRISPR-Cas systems. *Nat Rev Microbiol*. 2014;12(7):479-492. doi:10.1038/NRMICRO3279
21. Horvath P, Barrangou R. CRISPR/Cas, the immune system of Bacteria and Archaea. *Science (80-)*. 2010;327(5962):167-170. doi:10.1126/SCIENCE.1179555
22. Makarova KS, Wolf YI, Alkhnbashi OS, et al. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*. 2015;13(11):722-736. doi:10.1038/NRMICRO3569
23. Crawley AB, Henriksen JR, Barrangou R. CRISPRdisco: An Automated Pipeline for the Discovery and Analysis of CRISPR-Cas Systems. *Cris J*. 2018;1(2):171-181. doi:10.1089/CRISPR.2017.0022/ASSET/IMAGES/MEDIUM/FIGURE3.GIF
24. Makarova KS, Haft DH, Barrangou R, et al. Evolution and classification of the CRISPR-

Cas systems. *Nat Rev Microbiol*. 2011;9(6):467-477. doi:10.1038/NRMICRO2577

25. Koonin E V., Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol*. 2017;37:67-78. doi:10.1016/J.MIB.2017.05.008
26. Doudna JA, Charpentier E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-)*. 2014;346(6213). doi:10.1126/SCIENCE.1258096
27. Knott GJ, Doudna JA. CRISPR-Cas guides the future of genetic engineering. *Science (80-)*. 2018;361(6405):866-869. doi:10.1126/SCIENCE.AAT5011
28. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science (80-)*. 2012;337(6096):816-821. doi:10.1126/SCIENCE.1225829
29. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31(3):233-239. doi:10.1038/NBT.2508
30. Cong L, Ran FA, Cox D, et al. Multiplex genome engineering using CRISPR/Cas systems. *Science (80-)*. 2013;339(6121):819-823. doi:10.1126/SCIENCE.1231143
31. Nishimasu H, Ran FA, Hsu PD, et al. Crystal structure of Cas9 in complex with guide RNA and target DNA. *Cell*. 2014;156(5):935-949. doi:10.1016/J.CELL.2014.02.001
32. Ran FA, Hsu PD, Lin CY, et al. XDouble nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*. 2013;154(6). doi:10.1016/J.CELL.2013.08.021
33. Qi LS, Larson MH, Gilbert LA, et al. Repurposing CRISPR as an RNA-γuided platform for sequence-specific control of gene expression. *Cell*. 2013;152(5):1173-1183.
34. Bikard D, Jiang W, Samai P, Hochschild A, Zhang F, Marraffini LA. Programmable repression and activation of bacterial gene expression using an engineered CRISPR-Cas

system. *Nucleic Acids Res*. 2013;41(15):7429-7437. doi:10.1093/NAR/GKT520

35. Gilbert LA, Larson MH, Morsut L, et al. XCRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell*. 2013;154(2):442.

36. Gaudelli NM, Komor AC, Rees HA, et al. Programmable base editing of T to G C in genomic DNA without DNA cleavage. *Nature*. 2017;551(7681):464-471. doi:10.1038/NATURE24644

37. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. *Nature*. 2016;533:420-424. doi:10.1038/NATURE17946

38. Cobb RE, Wang Y, Zhao H. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth Biol*. 2015;4(6):723. doi:10.1021/SB500351F

39. Huang H, Zheng G, Jiang W, Hu H, Lu Y. One-step high-efficiency CRISPR/Cas9-mediated genome editing in Streptomyces. *Acta Biochim Biophys Sin (Shanghai)*. 2015;47(4):231-243. doi:10.1093/ABBS/GMV007

40. Tong Y, Charusanti P, Zhang L, Weber T, Lee SY. CRISPR-Cas9 Based Engineering of Actinomycetal Genomes. *ACS Synth Biol*. 2015;4(9):1020-1029. doi:10.1021/ACSSYNBIO.5B00038

41. Zeng H, Wen S, Xu W, et al. Highly efficient editing of the actinorhodin polyketide chain length factor gene in Streptomyces coelicolor M145 using CRISPR/Cas9-CodA(sm) combined system. *Appl Microbiol Biotechnol*. 2015;99(24):10575-10585. doi:10.1007/S00253-015-6931-4

42. Zhang MM, Wong FT, Wang Y, et al. CRISPR-Cas9 strategy for activation of silent

Streptomyces biosynthetic gene clusters. *Nat Chem Biol*. 2017;13(6):607-609. doi:10.1038/NCHEMBIO.2341

43. Liu Y, Wei W-P, Ye B-C. High GC Content Cas9-Mediated Genome-Editing and Biosynthetic Gene Cluster Activation in *Saccharopolyspora erythraea*. *ACS Synth Biol*. 2018;7(5):1338-1348. doi:10.1021/acssynbio.7b00448

44. Liu Y, Ren CY, Wei WP, You D, Yin BC, Ye BC. A CRISPR-Cas9 Strategy for Activating the Saccharopolyspora erythraea Erythromycin Biosynthetic Gene Cluster with Knock-in Bidirectional Promoters. *ACS Synth Biol*. 2019;8(5):1134-1143. doi:10.1021/ACSSYNBIO.9B00024

45. Mo J, Wang S, Zhang W, et al. Efficient editing DNA regions with high sequence identity in actinomycetal genomes by a CRISPR-Cas9 system. *Synth Syst Biotechnol*. 2019;4(2):86-91. doi:10.1016/J.SYNBIO.2019.02.004

46. Yeo WL, Heng E, Tan LL, et al. Characterization of Cas proteins for CRISPR-Cas editing in streptomycetes. *Biotechnol Bioeng*. 2019;116(9):2330-2338. doi:10.1002/BIT.27021

47. Li L, Wei K, Zheng G, et al. CRISPR-Cpf1-assisted multiplex genome editing and transcriptional repression in Streptomyces. *Appl Environ Microbiol*. 2018;84(18):827-845. doi:10.1128/AEM.00827-18

48. Zhang J, Zhang D, Zhu J, Liu H, Liang S, Luo Y. Efficient Multiplex Genome Editing in Streptomyces via Engineered CRISPR-Cas12a Systems. *Front Bioeng Biotechnol*. 2020;0:726. doi:10.3389/FBIOE.2020.00726

49. Zhao Y, Li L, Zheng G, et al. CRISPR/dCas9-Mediated Multiplex Gene Repression in Streptomyces. *Biotechnol J*. 2018;13(9). doi:10.1002/BIOT.201800121

50. Rock JM, Hopkins FF, Chavez A, et al. Programmable transcriptional repression in

mycobacteria using an orthogonal CRISPR interference platform. *Nat Microbiol*. 2017;2. doi:10.1038/NMICROBIOL.2016.274

51. DeLorenzo DM, Rottinghaus AG, Henson WR, Moon TS. Molecular Toolkit for Gene Expression Control and Genome Modification in Rhodococcus opacus PD630. *ACS Synth Biol*. 2018;7(2):727-738. doi:10.1021/ACSSYNBIO.7B00416

52. Tong Y, Whitford CM, Robertsen HL, et al. Highly efficient DSB-free base editing for streptomycetes with CRISPR-BEST. *Proc Natl Acad Sci U S A*. 2019;116(41):20366-20375. doi:10.1073/PNAS.1913493116

53. Zhao Y, Tian J, Zheng G, et al. Multiplex genome editing using a dCas9-cytidine deaminase fusion in Streptomyces. *Sci China Life Sci 2019 637*. 2019;63(7):1053-1062. doi:10.1007/S11427-019-1559-Y

54. Wang Y, Liu Y, Liu J, et al. MACBETH: Multiplex automated Corynebacterium glutamicum base editing method. *Metab Eng*. 2018;47:200-210.

55. Wang Y, Liu Y, Li J, et al. Expanding targeting scope, editing window, and base transition capability of base editing in Corynebacterium glutamicum. *Biotechnol Bioeng*. 2019;116(11):3016-3029. doi:10.1002/BIT.27121

56. Nah H-J, Pyeon H-R, Kang S-H, Choi S-S, Kim E-S. Cloning and Heterologous Expression of a Large-sized Natural Product Biosynthetic Gene Cluster in Streptomyces Species. *Front Microbiol*. 2017;0(MAR):394. doi:10.3389/FMICB.2017.00394

57. Jiang W, Zhu TF. Targeted isolation and cloning of 100-kb microbial genomic sequences by Cas9-assisted targeting of chromosome segments. *Nat Protoc*. 2016;11(5):960-975. doi:10.1038/NPROT.2016.055

58. Kang HS, Charlop-Powers Z, Brady SF. Multiplexed CRISPR/Cas9- and TAR-Mediated

Promoter Engineering of Natural Product Biosynthetic Gene Clusters in Yeast. *ACS Synth Biol*. 2016;5(9):1002-1010. doi:10.1021/ACSSYNBIO.6B00080

59. Kim H, Ji CH, Je HW, Kim JP, Kang HS. MpCRISTAR: Multiple Plasmid Approach for CRISPR/Cas9 and TAR-Mediated Multiplexed Refactoring of Natural Product Biosynthetic Gene Clusters. *ACS Synth Biol*. 2020;9(1):175-180. doi:10.1021/ACSSYNBIO.9B00382

60. Kudo K, Hashimoto T, Hashimoto J, et al. In vitro Cas9-assisted editing of modular polyketide synthase genes to produce desired natural product derivatives. *Nat Commun 2020 111*. 2020;11(1):1-12. doi:10.1038/s41467-020-17769-2

61. Li Q, Chen J, Minton NP, et al. CRISPR-based genome editing and expression control systems in Clostridium acetobutylicum and Clostridium beijerinckii. *Biotechnol J*. 2016;11(7):961-972. doi:10.1002/BIOT.201600053

62. X Song HHZXLASY. CRISPR-Cas 9(D10A) nickase-assisted genome editing in Lactobacillus casei. *Appl Env Microbiol*. 2017;83(22). doi:10.1128/aem.01259-17

63. T Xu YLZSCHYLYZJZ. Efficient genome editing in Clostridium cellulolyticum via CRISPR-Cas9 nickase. *Appl Env Microbiol*. 2015;81(13):4423-4431. doi:10.1128/aem.00873-15

64. Pyne ME, Bruder MR, Moo-Young M, Chung DA, Chou CP. Harnessing heterologous and endogenous CRISPR-Cas machineries for efficient markerless genome editing in Clostridium. *Sci Rep*. 2016;6. doi:10.1038/SREP25666

65. Chang Y, Su T, Qi Q, Liang Q. Easy regulation of metabolic flux in Escherichia coli using an endogenous type I-E CRISPR-Cas system. *Microb Cell Fact*. 15:195.

