UNIVERSITY OF CALIFORNIA SAN DIEGO

Exploring the Role of SeAgo in Host Defense and Gene Transfer Processes in *Synechococcus elongatus* PCC 7942

A thesis submitted in partial satisfaction of the
requirements for the degree
Master of Science
in

Biology

by

Tami Sofia Gilderman

Committee in charge:
Professor James Golden, Chair
Professor Susan Golden, Co-chair
Professor Stuart Brody

2020

The thesis of Tami Sofia Gilderman is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

Co-Chair

_____

Chair

University of California San Diego

2020

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest and most sincere appreciation to all members of my committee and the Golden Lab for providing me an exceptionally cordial place to complete my master's degree. I am particularly grateful for the assistance given by Jim and Arnaud; without the unwavering encouragement, patience, and guidance you both gave me throughout the duration of this project, the completion of my thesis would not have been possible. Thank you, Arnaud, for your invaluable insights, your unparalleled knowledge and mentorship throughout my time at the Golden Lab cannot be overestimated. Thank you, Jim, for your useful and constructive criticisms of this project, as well as, your considerate accommodations during the unprecedented COVID-19 pandemic.

Finally, I wish to thank all of my family for their unconditional love and emboldening support, and my extraordinary fiancé for always pushing me to strive higher. This tremendous accomplishment would not have been achieved without all of you.

ABSTRACT OF THE THESIS

Exploring the Role of SeAgo in Host Defense and Gene Transfer Processes in *Synechococcus elongatus* PCC 7942

by

Tami Sofia Gilderman
Master of Science in Biology
University of California San Diego, 2020
Professor James Golden, Chair
Professor Susan Golden, Co-chair

Widely distributed and highly conserved among all three domains of life, argonaute proteins generally function in binding and utilizing short nucleic acid guides, and some additionally perform guide-mediated cleavage of complementary DNA or RNA targets. Due to their functional analogy to CRISPR-Cas systems, catalytically active prokaryotic argonautes (pAgos) have been primarily highlighted in the field of biotechnology for applications in seamless genetic engineering, leading to the emergence of the first pAgo-based genetic modification technology. However, characterization of the *in vivo* role of pAgos is in its early stage and only a few catalytically active pAgos have been broadly analyzed such as the *T. thermophilus* TtAgo. Through bioinformatics, gene transfer assays, and RB-TnSeq, we identify and investigate a closely related homolog of TtAgo, SeAgo from the genetically tractable model cyanobacterium *Synechococcus elongatus* PCC 7942 (*S. elongatus)*.

In this study, we explored the effect of SeAgo on natural transformation and conjugation in *S. elongatus* as well as the potential relationship between SeAgo and the predicted associated nuclease SeCas4. For the purpose of this study, we designed improved CRISPR-Cpf1 genetic

engineering tools by constructing a CRISPR-CpfI module adapted for the CYANO-VECTOR platform in addition to five CRISPR-Cpf1 plasmids carrying an improved RSF1010 backbone which significantly reduced the difficulty of the subsequent cloning. Nevertheless, even with an improved RSF1010 backbone, the transfer of RSF1010 based plasmids in *S. elongatus* repeatedly fails or results in small unstable colonies. Here we show that loss of function of SeAgo drastically increases the efficiency of transfer of RSF1010-based plasmids. With new RSF1010-based CRISPR tools continuing to emerge, overcoming the challenges of utilizing these technologies in the model cyanobacterium *S. elong atus* could significantly advance our understanding of these photosynthetic prokaryotes and their defense mechanisms.

Chapter 1. Introduction

1.1 Cyanobacteria

1.1.1 General Characteristics of Cyanobacteria

As the sole prokaryotic autotrophs with the capacity to generate their own organic compounds through oxygenic photosynthesis, cyanobacteria are a diverse phylum of over 2,500 identified species from over 300 different genera [1,2]. The high levels of diversity of this phylum is demonstrated by the broad spectrum of reported physiologies, morphologies, habitats, and genomes [2,3]. Cyanobacteria perform a variety of roles in the global nitrogen and carbon cycles and have adapted to inhabit diverse ecosystems such as terrestrial, freshwater, and oceans [4]. One of the key features of cyanobacteria is their distinguishing ability to generate oxygen, leading to the oxygenation of the Earth's atmosphere [3].

Cyanobacteria have evolved several adaptations to perform specialized functions. Various filamentous cyanobacterial species have developed the ability to fix atmospheric nitrogen, as well as strategies for spatial and temporal separation of nitrogen fixation from oxygenic photosynthesis [5–7]. Another key feature of cyanobacteria is their role as microbial production factories [5]. Though many cyanobacterial species produce toxins found in harmful algal blooms (HABs), others generate natural products that can potentially have beneficial properties such as antibiotics, anti-inflammatory, and cancer treatment drugs [5,6]. In addition to their ability to fix carbon and nitrogen, generate oxygen, and produce a variety of secondary metabolites, cyanobacteria are studied for their wide range of applications made possible by their genetic tractability and other significant discoveries such as the circadian clock [8,9].

1.1.2 *Synechococcus elongatus* PCC 7942

A highly characterized cyanobacterial model organism is the genetically tractable, obligate phototroph, circadian clock-dependent cyanobacterium currently known as *Synechococcus elongatus* PCC 7942 (hereafter, *S. elongatus*) [10]. The unicellular cyanobacterium *S. elongatus* was first isolated from natural California aquatic habitats and was originally named *Anacystis nidulans* R-2 [10,11] (Figure 1.1). In the late 1970s, the newly discovered *Anacystis nidulans* R-2 was shown to be transformable by Galina Grigorieva in the Shestakov lab which revealed its potential for gene cloning and mutagenesis experiments[10]. Preliminary examination of this

strain showed the presence of two endogenous plasmids which differ greatly in size: pANS (8.0 kb) and pANL (48.7 kb) [12]. *S. elongatus* serves as a model organism for the study of photosynthesis in comparison to higher plants, signal-transduction and gene expression pathways, bioproduction of valuable compounds, and other biological investigations [6,9]. Furthermore, since the discovery of the circadian clock, *S. elongatus* was established as the premier prokaryotic model organism for the study of circadian rhythms and light-regulated gene expression [9].

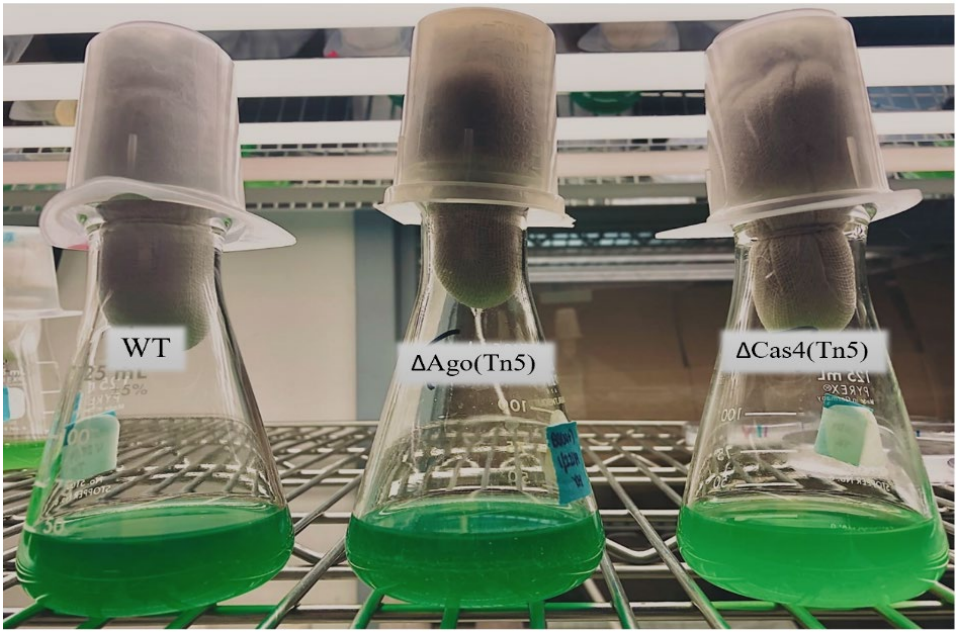


Figure 1.1 Standard cultures of Wildtype and Mutant *S. elongatus PCC 7942* strains at different growth stages as cultured in the lab. Strain and culture conditions are described in

1.1.3 Horizontal Gene Transfer in Cyanobacteria

The exchange or transfer of genetic material between organisms can occur through both sexual and asexual reproduction. While multicellular eukaryotic organisms acquire genetic material through sexual reproduction, also known as vertical gene transfer, prokaryotes obtain new genes through mechanisms of horizontal gene transfer (HGT) [13]. HGT results in the acquisition of foreign genetic material, potentially encoding beneficial genes, through the processes of transformation, conjugation, and transduction [14]. HGT events have influenced

evolution by generating significant genetic diversity across the three domains of life [13,14]. The ability to take up and integrate exogenous DNA into the genome provides an advantage in improving environmental adaptations, general resilience, and survival potential [15].

In addition to intraphylum and interphylum HGT, transfer of entire cells is possible and may result in endosymbiotic relationships [14–17]. In fact, photosynthetic eukaryotes originally acquired their photosynthetic abilities as a result of interphylum HGT leading to the symbiotic relationship between two cells of ancient cyanobacteria and unicellular eukaryotes [14]. About 50% of cyanobacterial extended gene families are presumed to have a history of HGT through either intraphylum, interphylum, or both gene transfer events [14].

The ability to take up exogenous DNA into the cell, also known as **natural transformation**, is a rare feature among living organisms. The first report of natural transformation in cyanobacteria was in the early 1970s in *S. elongatus* [10]. The simplicity of the original transformation procedure has maintained the process of transformation as the primary method of gene transfer in cyanobacteria, resulting in a wide range of applications in identifying, isolating, and characterizing cyanobacterial genes [9,18]. The extensive characterization of natural transformation in cyanobacteria was only recently detailed in the work done by Taton et al., 2020. Early studies of natural transformation described several protocol modifications to optimize the efficiency of the transformation process[9,18]. Increased transformation efficiency was observed by inhibition of photosynthesis and increased incubation period of the cells and DNA (up to 18 hours) [9,18]. The use of iron-deficient media or chemical inhibitors may improve transformation efficiency by inhibition of photosynthesis. However, the simplest means of significantly increasing the efficiency of the transformation process is the absence of light [9]. The impact of dark incubation on transformation efficiency in *S. elongatus* is related to the function of the internal clock which synchronizes physiological processes with the environmental changes between light and dark [9,19]. The circadian rhythm controls gene expression patterns, regulating photosynthetic metabolism and growth during daytime when the necessary light is available [9]. The circadian expression of genes that are essential for natural transformation is synchronized to the onset of night (darkness) thus providing a potential explanation for the peak in competence that is observed at dusk [9].

Although transformation is the primary method of HGT, the transfer of genetic material may also occur by direct cell-cell contact through the process of **bacterial conjugation**. Conjugation involves the physical contact of two cells, a donor and a recipient cell. The donor cell attaches to the recipient cell through conjugal pili, forming the mating bridge through which the conjugal DNA is transferred from the donor to the recipient cell. In addition, conjugation allows for gene transfer in a wide range of cyanobacteria because it does not involve the machinery utilized in natural transformation and thus does not require a naturally competent strain to efficiently produce transconjugants [20].

Transduction is HGT in which host DNA is transferred by bacteriophage [21]. Bacteriophages modulate bacterial evolution, geographical distribution, and population dynamics [22]. Cyanophages can be categorized into three characterized families of double-stranded DNA viruses based on variations in the tail's length and contractility[22]. Myoviridae possess contractile tails while Styloviridae and Podoviridae display long non-contractile or short tails, respectively[22]. The broadly distributed cyanophages undergo frequent mutational changes and are capable of attacking a wide variety of cyanobacteria [21]. Several observations of cyanophage fitness adaptations suggest specific fitness to cyanobacterial physiology [21]. For instance, the infection cycle of cyanophages occurs slower than other bacteriophages, likely correlated to the relatively slow growth rate of cyanobacteria[21–23]. Given that cyanobacteria symbolize a prokaryotic model organism for the study of plant photosynthesis, the discovery of cyanophages has advanced our understanding of photoautotrophic metabolism during viral infection which could previously only be investigated with tissue samples from higher plants [22,24,25].

Another observation of cyanophage adaptations to cyanobacterial physiology is the synchronization of the cyanophage infection cycle with the circadian rhythm of cyanobacteria [24,25]. While the cyanophage diurnal rhythm patterns display a degree of synchronization to the circadian rhythm of the cyanobacterial host, evidence suggests that the rhythmic transcript abundance and absorption patterns of the cyanophage are linked to the host's photosynthetic activity and the availability of light rather than the internal clock itself [25]. Laboratory cyanophage cultures have also been shown to replicate dependent on the amount and duration of light exposure where production of the cyanophage progeny adopts a diel pattern under

natural light conditions [24,25]. The exact mechanisms driving the cyanophage diurnal rhythm patterns have yet to be determined.

1.2 Prokaryotic Immune Systems

1.2.1 Common Prokaryotic Innate and Acquired Immunity Mechanisms

Whilst phages have evolved adaptations to increase their infection fitness, bacteria have not been helpless in defending themselves against infection or combating harmful HGT events. Over time, bacteria have developed immunity mechanisms and strategies for protection and self-preservation [26] (Figure 1.2). Despite the divergence in the specific mechanisms, bacterial defense strategies are allied in their goal of protecting the host from infection and have been shown to provide resistance throughout the various stages of the phage infection cycle [26,27].

The inhibition of phage absorption in the first stage of the phage infection cycle via cell surface modification is a consequence of mutation and selection rather than an innate defense strategy [26,28]. Even so, downregulation or complete loss of function of specific host surface receptors can prevent phage absorption through the physical blockage of the phage entry point [29] (Figure 1.2). The second stage of the phage infection cycle is the injection of the phage genome into the host cell, and can be hindered by superinfection exclusion (Sie) mechanisms where the antiviral resistance is encoded in the phage itself [26] (Figure 1.2). Because Sie can inhibit the injection of the phage DNA or prevent intracellular genome replication, a secondary phage infection is blocked. This strategy allows the primary infecting phage to survive and be maintained as a prophage [26–29]. In the last stage of the phage infection cycle, the phage genome is replicated within the host cells. Some bacteria prevent phage genome replication by the process of abortive infection (Abi) which triggers the process of cell suicide [27,30–32] (Figure 1.2). Other bacteria have developed complex systems that are capable of cleaving injected phage DNA and thus also inhibit phage replication [26,31,32]. For example, the diverse and widespread restriction modification (RM) systems can cleave foreign DNA at specific DNA sequences (restriction sites) [32]. To distinguish foreign DNA from the host genome, RM systems utilize a methyltransferase that methylates the host DNA restriction sites thus preventing the restriction endonuclease component of the RM system from recognizing and cleaving chromosomal restriction sites [32,33] (Figure 1.2).

Apart from Sie, the defense strategies described so far represent innate immunity mechanisms as they do not display invader-specificity [31,32,34]. However, the groundbreaking discovery of the CRISPR-Cas (clustered regularly interspaced short palindromic repeat–CRISPR-associated gene) system as a prokaryotic adaptive immune system has revolutionized biology [26,32,33] (Figure 1.2). Since its recent discovery, the CRISPR-Cas system has been studied at length in order to harness its capacity for precise, rapid, and efficient marker-less gene editing [26,32,34]. As an adaptive immune system, CRISPR provides heritable invader-specific defense through short RNA transcripts that guide CRISPR-associated (Cas) nucleases to bind and cleave the complementary target invader DNA[32]. The CRISPR-Cas system consists of a series of highly conserved short repeating DNA sequences separated by variable spacer sequences (originating from foreign DNA) that can be transcribed and utilized as guides by the Cas proteins encoded nearby [33,35].

Another prokaryotic immune system that can be exploited in the field of biotechnology for its ability to perform efficient and precise seamless gene editing involves the argonaute proteins. Unlike the CRISPR system, argonaute proteins are present in both prokaryotic and eukaryotic organisms and are capable of host defense through both DNA and RNA interference pathways [36]. Along with RM and CRISPR systems, argonautes are present in some species of cyanobacteria [37–40]. The research presented in this thesis focused on the active argonaute in *S. elongatus* that has been recently named SeAgo [41]. Argonaute proteins and SeAgo are discussed in further detail below.

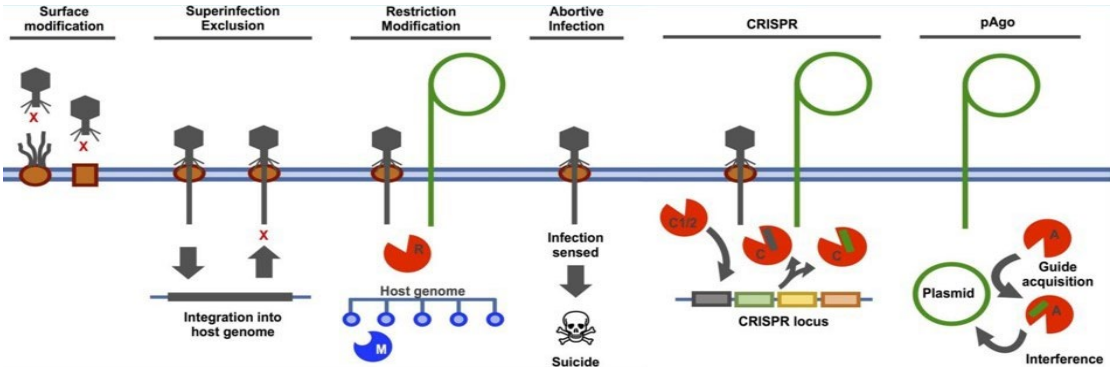


Figure 1.2 Various Prokaryotic Immunity Mechanisms. Figure source: Houte et al., 2016. See section 1.2 for a more detailed description of each of the defense strategies depicted

1.3 Argonautes

1.3.1 General Characteristics of Argonaute proteins

Argonaute (Ago) proteins, part of the PIWI protein superfamily, are characterized by the presence of a PAZ and a PIWI domain. Broadly, Ago proteins function in binding and utilizing small nucleic acids as guides to execute sequence-dependent target recognition and if possible proceed to guide-mediated cleavage of the complementary target strands [41–43] (Figure 1.3). By binding to short nucleic acid guides, Ago proteins are capable of mediating the repression of complementary targets by degradation or inhibition of translation [44]. Eukaryotic Ago (eAgo) proteins are typically involved in RNA interference, serving as highly specialized small-RNA-binding proteins in various RNA-silencing pathways [42] (Figure 1.3B).

