Synthesizing and Deploying an Improved CRISPR Cas12a Knockout System for *Synechococcus elongatus sp. PCC 7942*

John H. Dorlon IV

Author

Dr. Rob Burnap

Principal Investigator and Thesis Reader

Dr. Sabrina Beckmann

Thesis Reader

Abstract

CRISPR mediated genome editing may be one of the most profound biochemical breakthroughs of this century. As more research laboratories begin to incorporate CRISPR systems into their repertoire, having a CRISPR system that is easy to use becomes more important. In this study, we developed and demonstrated the efficacy of a CRISPR Cas12a system containing GFP whose implementation is more facile than its predecessors. This system was created primarily through restriction digestions and Gibson Assembly and was implemented via *E. coli* transformations and tri-parental mating. Using *Synechococcus elongatus sp. PCC 7942* as our research strain, we demonstrated the efficacy of this system by deleting *ndhF4*, a part of the cyanobacterial CO_2 concentrating mechanism. In addition to developing a powerful tool for other genome editing, our Δ*NdhF4* strain can be used as a genetic background for structure/function studies of NdhF4.

Introduction

CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) genome editing systems are a revolutionary molecular genetic technique for precise genome editing in model scientific organisms (Niu et al., 2019); (Ma & Liu, 2015); (Ungerer & Pakrasi, 2016). Specifically, using various RNA molecules, CRISPR systems can be "programmed" to target and cut specific DNA sequences of a target gene, thus facilitating the creation of everything from a gene deletion mutant to site-directed mutants. However, CRISPR technologies are still relatively new and not all species of model organisms have a CRISPR system designed to work with them. Furthermore, many existing systems have room for optimization.

CRISPR was characterized and hypothesized to function as an adaptive immune system in prokaryotes against foreign genetic material, such as that from bacteriophages (Mojica et al., 2005). In nature, the CRISPR mechanism excises fragments of the invading viral genome to incorporate into its own genome in a so-called CRISPR array. This preserves the foreign sequence, which can be expressed to identify other instances of the same foreign genetic material. Copies of the excised portions can then guide the CRISPR nucleases to other occurrences of the viral DNA or RNA in the cell. Because the CRISPR array is integral to the genome, subsequent generations are afforded a degree of protection to the cognate viral species. The viral DNA or RNA recognized by the CRISPR system is cut by the CRISPR nuclease and rendered unable to hijack the cell. It is this "guiding" mechanism that can be commandeered for use in biological research laboratories. This guiding mechanism proved to be an important part of CRISPR gene editing systems, and by 2013, CRISPR had been used to make accurate point mutations in mice and in human cells (Cong et al., 2013). However, not all CRISPR systems are universally effective across species. For example, the relatively well-known CRISPR Cas9 seems to be toxic in cyanobacteria, our research organism, resulting in cell death at high concentration levels (Wendt et al., 2016). More specifically, Wendt et al. were unable to produce conjugation mutants in *Synechococcus elongatus sp. PCC 2973* (a close relative of S7942) with a medium-copy number plasmid encoding *cas9*. Therefore, we utilized another CRISPR protein for this study. The protein CRISPR Cas12a, formerly known as Cpf1, has been shown to work in cyanobacteria for genome editing purposes (Ungerer & Pakrasi). While Cas12a differs from Cas9 in several ways, it is still capable of facilitating genomic manipulation in cyanobacteria.

CRISPR Cas12a, in nature, has three steps. The first of these steps is 'adaptation', in which Cas proteins (CRISPR-associated) excise a portion of target DNA known as a protospacer.

The protospacer is inserted into a CRISPR loci, where it becomes known as a spacer. In the second step, the spacer is expressed, and the resulting RNA molecules are processed several times, eventually resulting in mature crRNA. Finally, mature crRNA (gRNA) serves to "guide" CRISPR endonucleases to target DNA for cleavage. To discriminate between self and non-self, CRISPR endonucleases, such as Cas9 and Cas12a, recognize specific DNA sequences known as PAM sites (Protospacer Adjacent Motif). PAM sites are an essential part of identifying non-self CRISPR targets and do not occur in the self-CRISPR loci but do occur preceding other instances of foreign DNA/RNA (Zetsche et al., 2015).

The first goal of this study is to improve upon the cyanobacterial CRISPR editing system described by a former colleague (Fifield et al., 2020). Specifically, this CRISPR-containing plasmid vector could be improved by changing the restriction enzyme sites where the so-called "repair template" is inserted into the plasmid. Currently, one of the enzymes at this location cannot be heat inactivated, which can complicate working with this large and low-copy number plasmid. Additionally, to facilitate screening for successful transformants during cloning steps, we will add a gene for GFP (green fluorescent protein), whose presence or absence will help indicate whether a colony contains a successfully constructed product or simply a product of re-legation of the plasmid without incorporation of the desired insert DNA fragment. A related second goal of this study is to test the efficacy of this system by using it to delete the n*dhF4* gene encoding the cyanobacterial NdhF4 protein of *Synechococcus elongatus sp. PCC 7942* (S7942), a hypothesized proton pumping subunit in the cyanobacterial CO_2 concentrating mechanism. The final goal of this study is to determine whether or not a second gRNA targeting sequence improves the rate at which *ndhF4* is deleted in S7942.

We chose *ndhF4* as a target for gene deletion because its gene product, the NdhF4 protein, it is hypothesized to pump protons across the thylakoid membrane and helps establish a proton gradient across the thylakoid membrane. The NdhF4 polypetide is a membrane-intrinsic subunit of the cyanobacterial NDH-1$_4$ complex—a large (~500 kDA) thylakoid membrane complex (Fig. 1)—which plays an important role in cyclic electron flow and the cells CO_2 concentrating mechanism (CCM). The construction of a *ndhF4* mutant (*ΔndhF4*) mutant in wild-type cells and in other genetic background is an important part of characterizing this protein.

To fulfill the goals of this study, we designed and constructed a CRISPR Cas12a plasmid vector containing GFP. We then used this plasmid to construct a CRISPR Cas12a vector capable of inducing markerless deletions of *ndhF4* and introduced it to S7942 via conjugation with *E. coli*. The procedure and results are described below.