66. Zhang J, Zong W, Hong W, Zhang ZT, Wang Y. Exploiting endogenous CRISPR-Cas

system for multiplex genome editing in Clostridium tyrobutyricum and engineer the strain for high-level butanol production. *Metab Eng*. 2018;47:49-59. doi:10.1016/J.YMBEN.2018.03.007

67. Gong T, Tang B, Zhou X, et al. Genome editing in Streptococcus mutans through self-targeting CRISPR arrays. *Mol Oral Microbiol*. 2018;33(6):440-449. doi:10.1111/OMI.12247

68. Hidalgo-Cantabrana C, Goh YJ, Pan M, Sanozky-Dawes R, Barrangou R. Genome editing using the endogenous type I CRISPR-Cas system in Lactobacillus crispatus. *Proc Natl Acad Sci*. 2019;116(32):15774-15783. doi:10.1073/PNAS.1905421116

69. Stachler AE, Marchfelder A. Gene repression in haloarchaea using the CRISPR (clustered regularly interspaced short palindromic repeats)-cas I-B system. *J Biol Chem*. 2016;291(29):15226-15242. doi:10.1074/jbc.m116.724062

70. Zhang J, Li X, Deng Z, Ou H-Y. Comparative Analysis of CRISPR Loci Found in Streptomyces Genome Sequences. *Interdiscip Sci Comput Life Sci 2018 104*. 2018;10(4):848-853. doi:10.1007/S12539-018-0301-8

71. Qiu Y, Wang S, Chen Z, Guo Y, Song Y. An Active Type I-E CRISPR-Cas System Identified in Streptomyces avermitilis. *PLoS One*. 2016;11(2):e0149533. doi:10.1371/JOURNAL.PONE.0149533

72. Barrangou R, Fremaux C, Deveau H, et al. CRISPR Provides Acquired Resistance Against Viruses in Prokaryotes. *Science (80-)*. 2007;315(5819):1709-1712. doi:10.1126/SCIENCE.1138140

73. Karginov F V., Hannon GJ. The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea. *Mol Cell*. 2010;37(1):7-19. doi:10.1016/J.MOLCEL.2009.12.033

74. Marraffini LA, Sontheimer EJ. CRISPR interference: RNA-directed adaptive immunity in bacteria and archaea. *Nat Rev Genet*. 2010;11(3):181-190. doi:10.1038/nrg2749

75. Horvath P. Diversity, activity, and evolution of CRISPR loci in Streptococcus thermophilus. *J Bacteriol*. 2008;190(4):1401-1412. doi:10.1128/jb.01415-07

76. Brouns SJ. Small CRISPR RNAs Guide Antiviral Defense in Prokaryotes. *Science (80-)*. 2008;321(5891):960-964. doi:10.1126/science.1159689

77. Carte J, Wang R, Li H, Terns RM, Terns MP. Cas6 is an endoribonuclease that generates guide RNAs for invader defense in prokaryotes. *Genes Dev*. 2008;22(24):3489-3496. doi:10.1101/GAD.1742908

78. Deltcheva E, Chylinski K, Sharma CM, et al. CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nat 2011 4717340*. 2011;471(7340):602-607. doi:10.1038/nature09886

79. Crawley AB, Henriksen JR, Barrangou R. CRISPRdisco: An automated pipeline for the discovery and analysis of CRISPR-cas systems. *Cris J*. 2018;1(2):171-181. doi:10.1089/crispr.2017.0022

80. K.S. M, Y.I. W, J. I, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 18:67-83.

81. Makarova KS. An updated evolutionary classification of CRISPR-Cas systems. *Nat Rev Microbiol*. 2015;13(11):722-736. doi:10.1038/nrmicro3569

82. J.E. G, M.E. D, M. V, et al. The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA. *Nature*. 468:67-71.

83. Zetsche B. Cpf1 is a single RNA-guided endonuclease of a class 2 CRISPR-Cas system. *Cell*. 2015;163(3):759-771. doi:10.1016/j.cell.2015.09.038

84. Hochstrasser ML, Taylor DW, Bhat P, et al. CasA mediates Cas3-catalyzed target degradation during CRISPR RNA-guided interference. *Proc Natl Acad Sci U S A*. 2014;111(18):6618-6623. doi:10.1073/PNAS.1405079111

85. Tamulaitis G, Venclovas Č, Siksnys V. Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends Microbiol*. 2017;25(1):49-61. doi:10.1016/j.tim.2016.09.012

86. Staals RHJ, Zhu Y, Taylor DW, et al. RNA Targeting by the Type III-A CRISPR-Cas Csm Complex of Thermus thermophilus. *Mol Cell*. 2014;56(4):518-530. doi:10.1016/j.molcel.2014.10.005

87. Tamulaitis G, Kazlauskiene M, Manakova E, et al. Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Mol Cell*. 2014;56(4):506-517. doi:10.1016/j.molcel.2014.09.027

88. C.R. H, A. C, H. L, R.M. T, M.P. T. Target RNA capture and cleavage by the Cmr type III-B CRISPR–Cas effector complex. *Genes Dev*. 28:2432-2443.

89. J.R. E, N.F. S, N. R, et al. Bipartite recognition of target RNAs activates DNA cleavage by the Type III-B CRISPR–Cas system. *Genes Dev*. 30:447-459.

90. M. K, G. T, G. K, C. V, V. S. Spatiotemporal control of Type III-A CRISPR–Cas immunity: coupling DNA degradation with the target RNA recognition. *Mol Cell*. 62:295-306.

91. Jung TY, An Y, Park KH, Lee MH, Oh BH, Woo E. Crystal structure of the Csm1 subunit of the Csm complex and its single-stranded DNA-specific nuclease activity. *Structure*. 2015;23(4):782-790. doi:10.1016/J.STR.2015.01.021

92. W. H, Y. L, L. D, et al. A type III-B CRISPR–Cas effector complex mediating massive

target DNA destruction. *Nucleic Acids Res*. 45:1983-1993.

93. Goldberg GW, Jiang W, Bikard D, Marraffini LA. Conditional tolerance of temperate phages via transcription-dependent CRISPR-Cas targeting. *Nature*. 2014;514(7524):633-637. doi:10.1038/nature13637

94. Kazlauskiene M, Kostiuk G, Venclovas Č, Tamulaitis G, Siksnys V. A cyclic oligonucleotide signaling pathway in type III CRISPR-Cas systems. *Science (80-)*. 2017;357(6351):605-609. doi:10.1126/SCIENCE.AAO0100

95. Niewoehner O, Garcia-Doval C, Rostøl JT, et al. Type III CRISPR-Cas systems produce cyclic oligoadenylate second messengers. *Nature*. 2017;548(7669):543-548. doi:10.1038/NATURE23467

96. Foster K, Kalter J, Woodside W, Terns RM, Terns MP. The ribonuclease activity of Csm6 is required for anti-plasmid immunity by Type III-A CRISPR-Cas systems. *RNA Biol*. 2019;16(4):449-460. doi:10.1080/15476286.2018.1493334

97. Sheppard NF, Glover CVC, Terns RM, Terns MP. The CRISPR-associated Csx1 protein of Pyrococcus furiosus is an adenosine-specific endoribonuclease. *RNA*. 2016;22(2):216-224. doi:10.1261/RNA.039842.113

98. Koonin E V., Makarova KS. Discovery of Oligonucleotide Signaling Mediated by CRISPR-Associated Polymerases Solves Two Puzzles but Leaves an Enigma. *ACS Chem Biol*. 2018;13(2):309-312. doi:10.1021/ACSCHEMBIO.7B00713

99. Maniv I, Jiang W, Bikard D, Marraffini LA. Impact of Different Target Sequences on Type III CRISPR-Cas Immunity. *J Bacteriol*. 2016;198(6):941. doi:10.1128/JB.00897-15

100. Oliynyk M, Samborskyy M, Lester JB, et al. Complete genome sequence of the erythromycin-producing bacterium Saccharopolyspora erythraea NRRL23338. *Nat*

Biotechnol. 2007;25(4):447-453. doi:10.1038/nbt1297

101. Bland C, Ramsey TL, Sabree F, Lowe M, Brown K, Kyrpides NC HP. CRISPR Recognition Tool (CRT): a tool for automatic detection of clustered regularly interspaced palindromic repeats. *BMC Bioinformatics*.

102. Kearse M, Moir R, Wilson A, et al. Geneious Basic: An integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647-1649. doi:10.1093/bioinformatics/bts199

103. Russel J, Pinilla-Redondo R, Mayo-Muñoz D, Shah SA, Sørensen SJ. CRISPRCasTyper: Automated Identification, Annotation, and Classification of CRISPR-Cas Loci. *https://home.liebertpub.com/crispr*. 2020;3(6):462-469. doi:10.1089/CRISPR.2020.0059

104. Burstein D, Harrington LB, Strutt SC, et al. New CRISPR-Cas systems from uncultivated microbes. *Nature*. 2017;542(7640):237-241. doi:10.1038/NATURE21059

105. Koonin E V., Makarova KS, Zhang F. Diversity, classification and evolution of CRISPR-Cas systems. *Curr Opin Microbiol*. 2017;37:67-78. doi:10.1016/j.mib.2017.05.008

106. Couvin D, Bernheim A, Toffano-Nioche C, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. *Nucleic Acids Res*. 2018;46(W1):W246-W251. doi:10.1093/NAR/GKY425

107. Crooks GE, Hon G, Chandonia J-M, Brenner SE. WebLogo: A Sequence Logo Generator. http://www.genome.org/cgi/doi/10.1101/. Accessed April 25, 2018.

108. Zuker M, Stiegler P. Optimal computer folding of large RNA sequences using thermodynamics and auxiliary information. *Nucleic Acids Res*. 1981;9(1):133-148. http://www.ncbi.nlm.nih.gov/pubmed/6163133. Accessed April 25, 2018.

109. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. *J Mol Biol*. 1990;215(3):403-410. doi:10.1016/S0022-2836(05)80360-2

110. Jiang W, Maniv I, Arain F, Wang Y, Levin BR, Marraffini LA. Dealing with the Evolutionary Downside of CRISPR Immunity: Bacteria and Beneficial Plasmids. *PLOS Genet*. 2013;9(9):e1003844. doi:10.1371/JOURNAL.PGEN.1003844

111. Wimmer F, Beisel CL. CRISPR-Cas Systems and the Paradox of Self-Targeting Spacers. *Front Microbiol*. 2020;10. doi:10.3389/FMICB.2019.03078

112. Jackson SA, McKenzie RE, Fagerlund RD, Kieper SN, Fineran PC, Brouns SJJ. CRISPR-Cas: Adapting to change. *Science (80-)*. 2017;356(6333):eaal5056. doi:10.1126/science.aal5056

113. Liu W-B, Shi Y, Yao L-L, Zhou Y, Ye B-C. Prediction and Characterization of Small Non-Coding RNAs Related to Secondary Metabolites in Saccharopolyspora erythraea. *PLoS One*. 2013;8(11):80676. doi:10.1371/JOURNAL.PONE.0080676

114. Marcellin E, Mercer TR, Licona-Cassani C, et al. Saccharopolyspora erythraea'sgenome is organised in high-order transcriptional regions mediated by targeted degradation at the metabolic switch. *BMC Genomics 2013 141*. 2013;14(1):1-9. doi:10.1186/1471-2164-14-15

115. Y Z. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics*. 2008;9. doi:10.1186/1471-2105-9-40

116. Burroughs AM, Zhang D, Schäffer DE, Iyer LM, Aravind L. Comparative genomic analyses reveal a vast, novel network of nucleotide-centric systems in biological conflicts, immunity and signaling. *Nucleic Acids Res*. 2015;43(22):10633-10654. doi:10.1093/NAR/GKV1267

117. Hornung V, Hartmann R, Ablasser A, Hopfner K-P. OAS proteins and cGAS: unifying concepts in sensing and responding to cytosolic nucleic acids. *Nat Rev Immunol 2014 148*. 2014;14(8):521-528. doi:10.1038/nri3719

118. Zhang X, Wu J, Du F, et al. The Cytosolic DNA Sensor cGAS Forms an Oligomeric Complex with DNA and Undergoes Switch-like Conformational Changes in the Activation Loop. *Cell Rep*. 2014;6(3):421-430. doi:10.1016/J.CELREP.2014.01.003

119. Shmakov SA, Makarova KS, Wolf YI, Severinov K V., Koonin E V. Systematic prediction of genes functionally linked to CRISPR-Cas systems by gene neighborhood analysis. *Proc Natl Acad Sci*. 2018;115(23):E5307-E5316. doi:10.1073/PNAS.1803440115

120. Estrella MA, Kuo F-T, Bailey S. RNA-activated DNA cleavage by the Type III-B CRISPR-Cas effector complex. *Genes Dev*. 2016;30(4):460-470. doi:10.1101/gad.273722.115

121. Rouillon C, Athukoralage JS, Graham S, Grüschow S, White MF. Control of cyclic oligoadenylate synthesis in a type III CRISPR system. *Elife*. 2018;7. doi:10.7554/ELIFE.36734

122. Marraffini LA, Sontheimer EJ. Self versus non-self discrimination during CRISPR RNA-directed immunity. *Nature*. 2010;463(7280):568-571. doi:10.1038/nature08703

123. Bierman M, Logan R, O'Brien K, Seno ET, Nagaraja Rao R, Schoner BE. Plasmid cloning vectors for the conjugal transfer of DNA from Escherichia coli to Streptomyces spp. *Gene*. 1992;116(1):43-49. doi:10.1016/0378-1119(92)90627-2

124. Cobb RE, Wang Y, Zhao H. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth Biol*. 2015;4(6):723-728.

doi:10.1021/sb500351f

125. Shao Z, Rao G, Li C, Abil Z, Luo Y, Zhao H. Refactoring the Silent Spectinabilin Gene Cluster Using a Plug-and-Play Scaffold. 2013. doi:10.1021/sb400058n

126. Jiang W, Bikard D, Cox D, Zhang F, Marraffini LA. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol*. 2013;31(3):233-239. doi:10.1038/nbt.2508

127. Pattanayak V, Lin S, Guilinger JP, Ma E, Doudna JA, Liu DR. High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity. *Nat Biotechnol 2013 319*. 2013;31(9):839-843. doi:10.1038/nbt.2673

128. Karvelis T, Gasiunas G, Young J, et al. Rapid characterization of CRISPR-Cas9 protospacer adjacent motif sequence elements. *Genome Biol 2015 161*. 2015;16(1):1-13. doi:10.1186/S13059-015-0818-7

129. Leenay RT, Maksimchuk KR, Slotkowski RA, et al. Identifying and visualizing functional PAM diversity across CRISPR-Cas systems. *Mol Cell*. 2016;62(1):137. doi:10.1016/J.MOLCEL.2016.02.031

130. Kasey CM, Zerrad M, Li Y, Cropp TA, Williams GJ. Development of Transcription Factor-Based Designer Macrolide Biosensors for Metabolic Engineering and Synthetic Biology. *ACS Synth Biol*. 2017;7(1):227-239. doi:10.1021/ACSSYNBIO.7B00287

131. Hale CR, Zhao P, Olson S, et al. RNA-Guided RNA Cleavage by a CRISPR RNA-Cas Protein Complex. *Cell*. 2009;139(5):945-956.

132. Schürer H, Lang K, Schuster J, Mörl M. A universal method to produce in vitro transcripts with homogeneous 3′ ends. *Nucleic Acids Res*. 2002;30(12):e56-e56. doi:10.1093/NAR/GNF055

133. Crawley AB, Henriksen JR, Barrangou R. CRISPRdisco: An Automated Pipeline for the Discovery and Analysis of CRISPR-Cas Systems. *Cris J*. 2018;1(2):171-181. doi:10.1089/CRISPR.2017.0022/ASSET/IMAGES/LARGE/FIGURE5.JPEG

134. Cobb RE, Wang Y, Zhao H. High-Efficiency Multiplex Genome Editing of Streptomyces Species Using an Engineered CRISPR/Cas System. *ACS Synth Biol*. 2015;4(6):723-728. doi:10.1021/SB500351F

135. Hidalgo-Cantabrana C, Goh YJ, Pan M, Sanozky-Dawes R, Barrangou R. Genome editing using the endogenous type I CRISPR-Cas system in Lactobacillus crispatus. *Proc Natl Acad Sci U S A*. 2019;116(32):15774-15783. doi:10.1073/PNAS.1905421116/-/DCSUPPLEMENTAL

136. Ichikawa HT, Cooper JC, Lo L, Potter J, Terns RM, Terns MP. Programmable type III-A CRISPR-Cas DNA targeting modules. *PLoS One*. 2017;12(4). doi:10.1371/JOURNAL.PONE.0176221

137. Cheng F. Harnessing the native type I-B CRISPR-Cas for genome editing in a polyploid archaeon. *J Genet Genomics*. 2017;44(11):541-548. doi:10.1016/j.jgg.2017.09.010

138. Zhang J, Zong W, Hong W, Zhang Z-T, Wang Y. Exploiting endogenous CRISPR-Cas system for multiplex genome editing in Clostridium tyrobutyricum and engineer the strain for high-level butanol production. *Metab Eng*. 2018;47:49-59. doi:10.1016/J.YMBEN.2018.03.007

139. S.J. B, M.M. J, M. L, et al. Small CRISPR RNAs guide antiviral defense in prokaryotes. *Science (80-)*. 321:960-964.

140. Sinkunas T. In vitro reconstitution of Cascade-mediated CRISPR immunity in Streptococcus thermophilus. *EMBO J*. 2013;32(3):385-394. doi:10.1038/emboj.2012.352

141. Westra ER, van Erp PBG, Künne T, et al. CRISPR Immunity Relies on the Consecutive Binding and Degradation of Negatively Supercoiled Invader DNA by Cascade and Cas3. *Mol Cell*. 2012;46(5):595-605.

142. Rouillon C, Zhou M, Zhang J, et al. Structure of the CRISPR interference complex CSM reveals key similarities with cascade. *Mol Cell*. 2013;52(1):124-134. doi:10.1016/J.MOLCEL.2013.08.020

143. Tamulaitis G, Venclovas Č, Siksnys V. Type III CRISPR-Cas Immunity: Major Differences Brushed Aside. *Trends Microbiol*. 2017;25(1):49-61. doi:10.1016/J.TIM.2016.09.012

144. Zhang J, Graham S, Tello A, Iu H, White MF. Multiple nucleic acid cleavage modes in divergent type III CRISPR systems. *Nucleic Acids Res*. 2016;44(4):1789-1799. doi:10.1093/NAR/GKW020

145. Benda C, Ebert J, Scheltema RA, et al. Structural model of a CRISPR RNA-silencing complex reveals the RNA-target cleavage activity in Cmr4. *Mol Cell*. 2014;56(1):43-54. doi:10.1016/J.MOLCEL.2014.09.002

146. Staals RHJ, Agari Y, Maki-Yonekura S, et al. Structure and Activity of the RNA-Targeting Type III-B CRISPR-Cas Complex of Thermus thermophilus. *Mol Cell*. 2013;52(1):135-145.

147. Spilman M, Cocozaki A, Hale C, et al. Structure of an RNA Silencing Complex of the CRISPR-Cas Immune System. *Mol Cell*. 2013;52(1):146-152. doi:10.1016/J.MOLCEL.2013.09.008

148. Osawa T, Inanaga H, Sato C, Numata T. Crystal structure of the crispr-cas RNA silencing cmr complex bound to a target analog. *Mol Cell*. 2015;58(3):418-430.

149. Hale CR, Cocozaki A, Li H, Terns RM, Terns MP. Target RNA capture and cleavage by the Cmr type III-B CRISPR–cas effector complex. *Genes Dev*. 2014;28(21):2432-2443. doi:10.1101/GAD.250712.114

150. Tamulaitis G, Kazlauskiene M, Manakova E, et al. Programmable RNA Shredding by the Type III-A CRISPR-Cas System of Streptococcus thermophilus. *Mol Cell*. 2014;56(4):506-517. doi:10.1016/J.MOLCEL.2014.09.027

151. M.A. E, F.T. K, S. B. RNA-activated DNA cleavage by the Type III-B CRISPR–Cas effector complex. *Genes Dev*. 30:460-470.

152. Kazlauskiene M, Tamulaitis G, Kostiuk G, Venclovas Č, Siksnys V. Spatiotemporal Control of Type III-A CRISPR-Cas Immunity: Coupling DNA Degradation with the Target RNA Recognition. *Mol Cell*. 2016;62(2):295-306. doi:10.1016/J.MOLCEL.2016.03.024

153. Anantharaman V, Makarova KS, Burroughs AM, Koonin E V., Aravind L. Comprehensive analysis of the HEPN superfamily: Identification of novel roles in intra-genomic conflicts, defense, pathogenesis and RNA processing. *Biol Direct*. 2013;8(1). doi:10.1186/1745-6150-8-15

154. Niewoehner O, Jinek M. Structural basis for the endoribonuclease activity of the type III-A CRISPR-associated protein Csm6. *RNA*. 2016;22(3):318-329. doi:10.1261/RNA.054098.115

155. Guan J, Wang W, Sun B. Chromosomal Targeting by the Type III-A CRISPR-Cas System Can Reshape Genomes in Staphylococcus aureus. *mSphere*. 2017;2(6). doi:10.1128/MSPHERE.00403-17

156. Makarova KS, Aravind L, Wolf YI, Koonin E V. Unification of Cas protein families and a

simple scenario for the origin and evolution of CRISPR-Cas systems. *Biol Direct*. 2011;6. doi:10.1186/1745-6150-6-38

157. Reeks J, Naismith JH, White MF. CRISPR interference: A structural perspective. *Biochem J*. 2013;453(2):155-166. doi:10.1042/BJ20130316

158. Nickel L, Ulbricht A, Alkhnbashi OS, et al. Cross-cleavage activity of Cas6b in crRNA processing of two different CRISPR-Cas systems in Methanosarcina mazei Gö1. *RNA Biol*. 2019;16(4):492-503. doi:10.1080/15476286.2018.1514234

159. Hale CR, Majumdar S, Elmore J, et al. Essential Features and Rational Design of CRISPR RNAs that Function with the Cas RAMP Module Complex to Cleave RNAs. *Mol Cell*. 2012;45(3):292-302.