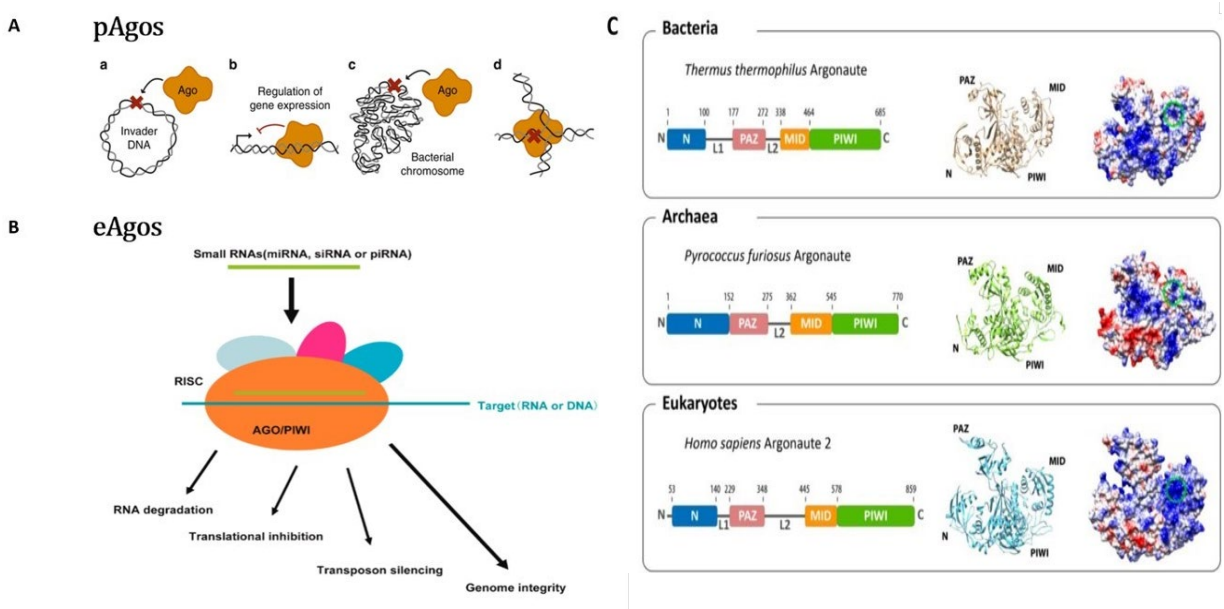


Figure 1.3 Argonautes across the three domains of life. Figure sources: (A) Lisitskaya et al., 2018; (B) Zhai et al., 2016; (C) Willkomm et al., 2015. Predicted functions of pAgos (A) and eAgos (B). (C) Conserved domains, alignment, and structure of argonaute proteins in bacteria, archea, and eukarya.

In contrast, prokaryotic Ago (pAgo) proteins have a significantly greater affinity for binding DNA guides and DNA targets. The role of pAgos in protecting prokaryotic hosts from invading plasmids and transposons involves DNA-guided DNA interference mechanisms [45] (Figure 1.3A). Some evidence also implicates pAgo proteins in suicidal systems, triggering the cleavage

of host DNA [44–46] (Figure 1.2A, 1.3A). The ability of some Ago proteins to perform guide-mediated sequence-specific target cleavage is conceptually analogous to the CRISPR-Cas system, therefore, catalytically active pAgos represent an alternative to CRISPR-based technology [37,40]. Prokaryotic argonaute studies in the field of biotechnology recently led to the first commercialized Ago gene editing kit sold by New England BioLabs (NEB) [37,40].

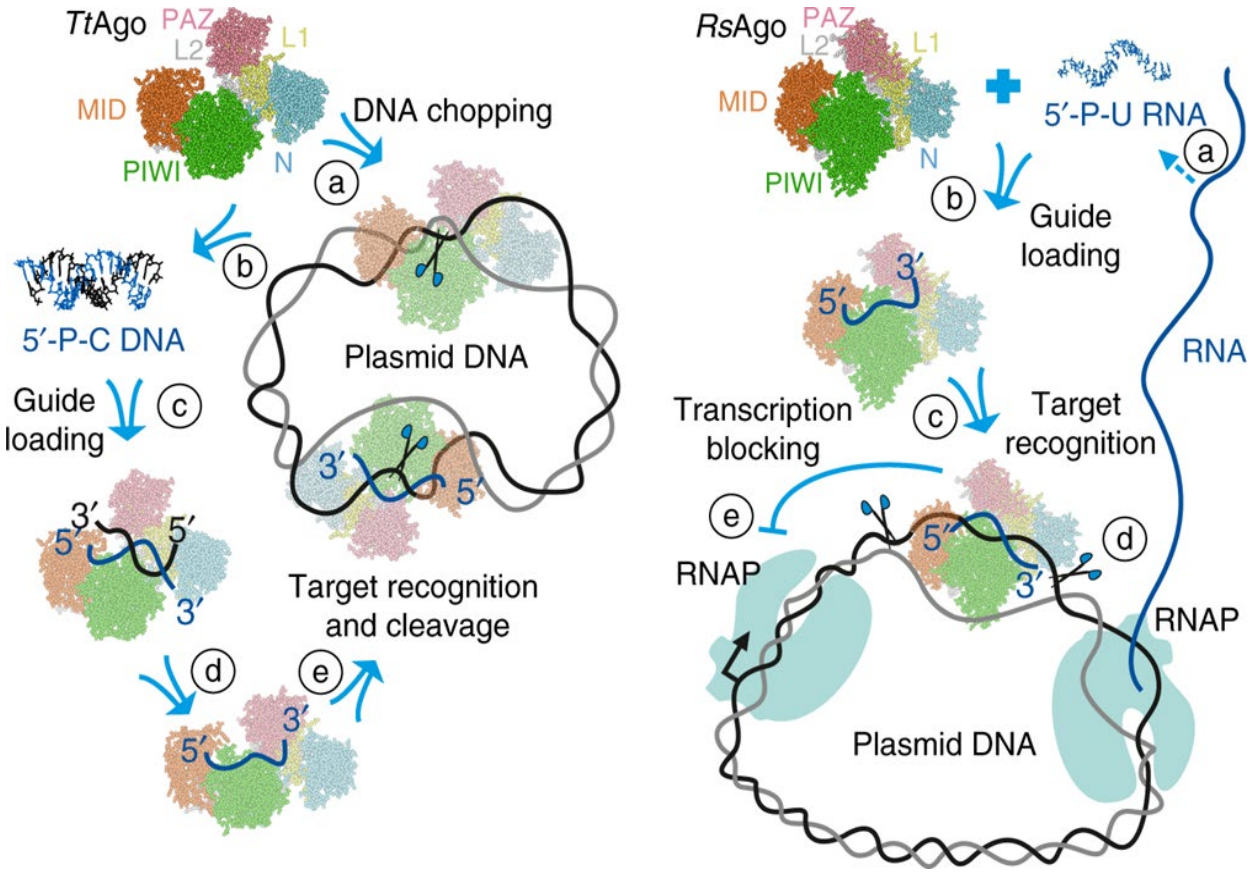


Figure 1.4 Varying functions of prokaryotic argonaute proteins. Figure source: Lisitskaya et al., 2018. Proposed mechanisms of DNA-guided DNA interference by TtAgo (left) and RNA-guided DNA interference by RsAgo (right).

1.3.2 Argonaute Evolution

The Ago protein family was identified in the study of mutant plants (*Arabidopsis thaliana*) [40]. While eAgo proteins mainly function in RNA interference pathways and pAgo proteins are implicated in processes of DNA interference, overall, Ago proteins are conserved between

organisms and are found in all three domains of life (Figure 1.3C) [40]. Ago proteins are present in about 65% of sequenced eukaryotic genomes, 32% of archaeal genomes, and in 9% of bacterial genomes (17 of 37 prokaryotic phyla) with many organisms expressing more than one Ago protein [40]. The phylogenetic analysis of *ago* genes demonstrates a patchy distribution, likely due to the history of HGT events [37,38]. The evolution of Ago proteins began in prokaryotes, most likely through a fusion of a PIWI-like (P element-induced wimpy testis) RNase H domain with a MID-like nucleic acid-binding domain, leading to the formation of the first guide-dependent pAgo [43,47]. Overall, only a small percentage of pAgo proteins are predicted to be catalytically active, forming a monophyletic group that often encodes *ago* genes alongside predicted helicases [37,38].

1.3.3 Structure, Composition, and Associated Proteins of Prokaryotic Argonautes

The bilobal structure of pAgos is comprised of the PIWI and MID domains within the PIWI lobe, connected to the PAZ (Piwi Argonaut and Zwille) lobe which contains the N- and PAZ domains [37] (Figure 1.3C). Based on the domain alignment and degree of conservation, pAgo proteins can be categorized into three clades: short, long-A, and long-B pAgo proteins [37,38,40] (Table 1). Short pAgos contain the MID and PIWI domains while lacking both the N- and PAZ domains [37,38]. About 1% are additionally missing the MID domain and are therefore presumed to have lost the ability to interact with short nucleic acid guides [37,40]. Most pAgos belong to the long-A and long-B clades, defined by the additional presence of either the PAZ domain, the N-domain, or both[41,43,47] (Table 1). Aside from the high degree of conservation observed in the PIWI domain of Ago proteins, the overall relationship between the three Ago clades remains unclear [48].

Only about 18% of pAgos are catalytically active which requires the presence of a conserved catalytic tetrad of amino acid residues within the PIWI domain [48]. In the absence of the catalytic tetrad, pAgo proteins rely on associated nucleases and other DNA binding proteins to complete guide-mediated target cleavage [37,38]. The majority of both catalytically active and inactive pAgos associate with helicases, DNA binding proteins, or cellular endonucleases [40]. The associated nucleases have been identified as members of the SIR2, TIR, PLD, and Cas4 protein families [44,47,49]. However, the specific roles, interactions, and relationship between pAgo proteins and their associated nucleases remains poorly understood. Below is a brief description

of each of the four domains; the general domain functions, categorization, and variations are summarized in table 1.

The **PIWI domain** is present in all Ago proteins indicating that it is essential for Ago function [41,47] (Figure 1.3C). Structurally, the PIWI domain adopts an RNase H-like fold and is the domain responsible for the nucleolytic function in all catalytically active Ago proteins [43,44,47]. Cleavage of the target nucleic acids is conditional upon the presence of a conserved tetrad of three catalytic aspartic acid residues constituting the active site [44,47]. All catalytically active pAgo proteins belong to the long-A clade, including all cyanobacterial argonautes, while the long-B and short clades are typically catalytically inactive pAgo proteins [37,38,40]. All short pAgos lack the catalytic tetrad, resulting in the loss of catalytic activity, and have therefore been categorized as PIWI* variants [37,46,50] (Table 1). Similarly, the catalytically inactive long-B pAgo proteins also lack the conserved active site, however, only short pAgo proteins are classified as PIWI* variants [46].

Alongside the PIWI domain, the PIWI lobe additionally features the **MID domain** in 99% of all identified Ago proteins [46]. The highly conserved MID domain is required for pAgo function as it provides all Ago proteins with the ability to bind short nucleic acid guides [37,46,50]. The structure of the MID domain consists of a basic nucleotide-binding pocket where a conserved group of amino acid residues interact with the 5'-end of the nucleic acid guide [46,50]. The basic pocket adopts a Rossman-like fold that allows Ago proteins to recognize specific 5'-end bases with a nucleotide specificity loop [44,47]. Overall, the MID domain facilitates the process of guide-mediated target cleavage by attaching the guide in a helical formation which promotes target binding [47].

The **N- domain** is the least understood domain, although it is presumed to be involved in target cleavage and the final dissociation of the cleaved strands in catalytically active Ago proteins [37]. The N- domain is reported to function as an active wedge that disrupts the guide-target base-pairing at the 3'-end of the guide. Other reports suggest that the N- domain may also work as a passive wedge by instead causing the disruption downstream of the guide [46,50].

Similar to the N- domain, the **PAZ domain** is absent in more than half of pAgo proteins (60%) and displays sequence variations between organisms [37]. The PAZ domain is not

necessary for the basic functions of Ago proteins and an incomplete PAZ domain is often sufficient for maintaining the capacity to bind the 3'-end of the guide[38,51]. Despite the low sequence conservation, the structural folding of PAZ domains is consistent in the formation of a hydrophobic nucleotide-binding pocket for the 3'-end of the guide molecule [37,40]. The PAZ domain facilitates guide-mediated target recognition which involves the release of the 3'-end of the guide to complete the guide-target base-pairing[46]. Lastly, 18% of the short pAgo clade contains an APAZ (analog of PAZ) domain which does not share sequence similarities with PAZ and is characterized by its consistent presence either within or around short pAgo proteins[46].

Table 1: Argonaute domains and variations between the three clades.

	General Function	Long-A	Long-B	Short
PIWI	Target cleavage	Catalytically active	Inactive PIWI*	Inactive PIWI*
PAZ	lipophilic pocket to anchor guide 3'-end.	Normal size	Shortened PAZ*	absent
N-	disruption of guide-target base pairs	Mostly present	Mostly absent	absent
MID	positions guide in a helical formation for target binding.	present	present	subset of canonical MID motifs (HK/RK)

1.3.4 *Thermus thermophilus* Argonaute (TtAgo)

The most structurally similar homolog of SeAgo from *S. elongatus* is the *T. thermophilus* argonaute protein, TtAgo [46,52]. As a member of the long-A pAgo clade, TtAgo possesses the characteristic bilobal structure described above [46,52]. Based on the crystal structure and experimental evidence, TtAgo contains the conserved catalytic tetrad in the PIWI domain and thus functions in the utilization of either DNA or RNA guides for guide-mediated target cleavage of invading foreign DNA [52]. With the capacity for endonucleolytic cleavage, TtAgo may serve as an immune system to provide host defense through DNA interference. Furthermore, *in vivo* studies suggest that TtAgo is capable of decreasing transformation

efficiency and is likely involved in maintenance of intracellular plasmid concentrations [52]. The general mechanism of action of TtAgo begins with a guide-free Ago (apo-Ago) that performs nonspecific DNA degradation (chopping) [46,52]. Apo-TtAgo generates small interfering DNAs (siDNAs) by preferentially binding and chopping unstable and partially unwound dsDNA[52] (Figure 1.4). The short DNA fragments, products of the chopping, are selectively loaded onto the Ago protein and are then utilized as guides for subsequent guide-mediated target cleavage which occurs with significantly greater efficiency than the guide-independent chopping[52].

1.4 Research Objectives

The field of DNA interference and prokaryotic argonautes has remained largely unexplored. Until earlier this year, there have been no publications on the catalytically active argonaute in *S. elongatus* [36]. Argonaute research in the field of biotechnology primarily highlights the applications of prokaryotic argonautes in efficient and precise marker-less genetic engineering. Recent advances in the field have led to the release of the first purified TtAgo kit for programmable guide-mediated DNA cleavage in a manner that is conceptually analogous to the CRISPR-Cas system.

The presence of a catalytically active argonaute in *S. elongatus* was detected through analysis of natural transformation fitness in a *S. elongatus* transposon mutant library (Figure 1.5). Considering the absence of published data and lacking genome annotations on SeAgo at the time, the identification of two distinct uncharacterized loci on a mutant fitness plot stimulated further investigation. This initial mutant fitness analysis, which measures the ability of a cell to thrive under specific selective conditions, demonstrated the potential for increased natural transformation efficiency upon loss of function in either of the two genes, Synpcc7942_1534 (SeAgo) and Synpcc7942_2273 (SeCas4). This study was aimed to investigate these uncharacterized loci to explore the effect of SeAgo and SeCas4 on natural transformation and conjugation in *S. elongatus*. Moreover, this study intended to explore the relationship or interactions between SeAgo and SeCas4.

Lastly, the emergence of new RSF1010-based genetic engineering tools, combined with the expanding application of marker-less gene editing in model cyanobacteria such as *S. elongatus*, illustrates the need for characterization of the interaction between SeAgo and RSF1010 plasmids.

We examined the role of SeAgo in the highly inefficient cloning process of RSF1010 and its derivative plasmids in *S. elongatus*. Unfortunately, due to the unexpected COVID-19 pandemic and the resulting mandatory shut-down, this project was forced to halt without the finalized results. Thus, portions of this study will be detailed as theoretical experimental plans that we hope to accomplish in the near future.

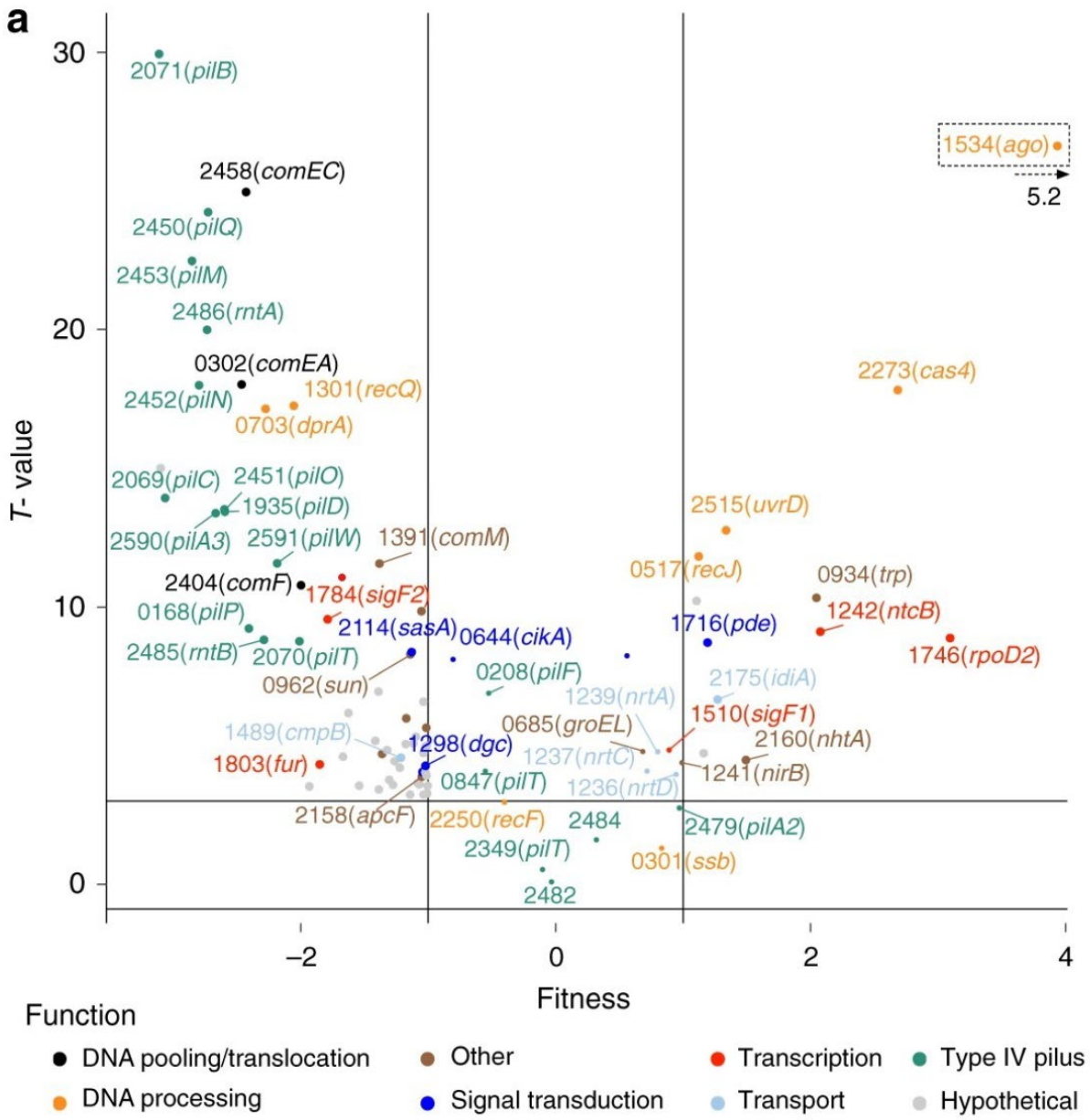


Figure 1.5 Natural transformation fitness analysis of *S. elongatus* transposon mutant library. Figure source: Taton et al., 2020. Aside from a few selected loci representing specific functional categories, only those that had demonstrated a significant fitness effect upon loss of function are labeled on this plot with the *S. elongatus* locus tag (Synpcc7942_) numbers and gene names in parentheses. High fitness and T-value characterize loci whose loss of function contribute to improved transformation while loci that are highly important for natural transformation are characterized by a low fitness with a high T-value. Fitness values were calculated from a representative total of 82,495 distinct mutants under selection and 90,872 mutants under control conditions.