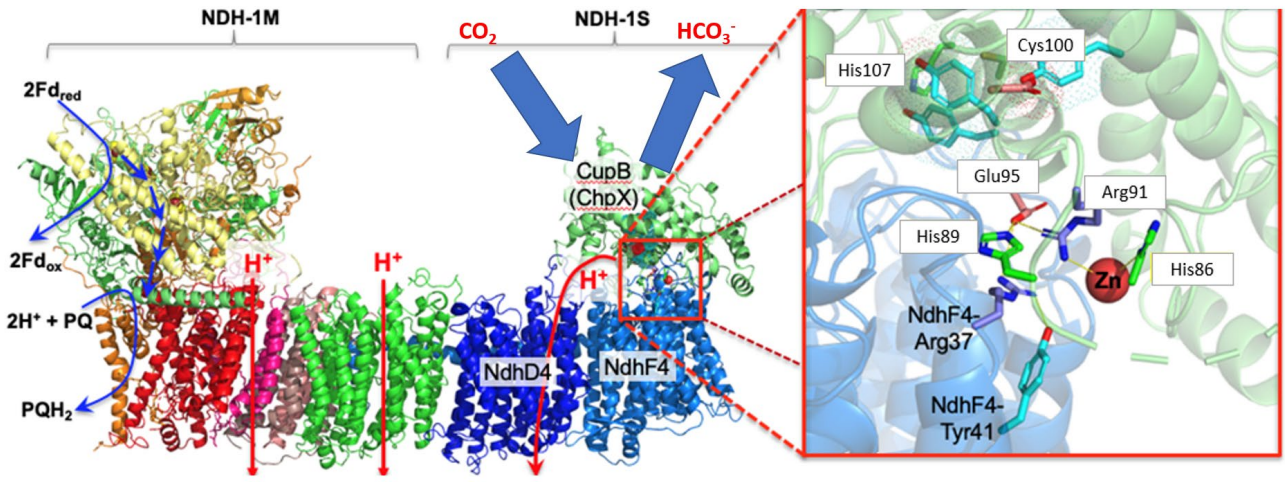

Fig. 1: Structure of the Cyanobacterial Ndh-1$_4$ Complex. A homology model of the Ndh-1$_4$ complex was created using a cryo-EM structure of Ndh-1$_3$ and sequence alignments. NdhF4 (light blue), our target for gene deletion, is hypothesized to pump protons into the thylakoid lumen. Additionally, it may help coordinate the Zn^{2+} active site of the carbonic anhydrase-like enzyme, CupB. Figure adapted from (Artier et al., 2022).

Materials and Methods

Plasmid Preparation

Plasmid DNA was isolated from the DH5α strain of *E. coli* containing the desired plasmid. All plasmids were isolated using Plasmid DNA Kits obtained from Omega Bio-Tek. Where applicable for low-copy number plasmids, the protocol was modified according to the Plasmid DNA Kit Protocol – Low Copy Number Plasmid and Cosmid DNA Protocol from Omega Bio-Tek. When possible, a Vac-Man Laboratory Vacuum Manifold was used instead of centrifuging the samples.

DNA Restriction Digestion

Restriction digestion reactions were set up using 500 ng of target plasmid DNA, 1 unit of each enzyme required for the digestion reaction, buffer, and nanopure water (nanoH_2O) to a volume of 20 μL. The sample was lightly mixed and placed in a hot water bath according to the enzyme manufacturer's instructions. When the reaction had run to completion, restriction enzymes were heat inactivated according to the enzyme manufacturer's instructions. After heat inactivation, the reaction was stored on ice for future use.

Polymerase Chain Reaction (PCR)

The target DNA, forward and reverse primers, Q5 High Fidelity 2X Master Mix (New England Biolabs), and nanoH_2O were collected on ice. The reaction was assembled on ice to

contain ~1 ng plasmid target DNA or ~25 ng chromosomal target DNA, 1.25 µL of each 10 µM primer, 12.5 µL of Q5 polymerase, and nanoH_2O to adjust the reaction volume to 25 µL. A negative control with nanoH_2O replacing the target DNA was also made for PCR reactions. For colony PCRs, cells were collected with a toothpick, lightly touched to a cyanobacterial colony, and mixed with 15 µL nanoH_2O in a PCR tube. The tube was vortexed, and 1 µL of this mixture was used in place of DNA for the PCR reaction. We did not perform colony PCRs with *E. coli*.

The PCR protocol was set up in a thermocycler. This protocol calls for an initial denaturing step at 98°C for 30 seconds that is not repeated. Then, there are 30 cycles of the following: a 95°C denaturing step for 30 seconds, an annealing step at the lowest primer T_m + 3°C for 20 seconds, and a 72°C extension step for 60 seconds per 1000 base pairs in the PCR product band. Finally, there is a 72°C final extension step that occurs one time for 3 minutes. The completed PCR product was kept at 15°C in the thermocycler until it was collected.

Gel Electrophoresis

We prepared 1% agarose gels by microwaving 30 mL 1X TAE buffer (40mM Tris, 20mM acetic acid, and 0.4mM EDTA which was diluted from a 50X stock solution. The pH was not adjusted but should range from 8.3-8.6) and 0.3g of agarose powder in a 250 mL Erlenmeyer flask at full power until the agarose powder had fully dissolved (78 seconds in the laboratory microwave unit). The solution was then poured into a gel chamber fitted with a comb. Gels were allowed to solidify over the course of 30-60 minutes. After solidifying, the comb was gently removed, and the gel was placed in an electrophoresis chamber containing 1X TAE buffer. Samples were prepared as a droplet on a strip of Parafilm and contained 1 µL gel loading dye (Thermo

Scientific; Ref R1161), DNA sample (either 2 µL PCR product or 5 µL restriction digestion product). When necessary, nanoH_2O was used to adjust the total volume to 6 µL. Similarly, the standard was made with 1 µL gel loading dye, 1 µL 1 kb Plus DNA Ladder (Thermo Scientific; Ref SM0311), and 4 µL nanoH_2O. The samples were then loaded into the gel, and once the lid was placed on the contained, the contained was connected to a voltmeter. The voltmeter was set at 75 volts for 30 minutes, after which, the gel was stained on a shaker (70 rpm) using a solution of 30 mL 1X TAE and 9 µL Gel Green Dye (Biotium, Ref 20G0121) for 30 minutes. Afterwards, the gel was imaged using an E-Gel Imager (Life Technologies).

Gibson Assembly

Gibson Assemblies (GAs) were performed using a PCR product and a restriction-digested plasmid with overlapping overhangs as described by (Gibson et al., 2009), which were both run on an agarose gel as described in the gel electrophoresis section. After imaging the gel, the band brightness of the restriction digested plasmid and the PCR product were compared to the brightness of the DNA Ladder to aid in calculating a 2:1 (insert:backbone) molar ratio. Both DNA samples were collected and placed on ice, along with 2X Gibson Assembly Master Mix (New England Biolabs) and nanoH_2O. The calculated volumes of backbone and insert were added to 10 µL GA master mix and the volume was adjusted to 20 µL using nanoH_2O. The GA reaction was run at 50°C for 15 minutes. GA products were immediately used for *E.* coli transformations.