160. Zhang J, Rouillon C, Kerou M, et al. Structure and Mechanism of the CMR Complex for CRISPR-Mediated Antiviral Immunity. *Mol Cell*. 2012;45(3):303-313. doi:10.1016/j.molcel.2011.12.013

161. S D, L K. Organization of the enzymatic domains in the multifunctional polyketide synthase involved in erythromycin formation in Saccharopolyspora erythraea. *Gene*. 1992;111(1):51-60. doi:10.1016/0378-1119(92)90602-L

162. Malico AA, Nichols L, Williams GJ. Synthetic biology enabling access to designer polyketides. *Curr Opin Chem Biol*. 2020;58:45-53. doi:10.1016/J.CBPA.2020.06.003

163. Liu Y, Wei W-P, Ye B-C. High GC Content Cas9-Mediated Genome-Editing and Biosynthetic Gene Cluster Activation in Saccharopolyspora erythraea. *ACS Synth Biol*. 2018;7(5):1338-1348. doi:10.1021/ACSSYNBIO.7B00448

164. Kasey CM, Zerrad M, Li Y, Cropp TA, Williams GJ. Development of Transcription Factor-Based Designer Macrolide Biosensors for Metabolic Engineering and Synthetic

Biology. doi:10.1021/acssynbio.7b00287

165. Flett F, Mersinias V, Smith CP. High efficiency intergeneric conjugal transfer of plasmid DNA from Escherichia coli to methyl DNA-restricting streptomycetes. *FEMS Microbiol Lett*. 1997;155(2):223-229. doi:10.1111/J.1574-6968.1997.TB13882.X

166. Wang Y, Wang Y, Zhang S. High frequency transformation of the industrial erythromycin-producing bacterium Saccharopolyspora erythraea. *Biotechnol Lett 2007 302*. 2007;30(2):357-361. doi:10.1007/S10529-007-9547-7

167. Jiang W, Samai P, Marraffini LA. Degradation of Phage Transcripts by CRISPR-Associated RNases Enables Type III CRISPR-Cas Immunity. *Cell*. 2016;164(4):710-721. doi:10.1016/j.cell.2015.12.053

168. Deng L, Garrett RA, Shah SA, Peng X, She Q. A novel interference mechanism by a type IIIB CRISPR-Cmr module in Sulfolobus. *Mol Microbiol*. 2013;87(5):1088-1099. doi:10.1111/MMI.12152

169. Rouillon C, Athukoralage JS, Graham S, Grüschow S, White MF. Investigation of the cyclic oligoadenylate signaling pathway of type III CRISPR systems. *Methods Enzymol*. 2019;616:191-218. doi:10.1016/BS.MIE.2018.10.020

170. Osawa T, Inanaga H, Sato C, Numata T. Crystal structure of the crispr-cas RNA silencing cmr complex bound to a target analog. *Mol Cell*. 2015;58(3):418-430. doi:10.1016/J.MOLCEL.2015.03.018

171. Taylor DW, Zhu Y, Staals RHJ, et al. Structures of the CRISPR-Cmr complex reveal mode of RNA target positioning. *Science (80-)*. 2015;348(6234):581-585. doi:10.1126/SCIENCE.AAA4535

172. Wiedenheft B, Lander GC, Zhou K, et al. Structures of the RNA-guided surveillance

complex from a bacterial immune system. *Nature*. 2011;477(7365):486-489. doi:10.1038/NATURE10402

173. Makarova KS, Wolf YI, Koonin EV. The basic building blocks and evolution of CRISPR–Cas systems. *Biochem Soc Trans*. 2013;41(6):1392-1400. doi:10.1042/BST20130038

174. Koonin E V., Makarova KS. Origins and evolution of CRISPR-Cas systems. *Philos Trans R Soc B*. 2019;374(1772). doi:10.1098/RSTB.2018.0087

175. Makarova KS, Wolf YI, Iranzo J, et al. Evolutionary classification of CRISPR–Cas systems: a burst of class 2 and derived variants. *Nat Rev Microbiol*. 2020;18(2):67-83. doi:10.1038/S41579-019-0299-X

176. Stern A, Keren L, Wurtzel O, Amitai G, Sorek R. Self-targeting by CRISPR: Gene regulation or autoimmunity? *Trends Genet*. 2010;26(8):335-340. doi:10.1016/J.TIG.2010.05.008

177. McDaniel R, Thamchaipenet A, Gustafsson C, Fu H, Betlach M, Ashley G. Multiple genetic modifications of the erythromycin polyketide synthase to produce a library of novel "unnatural" natural products. *Proc Natl Acad Sci U S A*. 1999;96(5):1846-1851. http://www.ncbi.nlm.nih.gov/pubmed/10051557. Accessed April 2, 2019.

178. Fang L, Guell M, Church GM, Pfeifer BA. Heterologous Erythromycin Production across Strain and Plasmid Construction. *Biotechnol Prog*. 2018;34(1):271. doi:10.1002/BTPR.2567

179. Koryakina I, Kasey C, McArthur JB, et al. Inversion of Extender Unit Selectivity in the Erythromycin Polyketide Synthase by Acyltransferase Domain Engineering. *ACS Chem Biol*. 2017;12(1):114-123. doi:10.1021/acschembio.6b00732

180. Mutka SC, Bondi SM, Carney JR, Da Silva NA, Kealey JT. Metabolic pathway engineering for complex polyketide biosynthesis in Saccharomyces cerevisiae. *FEMS Yeast Res*. 2006;6(1):40-47. doi:10.1111/J.1567-1356.2005.00001.X

181. Kealey JT, Liu L, Santi D V., Betlach MC, Barr PJ. Production of a polyketide natural product in nonpolyketide-producing prokaryotic and eukaryotic hosts. *Proc Natl Acad Sci*. 1998;95(2):505-509. doi:10.1073/PNAS.95.2.505

182. Zhang G-C, Kong II, Kim H, Liu J-J, Cate JHD, Jin Y-S. Construction of a Quadruple Auxotrophic Mutant of an Industrial Polyploid Saccharomyces cerevisiae Strain by Using RNA-Guided Cas9 Nuclease. *Appl Environ Microbiol*. 2014;80(24):7694. doi:10.1128/AEM.02310-14

183. Baker Brachmann C, Davies A, Cost GJ, et al. Designer Deletion Strains derived from Saccharomyces cerevisiae S288C: a Useful set of Strains and Plasmids for PCR-mediated Gene Disruption and Other Applications. *Yeast*. 1998;14:115-132. doi:10.1002/(SICI)1097-0061(19980130)14:2

184. Schiestl RH, Gietz RD. High efficiency transformation of intact yeast cells using single stranded nucleic acids as a carrier. *Curr Genet 1989 165*. 1989;16(5):339-346. doi:10.1007/BF00340712

APPENDICES

Appendix A

Table A1. Strains and plasmids used in Chapter 2 studies.

Strains/ Plasmids	Relevant characteristic	Sources
Strains		
E. coli		
BL21(DE3)	F $^-$ *ompT hsdSB (rB$^-$ mB$^-$) gal dcm* (DE3)	Lucigen
E. cloni 5-alpha	*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Lucigen
S. erythraea		
NRRL2338	wild type stain	ATCC
Plasmids		
pCRISPomyces-2	pSG5 ori, ColE1 ori, Apm^R, Cas9, sgRNA	[134]
pCRISPomyces-Target	pCRISPomyces-2 derivative; pSG5 ori, ColE1 ori, Apm^R, protospacers inserted at BbsI sites	This work
pCDFDuet-1	pCloDF13 ori, pT7, two multiple cloning sites, *lac* operator, *lacI* , streptomycin/spectinomycin resistance marker	Novagen
pCDFDuet-effector-complex-ancillary	pCDFDuet-1 derivative; pT7-Csm1-Cas10-Csm3-Csm4-Csm5-Cas6, pT7-SeryHelicase-SACE_RS24645-SACE_RS24720	This work
pCDFDuet-effector-complex	pCDFDuet-1 derivative; pT7-Csm1-Cas10-Csm3-Csm4-Csm5-Cas6	This work
pCDFDuet-effector-complex-ΔCsm1	pCDFDuet-1 derivative; pT7-Cas10-Csm3-Csm4-Csm5-Cas6	This work
pCDFDuet-effector-complex-ΔCas10	pCDFDuet-1 derivative; pT7-Csm1-Csm3-Csm4-Csm5-Cas6	This work
pCDFDuet-effector-complex-ΔCsm3	pCDFDuet-1 derivative; pT7-Csm1-Cas10-Csm4-Csm5-Cas6	This work
pCDFDuet-effector-complex-ΔCsm4	pCDFDuet-1 derivative; pT7-Csm1-Cas10-Csm3-Csm5-Cas6	This work
pCDFDuet-effector-complex-ΔCsm5	pCDFDuet-1 derivative; pT7-Csm1-Cas10-Csm3-Csm4-Cas6	This work
pCDFDuet-effector-complex-ΔCas6	pCDFDuet-1 derivative; pT7-Csm1-Cas10-Csm3-Csm4-Csm5	This work
pZZ04	pSC101 ori, Apm^R, AraC, P*araBAD* transcribing two direct repeats interspaced by BbsI-*lacZ*-BbsI cassette	This work
pZZ04-tgt-ftsZ	pZZ04 derivative; *lacZ* was replaced by *ftsZ* targeting spacer	This work
pZZ04-tgt-template	pZZ04 derivative; *lacZ* was replaced by template stranding targeting spacer of pSense-5N	This work
pZZ04-tgt-coding	pZZ04 derivative; *lacZ* was replaced by coding strand targeting spacer of pSense-5N	This work
pSense	CloE1 ori, Amp^R, P*lac*-mphO-*sfGFP*, PlacIQ-*mphR*	[130]
pSense-5	pSense derivative; 5nt library was installed at *mphR*	This work

Table A2. Identified Cas proteins and hypothetically related proteins in *S. erythraea*.

Locus tag	**Systematic name**	**Specialty name**	**Amino Acid Sequence**
SACE_RS24635	Cas4/Cas1		MDPPLPIGARALSDFRFCPRLFHLEHAEGYRFDSEEMRLGRHVHASVDVTSGRERPAQRIDPSNWRICALALSSDELGLVAVCDVVEAVGGNVRPVEYRRGAPQRDGQPWPNDRIQLLAQIVLLRHHGYHCTHGYLWYDSVRRRVPVPWSATAEAELRHHLFRARHVASQGQPPPPLRHSPKCPRCALLPICMPDEINELSERDSEKPRKLLARAPARDPVYVTEPGTTVGIRSERLAVRKDHEELLSCRLRDVLHLVAAGPVQVTSQAVHALAEQGSPVVWTSTTGRLKSVDIPTVGKHVELRRRQFTATPNTALDFARRIVGGKIRNARTLLRRNPHVEEPDLLNRLDADAVRAERAPTRSTLLGIEGAAARTYFAGLVETFRTDHRLPGPAFDTMGRTRRPPRDAVSCLLSFLYCLLIKDITTACYALGLDPYFGFYHQPRHGRPALTLDLAEEFRPLIADSTALTLINNLQADPAMFHVHSTAVALTSSGRRDVIDAYERRLTTEVIHPVFRYKTTYRRAIEIQARLFAAHLLDEIPHYTAFTSR*
	Cas2		MRRHRYLVAYDIRDPRRLRLVAKKMEDFGDRTQYSVFLCDLTRADVADMTRALLTVIDSTIDRILIVDLGSGDSDSRFEFLGRRHGLPTTGHRVL*
SACE_RS24645			MTANTDTTDFTPPYNITWRTFLGSLDRMASDASLPSVIDRSYLSWMPGSVQTSYLTLCRQFGLIDDDGSPMSLLSEIVYTPNSRPEVVARLLRVHYAAIVELGKTHATLQQLMDLWKETFGQNGETRRKAITFYMQAADFAQIPVSPLWERGVAKASRQATPARRARPRARRTKQKPTAPQSTSTPSDTVVLASGAGTLSISVDIDPLLLSEEDRGFVFGVIDSVHEYQNKHTSDTTNDPEDTTS*

Table A2 (continued).