Chapter 2. Materials and Methods

2.1 Strains and Culture Conditions

A comprehensive list of all bacterial strains utilized in this study is presented in Table 2.1. All assays and strain mutations were performed in *S. elongatus PCC7942*. All liquid cultures were grown in 250mL flasks (PYREX) in 100mL BG-11 medium maintained at 30°C and shaking at 150 rpm (Thermo Fischer MaxQ 2000 Orbital Shaker). All plated cells were grown on agarose plates (40 mL BG-11 media solidified with 1.5% agar) at 30°C. Both plated and liquid cultures were grown under continuous light with intensity ranging from 25-350 µmol photons m^{-2} s^{-1}. For engineered *S. elongatus* strains, appropriate antibiotics were included in the media or plate when necessary.

All *E. coli* strains were grown on either LB agar plates or in 5-mL LB liquid medium which includes the proper antibiotics. The cultures were incubated at 37°C with liquid cultures incubated overnight on a roller drum.

2.2 Genomic and Plasmid DNA Extraction Protocols

Genomic DNA extractions from genetically engineered *S. elongatus* was performed utilizing a chloroform:isoamyl alcohol extraction with cetyltrimethylammonium bromide (CTAB). Starting with a 10 mL culture, the cells were centrifuged and resuspended in 1000 µl of BG-11. The cells were then centrifuged again for 3-minute centrifugation at 16,000g and the resulting pellet was resuspended in 340 µl STET buffer (8% sucrose, 5% Triton X-100, 50 mM EDTA [pH 8.0]). Next, Lysozyme was added along with RNase A to result in a final concentration of 2 mg/mL and 0.1 mg/mL, respectively. Following a 45 minute incubation at 37°C, 2 µl of a 10 mg/mL proteinase K stock solution and 20 µl of 20% sarkosyl were added to the cell suspension. The new mixture was vortexed for 20 seconds and incubated at 55°C for 30 minutes before adding 57µl of 5M NaCl and 45µl of 10% CTAB in 0.7 M NaCl. The mixture was then vortexed again and incubated at 65°C for 10 minutes. Finally, each sample was extracted twice with a 500µl solution of 24:1 chloroform:isoamyl alcohol by vortexing the mixture for 10 seconds, centrifuging at 16,000 g for 10 minutes, and extracting the upper aqueous layer containing the DNA. After two subsequent extractions, the DNA was precipitated with 2 volumes of 100%

ethanol (16,000g for 10 minutes) and washed with 1 mL of 70% ethanol (16,000g for 5 minutes) before evaporation and resuspension in 50 µl sterile water.

Plasmid DNA extraction and purification from all *E. coli* cells was performed on 5-mL overnight cultures employing either ethanol precipitation or the Qiagen QIAprep Spin Miniprep Kit, according to the manufacturer's instructions.

Table 2.1 List of strains and plasmids used in this research

Strain or Plasmid	Description / Genotype	Antibiotic Resistance	Source/Creator
	Strains		
AMC06	*Wild-type strain of S. elongatus PCC 7942*		*Golden Lab Collection*
AMC2302	*S. elongatus PCC 7942 cured of pANS*		*Chen et al.*
DH5α	*Chemically competent E. coli cloning host*		*Golden Lab Collection*
DH10β	*Electrocompetent E. coli cloning host*		*Golden Lab Collection*
AM1359	*Electrocompetent conjugal donor E. coli strain containing pRL623 and pRL443*	*Ap, Cm, Tc*	*Golden Lab Collection*
AMT007	*ago(Tn5)-(NS2-P_{T7}Ago + NS1-T7RNAP)*	*Gm, Sp+Sm, Km*	*This Study*
AMT008	*cas4(Tn5)-(NS2-P_{T7}Cas4 + NS1-T7RNAP)*	*Gm, Sp+Sm, Km*	*This Study*
AMT009	*ago(Tn5)+cas4(Mu), (NS2-P_{T7}Cas4-Ago + NS1-T7RNAP)*	*Gm, Sp+Sm, Km, Cm*	*This Study*
AMT013	*AMC06, ΔAgo*		*This Study*
AMT014	*AMT013, (NS2::PT7-ago + NS1::PconII*-RSWB-T7RNAP)*	*Gm, Sp, Sm*	*This Study*
AMT015	*AMC06, (NS2::PT7-ago + NS1::PconII*-RSWB-T7RNAP)*	*Gm, Sp, Sm*	*This Study*
AMT016	*AMC06, Δcas4*		*This Study*
AMT017	*AMC06, (NS2::PT7-cas4+NS1::PconII*-RSWB-T7RNAP)*	*Gm, Sp, Sm*	*This Study*
AMT018	*AMC06, ΔAgo+ ΔCas4*		*This Study*
AMT019	*AMT018, (NS2::PT7-cas4+ago + NS1::PconII*-RSWB-T7RNAP)*	*Gm, Sp, Sm*	*This Study*
AMT020	*AMC06, (NS2::PT7-cas4+ago + NS1::PconII*-RSWB-T7RNAP)*	*Gm, Sp, Sm*	*This Study*
AMT021	*AMT013, NS1::yfp*	*Sp, Sm*	*This Study*
AMT022	*AMT013, NS1::bfp*	*Sp, Sm*	*This Study*
AMT023	*AMC06, NS1::yfp*	*Sp, Sm*	*This Study*
AMT024	*AMC06, NS1::bfp*	*Sp, Sm*	*This Study*
AMT026	*ago(Tn5[8S1-BB11])*	*Km*	*This Study*
AMT027	*ago(Tn5[8S1-HH1])*	*Km*	*This Study*
AMT028	*cas4(Tn5[2D5-E-L9])*	*Km*	*This Study*
AMT029	*cas4(Mu[2D4-B4])*	*Cm*	*This Study*
AMT030	*ago(Tn5[8S1-BB11]+cas4::Mu[2D4-B4])*	*Km, Cm*	*This Study*
AMT031	*ago(Tn5[8S1-HH1]+cas4::Mu[2D4-B4])*	*Km, Cm*	*This Study*
	Deletion KO expression vectors and final plasmids		
pCVD015	*ccdB-SwaI-pUC19ori*	*Ap*	*Taton et al. 2014.*
pSL2680	*Broad-host range replicon encoding CRISPR/CpfI*	*Km*	*Ungerer et al. 2016*
pCVD001	*Gm^R gene cassette (CYANO-VECTOR donor plasmid)*	*Ap, Gm*	*Taton et al. 2014*
pCVD002	*$SpSm^R$ gene cassette (CYANO-VECTOR donor plasmid)*	*Ap, Sp, Sm*	*Taton et al. 2014*
pCVD003	*Km^R gene cassette (CYANO-VECTOR donor plasmid)*	*Ap, Km*	*Taton et al. 2014*
pCVD007	*Cm^R gene cassette (CYANO-VECTOR donor plasmid)*	*Ap, Cm*	*Taton et al. 2014*
pCVD010	*Nt^R gene cassette (CYANO-VECTOR donor plasmid)*	*Ap, Nt*	*Taton et al. 2014*
pRL278	*Suicide plasmid with SacB-SacI for cloning*	*Km*	*Taton et al. 2014*
pAM5406	*RSF1010Y25F_RK2BOM_pUC19ori*	*Ap*	*Bishe et al. 2019*
pAMT001	*CRISPR-CpfI + pUC19 origin of replication.*	*Ap*	*This study*
pAMT002	*RSF1010Y25F++_ cat_S7942_CRISPR-CpfI*	*Cm*	*This study*

Table 2.1 List of strains and plasmids used in this research (Continued).

Strain or Plasmid	***Description / Genotype***	***Antibiotic Resistance***	***Source/Creator***
pAMT004	*RSF1010Y25F++_aph1_ CRISPR-Cpf1*	*Km*	*This study*
pAMT005	*RSF1010Y25F++_nat1_A7120_CRISPR-Cpf1*	*Nt*	*This study*
pAMT006	*RSF1010Y25F++_aacC1_CRISPR-Cpf1*	*Gm*	*This study*
pAMT010	*pAMT003_Ago guide + homologous arms*	*Sp, Sm*	*This study*
pAMT012	*PRL278_Cas4 Homologous Arms*	*Km*	*This study*
Complementation expression vectors and final plasmids			
pAM5467	*S7942_NS2::(aacC1-P_{T7}-yfp)*	*Tc, Gm*	*Roulet et al 2018*
pAM5470	*S7942_NS1::(aadA-PconII*- RSWB-T7RNAP)*	*Sp, Sm*	*Roulet et al 2018*
pAM5204	*RP27 NS2-Tc-Gm-ccdB-riboJ-yfp*	*Gm, Tc*	*Golden Lab Collection*
pAMT007	*S7942_NS2::(aacC1-P_{T7}-Ago)*	*Gm*	*This study*
pAMT008	*S7942_NS2::(aacC1-P_{T7}-Cas4)*	*Gm*	*This study*
pAMT009	*S7942_NS2::(aacC1-P_{T7}-Cas4-Ago)*	*Gm*	*This study*
Plasmids utilized in Assays			
pAM5407	*RSF1010-aadA-PconII-oRBS- yfp*	*Sp, Sm*	*Taton et al. 2014*
DE03	*NS1::yfp*	*Sp, Sm*	*Golden Lab Collection/D. Ernst*
DE17	*NS1::bfp*	*Sp, Sm*	*Golden Lab Collection/D. Ernst*
pAMT025	*pANS-aadA-PconII-oRBS- yfp*	*Sp, Sm*	*This Study*
AM2991	*NS1::(aadA-lacIq-Ptrc)*	*Sp, Sm*	*Golden Lab Collection*
pAM5328	*NS3::aacC1*	*Gm*	*Golden Lab Collection*

2.3 General PCR Protocols

PCR amplification for screening DNA was executed with Taq DNA Polymerase from New England BioLabs (NEB). Per NEB guidelines, the thermocycler setting for Taq Polymerase chain reactions started with a 30-second initial denaturation at 95°C followed by 30 cycles of 30 seconds at 95°C, 60 seconds at optimal annealing temperature (45-68°C), and 1 minute/kb extension at 68°C and a final extension period of 5 minutes at 68°C. PCR amplification for plasmid assembly was performed using Q5 High-fidelity polymerase from NEB according to the manufacturer's instructions. Also per NEB guidelines, the thermocycler was set to an initial denaturation of 30 seconds at 98°C followed 30 cycles of 10 seconds at 98°C, 30 seconds at the recommended annealing temperature (65-70°C), and 30 seconds/kb at 72°C and terminating with a final extension period of 2 minutes at 72°C. PCR products were cleaned and concentrated using the Zymo Research DNA Clean and Concentrator Kit for downstream use. Primers used in this research are listed in Table 2.2.

Table 2.2: List of primers designed or utilized by this study

Name	*Purpose*	*Sequence (5' to 3')*
Agodel_LA_F	*Amplify ago left homologous arm*	*agctttaatgcggtagttggtaccgcgataacccgtttcttcttc*
Agodel_LA_R	*Amplify ago left homologous arm*	*ggcaaggacgtttctgggtaagcgaaagcggcgaccccactcca*
Agodel_RA_F	*Amplify ago right homologous arm*	*tgtcgcccgagctatagctggccgatcggagtagtaaagag*
Agodel_RA_R	*Amplify ago right homologous arm*	*CCTGTTGAGGACAGACAGCACTTTG*
ago_gRNA	*CRISPR-Cpf1 ago guide RNA*	*AGATATGCTTTCACATGGATCTAC*
cas4_gRNA	*CRISPR-Cpf1 cas4 guide RNA*	*AGATatgcagcagcaggctttaggctac*
cpf1_F	*Amplify CRISPR array + cpf1*	*cgccggatCTGCAGgcatcggaaaaataccatgctcagaaaaggc*
cpf1_R	*Amplify CRISPR array + cpf1*	*cccccgatGTCGACggtaccgatcacgcatcttcccgacaac*
CAS4_LA_F	*Amplify cas4 left homologous arm*	*agctttaatgcggtagttggtaccgcgataacccgtttcttcttc*
CAS4_LA_R	*Amplify cas4 left homologous arm*	*ggcaaggacgtttctgggtaagcgaaagcggcgaccccactcca*
CAS4_RA_F	*Amplify cas4 right homologous arm*	*cttggagtggggtcgccgctttcgcttacccagaaacgtccttgc*
CAS4_RA_D1209R	*Amplify cas4 right homologous arm*	*gcccccgatgtcgacggtaccaggttggcgatttggtcatc*
CAS4_LA_U855F	*Amplify cas4 left homologous arm*	*catgcaagctttcgcgagctcgataacccgtttcttcttc*
CAS4_LA_R_AT	*Amplify cas4 left homologous arm*	*ggcaaggacgtttctgggtaagcgaaagcggcgaccccactcca*
CAS4_RA_F_AT	*Amplify cas4 right homologous arm*	*ggtcgccgctttcgcttacccagaaacgtccttgcc*
Synpcc7942_1534_U376F	*Amplify Synpcc7942_1534*	*TGTGTCGCTGCCTGTAAAAATGC*
Synpcc7942_1534_D286R	*Amplify Synpcc7942_1534*	*CGCGATCGTGGTTGAGTAGAGC*
Synpcc7942_2273_D1F	*Amplify Synpcc7942_2273*	*ATGAGCGACTATCCAGCCCTTTCTC*
Synpcc7942_2273_D768R	*Amplify Synpcc7942_2273*	*TTACAGCTCTACTACCGGAATCTCGTCC*
CAS4_2273_1F	*Amplify Synpcc7942_2273*	*gggagaatttaaataaaggaggtcttaagatgagcgactatccagccctttctc*
CAS4_2273_D773R	*Amplify Synpcc7942_2273*	*caggatggccttctcctgcaccagcttacagctctactaccggaatc*
ago_1534_1F	*Amplify Synpcc7942_1534*	*gggagaatttaaataaaggaggtcttaagatggatctactatcaaatttgcggcg*
ago_1534_D2228R	*Amplify Synpcc7942_1534*	*caggatggccttctcctgcacgagtactgagccatcacgcttagac*
cas4_2273_D773R_ago	*Double complementation plasmid*	*taagacctcctttgtccggttagccagcttacagctctactaccggaatc*
ago_1534_1F_cas4	*Double complementation plasmid*	*aaccggacaaaggaggtcttaagatggatctactatcaaatttgcggcg*

2.4 Plasmid Construction Methods

The in silico construction of all plasmids utilized in this study followed the streamlined assembly strategy from the previously published platform for broad-host-range vector systems which relies on the overlap of short GC-rich overhanging sequences to facilitate the process of

isothermal assembly [53]. Based on this strategy, plasmids with more common modules were constructed in silico through the virtual CYANO-VECTOR assembly portal with molecular devices from its library [53]. Plasmid DNA fragments were generated by restriction digestion of donor plasmids carrying the module or through the direct PCR amplification of the target sequence. Restriction enzyme digestions were incubated at 37°C for 3-5 hours with 5U of enzyme per ug of DNA in a solution that is 50 times greater in volume than the enzyme used. In acquiring DNA modules by PCR amplification, the overlapping GC-rich adaptor sequences were incorporated into the DNA fragment through the 5'-end of the PCR primers. Digest and PCR generated DNA fragment samples were cleaned and concentrated with the Zymo Clean and Concentrator TM-5 spin column kit, adhering to the protocol provided by the manufacturer. DNA concentrations were measured for each DNA fragment with a UV-Vis spectrophotometer Nanodrop 2000c followed by the plasmid assembly using the GeneArt Seamless Cloning and Assembly Kit from Life Technologies, according to the manufacturer's instructions. A complete list of the plasmids constructed and utilized in this study is detailed in Table 2.1

2.5 Gene Transfer in *S. elongatus* PCC 7942

Natural transformation was performed in *S. elongatus* to transfer 200 ng – 2 µg of recombinant DNA into a 5 mL liquid culture[9]. The starting 5 mL of culture was centrifuged at 3500 g for 5 minutes and the cell pellet was washed in 1 mL BG-11. The cells were then centrifuged again and the pellet was resuspended in 200 µl BG-11. Next, 200 ng of transforming DNA was added to the cells (increasing amount for bigger plasmids) and the mixture was left to incubate in the dark on a shaker at 30°C for 4-18 hours. Finally, the mixture was divided and various dilutions were plated on selective BG-11 agar plates. The plates were incubated at 30°C in continuous light (or a cycle of 16 hours with light and 8 hours of darkness) at 150-300 µmol photons $m^{-2}s^{-1}$ for a minimum of 5 days to allow time for colony growth.

The transfer of DNA by bacterial conjugation in *S. elongatus* was primarily performed through biparental mating using the conjugal donor *E. coli* strain AM1359 to deliver the shuttle vector. AM1359 cells were transformed with the shuttle vector of interest via electroporation and the resulting transformants were grown overnight in 2 x 5 mL LB with antibiotics at 37°C on a roller drum. The following day, the *E. coli* cultures were split into 2-mL aliquots (each

aliquot used for 1 conjugation) and centrifuged for 3 minutes at 3500 g followed by two consecutive washes with LB to clear any remaining antibiotics. The *E. coli* were then resuspended in 200 ul of LB without damaging the conjugal pili (avoiding vortex) and mixed with a 1-mL resuspension of *S. elongatus* cells prepared from 5-10 mL culture with an OD_{750} between 0.4-0.6 that was washed once with BG-11 to remove all antibiotics. The cell mixture was centrifuged for 5 minutes at 3500 g, resuspended in 200 µl of BG-11, plated on BG-11 plates containing 5% LB and no antibiotics, and incubated for 12-24 hours at 30°C with 10-20 µmol photons $m^{-2}s^{-1}$ of light. Then, the cells were scraped from the plates, resuspended in 1 mL of BG-11, and plated on BG-11 agar plates containing the appropriate antibiotics. The plates were incubated for 5-7 days at 30°C with approximately 300 µmol photons $m^{-2}s^{-1}$ of light to allow time for colony growth. Singular exconjugant colonies were selected and streaked on BG-11 agar plates with the proper antibiotics and independent colonies from each streak were picked and patched to maximize segregation for subsequent inoculation into a liquid culture.