E. Coli Transformations

Transformations were performed with two strains of chemically competent *E. coli*; DH5α (acquired from New England Biolabs) and HB101 (prepared in-house, see below). Both strains were allowed to thaw from -80°C on ice. After thawing, 10 ng of target plasmid (or 2 µL of GA product) was added to the thawed cells. Tubes were lightly flicked to mix, and then incubated on ice for 30 minutes. After incubating, the cells were heat shocked for 30 seconds at 42°C and then immediately placed back on the ice for 2 minutes. During the 2-minute ice incubation, 950 µL SOC outgrowth medium (New England Biolabs; B9020S) was added to a 50 mL round-bottomed polypropylene flask. After the 2-minute ice incubation, the cells were transferred into the appropriate polypropylene flask containing SOC outgrowth medium and placed on a shaker to incubate for 1 hour at 37°C at approximately 250 rpm. After incubation, various volumes of the sample (ranging from 100-200 µL for pCpf1RB vectors and 50-100 µL for all other plasmid vectors) were spread plated on plates containing the appropriate antibiotic. The plates were then incubated overnight at 37°C.

The protocol for using in-house competent cells (see below) is identical to the protocol for transforming cells acquired from NEB in all regards except the heat shock, which should instead be for 90 seconds at 42°C and the volume of SOC media added, which should be 900 µL.

Preparation of chemically competent *E. coli* cells

E. coli cells at 0.35 OD_{600} were collected and placed in a chilled, sterile 50 mL round-bottomed polypropylene tube on ice for 10 minutes. The cells were centrifuged for 10 minutes at 4°C and 2700 g. Once centrifugation was complete, the media was decanted, and the tube was

flipped and allowed to drain excess media onto a paper towel for 1 minute. The pellet was gently resuspended by swirling in 30 mL ice-cold $MgCl_2$-$CaCl_2$ (80 mM $MgCl_2$ and 20 mM $CaCl_2$). The cells were centrifuged again for 10 minutes at 4°C and 2700 g. The media was decanted, and excess media was again allowed to drain onto a paper towel for 1 minute. The cells were resuspended by gently swirling in 2 mL 0.1M $CaCl_2$ for each 50 mL original culture. The now-chemically competent cells were immediately used for transformation, with the remaining cells stored for future use.

For long-term storage of competent cells, 140 µL DMSO was added per 4 mL created competent cells. The cells were gently swirled to mix and then iced for 15 minutes. The same volume of DMSO from the previous step was added, and the cells were dived into 100 µL aliquots in 1.5 mL microcentrifuge tubes. The microcentrifuge tubes were flash-frozen with liquid nitrogen and stored at -80°C.

Conjugation with S7942

We began by transferring 350 µL of an overnight HB101 strain *E. coli* culture containing pRL623 and the CRISPR plasmid pCpf1RB-F4KO and 250 µL of an overnight culture of ED8654 strain *E. coli* culture containing the helper plasmid pRL443 to individual 50 mL round-bottomed polypropylene tubes containing 10 mL pre-warmed LB and the appropriate antibiotic(s). Helper plasmids, like pRL623, encode genes required for conjugation and DNA transfer while the shuttle vector pRL443 encodes genes that can mobilize plasmids (for example, pCpf1RB-F4KO) to cyanobacteria (Elhai et al., 1997). The *E. coli* cultures were incubated on a shaker for 2.5 hours at 37°C and approximately 250 rpm. The cells were spun down at 3000 g for 3 minutes and washed with 10 mL LB three times. After the third wash, the cells were pelleted again and resuspended in 60 µL LB. The *E. coli* cultures were then mixed and incubated at room

temperature for 1-2 hours. After this incubation, 100 mL WT S7942 grown in BG-11 media to an OD_{750} of 0.6-1.0 were centrifuged at 4000 rpm for 10 minutes and resuspended in 500 µL BG-11 media. From this 500 µL stock, a 1/100 cyanobacterial dilution was prepared in a plastic cuvette. To estimate the chlorophyll *a* concentration of this dilution, the OD of the sample was recorded at 620, 678, and 750 nm using the method described by Williams (1988) to estimate the chlorophyll *a* concentration directly from the suspension of cells. These wavelengths record the absorbance from phycobilisomes, chlorophyll *a*, and light scattering, respectively. The following formula was used to calculate the concentration of chlorophyll *a*:

$$[\text{Chlorophyll } a] = (14.96*(OD_{678}-OD_{620})-0.607*(OD_{620}-OD_{750})) * \text{dilution factor}$$

10 µg of chlorophyll *a* from the 500 µL cyanobacterial stock was mixed with the 120 µL of mixed and incubated *E. coli*. 30, 40, and 50 µL of cell mixture were plated on a nitrocellulose filter placed on a BG-11 plate with 5% LB (and no antibiotics) and stored under low light conditions for 3 hours (75-90 µE $m^{-2}s^{-1}$ of light; simulated by covering plates at the normal light intensity of 150-170 µE $m^{-2}s^{-1}$ in the lab growth chamber with a blank sheet of copy paper). After 3 hours, the copy paper was removed, and the plates were incubated under normal light conditions. The following day, the nitrocellulose filters were transferred to a BG-11 plate containing no antibiotics and were kept at normal light conditions. On the third day, the nitrocellulose filters were transferred to BG-11 plates containing 20 µg/mL spectinomycin and were incubated at normal light conditions. The filters were kept on these plates until the filters cleared (or were transferred to fresh spectinomycin plates if clearing was slow) and new colonies appeared.

Curing Cyanobacteria of Plasmids

To cure cyanobacterial transformants of CRISPR plasmids, an individual colony was picked from the selective plates and grown in 100 mL of BG-11 media without the antibiotic corresponding to the plasmid that is being cured from the strain (nAbx). The culture was grown at 30°C, shaking at 200 rpm, and, when necessary, with supplemental 5% CO_2. When the culture reached an OD_{750} of 0.6-1.0, 50 μL was plated onto nAbx BG-11 plates. The plates were incubated at 34°C and kept at 150-170 E $m^{-2}s^{-1}$ of light. When colonies appeared, they were struck out on nAbx BG-11 plates. These new plates were incubated until isolated colonies appeared. When isolated colonies appeared, half of each colony was struck out on nAbx BG-11 grid plates and the other half of each colony was struck out on BG-11 grid plates with the antibiotic of the plasmid being cured, being careful to keep track of the location of each individual colony on both plates. Colonies that grow on the nAbx plates but not on the antibiotic plates were assumed to be cured of the plasmid of interest.