SACE_RS24650		SeryHelicase	MSTYGFDPQRRPYQHDRVLLYVRAGGRCQRCGTE LGADYHAAHLVAWNNGGGTTLANMQAWCPRCN LGLGDHDVEPFGELALRPWQSQALPVIAERIFNYG TAVLHAAPGAGKTIFAAALFRILHDAGYVERMVV VVPNTALVRQWRSELAQLRIHLDPHPTDGALETTG LAGVIVTYQSLPTYAGVHATRIRQRPTLIVFDEVH HVATDASWGWAVRNMVGDVASGDIHAHAVLNM TGTLFRSDPRRRISTVNYDRLIVEGVPKLQAAADW SVPTRDLVGTELRAPDLYCYGGRAEYIDLTSAKAQ PRKQIADLDEQERVGILPQVFASPEWINGFAAEAV RLLAEQRAALDGAEHLKLLFVAPTIVAARHAAHA LNKATGTDFARLVVHDEPHALNTLRIAAEEPRSCG IVTVRMVTEGFDCPRIATIAYATNIVAPLFVAQMM ARAMRLTPTERDTGTHLPAKILIPDHPALRDAFVD AIRDTIPLTEDTEPDERGAAPETLGAHHYDLLGIDD PRLHNVNVLDQDEAVSTEELTEAQTLCEKLHIPTA FAARVAIGIRGRGT*
SACE_RS24690		SeryCsm1	MSGQFVNPYTFVPLPAVPPERSGPRGHSGEDPELL TGKLSVRIKAVSPLLIRGFNKRKRDDVADGRIPMR PDGTPFIPGSSLKGAVRSLHETLTGSCLRVFDTDFV PSYRESVLPRRQRRMAVVTAPPKDGRPPEIRLCHS GDSRKHRVHQDILRQYKGELVSGQRLNVLSWDK KGSPDKIELSEYGDWVLFLSDSGAREKDKAYRAHI RKLSQDMGVLDDDVWSTFLRVVEETDDRRTAQR ERYGDETTADVFFEYAPAKGKAEYLHIGRRYRASP HLVPGQPVWVDVTGLNRISWIGLAMNWRHAKSR TSAGQRATGYEPCAKSTELCPSCRLFGSIDPNERAP EERAVQSAYRGHVRFGDALAQHAVTGEEVTLPPM GAPHPGAGQAYLDNRQVAATAGDPPLREWGSSA DGRTPRRLRGRKHYWQAERNGREKAREHHSAEM VTTAHVFPEGTEFRATITFVDIDEAQLGGLLATLSP STALDAPDLRVHIGGGKPLGFGQCGIEVDLEHSEL DRSGSRYGVGGGVRDLEAEADALVESFRSSLNAE VTDLWPLVAKVLNPETVPGDLVWYPPGAEWAKR HSGNPDDVKEFDDGFEFWKRTSGASTSGEAKGVS DSPLVSLPHLTKDASETNQAMDIVVKSGEQDA*

Table A2 (continued).

SACE_RS24685		SeryCas10	MSDYMDIGVVRIQSWLTRTPKLRGRRGGSTLITQATDHDAIREVLSGFGETVEVHSEGGRIDGVVPLKLHDEAQAAAAEKAVVRHIRSRMPGVSLSVKRYSGDTYAVARSVGARQEHEWPAPAVDWPAGLPCQWCRTLPADPAPGARDEDGLRRCAECLARDAAAGRADGRDVPFAESELLHRLGIAEAKVPDEFKPMAELDRKGNRPDADGQTHLALVYADGNAIGKFITSALRHAPRKKRRSPDFAAAIDGSTWQALVTAVEQIRQEDQELLPVVPHFVGGDDVLVSVPACRAWRFTRALLGAFDEAIKEASGLHTDLPSLSAGVVFHHYTVPLYEVNELAKTALRRAKVHTSGEAASLAWHDTTHDGHDPVARSALTLADLEQKWDALHRLAKLPRSAHARLEAVTRAYGSRSSELLTHAIRVGSDQVIRQFQGSFDLADALGMARWWWHDDQA*
SACE_RS24680		SeryCsm3	MARGQGGGGSHDTIDERDQLPATSMAGVVRATAKKLLGSQNPVIDEVFGSDGNPSPWRWSPVRPDGDWHGAQPAARVRIDQHSRTAQGHMLILADHTGADQGRFSITKFRYVPRERLAVHQAVLKVAAQATCSLGAWRRRGSGWVGIRCAEEINEETVRIFLGLKR*
SACE_RS24675		SeryCsm4	MKDLARITVTLLEPMAAGKRLRADFLQDTYDHVPGTVVRGALAGRWINERGVEVTTTEEFLDVFEGDGAFGPLHNENSLPAPLSVKVHKYAAKDDCPRLWWDRAASDGEQEHCKCGQALEDSKGESTGSVARQRRTRAALDEDGVVREGQLFTQNSLQRDTRFSGWLHGPAVRALWIVDQPVDTVYLGGRLKHQGAAVVQVDTEVEPEPVEQRGNKLILRLLGPGAFVDKHGFPSVKPDLDELTDVLQVEAQAVENCWTRWCEVEGWHAASGLPKPVERAVQPGSTYVIRLERAADEQARKRLMARGIGLRRREGFGALYSVKQEPLTVQRLVQGSAPLRNPARLRTDVPLLRQRYEQMLRGEVDDTMFQESLTGDGGYEKALRRLLNVVEPTLFYAVLTDLESHR*

Table A2 (continued).

SACE_RS24670		SeryCsm5	MTLITARLRFETAGGVGGTEAGAGENGYLQPIRRD PAPKGKAKGKVSLPGTTIAGSLRAHCAAFDAGLG DLFGDTPEQLADKHGKAEAMRRSQKRNDEKPERA MPSAIQVLGTLLHRRGETFSHTRNSSDRKRGAART HHFHRVEMLETGTEFDVVLRWDNADEELLERFLD VLKQWHPTLGRGVTRGAGRCSLVGWGRVDYDLD SPDGLMAWLNTERRPNPAETPETRSAETYALDVD LSIVDAIHCGTELNSVNGPKHNTLTAFRHGEEFVV PGSTLKGVLRSRAEYICRVLGLLACKDQQCGECRP CLLFGFAGEENARRGRIAVDDAVIEDAQPDFRPHV AIDRFTGGARDQALYEHEVVAAGWFPMRVRWLG TETGAEAAQDARLLHAVVADLDARYVGIGARTTA GFGTVAVRECRSATFEVSELAGVLRDPVIASTEDV NA*
	Cas5U/6U	SeryCas6	MVARPGTRGTNCWNSPAGGRPRGPITSAFTRTSCP LSRRKPRTCGRGHDTGGCEFASITTTGGPHSSSRGT KSRDHPGSRRSPWTRRTSPRSRTGAVTVPSNSTDL PRERKTFSTVTTCSNSCHCARPPHCSSVTTRPAEHR KNPMTNTFSDHAGDSDVAVDSSLAERMVVFGFPS GIRVEAKRSGVVTAMLRKAIMSRMPEPLPPEVSGH NADELTHVAYLAVPDAGGVGARGDITSVAVWTP RGTPNLVSQISAALTTGKPLHLDLPGVRLRLRPQD AEPVTAGLSGPSRSWTTVTPMVLDRYPGKGREAA EIARTCIRSGLPSPIEVTTSRNPFVHGAADLGRSQLP RRDNRPYTHAKITFAEPATGPVLLGSQRYLGMGLF LPTHS*
SACE_RS24720			MAHLPTTGPTSVRTTGDYYQWLVAWEACLVLLR ETAARSHNPVRAVGVELDGVGNLDDVVLLRDVPP NTYKQVKYAVDSATPVNEEYLTKPSANGGPSILA KIARTWKGLTADGGSADLRLVTNRAADSEDGLM AGRDARTGLLMPRAAVGGARSDRGKARARWAQE AGLTEAELIDLLSVLRFDLPLDMVWYQENLRLLM AVTGLRHDQRALEEGAGWVAKMVREGQRELTLT MIETAVAELDLEAGPARAVLSIATLKPDPLASDAD HAINWVDRFDGDSPYTKRRPLPPNTWSQLQADIEA APGRLPAETTAVSVTGSIRLAPAFLVGTTFRMVTG TDLATVQRGRAGSQLWSTNDPFDTALVPGVSEDEI GQGDEIAVAIAVATDPTENVVEYLQEQNIPVRKLT VLTPPDGAANDGSLPDSTTANALAVGIRDHLRRST RRVRRLHLFLACPMGLAVLLGNRWNRLCQTVVY EDIKIDDGYEPAFTVEA*

Table A3. Primers and dsDNA fragments used in Chapter 2 studies.