The Δ*cas4* CRISPR KO applied a triparental mating strategy to simultaneously transfer the RSF1010 plasmid carrying cpfI and the crRNA sequence as well as the pRL278 plasmid carrying the template sequence for homology-directed repair. As previously published, this strategy follows the same protocol as that described for biparental mating with a single modification to allow for the transfer of two shuttle vectors instead of one [9]. AM1359 cells were transformed with each shuttle vector individually to obtain two independent conjugal donor *E. coli* strains. After the preparation of the cells as described for biparental mating, both *E. coli* strains were mixed along with the *S. elongatus* cells, and the mixture was processed and plated as previously stated for biparental mating.

2.6 RSF1010 Transformation in *S. elongatus* – Spot Assay

WT *S. elongatus* and SeAgo complementation and IKO strains were normalized to an OD_{750} of ~0.05 and grown in 100 mL of BG-11 media while shaking at 30°C and about 150-µmol photons m^{-2} s^{-1} until reaching an OD_{750} of ~0.4. 200 µl of cells, concentrated at approximately $1x10^9$ cells ml^{-1}, were transformed with 200 ng of pAM5407 for 16 hours in the dark and then serially diluted and spotted (4.5 µl per spot) on BG-11 agar plates with and without selection. Additionally, the serial dilutions were spotted on BG-11 agar plates with and without 1mM of

theophylline (from a stock solution of 200mM theophylline dissolved in 2 mL Of DMSO) which activates the T7 system in the SeAgo complementation strain.

2.7 Quantitative Circadian Transformation in *S. elongatus* PCC 7942

Duplicate cultures of each KO strains were grown in 100 mL of BG-11 medium bubbled with air at 30°C in 12 hr light dark cycles (LD 12:12) with 200 μmol photon m^{-1} s^{-1}. The cultures were started at an OD_{750} of ~0.1 and grown for 3 days until they reached an OD_{750} of ~0.5. For each culture, at both time point, 200 μl of cells concentrated at $5x10^8$ cells ml^{-1} were transformed with 200 ng of suicide plasmids for integration of either Sp+Sm resistance gene in NS1 (pAM2991) or Gm resistance gene in NS3 (pAM5328). The plasmid DNA was added two hours after the onset of darkness and two hours following the return of light. At each time point, the cells were incubated with the plasmid DNA in complete darkness while shaking for 3 hours at 30℃ after which the reactions were terminated by serially diluting and plating each transformation onto a 12-well plate (1 dilution/well) containing 3 mL of BG-11 agar in each well along with the appropriate antibiotics.

In order to quantify transformation efficiency, the plated cells were incubated at 30°C under constant illumination to allow for the colonies to grow (approximately 5 days). Each plate was photographed and the pictures were processed with ImageJ to produce binary images where the background has been removed and adjacent colonies were separated. The number of colonies was determined with ImageJ. Transformation efficiencies were calculated by first determining the number of antibiotic-resistant colonies per CFU without selection using ImageJ.

2.8 RSF1010 Conjugation in the *S. elongatus* RB-TnSeq Library

To screen the RB-TnSeq library, an aliquot was thawed from the −80 °C archive in a 37 °C water bath for 2 min, inoculated into 2 flasks to obtain 100 mL culture in BG-11 media with Km, and incubated at 30 °C in 20 μmol photons m^{-2} s^{-1} for 1 day. The cultures were then transferred onto an orbital shaker at 150 rpm under 70 μmol photons m^{-2} s^{-1} of light and allowed to grow until reaching an OD_{750} of ~0.3. Prior to the conjugation, T_0 was collected to establish the population baseline. Following the standard biparental conjugation protocol, the *S. elongatus* mutant population culture. The plated conjugations were incubated under continuous light of

~250 μmol photons m^{-2} s^{-1} for 5 days until the appearance of isolated colonies. To analyze the exconjugants, ~250,000 colonies were collected, pooled, and stored at −80 °C for subsequent genomic DNA extraction followed by amplification, sequencing and quantification in perl 5v18 using established BarSeq protocols. The analysis provided a fitness value to characterize the effect of loss of function on the ability to conjugate RSF1010, with the estimation of the corresponding statistics performed in R version 3.6.0 utilizing previously developed methods and scripts.

Chapter 3. Results and Discussion

3.1 Preliminary Analyses of Synpcc7942_1534 and Synpcc7942_2273

3.1.1 Synpcc7942_1534

Our knowledge on prokaryotic argonautes is limited and only a small number of pAgos have been characterized. Most of the examined pAgos originate from thermophilic bacteria such as TtAgo from *T. thermophilus*. Argonaute genes and proteins are frequently annotated as hypothetical proteins in the primary databases such as NCBI (National Center for Biotechnology Information). This is the case for *S. elongatus* Synpcc7942_1534 and other cyanobacterial pAgo orthologous genes. In an effort to characterize Synpcc7942_1534, sequence bioinformatics were conducted, including BLAST searches against all sequenced cyanobacterial genomes and genome sequences of other organisms. The BLAST analyses suggested that the Synpcc7942_1534 gene (*ago*) encodes the catalytically active argonaute enzyme, which has been demonstrated earlier this year[36].

To obtain a comprehensive picture of the distribution of argonautes across the cyanobacterial phylum, we constructed a phylogenetic tree (Figure 3.1B) including all cyanobacterial homologs of Synpcc7942_1534 identified by BLAST. The SeAgo protein displayed a sporadic distribution across the cyanobacterial phylum with only a few represented organisms in comparison to the phylogenetic analysis of the 16S rRNA gene in cyanobacteria (Figure 3.1). Most of the cyanobacterial relatives of SeAgo are members of the same family, *Synechococcaceae*, and about half of these members belong to thermophilic cyanobacteria (Figure 3.1B). The *Synechococcaceae* family also displays the highest degree of SeAgo sequence conservation in comparison to the remaining cyanobacterial relatives that display a scattered distribution in the rest of the cyanobacterial phylum. Overall, we determined that the closest phylogenetic relative of *S. elongatus* SeAgo is the hypothetical protein BRW62_09585 in *Synechococcus lividus* PCC 6715 (Figure 3.1B). Outside of cyanobacteria, the thermophilic bacterial family of *Thermacea*, containing the *T. thermophilus* TtAgo, represent the closest phylogenetic relatives of *S. elongatus* SeAgo.

We analyzed the protein sequence of SeAgo for the presence of any conserved regions, domains, and homologous protein superfamilies. Using NCBI's Conserved Domain (CD)

search feature and the InterPro Classification of Protein Families tool from the EMBL-European Bioinformatics Institute (EMBL-EBI), we identified a conserved PIWI domain as well as an Argonaute–PAZ domain, which characterize SeAgo as a long-A pAgo (Table 1, Figure 3.2B). The complete catalytic tetrad of the active site within the conserved PIWI domain was also identified, strongly suggesting an intrinsic nucleolytic capacity of SeAgo. Lastly, the NCBI Conserved Domain search located a conserved 5' RNA guide strand anchoring site which is indicative of a MID domain and the ability of SeAgo to perform guide-mediated sequence-specific target recognition.

Using the Phyre2 software, we examined the predicted protein structure of SeAgo and determined that it is most homologous to TtAgo (Figure 3.2D, 3.2E). Figure 3.2E displays six different arbitrary angles of the superposition of the predicted structure of SeAgo and the crystallized structure of TtAgo, demonstrating the structural similarity between these two long-A pAgos. The structural homology and domain conservation of SeAgo and TtAgo supports the prediction of the mechanism, structure, and overall function of SeAgo based on the experimentally characterized model of TtAgo.

Bioinformatics analyses of the *S. elongatus* genome and transcriptome determined that the *ago* gene is located between two highly expressed genes, succinate dehydrogenase subunit B and Rubisco chaperonin that are encoded on different transcripts (Figure 3.2A). Although neither of these genes is considered essential[54], their function and level of expression suggest that they are likely important for electron transfer and Rubisco activity of *S. elongatus*. While in our preliminary work we used Tn5 insertion knock-outs (IKOs), we concluded that an in-frame deletion strain was required to prevent possible polar effects and better assess the role of the argonaute system in *S. elongatus* (discussed in further detail in sections 3.2 and 3.3).

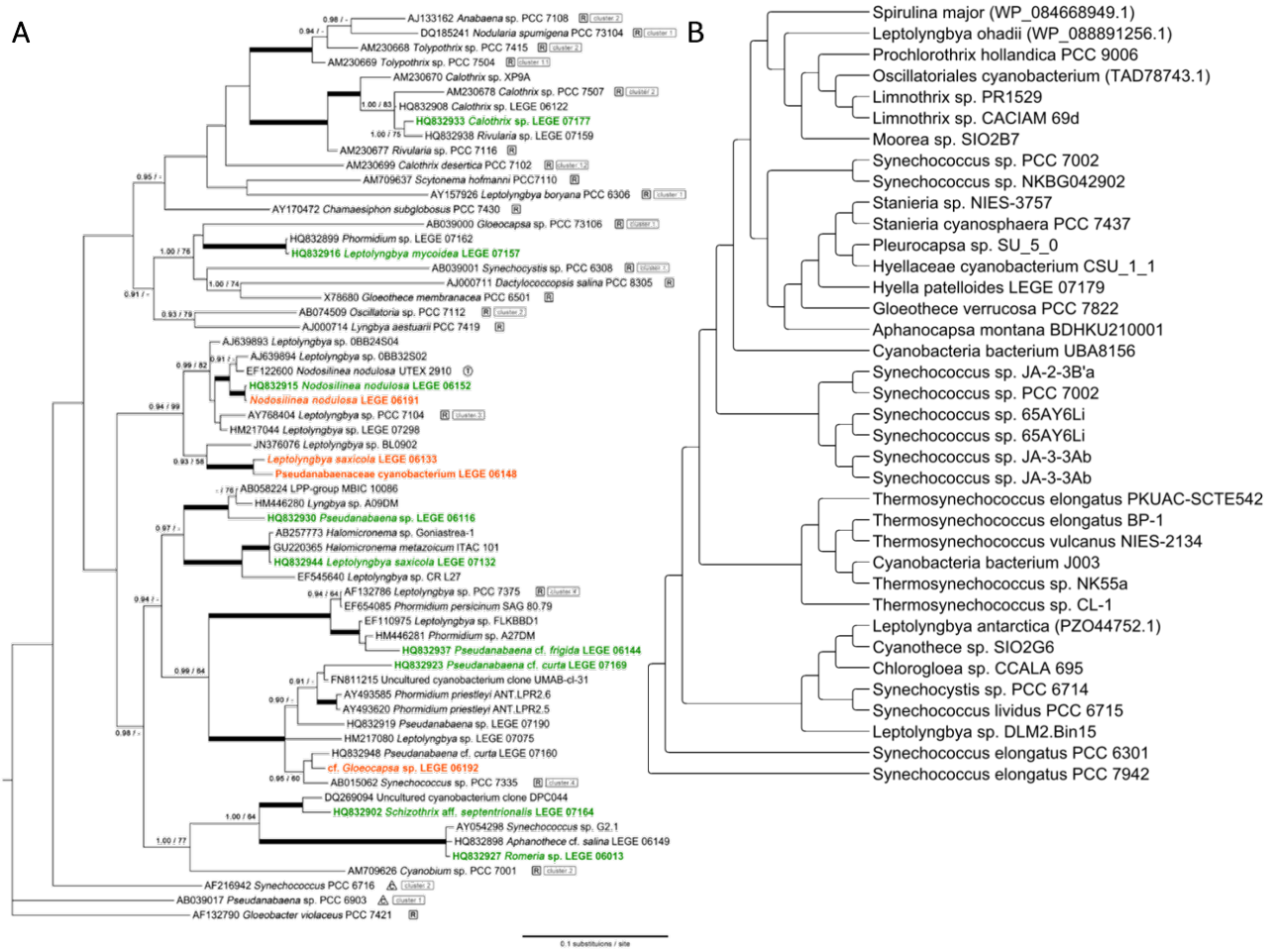


Figure 3.1 Phylogenetic distribution of Ago proteins in the cyanobacterial phylum. (A) Phylogenetic distribution of 16S rRNA, representing the phylogenetic relationships of cyanobacterial species. (Figure source: Leão et al. 2013). (B) Phylogenetic distribution of SeAgo homologs across the cyanobacterial phylum, constructed by BLAST sequence analysis from the NCBI database.

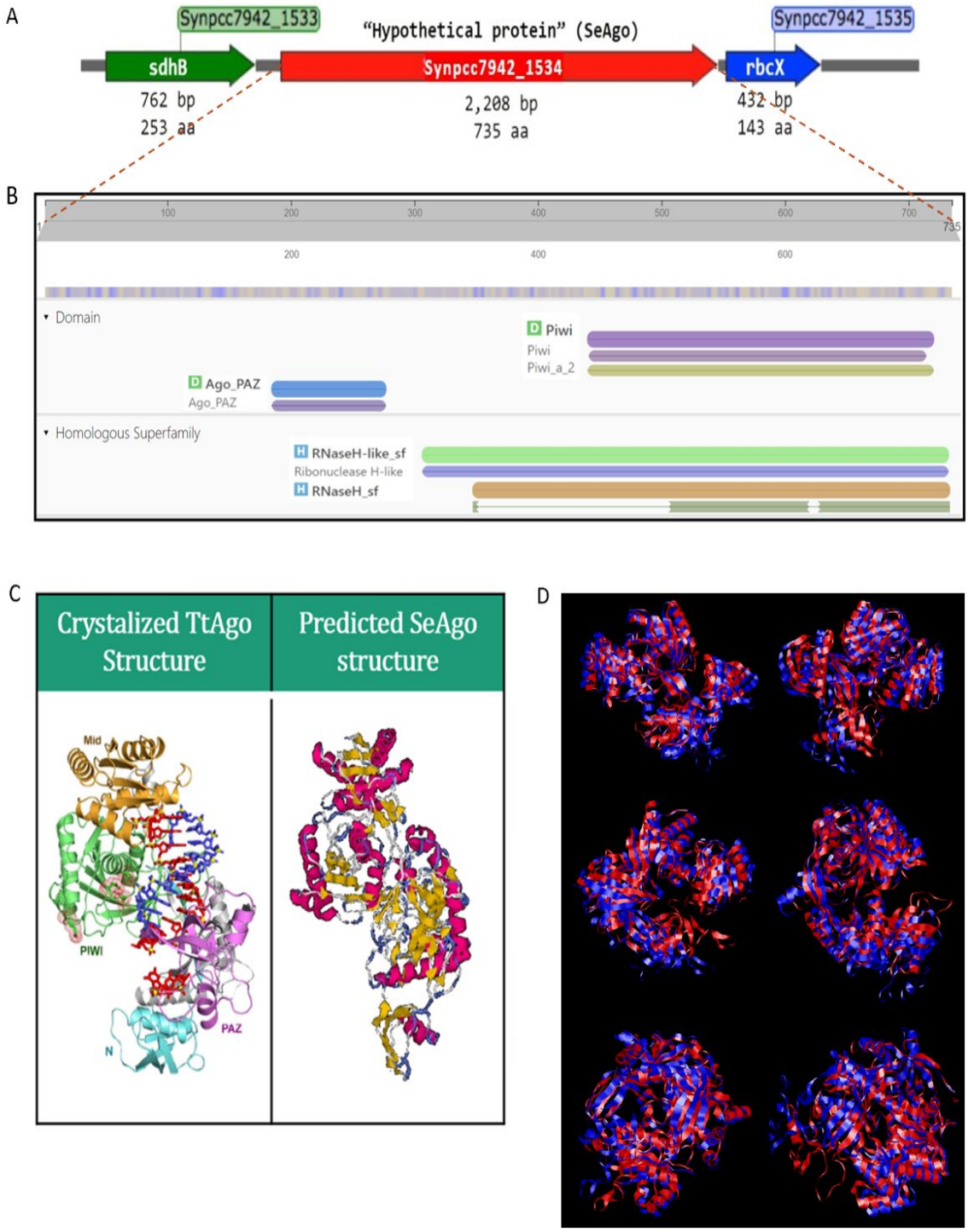


Figure 3.2 Bioinformatics of SeAgo. (A) Context of *ago* in *S. elongatus* chromosome. (B) Homologous superfamilies and conserved domains found in Synpcc7942_1534 (*ago)*. (C) Predicted protein structure of SeAgo and the crystallized structure of TtAgo. (E) Superposition of SeAgo predicted protein structure and crystallized TtAgo structure, constructed with Phyre2 software and visualized with RasWin. Shown are six arbitrary angles.

3.1.2 Synpcc7942_2273

Synpcc7942_2273 is the second gene, in addition to Synpcc7942_1534, which presented positive fitness for natural competence from the previously described data. We performed gene and protein sequence analyses to characterize the annotated “hypothetical protein” in the NCBI database. We examined the surrounding gene context which showed that Synpcc7942_2273 was closely surrounded by two essential genes, *folK* and *chID*. Using NCBI’s Conserved Domain (CD) search feature and the InterPro Classification of Protein Families tool from the EMBL-European Bioinformatics Institute (EMBL-EBI), we found that Synpcc7942_2273 protein consists of the conserved domain PDDEXK_AddAB-type and is homologous to the Exonuc_phg/RecB_C superfamily which includes CRISPR-associated nuclease Cas4. We name the Synpcc7942_2273 protein SeCas4 because the conserved domains and homologous protein families suggest that it is a Cas4-like nuclease.

Utilizing the Phyre2 software, we analyzed the predicted protein structure of SeCas4 which showed close homology to the Cas4 from *Pyrobaculum calidifontis* jcm 11548. Using BLAST pairwise sequence alignment between Synpcc7942_2273 (*cas4*) and all sequenced cyanobacterial genomes in the NCBI sequence database we examined the distribution of cas4-like proteins across the cyanobacterial phylum. The phylogenetic analysis showed a wide distribution with a high degree of conservation and determined the closest relative of *S. elongatus* SeCas4 to be PD-(D/E)XK nuclease family protein from *Synechococcus elongatus* PCC 11802.