Should no colonies be cured, nAbx colonies should be struck out additional times until the cells do not grow on antibiotic plates.

Results and Discussion

Construction of an Improved Cyanobacterial CRISPR Plasmid

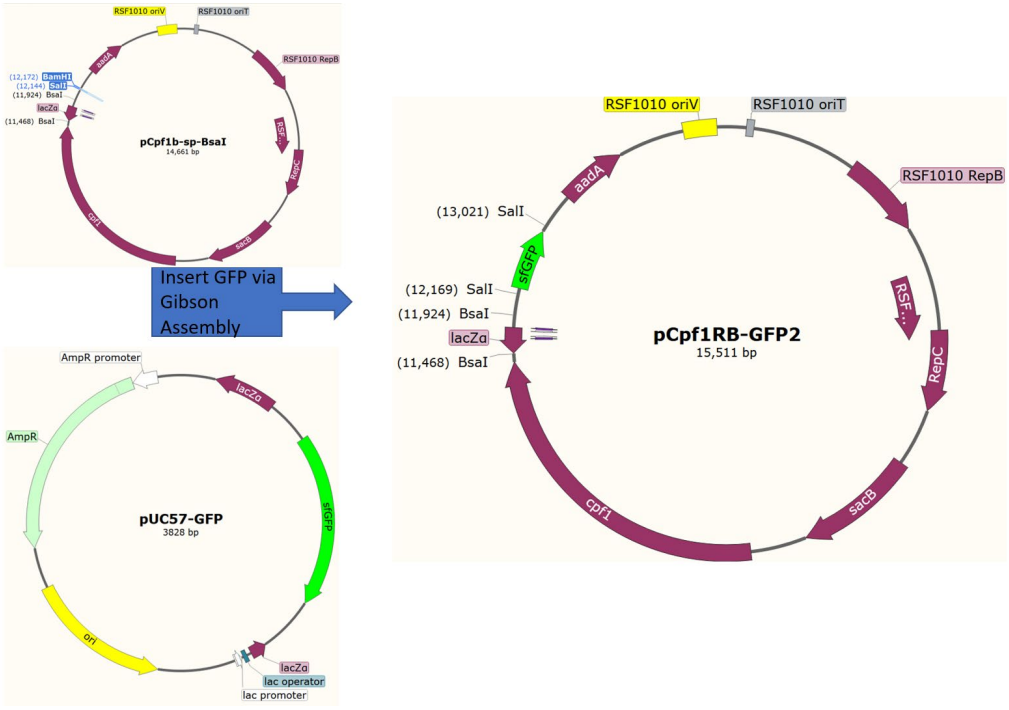

Fig. 2: Plasmid Construction Scheme of pCpf1RB-GFP2. pCpf1b-sp-BsaI was digested with the highlighted enzymes (BamHI and SalI). PCR was utilized to amplify the GFP region of pUC57-GFP, which was then assembled into the plasmid vector pCpf1RB-GFP2, shown on the right.

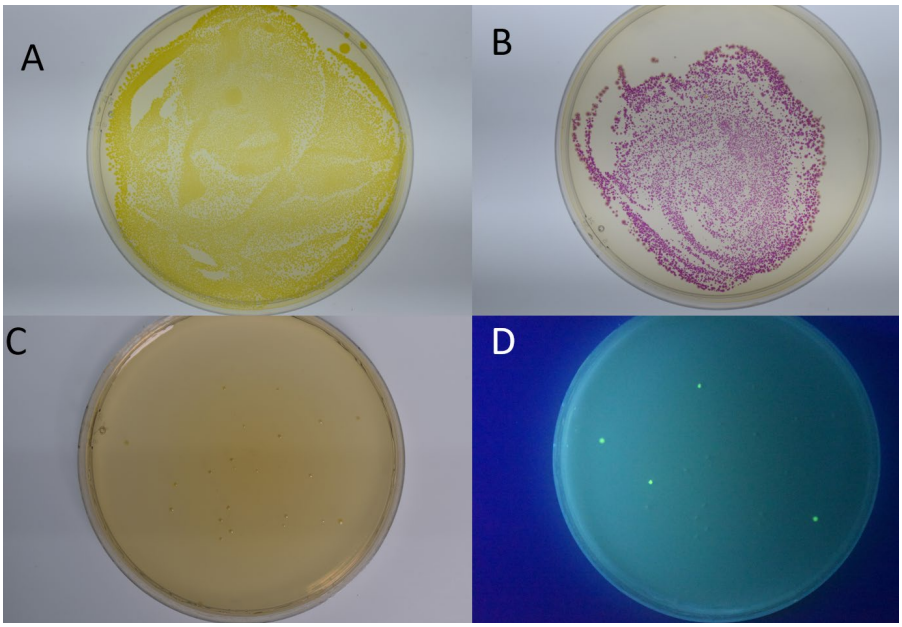

Fig. 3: Phenotypes of *GFP* and *spisPink*. A) *E. coli* cells transformed with the pUC57-GFP plasmid photographed under normal light conditions. **B)** *E. coli* cells transformed with the pUC57-Pink plasmid. **C)** A plate of *E. coli* transformed with pCpf1RB-GFP2 under normal light conditions. **D)** The same plate from C photographed under UV light conditions. Of the 21 colonies present on the plate, 4 of them are clearly fluorescing.

To create a CRISPR Cas12a plasmid vector, we utilized the pCpf1b-sp-BsaI plasmid vector originally created by a former colleague for an OSU Honors Thesis (Fifield et al., 2020). We had two proposed changes to this plasmid; the first was replacing the BamHI restriction enzyme site with a SalI restriction enzyme site. SalI was chosen because it was already used for inserting the repair template into the plasmid. SalI has several advantages over BamHI, most importantly, SalI can be heat inactivated, while BamHI cannot. Additionally, we decided to insert a *GFP* region into this plasmid between the two SalI restriction enzyme sites. The presence or absence of *GFP* would allow us to visually screen for green fluorescent colonies that are non-recombinants resulting from carry-over from the original vector or containing a re-ligation

product versus the desired plasmid of interest have the desired insert replacing the GFP gene. A second variation of this plasmid was also planned, using *spisPink* instead of *GFP*.