Primers (pair)	Sequences
construction of pCRISPomyces-Target	
primers to delete *cas9*	GGATCCGAATTCCATATGGCTAGCGCTGCTCCTTCGGTCGGACGTG CGT GCTAGCCATATGGAATTCGGATCCCAGCTCGCGGACGTGCTCATAG TCCACGAC
Primer pairs annealed to make protospacer inserts into pCRISPomyces-Target	
A1.01	acgcCTCCAGCATCGCCGGGTCCTTGAACAGCGCCTCGATC
	aaacGATCGAGGCGCTGTTCAAGGACCCGGCGATGCTGGAG
A1.10	acgcATTCGGCCTGGGGATGGATCTCAGCCTGCTCGCGTAC
	aaacGTACGCGAGCAGGCTGAGATCCATCCCCAGGCCGAAT
A1.19	acgcCCACATCGAGAGATCCCCTACTACAGCGTGTAGCCCGGCAG
	aaacCTGCCGGGCTACACGCTGTAGTAGGGGATCTCTCGATGTGG
A2.01	acgcCTCGTCGGTCATGGATTTGCGCGCTTCTTCGCGC
	aaacGCGCGAAGAAGCGCGCAAATCCATGACCGACGAG
A2.10	acgcTTTCCGAGGCGAAGGTTTCTGGTGTTCGCCATGTCGC
	aaacGCGACATGGCGAACACCAGAAACCTTCGCCTCGGAAA
A2.15	acgcACGTCCACGCCGAGGTAGTCGGTGCCGAGGCGATCGTCGGTG
	aaacCACCGACGATCGCCTCGGCACCGACTACCTCGGCGTGGACGT
reverse A1.01	aaacCTCCAGCATCGCCGGGTCCTTGAACAGCGCCTCGATC
	acgcGATCGAGGCGCTGTTCAAGGACCCGGCGATGCTGGAG
reverse A1.10	aaacATTCGGCCTGGGGATGGATCTCAGCCTGCTCGCGTAC
	acgcGTACGCGAGCAGGCTGAGATCCATCCCCAGGCCGAAT
reverse A1.19	aaacCCACATCGAGAGATCCCCTACTACAGCGTGTAGCCCGGCAG
	acgcCTGCCGGGCTACACGCTGTAGTAGGGGATCTCTCGATGTGG
reverse A2.01	aaacCTCGTCGGTCATGGATTTGCGCGCTTCTTCGCGC
	acgcGCGCGAAGAAGCGCGCAAATCCATGACCGACGAG
reverse A2.10	aaacTTTCCGAGGCGAAGGTTTCTGGTGTTCGCCATGTCGC
	acgcGCGACATGGCGAACACCAGAAACCTTCGCCTCGGAAA
reverse A2.15	aaacACGTCCACGCCGAGGTAGTCGGTGCCGAGGCGATCGTCGGTG
	acgcCACCGACGATCGCCTCGGCACCGACTACCTCGGCGTGGACGT
reverse repeat-A1.01	aaacattgcggcCTCCAGCATCGCCGGGTCCTTGAACAGCGCCTCGATC acgcGATCGAGGCGCTGTTCAAGGACCCGGCGATGCTGGAGgccgcaat
reverse repeat-A1.10	aaacattgcggcATTCGGCCTGGGGATGGATCTCAGCCTGCTCGCGTAC acgcGTACGCGAGCAGGCTGAGATCCATCCCCAGGCCGAATgccgcaat
reverse repeat-A1.19	aaacattgcggcCCACATCGAGAGATCCCCTACTACAGCGTGTAGCCCGG CAG acgcCTGCCGGGCTACACGCTGTAGTAGGGGATCTCTCGATGTGGgcc gcaat

Table A3 (continued).

reverse repeat-A2.01	**aaacattgcggcCTCGTCGGTCATGGATTTGCGCGCTTCTTCGCGC acgcGCGCGAAGAAGCGCGCAAATCCATGACCGACGAGgccgcaat**
reverse repeat-A2.10	**aaacattgcggcTTTCCGAGGCGAAGGTTTCTGGTGTTCGCCATGTC GC**
reverse repeat-A2.15	**acgcGCGACATGGCGAACACCAGAAACCTTCGCCTCGGAAAgccg caat**
	aaacattgcggcACGTCCACGCCGAGGTAGTCGGTGCCGAGGCGAT CGTCGGTG
reverse A1.01-repeat	**acgcCACCGACGATCGCCTCGGCACCGACTACCTCGGCGTGGAC GTgccgcaat**
reverse A1.10-repeat	**aaacCTCCAGCATCGCCGGGTCCTTGAACAGCGCCTCGATCccgtc acc**
reverse A1.19-repeat	**acgcggtgacggGATCGAGGCGCTGTTCAAGGACCCGGCGATGCTG GAG**
	aaacATTCGGCCTGGGGATGGATCTCAGCCTGCTCGCGTACccgtc acc
reverse A2.01-repeat	**acgcggtgacggGTACGCGAGCAGGCTGAGATCCATCCCCAGGCCG AAT**
reverse A2.10-repeat	**aaacCCACATCGAGAGATCCCCTACTACAGCGTGTAGCCCGGCA Gccgtcacc**
reverse A2.15-repeat	**acgcggtgacggCTGCCGGGCTACACGCTGTAGTAGGGGATCTCTCG ATGTGG**
	aaacCTCGTCGGTCATGGATTTGCGCGCTTCTTCGCGCctgtcacc
	acgcggtgacagGCGCGAAGAAGCGCGCAAATCCATGACCGACGAG
	aaacTTTCCGAGGCGAAGGTTTCTGGTGTTCGCCATGTCGCctgtc acc
	acgcggtgacagGCGACATGGCGAACACCAGAAACCTTCGCCTCGG AAA
	aaacACGTCCACGCCGAGGTAGTCGGTGCCGAGGCGATCGTCG GTGctgtcacc
	acgcggtgacagCACCGACGATCGCCTCGGCACCGACTACCTCGGC GTGGACGT

Table A3 (continued).

construction of pCDFDuet-effector-complex-ancillary	
SeryCsm1 insert	GATATACCATGGGCAGCAGCCATCACCATCATCACCACAGCGTCGT GTACGGTACGGCAC
	GTAGTCGCTCATGTTTTAACCTCCTTATTGATGGGACTTATTATCAC GCGTCCTGCTCCC
SeryCas10 insert	GCGTGATAATAAGTCCCATCAATAAGGAGGTTAAAACATGAGCGAC TACATGGACATCGG
	CGCGCCATTAGTTACCTCCTTATATTAAATTTGGATCGGGTCAAGCC TGGTCATCGTGCC
SeryCsm3 insert	GACCAGGCTTGACCCGATCCAAATTTAATATAAGGAGGTAACTAAT GGCGCGCGGGCAGG
	GGTGTACCTCCTTATGACCTTAGTTTACGTATGTCACCGTTTCAACC CCAGGAAAATGCG
SeryCsm4 insert	GTGACATACGTAAACTAAGGTCATAAGGAGGTACACCATGAAGGA CTTGGCCAGGATCAC
	CACTTTCATTTTTATCCTCCTTATTAGTTTTTAATAGTCACCTGTGTG ACTCCAGATCGG
SeryCsm5 insert	ACAGGTGACTATTAAAAACTAATAAGGAGGATAAAAATGAAAGTG ACACTGATCACCGCG
	TTGCGACCATGTATAACCGTAAAAAACGTTTTCGAACGTCTCATGC GTTCACGTCCTCGG
SeryCas6 insert	GACGTTCGAAAACGTTTTTTACGGTTATACATGGTCGCAAGACCCG GAAC
	TGCGGCCGCAAGCTTGTCGACCTGCAGGCGCGCCGAGCTCGATCA GCTATGGGTCGGCAG
SeryHelicase insert	CTTAGTATATTAGTTAAGTATAAGAAGGAGATATACATATGTCCACCT ACGGCTTCGAC
	TGGTGGGGAGATGGGCCATATAGTACCTCCTTAGTTTCTTCGATCA GGTTCCTCGCCCGC
SACE_RS24720 insert	CGGGCGAGGAACCTGATCGAAGAAACTAAGGAGGTACTATATGGC CCATCTCCCCACCAC
	TTGCGGTCATGTTATACCTCCTTAAATGTGTTCCTATTTTATGCCTCC ACGGTGAATGCC
SACE_RS24645 insert	TGGAGGCATAAAATAGGAACACATTTAAGGAGGTATAACATGACCG CAAACACCGACACC
	GCTAGTTATTGCTCAGCGGTGGCAGCAGCCTAGGTTAATCAGGAGG TTGTGTCCTCAGGG

Table A3 (continued).

construction of pZZ04	
gBlocks for repeat-lacZ-repeat cassette	5'- CTTAGTGCGAGTATCTGAAAGGGGATACGCCTGTCACCGGCGCCA AGCGTCGGTCCTCATTGCGGCAAGTCTTCTCAGCCGCTACAGGGC GCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGA TCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGA TGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAG TCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC TCACTATAGGGCGAATTGGGTACCGGGCCCCCCCTCGAGG-3' 5'- GCGAATTGGGTACCGGGCCCCCCCTCGAGGTCCTCCAGCTTTTGTT CCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATA GCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAAC ATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA GTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC ACCGGTGAAGACATCTGTCACCGGCGCCAAGCGTCGGTCCTCATT GCGGCTTTTTACTCCATCTGGATTTGTTCAGAACG-3'
primers for repeat-lacZ-repeat cassette	caactctctactgtttctccatacccgtttttttgggaattcgagcgcCTGTCACCGGCG tacatatcaaagggaaaactgtccatacccatggacctttttacggttcctggcctctag
Primers for *araC*	tgcacgacattgcactccaccgctgatgacatactcttccttttcaatattattgaagc CCGCAATGAGGACCGACGCTTGGCGCCGGTGACAGgcgctcgaattcccaa aaaaacggg
Primers for Apm^R	taaatcaatctaaagtatatatgagtaaacttggtctgacagtcagccaatcgactggc aataaccctgataaatgcttcaataatattgaaaaaggaagagtatgtcatcagcggtgg
Primers for pSC101 ori	aatgcgatgccgctcgccagtcgattggctgactgtcagaccaagtttactcatatatac gcgtttttatctagaggccaggaaccgtaaaaaggtccatgggtatggacagttttccc
Primer pairs annealed to make spacer inserts into pZZ04	
tgt-ftsZ	CGGCCGCCGATGACTTTAATCACCGCGTCATTGGTAAGTT ACAGAACTTACCAATGACGCGGTGATTAAAGTCATCGGCG
mphR-template	CGGCCGCTAATCACCGCGTCATCGATGACTTTGGTAAGTT ACAGAACTTACCAAAGTCATCGATGACGCGGTGATTAGCG
mphR-coding	CGGCCCATACAGAAGGTGAACACTGATGCCCCGCCCCAAG ACAGCTTGGGGCGGGGCATCAGTGTTCACCTTCTGTATGG
construction of pSense-5N	
primers to introduce protospacer and 5Ns	CGCTAATCACCGCGTCATCGATGACTTTGGTAAGTTGGGGCGCTAT CATGCCATAC NNNNNCCATACAGAAGGTGAACACTGATGC

Figure A1. Transcription profile of *S. erythraea* CRISPR loci.