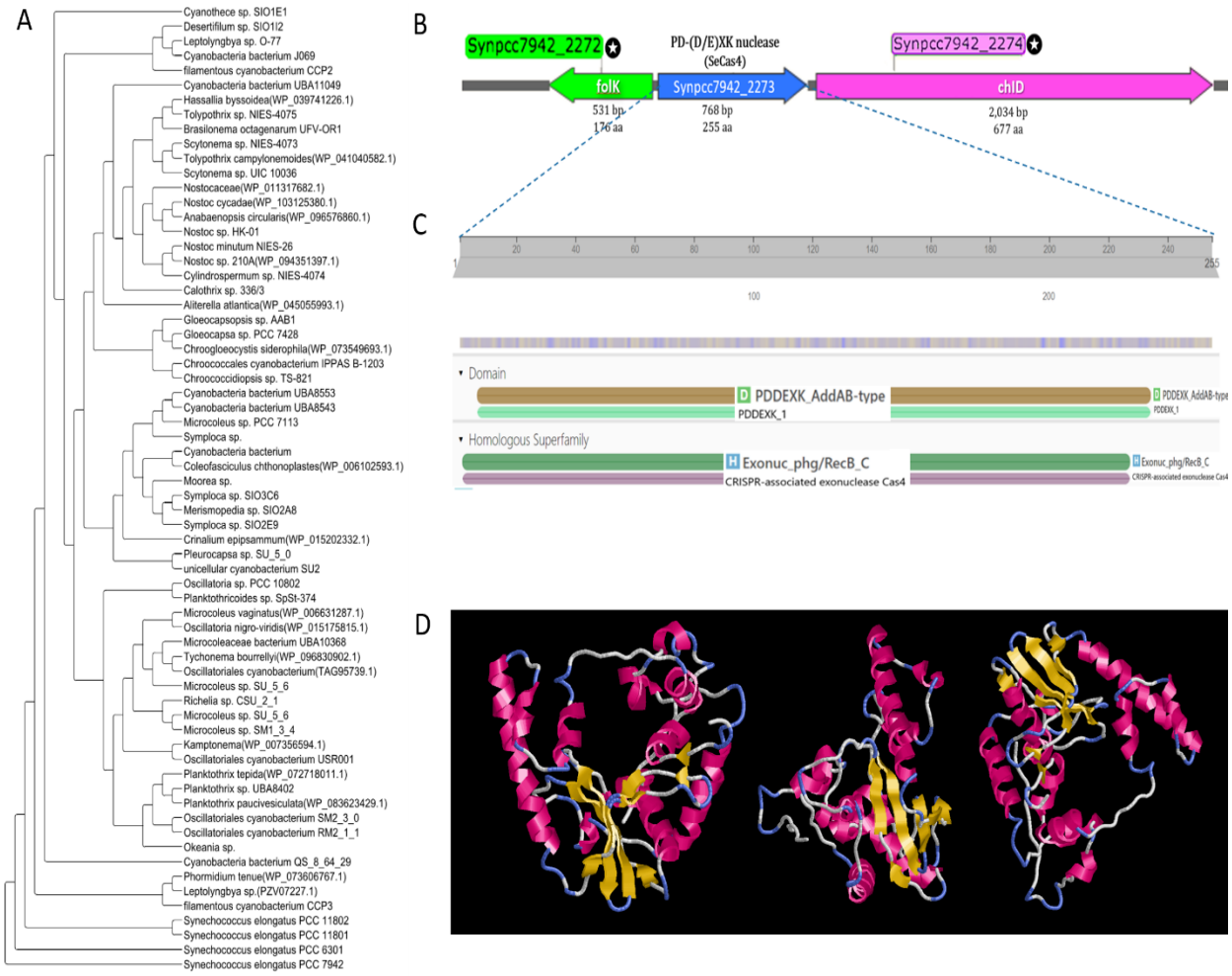


Figure 3.3 Bioinformatics of SeCas4. (A) Distribution of cas4-like proteins in the cyanobacterial phylum. The phylogenetic tree includes all cyanobacterial homologs of SeCas4 identified by BLAST. (B) Context of *cas4* in *S. elongatus* chromosome. (C) Homologous superfamilies and conserved domains in Synpcc7942_2273 (*cas4)*. (D) Predicted protein structure of SeCas4, constructed with Phyre2 software. Shown are three arbitrary angles.

3.2 Preliminary Analysis of Ago and Cas4 Insertion Knock-out Mutants

The fitness analysis of the *S. elongatus* transposon mutant library upon natural transformation[9], suggested that loss of function of SeAgo or SeCas4 would result in strains of *S. elongatus* with higher natural transformation efficiencies. To test this hypothesis, we conducted quantitative transformation assays on insertion KOs of *ago*, *cas4*, or both genes. The investigation of a double KO relied on the predicted nucleolytic function of both SeAgo and SeCas4 and the hypothesis that there may be a synergistic relationship between the two proteins. In addition, because circadian transcriptomics indicated that these genes may be differently

expressed during a circadian time course [9], these assays were performed at dawn (early morning) and dusk (early night), when the efficiency of natural transformation is either minimal or maximal, respectively.

Insertion KO strains of *S. elongatus* were engineered using uni-gene set (UGS) library plasmids[54]. Two distinct insertion KO strains were constructed for both the *ago* (UGS 8S1-HH1, 8S1-BB11) and *cas4* (2D4-B4, 2D5-E-L9) loci. While both UGS insertions for *ago* provided kanamycin resistance, different insertions in *cas4* provided resistance to either kanamycin or chloramphenicol, thereby enabling the construction of a double KO strain (*ago*::Tn5[8S1-HH1], *cas4*::Mu[2D4-B4]). To determine transformation efficiency, duplicate cultures of each KO strain were transformed with suicide plasmids that integrate either a Sp+Sm resistance gene into NS1 (NS, neutral site) (pAM2991) or a Gm resistance gene into NS3 (pAM5328). The plasmid DNA was added two hours after the onset of darkness and two hours following the return of light. The transformation protocol and calculation of transformation efficiencies are described in the Materials and Methods section.

Based on the results of the quantitative transformation assay at the timepoint corresponding to the peak in natural competence (dusk), the Ago IKOs transformed with higher efficiency than the WT and control strains (Figure 3.4). The SeCas4 IKOs and the double insertion knockouts transformed with similar efficiency to the WT and control strains. At the second timepoint (dawn), corresponding to the lowest competence state, the transformation efficiency of SeAgo IKOs and SeCas4 IKOs was orders of magnitude higher in comparison to the WT and control strains (Figure 3.4). However, the low competence state is characterized by a drastic reduction of transformation efficiency that is controlled by the internal clock and thus overshadows the potential effect of SeAgo and SeCas4 and prohibits the determination of their role at dawn.

The results obtained from biological replicates and the different strains for each loci were inconsistent. However, the *ago* 8S1-BB11 mutant had higher transformation efficiency in all experiments. The loss of SeAgo resulted in increased natural transformation efficiencies which was also observed for the SeCas4 Tn5 mutant and in two instances for the SeCas4 Mu mutant. Conversely, the double KO did not transform better than the WT. Transformation efficiency in the double KO strain may have been affected by the presence of two antibiotic resistance markers or the addition of these antibiotics to the BG-11 media. We hypothesized that the Tn5 or

Mu insertions resulted in polar effects on the highly expressed genes flanking *ago* or essential genes flanking *cas4*. To reduce such polar effects, we decided to construct marker-less in frame deletions of SeAgo and SeCas4.

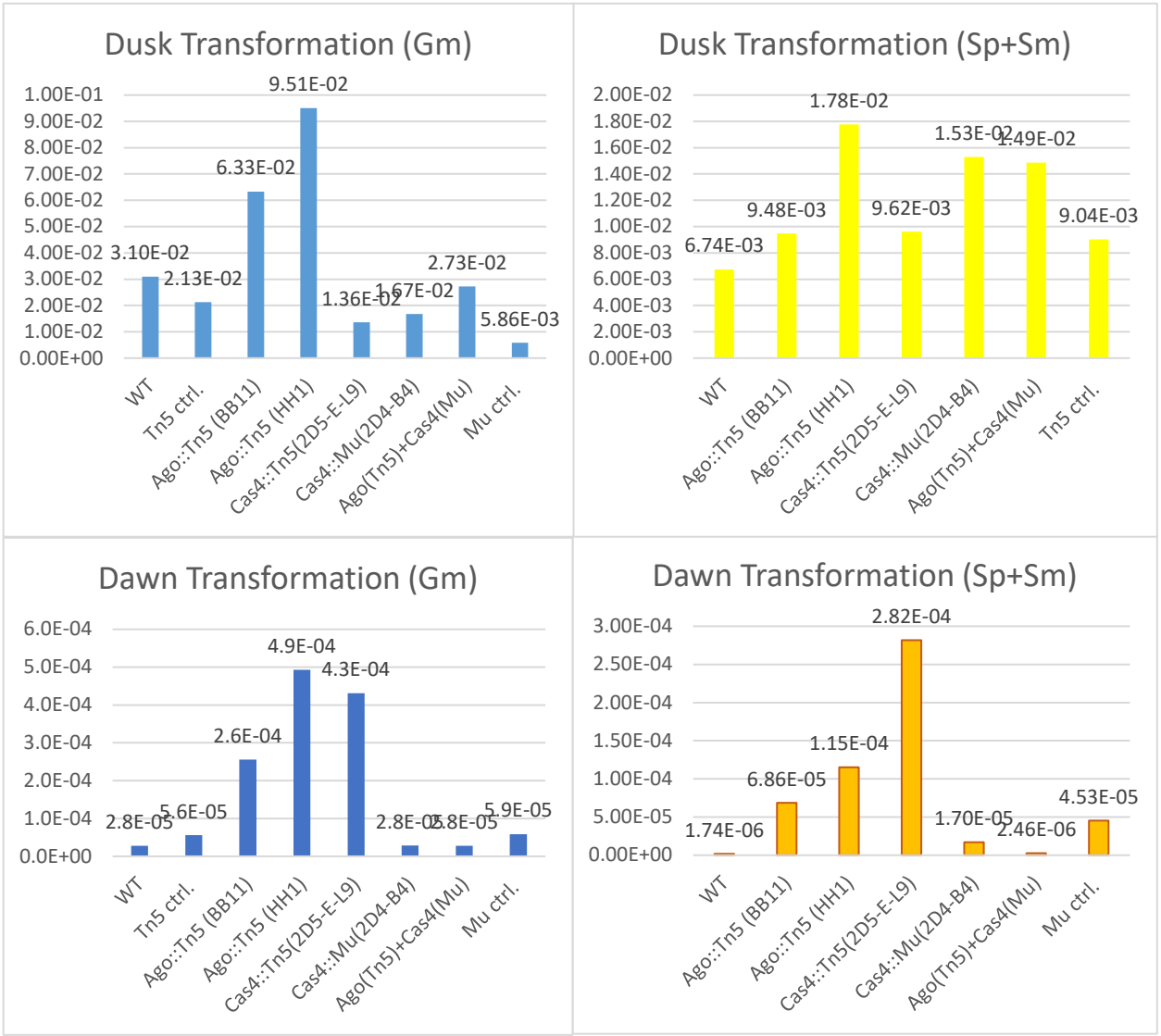


Figure 3.4 Quantitative transformation assay at dawn and dusk of two IKO strains of SeAgo, SeCas4, and both genes compare to WT and control strains. Cells were transformed with suicide plasmids carrying Gm or Sp+Sm resistance genes. Transformation efficiencies were calculated by determining the number of antibiotic-resistant colonies per CFU without selection.

3.3 Tool Development: Improved CRISPR-Cpf1 Gene Editing in Cyanobacteria

3.3.1 Marker-less Deletions in Cyanobacteria

The development of CRISPR/Cas-based tools for genetic engineering in cyanobacteria facilitates construction of marker-less mutant strains[55,56]. Because of the high toxicity of the Cas9 enzyme in cyanobacteria, a significant improvement to previously developed systems was the use of the alternative CRISPR nuclease, Cpf1, which is better suited for gene editing and cyanobacteria[55]. Nevertheless, the assortment of CRISPR tools optimized for genetic engineering in cyanobacteria is still limited. The plasmid pSL2680 was previously developed for expression of a CRISPR array along with *cpf1*, both cloned in an RSF1010 backbone carrying a kanamycin resistance gene[55]. While the construction of pSL2680 advanced the genetic engineering of cyanobacteria by replacing the toxic Cas9, there are difficulties associated with genetic manipulation of RSF1010-based plasmids such as its low copy number and self-mobilization, which involves nicking of the plasmid at the oriT and results in low amounts of useable DNA for cloning[53]. Therefore, to improve this system we constructed a CRISPR/Cpf1 module, compatible with CYANO-VECTOR[53], enabling the construction of different CRISPR/Cpf1 plasmids with any available CYANO-VECTOR antibiotic devices and replicons. The specific plasmid design and construction are described in the following section.

After construction of the CRISPR/Cpf1 device, we assembled five CRISPR-Cpf1 plasmids containing five different antibiotic resistance markers assembled using the GC-adapter-based CYANO-VECTOR assembly strategy[53]. The improved CRISPR plasmids carry a modified RSF1010 backbone in which MobA was inactivated to prevent the plasmid from getting nicked and a different base of mobilization was added, RK2BOM, to enable high mobilization efficiency in the presence of the RK2 conjugal plasmid [53,57]. In addition, the modified RSF1010 backbone encodes a pUC19 origin of replication that increases the plasmid copy number and thus facilitates plasmid purification and cloning procedures[53,57]. The constructed CRISPR plasmids with the improved RSF1010 backbone yielded significantly higher DNA concentrations from standard plasmid extractions procedures. Using the Qiagen QIAprep Spin Miniprep Kit, plasmid DNA extracted from a 2 mL overnight *E. coli* culture carrying the improved CRISPR plasmids yielded up to 10-times higher DNA concentrations than from a 2-

mL overnight *E. coli* culture carrying pSL2680 (Figure 3.5F). The increased yields reduce the challenges of further genetic manipulations of the improved CRISPR plasmids.

3.3.2 Tool Design and Construction

The CRISPR array and cpf1 were amplified by PCR from pSL2680 and assembled with linearized pCVD015 vector carrying a pUC19 origin of replication and an ampicillin resistance gene (Figure 3.5A). By adhering to the design principles of the CYANO-VECTOR platform, the resulting pUC19-CRISPR-Cpf1 (AMT001) plasmid is adapted to function as a novel tool for the CYANO-VECTOR assembly system for performing marker-less gene editing. To obtain the final enhanced CRISPR plasmids, the CRISPR-cpf1 module was released from the pCVD015 vector by restriction digestion with ZraI. Five antibiotic resistance devices, compatible with GC-rich adaptors, were digested from donor plasmids with EcoRV to produce five separate CRISPR plasmids, each carrying a different antibiotic-resistance marker (Figure 3.5B). Finally, the improved RSF1010 plasmid, described in section 3.3.1, was linearized with ZraI from NEB and assembled with the CRISPR-Cpf1 vector and each of the antibiotic resistance devices (Figure 3.5E). All the assembled fragments contained GC-adaptors to ensure compatibility to the CYANO-VECTOR system and streamline the assembly process. A schematic representation of the CRISPR plasmid construction process is diagramed in Figure 3.5A. Refer to the materials and methods section for a complete list of the constructed plasmids.

3.4 Deletion and complementation of genes encoding SeAgo and SeCas4 in *S. elongatus*

3.4.1 Marker-less deletion of genes encoding SeAgo and SeCas4 – Plasmid Construction

To obtain the marker-less deletion of *ago* and *cas4*, we engineered the modified CRISPR plasmid (AMT003) shown in Figure 3.5E. For both plasmids, the guide RNA coding sequences were designed to result in cleavage 166 nt downstream of the *cas4* start codon and 17 nt downstream of the *ago* start codon to avoid deletion of any regulatory regions of the important upstream genes. The guide was constructed by phosphorylating and annealing complementary oligonucleotides 24 nt in length (gRNA_F and gRNA_R) and then cloning the double-stranded fragment into the AarI restriction sites flanking *lacZ*-alpha gene within the first spacer in the CRISPR array. To introduce the deletion of either SeAgo or SeCas4, homologous repair templates were designed and amplified by PCR for each gene. The repair templates consisted of

two DNA fragments, 600-1,000 bp in length, homologous to the sequence upstream and downstream of the desired deletion region.

For reasons still unclear, seamless assembly of the CRISPR-guide vector, linearized with KpnI, and the two fragments of the homologous repair template, was difficult. While cloning of the *ago* deletion repair template and sequence encoding for the guide RNA into a single CRISPR deletion plasmid eventually succeeded (Figure 3.6C), cloning the repair template for the *cas4* deletion into the CRISPR deletion plasmid did not succeed and necessitated modifications in our approach. In this modified approach, cloning of the sequence encoding for the *cas4* guide RNA into the CRISPR plasmid was conducted as previously described, however, the *cas4* deletion repair template was instead cloned into a separate suicide plasmid, pRL278, linearized with SacI (Figure 3.6D). To account for the change to the pRL278 backbone, the repair template fragments for the deletion of *cas4* were resynthesized with the appropriate overhang sequences (See Table 2.2 for the detailed primers). Both amplified fragments were then cloned by seamless assembly into pRL278.