The *GFP* and *spisPink* sequences were generously provided by Dr. Dennis Nürnberg at the Free University of Berlin, and the DNA sequences were ordered from Genscript in the form of pUC57-GFP and pUC57-Pink, respectively (see Fig. 3C and D for GFP phenotype and 3B for pink phenotype). The plasmids were transformed into chemically competent *E. coli* obtained from NEB, and the plasmid was extracted via a plasmid preparation kit. Similarly, the pCpf1b-sp-BsaI was extracted via a plasmid preparation kit. pCpf1b-sp-BsaI was digested with the restriction enzymes BamHI-HF and SalI while *GFP* was amplified via PCR. After analyzing the products via gel electrophoresis, the products were assembled via a GA reaction and transformed into chemically competent *E. coli* from NEB. This resultant plasmid, pCpf1RB-GFP2, was analyzed via a restriction digest and confirmed to be the intended product.

After creating pCpf1RB-GFP2, we decided not to proceed with synthesizing a CRISPR plasmid containing *spisPink* because the *GFP* phenotype was much stronger and appeared earlier. Additionally, the pink phenotype did not evenly distribute to the peripheries of some colonies (see Fig. 3B), while the GFP phenotype was consistent across the entire colony.

Due to a cloning error, the first plasmid constructed retained the BamHI restriction enzyme site that was intended for deletion. This plasmid was named pCpf1RB-GFP. This mistake has since been corrected and pCpf1RB-GFP2 has been synthesized correctly.

Designing a CRISPR System to Delete *ndhF4*

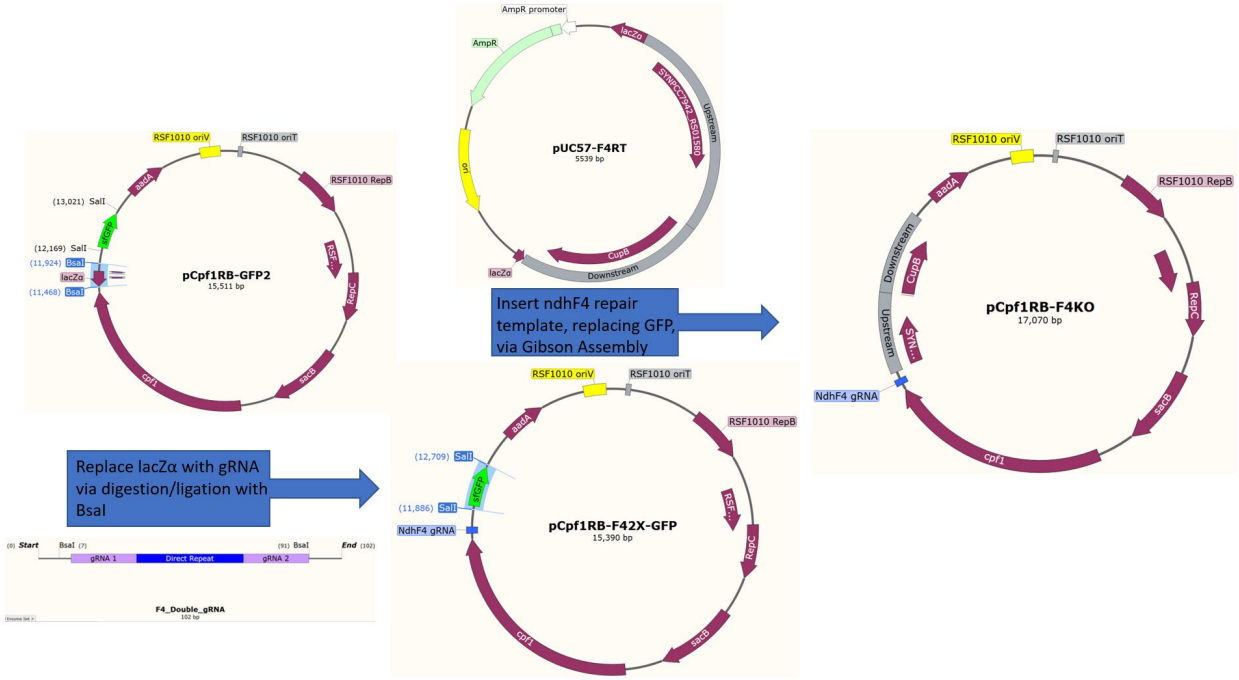

Fig. 4: Plasmid Construction Scheme of pCpf1RB-F4KO. pCpf1RB-GFP-2SalI was digested with the highlighted restriction enzymes (BsaI) and a DNA insert encoding gRNA targeting *ndhF4* featuring two spacers (or the targeting sequences) was inserted to the digested backbone via a ligation reaction. This step was outsourced to Genscript. We named the resultant product pCpf1RB-F42X-GFP. We then digested pCpf1RB-F42X-GFP with SalI to excise the highlighted portion. A PCR reaction amplified portions of the Upstream and Downstream repair templates of pUC57-F4RT and was assembled via a GA reaction, creating the plasmid vector pCpf1RB-F4KO, which is capable of inducing markerless deletions of *ndhF4*.

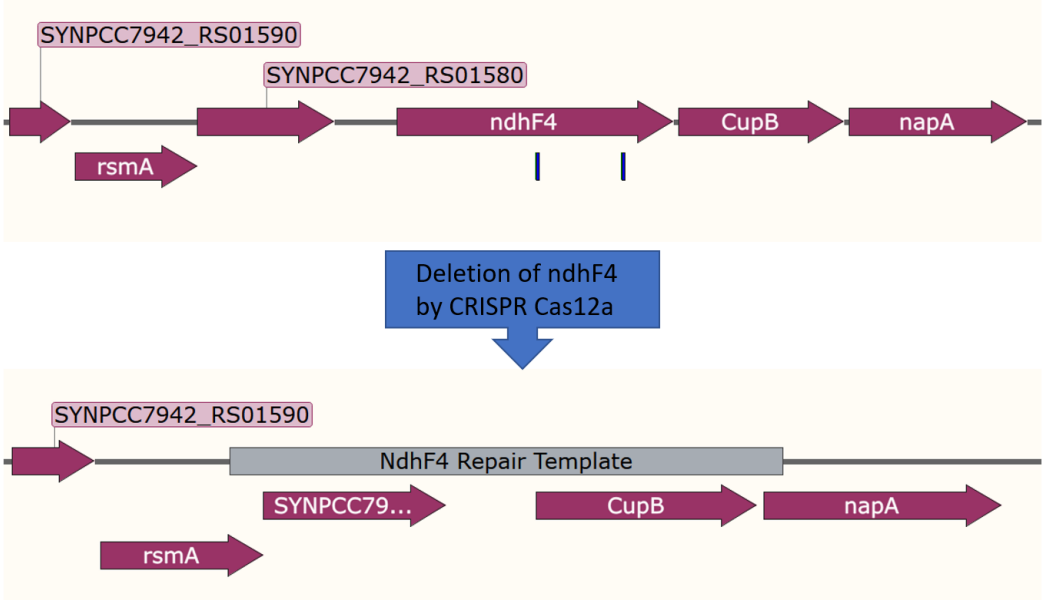

Fig. 5: S9742 Genome Before and After *ndhF4* Deletion. S7942's genome was edited with our CRISPR Cas12a system deleting *ndhF4* and replacing it with the providing the NdhF4 repair template, which consists of S7942 genomic DNA flanking *ndhF4*. The two blue bands in *ndhF4* are gRNA binding sites.