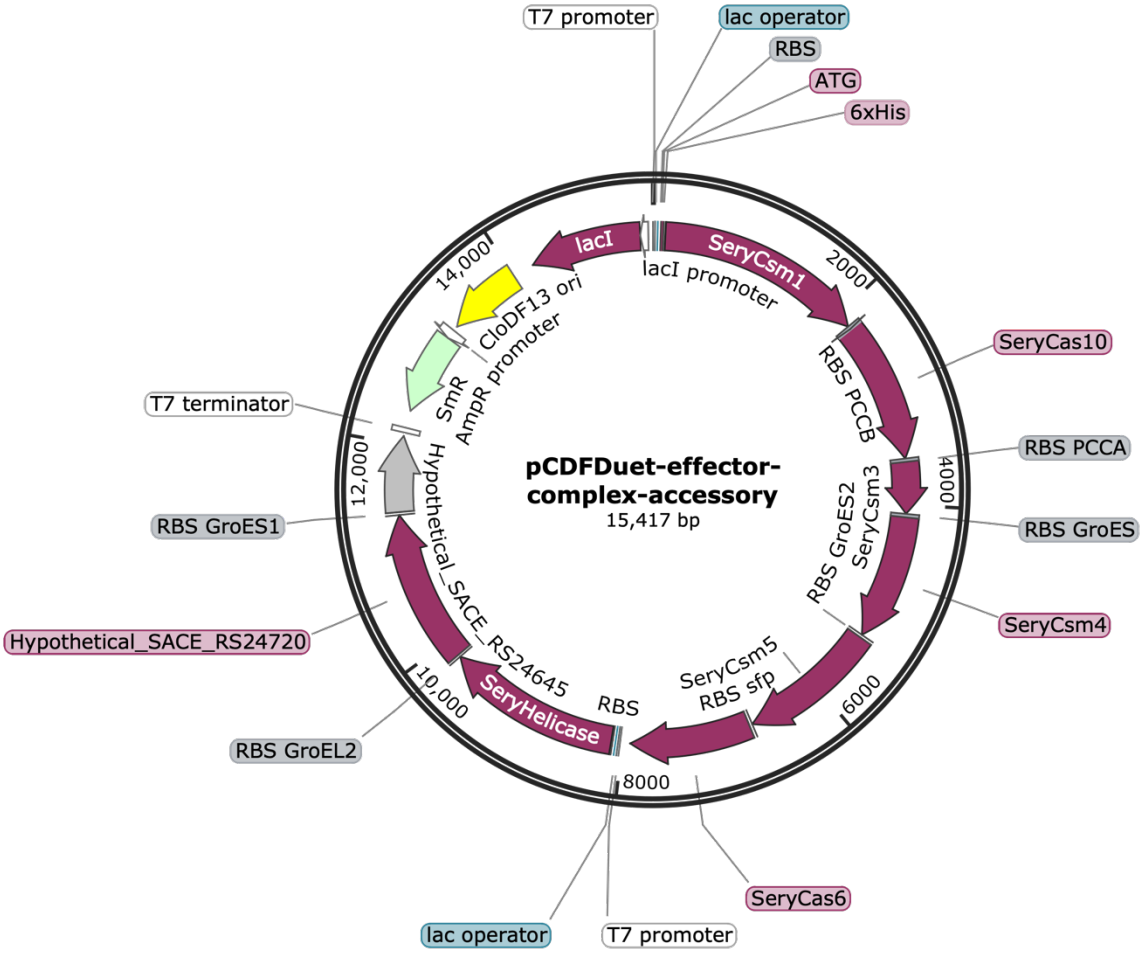


Figure A2. Plasmid map of pCDFDuet-effector-complex-ancillary

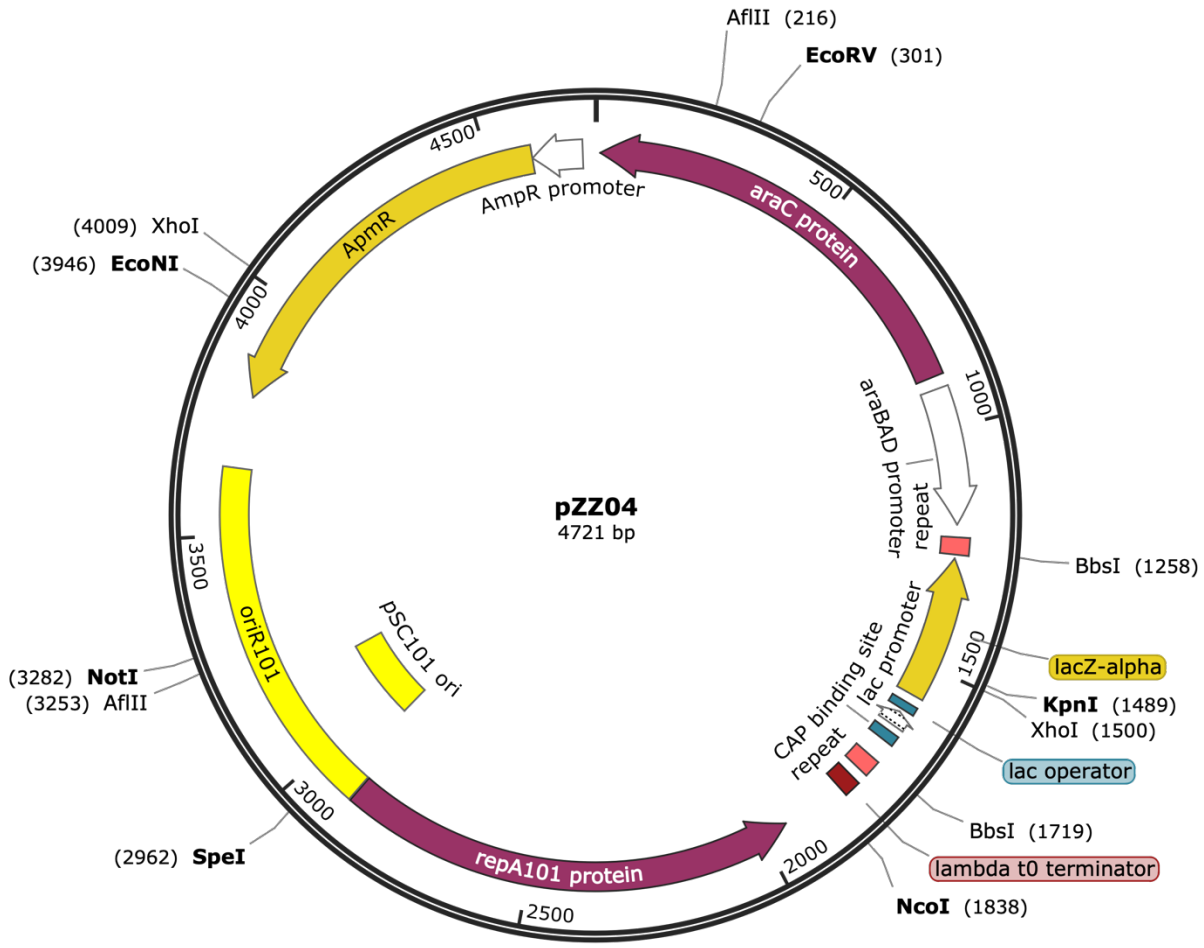


Figure A3. Plasmid map of pZZ04

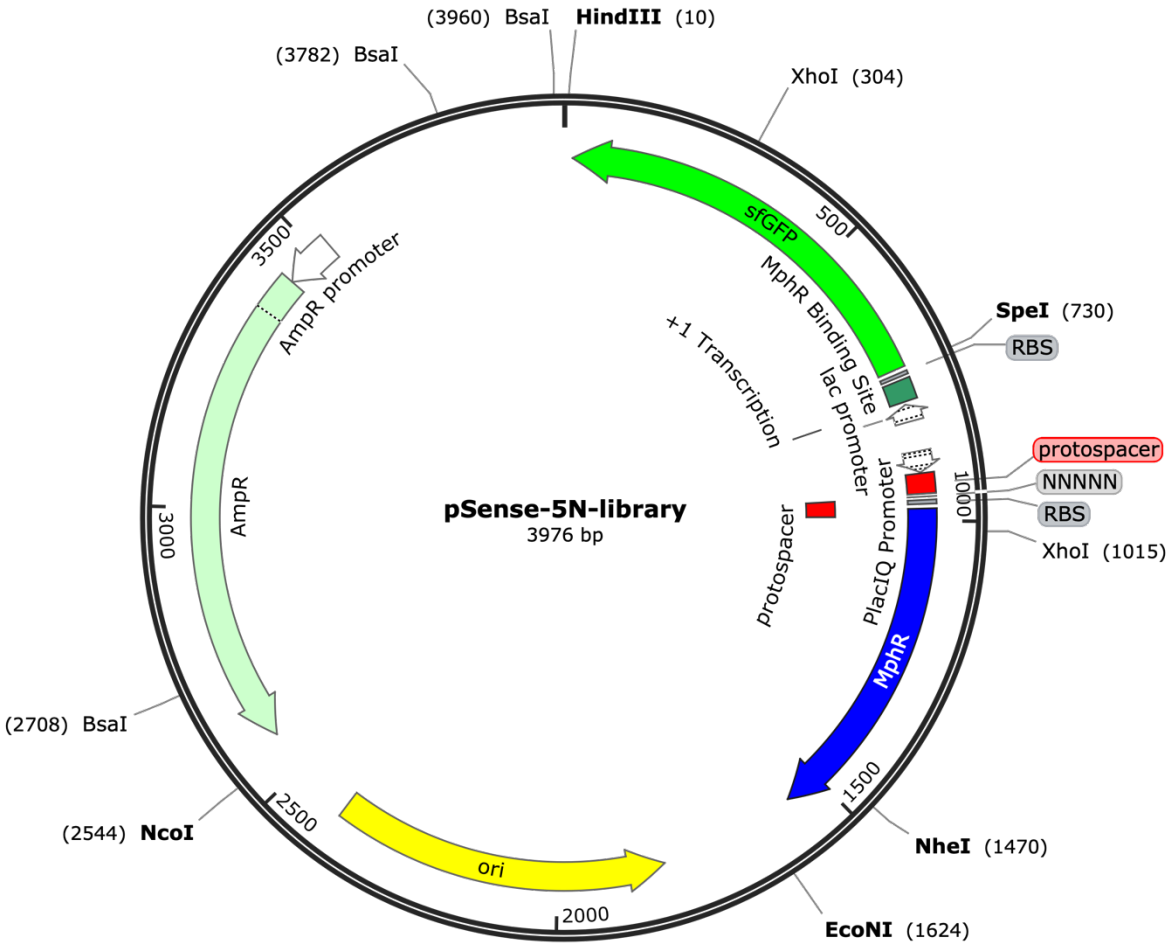


Figure A4. Plasmid map of pSense-5N library

Appendix B

Table B1. Strains and plasmids used in Chapter 3 studies.

Strains/ Plasmids	Relevant characteristic	Sources
Strains		
E. coli		
ET12567 (pUZ8002)	*F- dam-13::Tn9 dcm-6 hsdM hsdR zjj-202::Tn10 recF143 galK2 galT22 ara-14 lacY1 xyl-5 leuB6 thi-1 tonA31 rpsL136 hisG4 tsx-78 mtl-1 glnV44*	[165]
E. cloni 5-alpha	*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Lucigen
S. erythraea		
NRRL2338	wild type stain	ATCC
Plasmids		
pCDFDuet-effector-complex	pCDFDuet-1 derivative; pT7-Csm1-Cas10-Csm3-Csm4-Csm5-Cas6	This work
pZZ04-tgt-ftsZ variable spacer lengths	pZZ04 derivative; *lacZ* was replaced by *ftsZ* targeting spacers of variable lengths	This work
pZZ04-tgt-sfGFP	pZZ04 derivative; *lacZ* was replaced by *sfGFP* targeting spacer	This work
pZZ03	pSG5 ori, ColE1 ori, Apm^R, P*gapdh* transcribing two direct repeats interspaced by BbsI-*lacZ*-BbsI cassette	This work
pZZ03-tgt1765-mCherryE	pZZ03 derivative; SACE_1765 targeting spacer inserted at BbsI sites, homology arms and *mCherry* inserted at XbaI site	This work

Table B2. Primers and dsDNA fragments used in Chapter 3 studies.