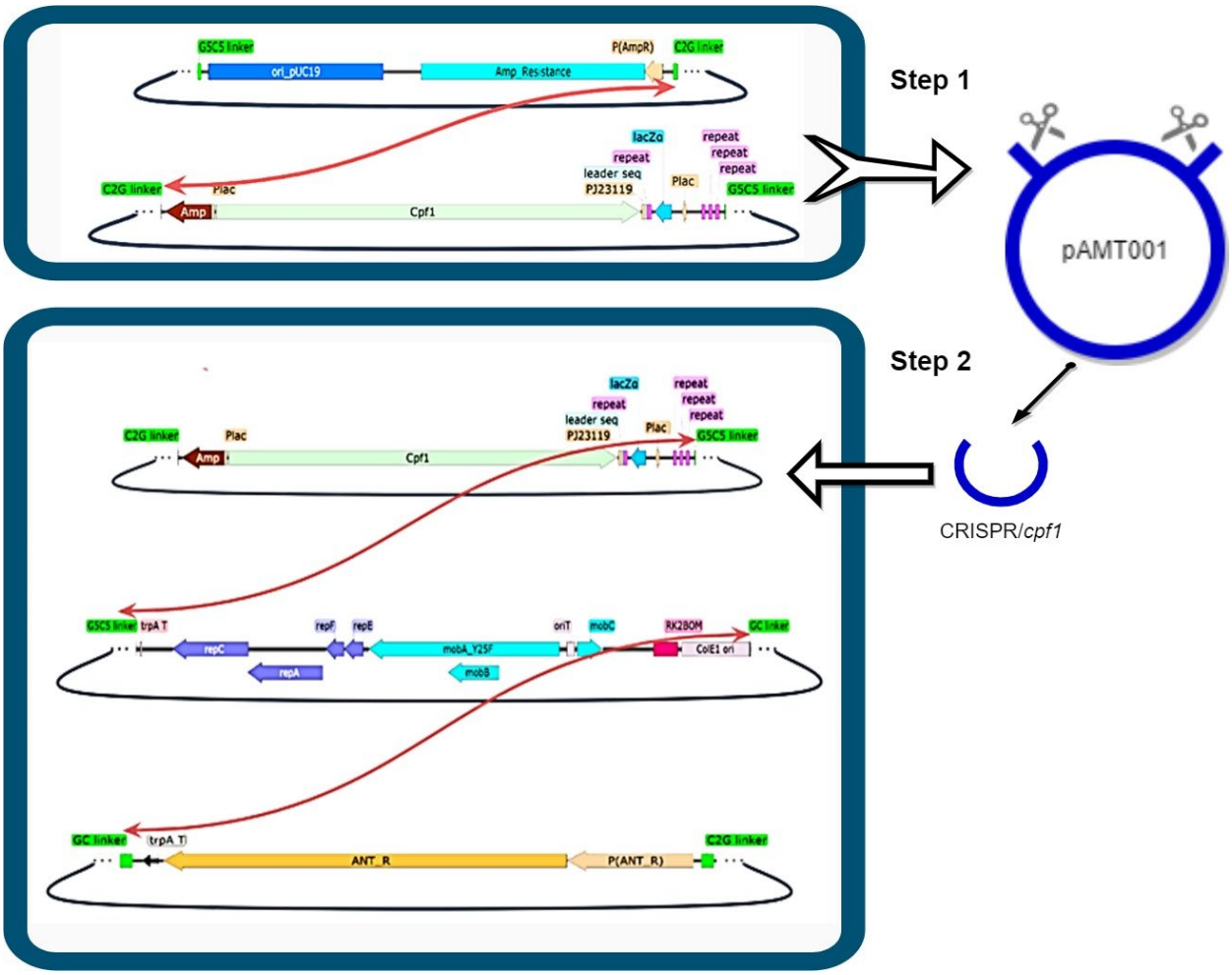


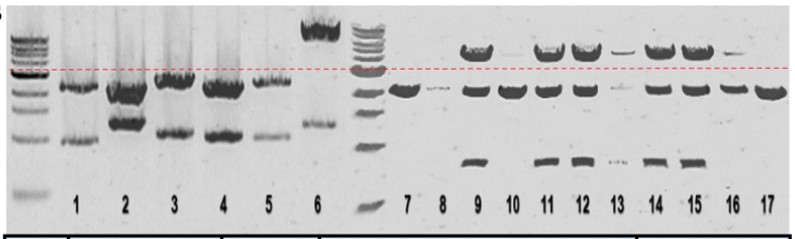
Figure 3.5A Schematic representation of the plasmid construction process, starting with assembly of the CRISPR-CpfI module with pCVD015 backbone, followed by release of the module and reassembly with linearized pAM5406 and 5 different antibiotic resistance markers isolated from pCVD001, pCVD002, pCVD003, pCVD007, and pCVD010. Plasmid map of one of the final constructs, pAMT003, carrying the CRISPR-CpfI module in an improved RSF1010 backbone is depicted in figure 3.5E.

3.4.2 Marker-less deletion of *ago* and *cas4* – Strain Construction

ago deletion plasmid (pAMT010) was transformed into the conjugal donor *E. coli* strain AM1359. To introduce the deletion, we performed biparental conjugation with WT *S. elongatus* which were plated on BG-11 plates with Sp + Sm (Figure 3.6A). Isolated exconjugants were first streaked on selective medium to force segregation, picked, and then patched and inoculated in 5 mL of BG-11 media. After approximately 4 days, the liquid cultures were PCR screened for the segregated SeAgo deletion. PCRs were carried out on genomic DNA as described in the materials and methods section with primers that annealed upstream and downstream of the homologous repair template fragments (primers detailed in Table 2.2). Three segregated clones

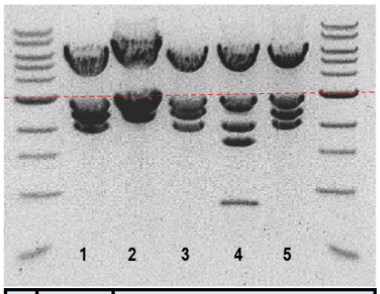
were selected and dilutions were plated on BG-11 agar to obtain single colonies that were picked and streaked on both selective and non-selective BG-11 agar plates to obtain an *ago* deletion clone that has been cured of the CRISPR plasmid. We performed triparental conjugation to introduce a deletion of *cas4* in *S. elongatus* WT and *ago* deletion backgrounds, we performed a double biparental conjugation in either *S. elongatus* WT or Δ*ago* with two distinct conjugal donor *E. coli* strains, carrying either the RSF1010-based CRISPR plasmid encoding the RNA guide or the pRL278 plasmid encoding the homologous repair template. The two conjugal donor *E. coli* strains also carry the helper and conjugal plasmids pRL623 and pRL443 (Figure 3.6B). Following the mating process, in which both conjugal donors are conjugated simultaneously, isolated exconjugants were grown and screened as previously detailed for SeAgo. The triparental conjugation protocol for the deletion of *cas4* is described in the Materials and Methods section.

B



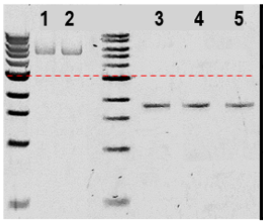
#	Plasmid	AR	Size (bp)	R.E.
1	pCVD001	Gm	2347; 966	EcoRV
2	pCVD002	Sp, Sm	2347; 1276	EcoRV
3	pCVD003	Km	2692; 1053	EcoRV
4	pCVD007	Cm	2347; 972	EcoRV
5	pCVD010	Nt	2347; 936	EcoRV
6	pAM5406	Ap	6880; 1117; 115	ZraI
7-17	AMT001	Ap	4481; 2196; 807	XmnI

D



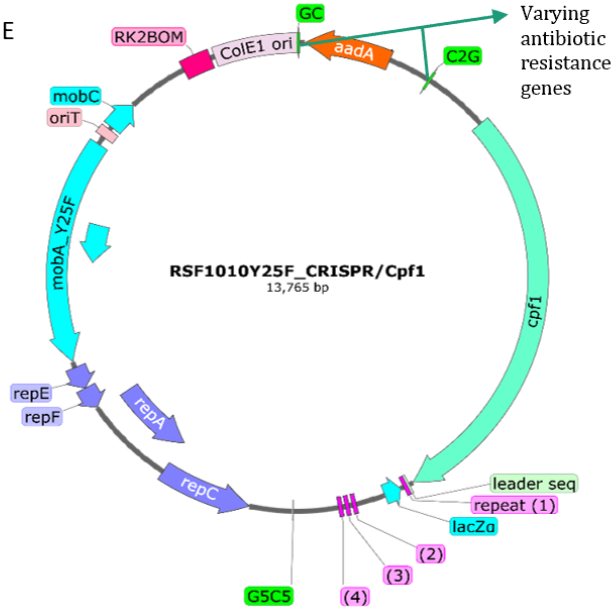
#	AR	XmnI - Size (bp)
1	Gm	5715; 3033; 2511; 2196.
2	Sp, Sm	5715; 3033; 2821; 2196.
3	Km	5715; 3033; 2598; 2196.
4	Cm	5715; 3033; 2196; 1714; 803.
5	Nt	5715; 3033; 2481; 2196.

C



#	DNA	Product	Size (bp)
1-2	pSL2680	CRISPR +Cpf1	5526
3-5	pCVD015	pUC19 ori	1878

E



F

Plasmid ID	[DNA] ng/µl	AR	R.E. 1	R.E. 2
AMT002	165.9	Cm	AarI	KpnI
AMT003	439.0	Sp, Sm	AarI	KpnI
AMT004	381.6	Km	AarI	KpnI
AMT005	134.5	Nt	AarI	KpnI
AMT006	499.4	Gm	AarI	KpnI
pSL2680	53.6	Sp, Sm	AarI	KpnI

Figure 3.5(B-F) Construction of improved CRISPR-CpfI gene editing plasmids. (B, C) Gel images of the tool assembly fragments either digested (B) or amplified by PCR (C). (D) gel image of restriction digest screening with XmnI to confirm the identity of the assembled plasmids. (E) Plasmid map of AMT003 carrying the CRISPR-CpfI module in an improved RSF1010 backbone and encoding the Sp+Sm resistance gene. (F) DNA concentrations resulting from plasmid extractions from a 2 mL overnight culture of the improved CRISPR plasmids and the alternative CRISPR tool pSL2680.

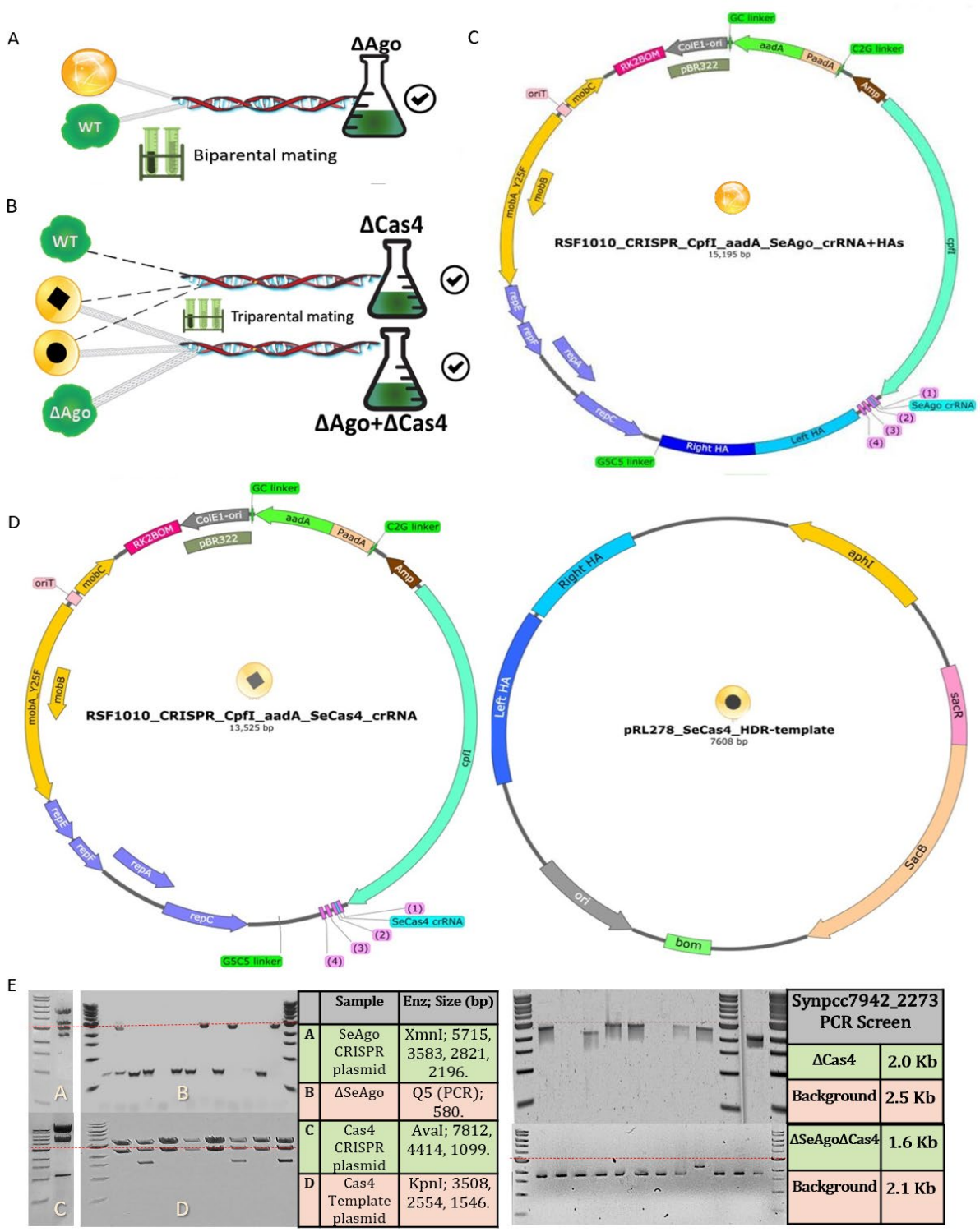


Figure 3.6 Marker-less deletion of SeAgo and SeCas4. (A-B) Schematic representation of the biparental and triparental conjugations performed to obtain Δ*ago* (A), Δ*cas4,* and Δ*ago*Δ*cas4* (B). (C-D) Plasmid map of CRISPR plasmids carrying either the SeAgo (C) or SeCas4 (D) guide RNA within the CRISPR array, *cpf1* gene, and the homologous repair template (C). (D) The plasmid map on the right displays a modified pRL278 containing the SeCas4 homologous repair template and Km resistance gene. (E) Gel images of the digests (left) confirming the assembly of pAMT010, pAMT011, and pAMT012 as well as a PCR screen (right) of the _1534 and _2273 loci used to verify segregation.

3.4.3 Complementation of SeAgo and SeCas4

To complement *ago* and *cas4* deletions in *S. elongatus* we initially attempted to construct plasmids in which the genes were driven by a Ptrc promoter. However, these cloning experiments repeatedly failed, which we assumed stemmed from the leaky expression of the Ptrc promoter system with the potential toxicity of SeAgo and SeCas4 in *E. coli*. Therefore, we decided to clone *ago* and *cas4* individually and together (for complementation of double IKO) into pAM5467 under control of a T7 promoter, which is inactive in the absence of the T7 RNA polymerase. We used a T7 expression system previously established for *S. elongatus* in the work by Roulet et al. 2017, that relies on the expression of the T7 RNA polymerase from a weak promoter at NS1, while the gene of interest driven by the T7 promoter is integrated at NS2. The translation of the T7 RNA polymerase is controlled by a minimally leaky theophylline inducible riboswitch[58] . The addition of theophylline results in toxicity and thus for the experiments presented in this study, we relied on the minimal leaky expression of T7 RNA polymerase to sufficiently express *ago.* Using this strategy, we were able to construct the desired plasmids (pAMT007, pAMT008, pAMT009) and the corresponding *S. elongatus* complementation and over-expression strains (AMT007, AMT008, AMT009, AMT014, AMT015, AMT017, AMT019, AMT020).

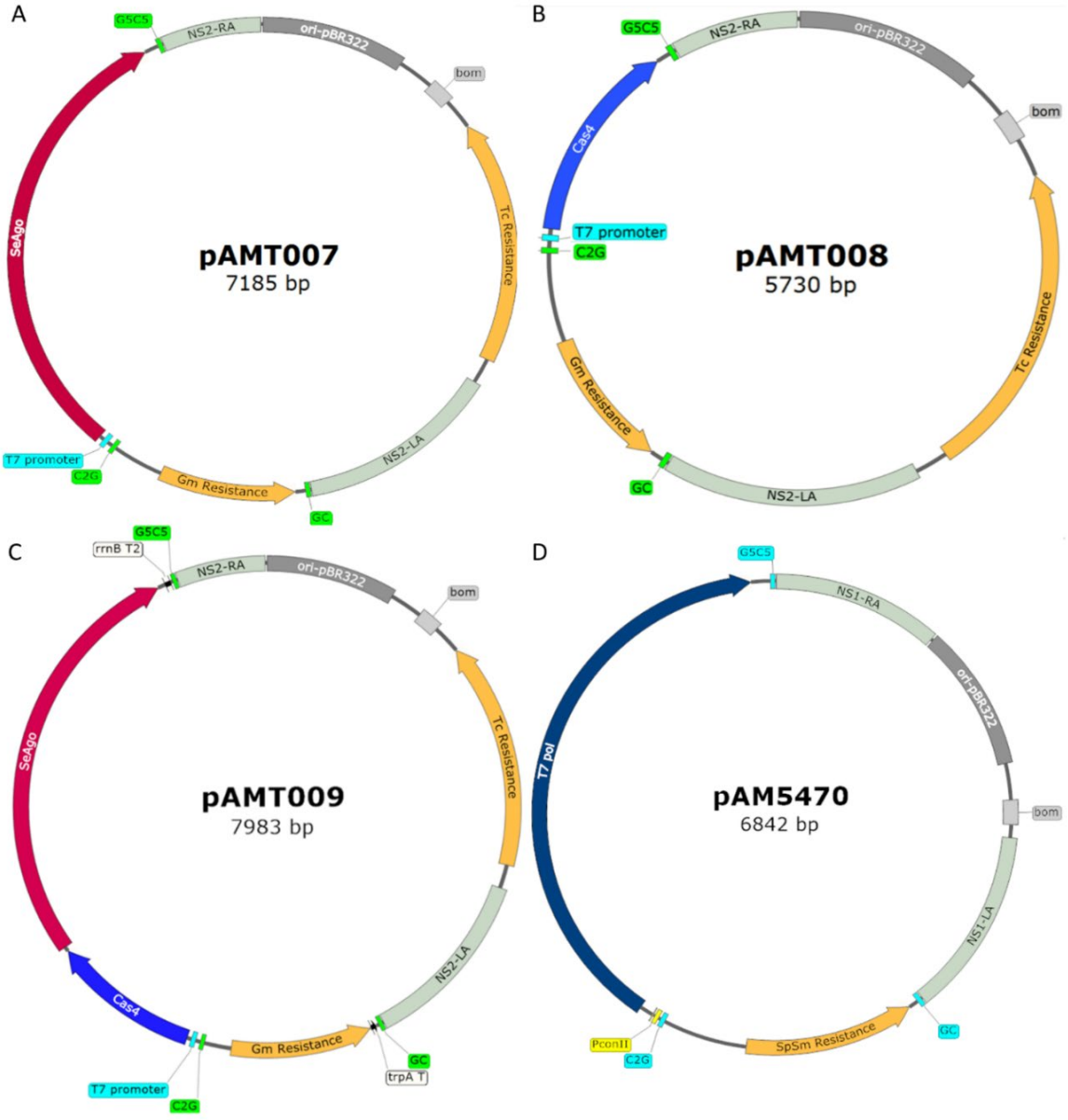


Figure 3.7 *ago* and *cas4* T7 Complementation Plasmid Maps. (A-C) Plasmid map of modified pAM5204 carrying Gm resistance and either *ago* (A), *cas4* (B), or both (C) driven by a T7 promoter for integration in NS2. (D) Plasmid map of pAM5470 carrying Sp+Sm resistance gene and T7 RNA polymerase driven by conII promoter for integration in NS1. Plasmid maps were generated with SnapGene.

3.5 SeAgo and RSF1010

3.5.1 RSF1010 in *S. elongatus*

Isolated from *E. coli*, the self-mobilizable RSF1010 is an IncQ-group plasmid, featuring autonomous replication in a broad range of hosts, including some cyanobacteria[53,57]. Many

RSF1010-based gene editing tools have emerged in recent years due to its broad-host range[53,57]. However, the transfer and maintenance of RSF1010-based plasmids in *S. elongatus* is normally unsuccessful or results in the formation of only a few small colonies. The mechanism which prevents the stable transfer of RSF1010-based plasmids in *S. elongatus* is unknown. However, based on results that suggested that prokaryotic argonautes may control intracellular plasmid concentrations [59], alongside we the idea that pAgos comprise an immune system against mobile elements [59], we hypothesized that SeAgo may be targeting RSF1010, thus preventing either its stable transfer, replication, or both.