We utilized the plasmid vector pCpf1RB-GFP2 created in the previous step as a template to create a plasmid, pCpf1RB-F4KO, capable of deleting genomic *ndhF4*. Creating this plasmid required two main cloning steps: insertion of gRNA and replacement of GFP with a *ndhF4* repair template.

The gRNA is a 20-24 base pair sequence that follows a protospacer adjacent motif site (PAM site), a short sequence (TTTV, where V is interchangeably A, C, or G) that helps the CRISPR Cas12a system determine self from non-self. CRISPR Cas12a screens cellular DNA for a gRNA spacer that follows a PAM site. When such a sequence is detected, CRISPR Cas12a introduces a cut to the DNA sequence. In nature, this cutting action breaks viral genomes and prevents them from hijacking the host cell's replication machinery. When used as a molecular genetics tool, the cut forces the host organism to repair the damage using a provided repair template.

Because our goal is to create a markerless deletion, our repair template consists of roughly 1000 base pairs immediately upstream and downstream relative to *ndhF4* in the S7942 genome and contains none of *ndhF4* except for the stop codon. We decided to leave the stop codon in the repair template because we were unable to conclusively determine if it was involved in the termination of other genes expressed in an operon.

We began construction of pCp1fRB-F4KO by inserting the DNA encoding our gRNA into pCp1RB-GFP. To reduce the number of cloning steps required, we outsourced this work to GenScipt, who synthesized the gRNA and cloned it directly into the plasmid vector. In non-industrial settings, the screening of this step can be simplified using a blue/white selection. We named the plasmid synthesized by Genscript containing CRISPR Cas12a, our *ndhF4* gRNA, and GFP pCpf1RB-F42X-GFP (where F42X refers to the two individual gRNA sequences targeting *ndhF4*).

Verifying *ndhF4* Deletion

After introducing pCpf1RB-F4KO to WT S7942 cells via conjugation, we needed to verify that the deletion was successful via colony PCR (Fig. 6b). Previous work in our lab involving a CRISPR Cas12a system that utilized one gRNA yielded a transformation efficiency of approximately 1/30 (Burnap lab, unpublished). This is complicated because S7942 has several copies of its chromosome, typically 4 (Griese et al., 2011), that must be edited. In addition to serving as templates for repairing the native gene function after CRSIPR Cas12a deletion, all copies of the chromosome likely need to have *ndhF4* deleted to have the phenotype of Δ*ndhF4*.

Attempting to improve this poor transformation efficiency was a contributing factor to our decision to design pCpf1RB-F4KO to contain 2 gRNAs.

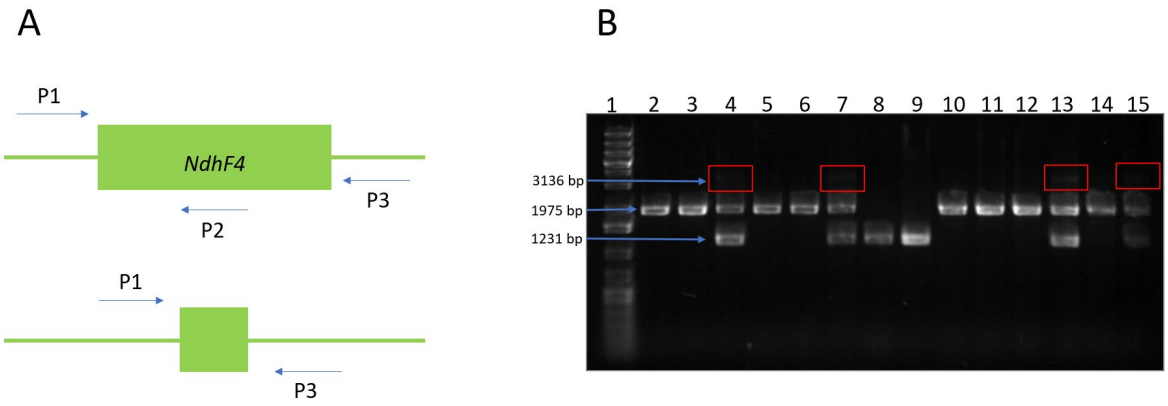

Fig. 6: Screening for *ΔndhF4* mutants by gel electrophoresis. A) Primer design for verifying deletion genotype. Primer pair P1 and P2 will only generate a PCR product if *ndhF4* is present because P2 binds inside the gene. Primer pair P1 and P3 can produce two different-sized bands depending on if *ndhF4* is present or absent. **B)** A gel electrophoresis showing the products of a colony PCR. Lane 1 contains the ladder, lanes 2 and 3 contain PCR products from WT S7942, and lanes 4-15 contain PCR products from cells conjugated with pCpf1RB-F4KO. Lanes 8 and 9 show full deletions of *ndhF4*. Lanes 4, 7, 13, and 15 show partial deletions of *ndhF4* (as evidenced by all three possible bands appearing). Lanes 5, 6, 10, 11, 12, and 14 show failed deletions of *ndhF4*.

Fig. 6b shows the results of 12 colonies that were introduced to pCpf1RB-F4KO via conjugation. Of the 12 colonies, 2 had *ndhF4* fully deleted from all their chromosomes, 4 were partial deletions of *ndhF4* (i.e.: some chromosomes had *ndhF4* deleted, but not all of them), and 6 colonies showed no evidence of *ndhF4* being deleted. Interestingly, and for unknown reasons, only the partial deletions showed the larger 3136 base pair band produced by the P1 and P3 primer pair when the primers sequenced over *ndhF4*. It is unknown why this band is absent in WT samples or failed deletion mutants. Regardless, the presence of the smaller P1/P3 1231 base pair band and the absence of the P1/P2 1975 base pair band provide sufficient evidence for the deletion of *ndhF4* in 2 of these colonies. It is possible that if kept on selection media for longer, the increased time would allow partial deletion mutants to become full deletion mutants.

Curiously, lanes 7-9 came from relatively small colonies (roughly the size of a toothpick tip) at harvesting. Although we do not currently know if the improved deletion efficiency exhibited in lanes 7-9 is coincidental, it may reflect that the loss of the NdhF4 protein is detrimental to cell growth. Although the overall deletion efficiency is still relatively poor, this shows that additional gRNA targeting sequences can significantly improve the efficiency of gene deletion.