Primers (pair)	**Sequences**
construction of pCDFDuet-effector-complex	
primers to delete ancillary genes	AGCGATACGACGAACGACCCTGAGGACACAACCTCCTGATTAACC TAGGCTGCTGCCACC TGCGGCCGCAAGCTTGTCGACCTGCAGGCGCGCCGAGCTCGATCA GCTATGGGTCGGCAG
Primer pairs annealed to make protospacer inserts into pZZ04	
***ftsZ*-36nt**	cggcCGCCGATGACTTTAATCACCGCGTCATTGGTAAGTT acagAACTTACCAATGACGCGGTGATTAAAGTCATCGGCG
***ftsZ*-31nt**	cggcCGCCGATGACTTTAATCACCGCGTCATTGGT acagACCAATGACGCGGTGATTAAAGTCATCGGCG
***ftsZ*-26nt**	cggcCGCCGATGACTTTAATCACCGCGTCA acagTGACGCGGTGATTAAAGTCATCGGCG
***ftsZ*-21nt**	cggcCGCCGATGACTTTAATCACCG acagCGGTGATTAAAGTCATCGGCG
sfGFP	cggcCACCCTCTCCACGGACAGAAAATTTGTGCCCATTAA acagTTAATGGGCACAAATTTTCTGTCCGTGGAGAGGGTG
construction of pZZ03	
gBlocks for repeat-lacZ-repeat cassette	5’- CTTAGTGCGAGTATCTGAAAGGGGATACGCCTGTCACCGGCGCCA AGCGTCGGTCCTCATTGCGGCAAGTCTTCTCAGCCGCTACAGGGC GCGTCCCATTCGCCATTCAGGCTGCGCAACTGTTGGGAAGGGCGA TCGGTGCGGGCCTCTTCGCTATTACGCCAGCTGGCGAAAGGGGGA TGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCCAG TCACGACGTTGTAAAACGACGGCCAGTGAGCGCGCGTAATACGAC TCACTATAGGGCGAATTGGGTACCGGGCCCCCCCTCGAGG-3’ 5’- GCGAATTGGGTACCGGGCCCCCCCTCGAGGTCCTCCAGCTTTTGTT CCCTTTAGTGAGGGTTAATTGCGCGCTTGGCGTAATCATGGTCATA GCTGTTTCCTGTGTGAAATTGTTATCCGCTCACAATTCCACACAAC ATACGAGCCGGAAGCATAAAGTGTAAAGCCTGGGGTGCCTAATGA GTGAGCTAACTCACATTAATTGCGTTGCGCTCACTGCCCGCTTTCC ACCGGTGAAGACATCTGTCACCGGCGCCAAGCGTCGGTCCTCATT GCGGCTTTTTACTCCATCTGGATTTGTTCAGAACG-3’

Table B2 (continued).

	agaacgctcggttgccgccgggcgtttttttatctagaATGGGCCACGTTCGGATGATGGC
primers for up homology arm	**ATGATGGCCATGTTGTCCTCCTCGCCCTTGGAGACCATCGCCCAGTCCAGTCCTTCCAGG**
Primers for mCherry	**TCCCCTCTCGCATCGGCCTGGAAGGACTGGACTGGGCGATGGTCTCCAAGGGCGAGGAGG** **TGCAGGTCCGAGTCGTCACCAGCGGCGGGGTCCATCACTTGTACAGCTCGTCCATGCCGC**
Primers for down homology arm	**GGCCGGCACTCGACCGGCGGCATGGACGAGCTGTACAAGTGATGGACCCCGCCGCTGGTG** **aaaacgccagcaacgcggccttttttacggttcctggcctctagaTCCGGCGGGATGCGGC**
cPCR flanking SACE_1765	**ACCAGCAGGAGCAGATCGAGGC** **ACGGCCGCCCTCGTAGACCT**

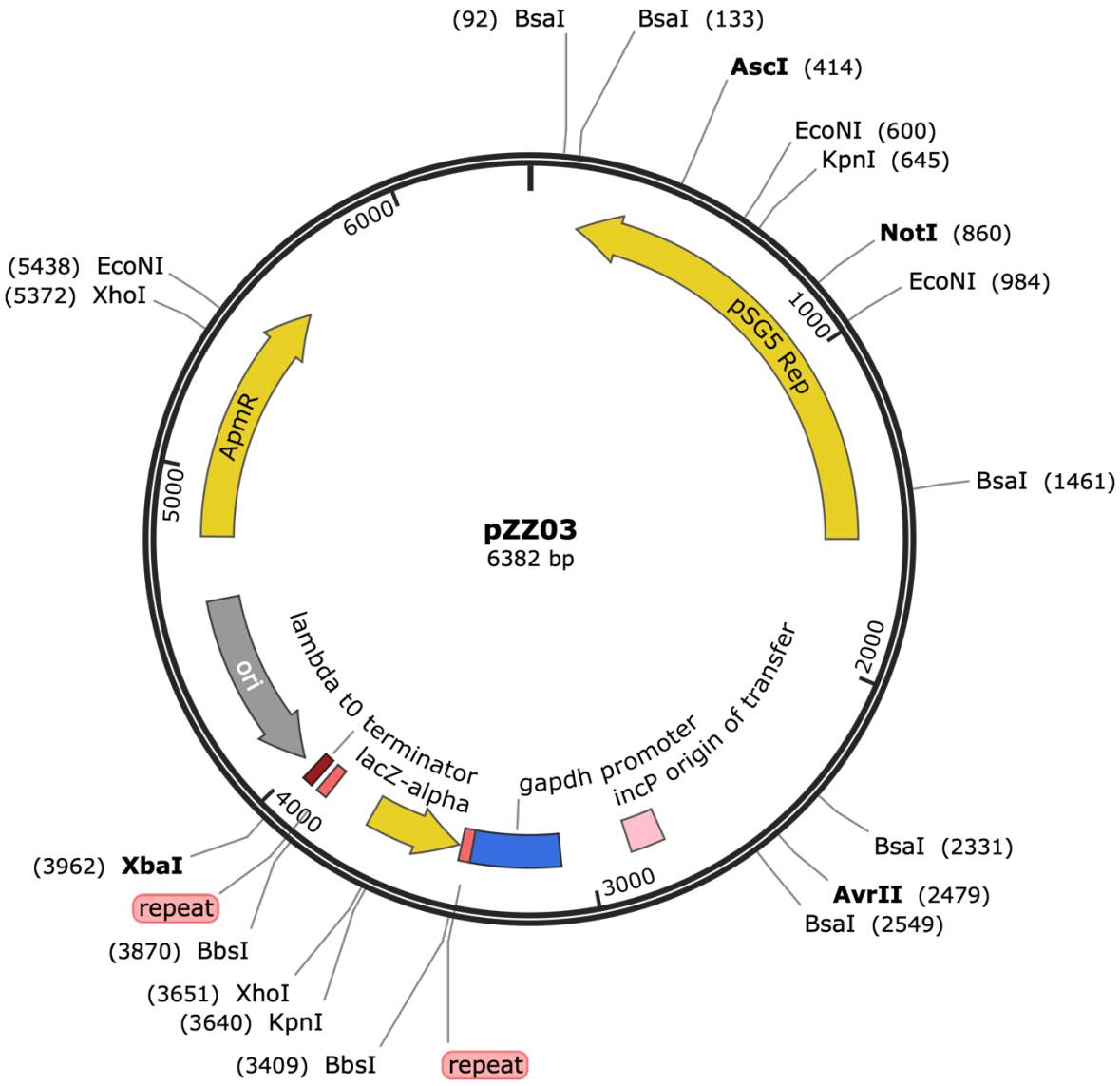


Figure B1. Plasmid map of pZZ03

Appendix C

Table C1. Strains and plasmids used in Chapter 4 studies.

Strains/ Plasmids	Relevant characteristic	Sources
Strains		
E. coli		
LF01	*F− ompT hsdSB (rB, mB) gal dcm (DE3) ΔprpRBCD::T7prom-sfp-T7prom-prpE ΔygfH ΔvioAB Δwzx ΔwecDE* (pDEBS/pTailoring)	[178]
E. cloni 5-alpha	*fhuA2Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*	Lucigen
Saccharomyces cerevisiae		
BY4742	S288C-derivative laboratory strain; *MATα his3Δ1 leu2Δ0 lys2Δ0 ura3Δ0*	[183]
Plasmids		
Cas9-NAT	Cas9 expression plasmid for *S. cerevisiae*	[182]
gRNA-ura-HYB	sgRNA cassette plasmid for *S. cerevisiae*	[182]
gRNA-GRX5	gRNA-ura-HYB derivative; 20nt targeting on GRX5	This work

Table C2. Primers and dsDNA fragments used in Chapter 4 studies.

Primers (pair)	Sequences
construction of gRNA-GRX5 plasmid	
	TACTGTTACCAGTTTGCTTTGATCATTTATCTTTCACTGCGGAGAAG TTTC AAAGCAAACTGGTAACAGTAGTTTTAGAGCTAGAAATAGCAAGTT AAAATAAGGCTAGTC
Primer pairs used to amplify editing template fragments DEBS1-TE	
up-200-arm	GCCTTTACGATCCATATTTATATATGCATGTACATTTG
	CTCGGCGGCTTCTAATCCGTCAAGTCTATGCGGGGAATATTAAATAT ATTCATAATTCTC
GAL1 promoter	GAGAATTATGAATATATTTAATATTCCCCGCATAGACTTGACGGATTA GAAGCCGCCGAG
	gacagtttggacagatcagccatCTCCTTGACGTTAAAGTATAGAGGTATATTAA CAATT
DEBS1-1st half	AATTGTTAATATACCTCTATACTTTAACGTCAAGGAGatggctgatctgtccaa actgtc
	aactttagcacgagaagcac
DEBS1-2nd half	ccgaccgtgaatctgttc
	gattacgcaggcctacagcggttaaggaatcgaagcccagttctttgaacggggtggtag
DEBS3 TE	cgaaagctgttcgtgctaccaccccgttcaaagaactgggcttcgattccttaaccgctg
	GGGCGTGAATGTAAGCGTGACATAACTAATTACATGAttaagagttaccacc acccagcc
CYC1 term	gcttggctgggtggtggtaactcttaaTCATGTAATTAGTTATGTCACGCTTACATTC AC
	TTGTGTGTTCATTATATCCCTATCACCTTCAAAAGGAGCAAATTAAA GCCTTCGAGCGTC
down-200-arm	TTGGGACGCTCGAAGGCTTTAATTTGCTCCTTTTGAAGGTGATAGG GATATAATGAACAC
	TTTTACATTTGATTTTACTTTCTATTATTATTTTCCGAAAATTTCACAC
construction of pDEBS RASH	
gBlocks sgRNA	5’- AAGCTAATACGACTCACTATAGGGAGTGGGATGCATAGTCAACGTT TTAGAGCTAGAAATAGCAAGTTAAAATAA-3’