In preliminary investigations, we compared the conjugation efficiency of RSF1010-based plasmids in 3 strains of *S. elongatus* including the WT and insertion knockouts of *ago* and *cas4* (*ago*::Tn5[8S1-HH1] and *cas4*::Mu[2D4-B4]). *ago* IKO, *cas4* IKO, and WT strains were grown for several days in the same conditions, aliquots of these cultures were then concentrated and adjusted to the same amount of cells and then conjugated with AM1359, carrying the RSF1010-based plasmid pAM5407 (RSF1010-AADA-PconII-oRBS-YFP). The cell mixtures were then serially diluted, plated on BG-11 medium supplemented with Sp+Sm and incubated at 30°C and 150-µmol photons m^{-2} s^{-1} for 5 days.

Unexpectedly, these experiments did not result in isolated colonies but lawns that covered the plates. For the *ago* and *cas4* IKOs, lawns were observed until the highest dilutions, and the lawns appeared thicker for the *ago* IKO. In contrast, the conjugation of pAM5407 in the WT resulted in lawns for the most concentrated samples only (Figure 3.8A). To obtain isolated clones, material from cyanobacterial lawns were streaked on selective plates. Each re-streak of the *ago* and *cas4* IKOs led to isolated colonies (Figure 3.8A) and while this approach was not quantitative, it appears that the conjugation of pAM5407 in the *ago* IKO led to more exconjugants. After streaking the WT exconjugants, one of the inoculated isolated colonies appeared to stabilize in liquid culture and began to grow in manner similar to the *ago* IKO. We hypothesize that this clone is likely a result of an acquired secondary mutation which allowed for the intracellular stabilization of pAM5407. Thus, we designated this clone as WT* (Figure 3.8C).

To examine whether pAM5407 is maintained as a stable independent replicon or integrated into *S. elongatus* chromosome, isolated exconjugants were inoculated in liquid cultures and

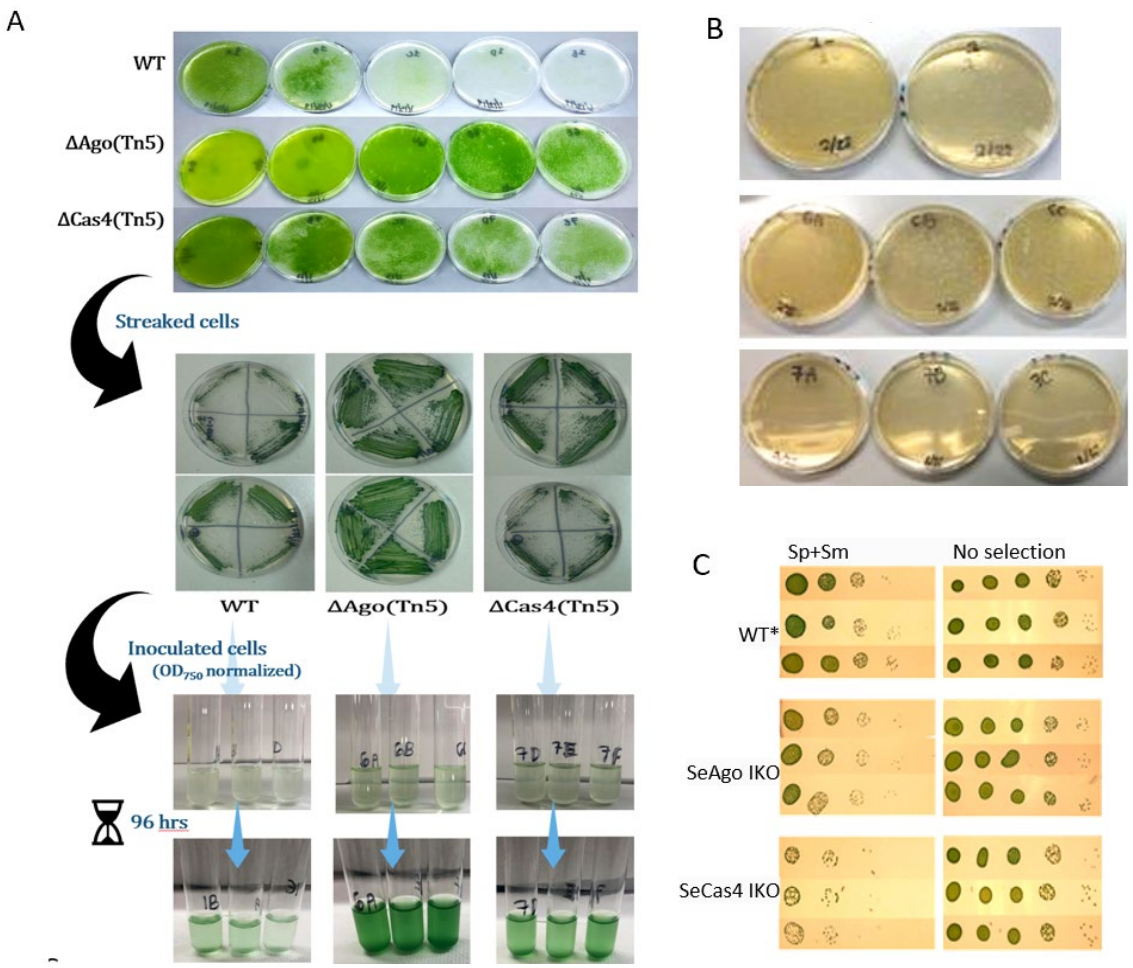


Figure 3.8 RSF1010 replication and stability in *S. elongatus.* (A) schematic representation of the conjugation of pAM5407 into WT *S. elongatus*, SeAgo IKO, and SeCas4 IKO. Exconjugants from each conjugation were streaked, normalized according to cell number, and inoculated in BG-11 for 4 days. (B) Transformation spot-assay of pAM5407 in triplicate cultures of WT, *ago* IKO, and *ago* complementation strains. (C) Counter screen of exconjugants to evaluate plasmid loss after two weeks in non-selective media. (D) Retransformation of genomic DNA extracted from 3 isolated exconjugants of each strain into DH5a to assess the intracellular destination of pAM5407 in *S. elongatus.* All transformations produced colonies except with DNA from the single WT clone.

genomic DNA was extracted followed by transformation into *E. coli* strain DH5a (Figure 3.8B). We observed the most *E. coli* transformants with DNA from exconjugants of the *ago* IKO, some growth with the *cas4* IKO, and little to no growth with the WT strain. Retransformation of the genomic DNA extracted from 3 isolated clones of each strain resulted in *E. coli* colonies for all tested clones. The integrity of pAM5407 re-isolated from *E. coli* clones was verified by restriction digest (Figure 3.8B).

3.5.2 Conjugation of RSF1010 in the RB-TnSeq Library

To identify all *S. elongatus* mutants that can receive and maintain RSF1010-based plasmids, we conjugated pAM5407 into the *S. elongatus* RB-TnSeq (random bar code transposon-site sequencing) library, comprised of ~250,000 transposon mutants[54]. RB-TnSeq combines bar code sequencing (BarSeq) and transposon mutagenesis with next-generation sequencing (TnSeq) for high throughput assays of mutant fitness using Tn5 transposons with integrated random DNA bar codes [60] (Figure 3.9). Analysis of mutant fitness through RB-TnSeq includes the assessment of DNA bar code abundance (BarSeq counts) at time zero (T_0), prior to selective conditions, in comparison to post selective growth [60] (Figure 3.9C). A schematic representation of RB-TnSeq is diagramed in Figure 3.9 [60]. For the purpose of this study, conjugation of the *S. elongatus* RB-TnSeq library was performed with pAM5407 based on RSF1010 as well as pAMT025 based on the endogenous replicon pANS as a positive control. The experimental procedure is described in the Material and Methods section. The analysis provided fitness values to characterize the effect of loss of function in individual loci on the ability to conjugate and replicate RSF1010-based plasmids.

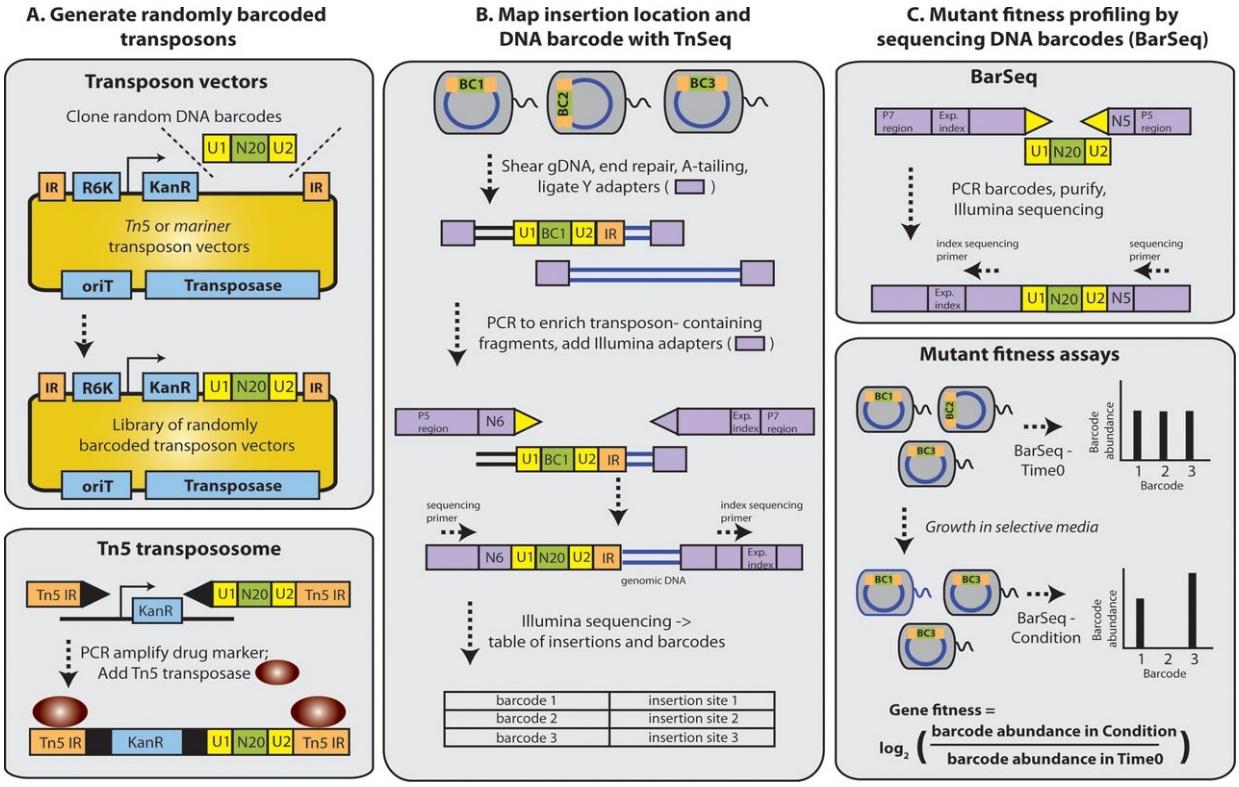


Figure 3.9 Schematic representation of the RB-TnSeq approach for the evaluation of gene fitness in bacteria. Figure source: Wetmore et al., 2015.

Similar to the previously described RB-TnSeq screen of natural competence fitness in *S. elongatus,* SeAgo displayed the highest fitness for conjugation with pAM5407 (Figure 3.10). As with natural competence, SeCas4 showed a positive fitness, though not as high as SeAgo (Figure 3.10). Three other genes had positive fitness after conjugation with pAM5407, including two genes involved in DNA processing and repair (*uvrD* and *recJ)* and the *rpoD2* gene which encodes an RNA polymerase sigma factor (Figure 3.10). Previous studies showed that TtAgo co-purified with both RecJ and UvrD[61] and predicted that that the interaction between these three proteins involves the generation of TtAgo guides at the end of circular chromosome replication from stalled replication forks that accumulate at this region [61]. As TtAgo represents the closest homolog of SeAgo, we hypothesize that RecJ and UvrB may also interact with SeAgo in the guide acquisition process.

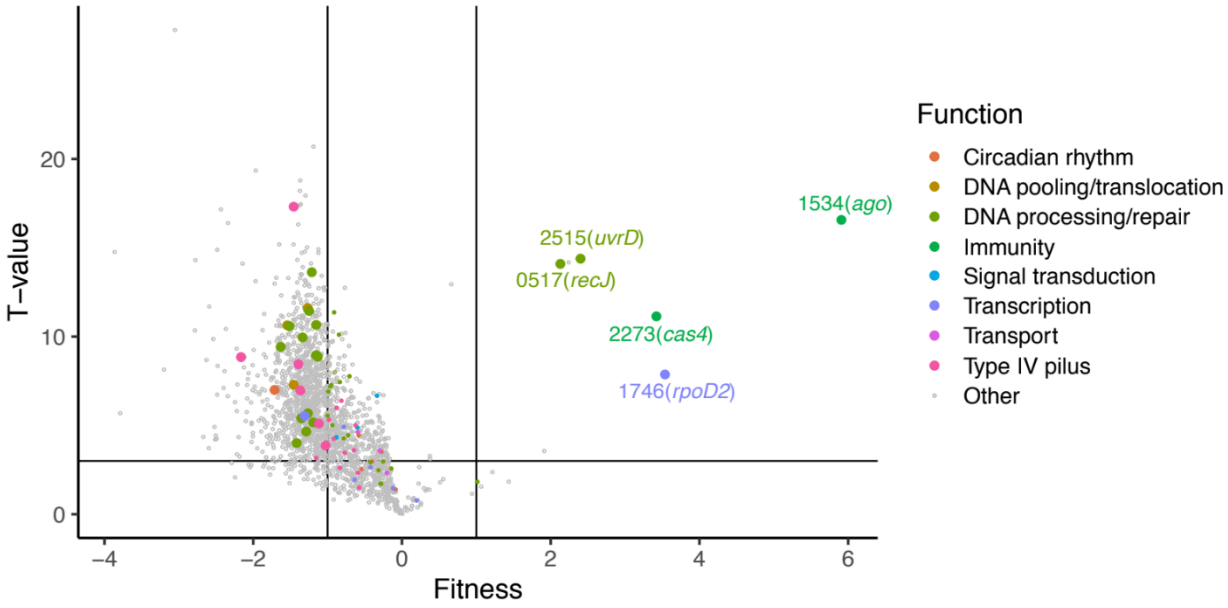


Figure 3.10 RB-TnSeq library analysis of fitness values and confidence levels associated with barcoded transposon mutants of *S. elongatus* for conjugation with pAM5407. A low fitness with a high T-value characterizes mutants that are deficient for conjugation with pAM5407 while a high fitness with high T-value describes mutants that improve conjugation with pAM5407. The data points were labeled the gene names in parentheses in addition to the *S. elongatus* locus tag (Synpcc7942_) numbers.

Chapter 4. Conclusion

Prokaryotic argonaute proteins are involved in mechanisms of host defense through the processing of foreign DNA by binding short nucleic acid guides that allow for sequence specific recognition of complementary targets [47]. The argonaute proteins are widely distributed across the three domains of life[40]. While usually annotated as "hypothetical proteins", argonaute homologs are found in cyanobacterial genomes. Phylogenetic analysis showed an infrequent and sporadic distribution of argonaute proteins within the cyanobacterial phylum. Previously, a study of natural competence in *S. elongatus* using a RB-TnSeq library revealed that the loss of function of the Synpcc7942_1534 gene [9,54], identified as an argonaute homolog, positively affected natural transformation with the highest fitness score. Through sequence and structural analyses, we determined that SeAgo is structurally and functionally homologous to the widely characterized TtAgo from *T. thermophilus*[36,52]. Both proteins present a conserved catalytic tetrad within the PIWI domain, providing evidence for intrinsic nucleolytic function. Our hypothesis for the catalytic activity of SeAgo was recently confirmed by Olina et al. [36] through the examination of the *in vitro* and *in vivo* activity of SeAgo. SeAgo autonomously generates short DNA guides and preferentially cleaves complementary DNA targets.[44,48] SeAgo has been shown to affect gene transfer processes and is predicted to function as an immune system in *S. elongatus* via DNA binding and cleavage pathways. This DNA processing implies the capacity to defend against mobile genetic elements such as phage infection by destruction of the invader DNA to prevent phage multiplication or through the processing of the host chromosome to initiate cell suicide [44–46].

The discovery of a catalytically active pAgo in *S. elongatus* has potential biotechnological applications. Although distinct in their mechanism of action, the functional analogy of catalytically active pAgos to CRISPR-Cas systems [45,46] has recently begun to attract considerable interest in the field of biotechnology and genetic engineering. The first commercialized Ago gene editing kit, which uses TtAgo, is now sold by New England BioLabs (NEB) [36]. TtAgo ago activity ranges from 75 to 80°C and is drastically reduced at temperatures below 70°C [36]. *S. elongatus* grows at much lower temperatures than the thermophilic strain *T. thermophilus* and cleavage of plasmid DNA by SeAgo was observed with similar efficiencies at 37°C, 45°C, and 53°C [36]. Thus, SeAgo may be a better substitute to TtAgo for *in vivo* and *in*

vitro experimental applications [55,56]. Argonaute-based marker-less gene editing technologies are a novel alternative to CRISPR-based genetic engineering. The ability to utilize pAgos for sequence-specific DNA cleavage *in vitro* is based on evidence of pAgos function in the processing of genetic elements such as plasmid DNA[53]. The capacity for DNA cleavage *in vivo* implies the involvement of pAgos in DNA transfer processes such as natural transformation. Quantitative transformation assays of *S. elongatus* at circadian times when the cells are the most or the least competent confirmed that the loss of function of SeAgo produced a substantial increase in the transformation efficiency. Therefore, pAgo innate immune systems provide host defense not only from phage infection but also plasmid DNA.

Broad-host range self-replicating plasmids such as RSF1010 are important vectors of gene transfer. However, the transfer of RSF1010-based plasmids in *S. elongatus* is not efficient. This guided us to investigate whether the predicted function of SeAgo in processing of plasmid DNA inhibits the stable transfer of broad host range RSF1010-based plasmids in *S. elongatus*. Through transformation and conjugation assays of RSF1010-based plasmids in knockout and complemented strains of SeAgo, evidence of increased conjugation efficiency, exconjugant stability, and highly efficient transformation of RSF1010 led us to conclude that SeAgo blocks the maintenance of RSF1010-based plasmids in *S. elongatus*. The identification of an *S. elongatus* mutant capable of the transfer and stable replication of RSF1010 is significant for the advancement of CRISPR-based gene editing in this model cyanobacterium.

Due to its broad range of hosts, the RSF1010 backbone has been used in development of CRISPR-based genetic engineering tools to allow marker-less gene editing in cyanobacteria [9]. However, the assortment of CRISPR tools is still limited and their RSF1010 backbone complicates genetic manipulation of this plasmid, due to low copy numbers and DNA-nicking activity involved in self-mobilization, which results in low amounts of DNA usable for cloning[53]. The need for marker-less deletion strains and the identification of *S. elongatus* proteins that hinder the transfer of RSF1010-based plasmids led us to construct an improved CRISPR-CpfI system. These novel tools feature a RSF1010 backbone modified to significantly increase the yield of usable DNA from plasmid purification procedures thereby facilitating subsequent cloning. As another improvement, these CRISPR tools were designed for

compatibility with the CYANO-VECTOR assembly system which enables the construction of CRISPR-Cpf1 plasmids with a set of available antibiotic devices and replicons.