We have three proposed mechanisms for why this plasmid more efficiently deleted *ndhF4* from S7942s genome than previous CRISPR plasmids. Mechanism one is that the addition of a second CRISPR target sequence in *ndhF4* allowed for the excision of a DNA fragment, thus preventing a re-ligation reaction from occurring to repair the gene and chromosome. This would instead force chromosomal repair to occur via homologous recombination from either pCpf1RB-F4KO or other chromosomes. Mechanism two is that the specific individual gRNA targets selected introduced cuts more reliably than cuts of previous CRISPR Cas12a systems. In simpler terms, certain regions of the genome could make better targets for CRISPR Cas12a. The third proposed mechanism is that the cells were exposed to the CRISPR plasmid for a longer duration before screening for mutations via colony PCR. This would result in more opportunities to replace *ndhF4*, and thus could appear as a higher transformation efficiency. For example, further propagation of partial deletion mutants which clearly contain pCpf1RB-F4KO may result in the creation of full deletion mutants. Determining which of these mechanisms is causing improved efficiency would necessitate further study. Future work will be needed to determine whether the newly constructed *ΔndhF4* strains have a phenotype such as slower growth under different inorganic carbon regimes.

Construction of a pSE4 *ndhF4* Complementation Vector

After confirming the successful deletion of *ndhF4*, the next step was to prepare a complementation vector that can restore the WT phenotype to our *ΔndhF4* S7942 strain by re-introducing the *ndhF4* gene using a recombinant plasmid vector capable of replication in S7942 As noted, we have not yet established whether there is indeed a discernible phenotype for *ΔndhF4* but the complementation vector described here can be used to test this hypothesis. This is part of confirming that the correct gene was deleted, but it also plays a role in establishing a genetic basis for future mutations. As mentioned previously, part of the reason that *ndhF4* was selected for gene deletion was our intention to study NdhF4's structure and its role in the Ndh-1_4 complex.

To construct this plasmid vector, we performed a restriction digestion reaction on the plasmid pSE4-H89E using SacI and NcoI and a simultaneous PCR reaction that amplified WT 7942 *ndhF4*. Then, a GA reaction using the digested plasmid and PCR product was performed to create a single plasmid vector, which we named pSE4-*ndhF4*, which was then transformed into *E. coli*. Because we did not have the powerful phenotype of GFP to help screen for successfully assembled plasmid vectors, we digested plasmid samples of the transformed *E. coli* and digested the product with SacI-HF and EcoRI-HF and ran it on gel with pSE4-H89E also digested with SacI-HF and EcoRI-HF (Fig. 8b). The results of this digestion and gel electrophoresis show that we achieved our desired plasmid construct.

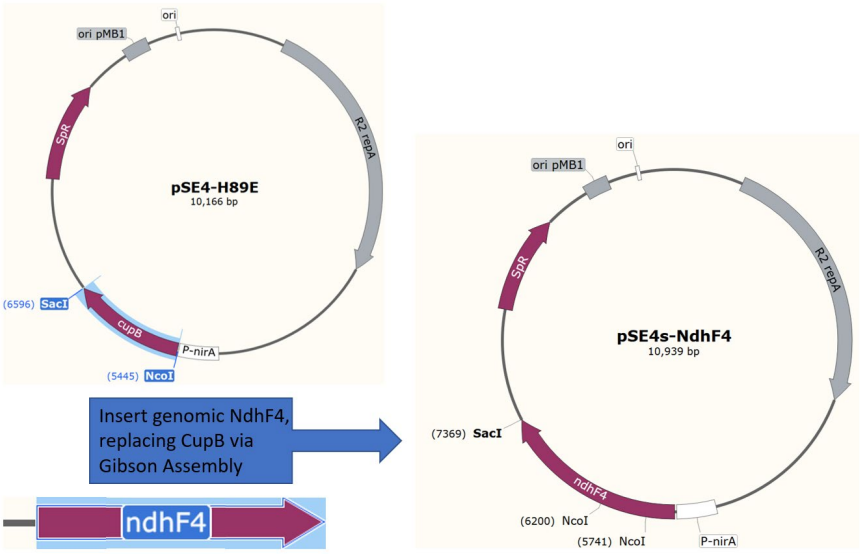

Fig. 7: Plasmid Construction Scheme of pSE4s-n*dhF4*. pSE4-H89E was digested with SacI and NcoI while genomic *ndhF4* was amplified via PCR. The restriction-digested product and PCR product were combined via a GA reaction, producing the plasmid vector pSE4s-*ndhF4*. Of note is the P-nirA promoter, which is activated by the binding of NO_3^-. Standard BG-11 growth media contains high concentrations of NO_3^-, making this promoter effectively constitutive.