Along with SeAgo, our study focused on a Cas4 homolog, which was also identified as a determinant for positive transformation in *S. elongatus* [9] . Bioinformatics of SeCas4 indicate catalytic activity targeting dsDNA, as it belongs to the RecB_C superfamily [48]. The functional homology between SeAgo and SeCas4 may be indicative of a synergistic relationship between the two proteins [36]. Preliminary analyses of SeCas4 KOs were not conclusive in terms of transformation efficiency or a possible synergistic role of Cas4 with pAgos. Although we believe these results may stem from technical problems, a set of in-frame marker-less deletion strains for both genes and their complementation strains was constructed to enable future studies. Interestingly, a SeCas4 KO strain also appears to better receive RSF1010-based plasmids and will be discussed later. One of the least understood aspects of prokaryotic argonautes is their association and interactions with predicted cellular nucleases, helicases, and other accessory factors and proteins. Studies of prokaryotic argonautes have previously established that pAgos typically associate with helicases, DNA binding proteins, and nucleases from the SIR2, TIR, PLD, Mrr, or Cas4 families [37,40]. SeAgo has specifically been shown to associate with several unidentified accessory factors and proteins [9]. Along with bioinformatic studies on the distribution of the ago gene in the prokaryotic kingdom, a previous study on natural transformation suggests that SeCas4 could potential associate with SeAgo[9]. Preliminary experiments could not resolve a relationship between SeAgo and SeCas4 but did indicate a role of SeCas4 in preventing the stable transfer of RSF1010-based plasmid in *S. elongatus*.

To identify all genes whose loss of function could enable the stable transfer of RSF1010-based plasmids in *S. elongatus* and thereby be candidate proteins that could interact with SeAgo, we conjugated an RSF1010-based plasmid into the RB-TnSeq library [54]. This analysis confirmed the critical role of SeCas4 but also revealed three additional proteins, including UvrD and RecJ, which are involved in DNA processing and repair, as well as the sigma factor RpoD2[61]. Based on previous reports, TtAgo co-purified with both RecJ and UvrB and in vitro studies found that the helicase UvrD increased the catalytic activity of TtAgo[61]. Considering the significant structural homology and domain conservation between TtAgo and SeAgo, our

results suggest that UvrB and RecJ are likely involved in the generation and loading of guides for SeAgo in a similar manner to *T. thermophilus*.

The experimental characterization of prokaryotic argonautes is in its early stage, and the lack of published studies and poor annotations in sequence databases pose significant challenges to advancing our scientific understanding of these ancient, diverse, and complex *defense* systems. The association of some pAgos with other proteins, such as the Cas4-like nuclease, has been demonstrated in several studies of prokaryotic argonaute proteins, though most of these associated proteins have not been identified and the specific interaction mechanisms are unknown [36]. In the recent study of Olina et al.[36], the growth rate of a SeAgo KO mutant and complemented strains was investigated leading to the conclusion that SeAgo does not significantly affect cell physiology [36]. While this initial study presents the first characterization of the catalytically active SeAgo argonaute in the model cyanobacterium *S. elongatus* PCC 7942 [36], we believe that further investigations is required to assess the role of SeAgo on cell physiology. For example, when cells are plated after transformation procedures, we consistently observed that SeAgo KO strains produced visible colonies earlier than the *S. elongatus* WT strain, in addition to an increased number of transformants. The identification of a catalytically functional argonaute protein in the well-established model organism *S. elongatus*, along with our preliminary results and observations and improved CRISPR tools reveal new information and opportunities for the further study of argonaute proteins in prokaryotes.

BIBLIOGRAPHY

1. Nabout JC, da Silva Rocha B, Carneiro FM, Sant'Anna CL. How many species of Cyanobacteria are there? Using a discovery curve to predict the species number. *Biodiversity and Conservation*. 2013;22(12):2907-2918. doi:10.1007/s10531-013-0561-x

2. Hamilton TL, Bryant DA, Macalady JL. The role of biology in planetary evolution: cyanobacterial primary production in low-oxygen Proterozoic oceans. 18(2):325-340. doi:10.1111/1462-2920.13118

3. Schirrmeister BE, Gugger M, Donoghue PC. Cyanobacteria and the Great Oxidation Event: evidence from genes and fossils. 58(5):935-936. doi:10.1111/pala.12193

4. Brauer VS, Stomp M, Bouvier T, Fouilland E, Leboulnger C, Confurius-Guns V, Weissing F, StalL, Huisman J. Competition and facilitation between the marine nitrogen-fixing cyanobacterium Cyanothece and its associated bacterial community. *Frontiers in Microbiology*. 2014;5(DEC):795. doi:10.3389/fmicb.2014.00795

5. Demay J, Bernard C, Reinhardt A, Marie B. Natural Products from Cyanobacteria: Focus on Beneficial Activities. 17(6):320. doi:10.3390/md17060320

6. Shah S, Akhter N, Auckloo B, Khan I, Lu Y, Wang K, Wu B, Guo Y. Structural Diversity, Biological Properties and Applications of Natural Products from Cyanobacteria. A Review. 15(11):354. doi:10.3390/md15110354

7. Wood R. Acute animal and human poisonings from cyanotoxin exposure - A review of the literature. 91(26995270):276-282. doi:10.1016/j.envint.2016.02.026

8. Broddrick JT, Rubin BE, Welkie DG, Du N, Mih N, Diamond S, Lee J, Golden J, Susan S

   Palsson, Bernhard O. Unique attributes of cyanobacterial metabolism revealed by improved genome-scale metabolic modeling and essential gene analysis. 113(51):E8344–E8353. doi:10.1073/pnas.1613446113

9. Taton A, Erikson C, Yang Y, Rubin, Benjamin E, Rifkin S, Golden JW, Golden SS. The circadian clock and darkness control natural competence in cyanobacteria. 11(1). doi:10.1038/s41467-020-15384-9

10. Golden SS. The international journeys and aliases of Synechococcus elongatus. *New Zealand Journal of Botany*. 2019;57(2):70-75. doi:10.1080/0028825X.2018.1551805

11. Williams JGK, Szalay AA. Stable integration of foreign DNA into the chromosome of the cyanobacterium Synechococcus R2. 24(1):37-51. doi:10.1016/0378-1119(83)90129-4

12. Tsinoremas NF, Kutach AK, Strayer CA, Golden SS. Efficient gene transfer in Synechococcus sp. strains PCC 7942 and PCC 6301 by interspecies conjugation and chromosomal recombination. *Journal of Bacteriology*. 1994;176(21):6764-6768. doi:10.1128/jb.176.21.6764-6768.1994

13. Villa TG, Feijoo-Siota L, Sánchez-Pérez A, Rama JLR, Sieiro C. Horizontal Gene Transfer in Bacteria, an Overview of the Mechanisms Involved. In: Villa TG, Viñas M, eds. *Horizontal Gene Transfer*. Springer International Publishing; 2019:3-76. doi:10.1007/978-3-030-21862-1_1

14. Zhaxybayeva O. Phylogenetic analyses of cyanobacterial genomes: Quantification of horizontal gene transfer events. 16(9):1099-1108. doi:10.1101/gr.5322306

15. Ponce-Toledo RI, López-García P, Moreira D. Horizontal and endosymbiotic gene transfer in early plastid evolution. *New Phytologist*. 2019;224(2):618-624. doi:10.1111/nph.15965

16. Villa TG, Feijoo-Siota L, Sánchez-Pérez A, Rama JLLR, Sieiro C. Horizontal Gene Transfer in Bacteria, an Overview of the Mechanisms Involved. In: Villa TG, Viñas M, eds. *Horizontal Gene Transfer*. Springer International Publishing; 2019:3-76. doi:10.1007/978-3-030-21862-1_1

17. Moreira D, Philippe H. Sure facts and open questions about the origin and evolution of photosynthetic plastids. 152(9):771-780. doi:10.1016/s0923-2508(01)01260-8

18. Golden SS, Sherman LA. Optimal conditions for genetic transformation of the cyanobacterium Anacystis nidulans R2. 158(1):36. http://jb.asm.org/content/158/1/36.abstract

19. Chen Y, Holtman CK, Taton A, Golden SS. Functional Analysis of the Synechococcus elongatus PCC 7942 Genome. In: Burnap R, Vermaas W, eds. Springer Netherlands; 2012:119-137. doi:10.1007/978-94-007-1533-2_5

20. Elhai J, Wolk CP. Conjugal transfer of DNA to cyanobacteria. 167:747-754. doi:10.1016/0076-6879(88)67086-8

21. Padan E, Shilo M. Cyanophages-viruses attacking blue-green algae. 37(3):343-370. doi:10.1128/mmbr.37.3.343-370.1973

22. Suttle CA. Cyanophages and Their Role in the Ecology of Cyanobacteria. In: *The Ecology of Cyanobacteria*. ; 2006. doi:10.1007/0-306-46855-7_20

23. Gao E bin, Huang Y, Ning D. Metabolic genes within cyanophage genomes: Implications for diversity and evolution. *Genes*. Published online 2016. doi:10.3390/genes7100080

24. Liu R, Liu Y, Chen Y, Zhan Y, Zeng Q. Cyanobacterial viruses exhibit diurnal rhythms during infection. *Proceedings of the National Academy of Sciences of the United States of America*. Published online 2019. doi:10.1073/pnas.1819689116

25. Kao CC, Green S, Stein B, Golden SS. Diel infection of a cyanobacterium by a contractile bacteriophage. 71(816085814):4276-4279. doi:10.1128/AEM.71.8.4276-4279.2005

26. van Houte S, Buckling A, Westra ER. Evolutionary Ecology of Prokaryotic Immune Mechanisms. *Microbiology and Molecular Biology Reviews*. Published online 2016. doi:10.1128/mmbr.00011-16

27. Bikard D, Marraffini LA. Innate and adaptive immunity in bacteria: Mechanisms of programmed genetic variation to fight bacteriophages. *Current Opinion in Immunology*. Published online 2012. doi:10.1016/j.coi.2011.10.005

28. Bondy-Denomy, J., Qian, J., Westra, E. R., Buckling, A., Guttman, D. S., Davidson, A. R., & Maxwell, K. L. (2016). Prophages mediate defense against phage infection through diverse mechanisms. ISME Journal, 10(12), 2854–2866. https://doi.org/10.1038/ismej.2016.79

29. Labrie SJ, Samson JE, Moineau S. Bacteriophage resistance mechanisms. *Nature Reviews Microbiology*. Published online 2010. doi:10.1038/nrmicro2315

30. Oliveira PH, Touchon M, Rocha EPC. The interplay of restriction-modification systems with mobile genetic elements and their prokaryotic hosts. *Nucleic Acids Research*. Published online 2014. doi:10.1093/nar/gku734

31. Jeltsch A, Pingoud A. Horizontal gene transfer contributes to the wide distribution and evolution of type II restriction-modification systems. *Journal of Molecular Evolution*. Published online 1996. doi:10.1007/BF02198833

32. Gasiunas G, Sinkunas T, Siksnys V. Molecular mechanisms of CRISPR-mediated microbial immunity. *Cellular and Molecular Life Sciences*. Published online 2014. doi:10.1007/s00018-013-1438-6

33. Cai F, Axen SD, Kerfeld CA. Evidence for the widespread distribution of CRISPR-Cas system in the Phylum Cyanobacteria. *RNA Biology*. Published online 2013. doi:10.4161/rna.24571

34. Bradde S, Vucelja M, Teşileanu T, Balasubramanian V. Dynamics of adaptive immunity against phage in bacterial populations. *PLoS Computational Biology*. Published online 2017. doi:10.1371/journal.pcbi.1005486

35. Stucken K, Koch R, Dagan T. Cyanobacterial defense mechanisms against foreign DNA transfer and their impact on genetic engineering. *Biological Research*. Published online 2013. doi:10.4067/S0716-97602013000400009

36. Olina A, Kuzmenko A, Ninova M, Aravin AA, Kulbachinskiy A, Esyunina D. Genome-wide DNA sampling by Ago nuclease from the cyanobacterium Synechococcus elongatus. *RNA Biology*. Published online 2020. doi:10.1080/15476286.2020.1724716

37. Ryazansky S, Kulbachinskiy A, Aravin AA. The Expanded Universe of Prokaryotic Argonaute Proteins. Papoutsakis ETE, ed. 9(6). doi:10.1128/mbio.01935-18

38. Höck J, Meister G. The Argonaute protein family. 9(2):210. doi:10.1186/gb-2008-9-2-210

39. Ender C, Meister G. Argonaute proteins at a glance. 123(11):1819-1823. doi:10.1242/jcs.055210

40. Swarts, D. C., Makarova, K., Wang, Y., Nakanishi, K., Ketting, R. F., Koonin, E. V., Patel, D. J., & Van Der Oost, J. (2014). The evolutionary journey of Argonaute proteins. Nature Structural and Molecular Biology, 21(9), 743–753. https://doi.org/10.1038/nsmb.2879

41. Olina A v., Kulbachinskiy A v., Aravin AA, Esyunina DM. Argonaute Proteins and Mechanisms of RNA Interference in Eukaryotes and Prokaryotes. *Biochemistry (Moscow)*. 2018;83(5):483-497. doi:10.1134/S0006297918050024

42. Zhai, L., Wang, L., Teng, F., Zhou, L., Zhang, W., Xiao, J., Liu, Y., & Deng, W. (2016). Argonaute and argonaute-bound small RNAs in stem cells. In International Journal of Molecular Sciences. https://doi.org/10.3390/ijms17020208

43. van der Oost J, Swarts DC, Jore MM. Prokaryotic argonautes – Variations on the rna interference theme. *Microbial Cell*. 2014;1(5):158-159. doi:10.15698/mic2014.05.144

44. Swarts, D. C., Jore, M. M., Westra, E. R., Zhu, Y., Janssen, J. H., Snijders, A. P., Wang, Y., Patel, D. J., Berenguer, J., Brouns, S. J. J., & Van Der Oost, J. (2014). DNA-guided DNA interference by a prokaryotic Argonaute. Nature, 507(7491), 258–261. https://doi.org/10.1038/nature12971

45. Hegge JW, Swarts DC, van der Oost J. Prokaryotic argonaute proteins: Novel genome-editing tools? *Nature Reviews Microbiology*. Published online 2018. doi:10.1038/nrmicro.2017.73

46. Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D. C., Van Der Oost, J., Patel, D. J., & Wang, Y. (2014). Structure-based cleavage mechanism of Thermusthermophilus argonaute DNA guide strand-mediated DNA target cleavage. Proceedings of the National Academy of Sciences of the United States of America, 111(2), 652–657. https://doi.org/10.1073/pnas.1321032111

47. Lisitskaya L, Aravin AA, Kulbachinskiy A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. *Nature Communications*. Published online 2018. doi:10.1038/s41467-018-07449-7

48. Makarova KS, Wolf YI, van der Oost J, Koonin E v. Prokaryotic homologs of Argonaute proteins are predicted to function as key components of a novel system of defense against mobile genetic elements. *Biology Direct*. Published online 2009. doi:10.1186/1745-6150-4-29

49. Swarts DC, Koehorst JJ, Westra ER, Schaap PJ, van der Oost J. Effects of argonaute on gene expression in Thermus thermophilus. *PLoS ONE*. 2015;10(4):e0124880. doi:10.1371/journal.pone.0124880

50. Hall TMT. Structure and function of argonaute proteins. *Structure*. 2005;13(10):1403-1408. doi:10.1016/j.str.2005.08.005

51. Sheng, G., Zhao, H., Wang, J., Rao, Y., Tian, W., Swarts, D. C., Van Der Oost, J., Patel, D. J., & Wang, Y. (2014). Structure-based cleavage mechanism of Thermusthermophilus argonaute DNA guide strand-mediated DNA target cleavage. Proceedings of the National Academy of Sciences of the United States of America, 111(2), 652–657. https://doi.org/10.1073/pnas.1321032111

52. Swarts, D. C., Szczepaniak, M., Sheng, G., Chandradoss, S. D., Zhu, Y., Timmers, E. M., Zhang, Y., Zhao, H., Lou, J., Wang, Y., Joo, C., & van der Oost, J. (2017).

Autonomous Generation and Loading of DNA Guides by Bacterial Argonaute. Molecular Cell. https://doi.org/10.1016/j.molcel.2017.01.033

53. Taton, A., Unglaub, F., Wright, N. E., Zeng, W. Y., Paz-Yepes, J., Brahamsha, B., Palenik, B., Peterson, T. C., Haerizadeh, F., Golden, S. S., & Golden, J. W. (2014). Broad-host-range vector system for synthetic biology and biotechnology in cyanobacteria. Nucleic Acids Research, 42(17), e136–e136. https://doi.org/10.1093/nar/gku673

54. Rubin, B. E., Wetmore, K. M., Price, M. N., Diamond, S., Shultzaberger, R. K., Lowe, L. C., Curtin, G., Arkin, A. P., Deutschbauer, A., & Golden, S. S. (2015). The essential gene set of a photosynthetic organism. Proceedings of the National Academy of Sciences of the United States of America. https://doi.org/10.1073/pnas.1519220112

55. Ungerer J, Pakrasi HB. Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. *Scientific Reports*. Published online 2016. doi:10.1038/srep39681

56. Behler J, Vijay D, Hess WR, Akhtar MK. CRISPR-Based Technologies for Metabolic Engineering in Cyanobacteria. *Trends in Biotechnology*. Published online 2018. doi:10.1016/j.tibtech.2018.05.011

57. Bishé B, Taton A, Golden JW. Modification of RSF1010-Based Broad-Host-Range Plasmids for Improved Conjugation and Cyanobacterial Bioprospecting. *iScience*. Published online 2019. doi:10.1016/j.isci.2019.09.002

58. Ma AT, Schmidt CM, Golden JW. Regulation of gene expression in diverse cyanobacterial species by using theophylline-responsive riboswitches. *Applied and Environmental Microbiology*. Published online 2014. doi:10.1128/AEM.01697-14

59. Lisitskaya L, Aravin AA, Kulbachinskiy A. DNA interference and beyond: structure and functions of prokaryotic Argonaute proteins. *Nature Communications*. 2018;9(1):5165. doi:10.1038/s41467-018-07449-7

60. Wetmore, K. M., Price, M. N., Waters, R. J., Lamson, J. S., He, J., Hoover, C. A., Blow, M. J., Bristow, J., Butland, G., Arkin, A. P., & Deutschbauer, A. (2015). Rapid

quantification of mutant fitness in diverse bacteria by sequencing randomly bar-coded transposons. MBio. https://doi.org/10.1128/mBio.00306-15

61. Hunt EA, Evans TC, Tanner NA. Single-stranded binding proteins and helicase enhance the activity of prokaryotic argonautes in vitro. *PLoS ONE*. Published online 2018. doi:10.1371/journal.pone.0203073