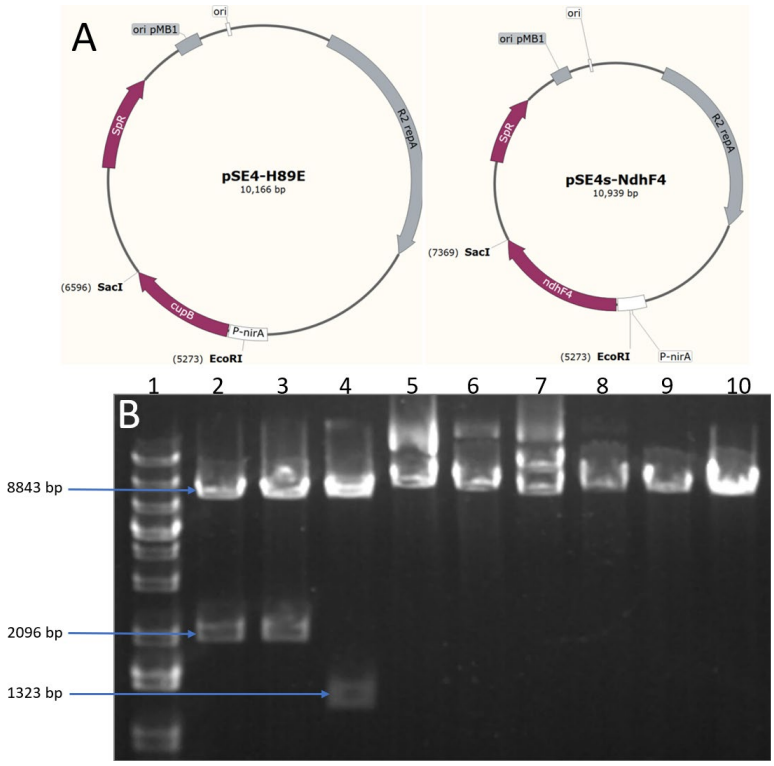

Fig. 8: Restriction digestion verification of pSE4s-*ndhF4*. A) Maps of digested plasmids. When digested with SacI-HF and EcoRI-HF, pSE4-H89E produces a 1323 bp band whereas pSE4s-*ndhF4* produces a 2096 bp band when digested with the same enzymes. **B)** A gel showing pSE4-H89E and pSE4s-*ndhF4* digested with combinations of SacI-HF and EcoRI-HF. Lane 1 shows the DNA marker. Lanes 2 and 3 show plasmid preps from two separate isolated colonies both digested with SacI-HF and EcoRI-HF. Lane 4 shows pSE4-H89E double digested with SacI-HF and EcoRI-HF. Lanes 5 and 6 show the isolated colony plasmid prep products linearized by SacI-HF. Lane 7 shows pSE4-H89E linearized by SacI-HF. Lanes 8 and 9 show the isolated colony plasmid prep products linearized by EcoRI-HF. Lane 10 shows pSE4-H89E linearized by EcoRI-HF. Note that the larger band present in lanes 2-4 is of a consistent size of 8843 bp (the size of the plasmid backbone) while the smaller band in these lanes is either 2096 bp (lanes 2 and 3) vs 1323 bp (lane 4). Also note that due to the similar sizes of the linearized fragments in lanes 5-10, it is difficult to see differences in plasmid sizes.

Conclusions

The purpose of this work was to create a GFP version of our CRISPR Cas12a plasmid vector that also contained 2 SalI cutsites. We then intended to use this plasmid construct as a

framework for the deletion of *ndhF4*, with a final goal of determining whether or not additional gRNA sequences improve the efficiency of CRISPR-mediated gene deletion. We were successful in all of these goals: we synthesized plasmid vectors pCpf1RB-GFP2 and pCpf1RB-F4KO, which can be used for other CRISPR-mediated gene deletion and the deletion of *ndhF4*, respectively. We also anecdotally showed that additional gRNA targeting sequences do improve the efficacy of gene deletion.

Future work could include utilizing the Δ*ndhF4* strain as a genetic background for structure/function studies of NdhF4. We plan on creating other pSE4 vectors that introduce modified versions of *ndhF4* to Δ*ndhF4*. Specific target residues for this modified *ndhF4* include the hypothetical shared NdhF4/CupB active site (Fig. 1), as well as residues that form a hypothetical "water wire;" or a series of residues that are responsible for shuttling the protons away from the shared active site and into the thylakoid lumen. This has the effect of not only contributing to the pmf gradient that generates ATP, but it also prevents CupB from catalyzing the reverse reaction ($HCO_3^- + H^+ \rightarrow CO_2 + H_2O$). Other work could include more studies on the additional gRNA sequences to determine which mechanism(s) proposed in this paper result in improved deletion efficiency.

Acknowledgements

I've received a lot of help over my years in the Burnap lab. I'd especially like to thank Dr. Anton Avramov, a former graduate student mentor of mine, for developing this CRISPR system and teaching me many fundamentals of research. I'd like to thank Ross Walker, PhD candidate, for putting up with all of my stupid questions and for always letting me bounce ideas

off of him. I'd also like to thank Swarnali Mukherjee, PhD student, for continuing my work on the Δ*ndhF4* strain and for collaborating with me on the construction of vectors to introduce point mutants. Dr. Rob Burnap was a fantastic and patient mentor who has guided me through this research and provided valuable advice and support as I look towards graduate school. Finally, I'd like to thank Dr. Sabrina Beckmann for being an external reader of my thesis and a phenomenal professor. It's been a pleasure to work with and get to know these people.

Citations

Artier, J., Walker, R. M., Miller, N. T., Zhang, M., Price, G. D., & Burnap, R. L. (2022). Modeling and mutagenesis of amino acid residues critical for CO(2) hydration by specialized NDH-1 complexes in cyanobacteria. *Biochim Biophys Acta Bioenerg*, *1863*(1), 148503. https://doi.org/10.1016/j.bbabio.2021.148503

Cong, L., Ran, F. A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P. D., Wu, X., Jiang, W., Marraffini, L. A., & Zhang, F. (2013). Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science*, *339*(6121), 819-823. https://doi.org/doi:10.1126/science.1231143

Elhai, J., Vepritskiy, A., Muro-Pastor, A. M., Flores, E., & Wolk, C. P. (1997). Reduction of conjugal transfer efficiency by three restriction activities of Anabaena sp. strain PCC 7120. *J Bacteriol*, *179*(6), 1998-2005. https://doi.org/10.1128/jb.179.6.1998-2005.1997

Fifield, D., Burnap, R., & Gustafson, J. (2020). CRISPR in cyanobacteria: A system for genomic editing in Synechocystis. https://shareok.org/handle/11244/329424

Gibson, D. G., Young, L., Chuang, R.-Y., Venter, J. C., Hutchison, C. A., & Smith, H. O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nature Methods*, *6*(5), 343-345. https://doi.org/10.1038/nmeth.1318

Griese, M., Lange, C., & Soppa, J. (2011). Ploidy in cyanobacteria. *FEMS Microbiology Letters*, *323*(2), 124-131. https://doi.org/10.1111/j.1574-6968.2011.02368.x

Ma, D., & Liu, F. (2015). Genome Editing and Its Applications in Model Organisms. *Genomics Proteomics Bioinformatics*, *13*(6), 336-344. https://doi.org/10.1016/j.gpb.2015.12.001

Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J., & Soria, E. (2005). Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol*, *60*(2), 174-182. https://doi.org/10.1007/s00239-004-0046-3

Niu, T. C., Lin, G. M., Xie, L. R., Wang, Z. Q., Xing, W. Y., Zhang, J. Y., & Zhang, C. C. (2019). Expanding the Potential of CRISPR-Cpf1-Based Genome Editing Technology in the Cyanobacterium Anabaena PCC 7120. *ACS Synth Biol*, *8*(1), 170-180. https://doi.org/10.1021/acssynbio.8b00437

Ungerer, J., & Pakrasi, H. B. (2016). Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. *Sci Rep*, *6*, 39681. https://doi.org/10.1038/srep39681

Wendt, K. E., Ungerer, J., Cobb, R. E., Zhao, H., & Pakrasi, H. B. (2016). CRISPR/Cas9 mediated targeted mutagenesis of the fast growing cyanobacterium Synechococcus elongatus UTEX 2973. *Microb Cell Fact*, *15*(1), 115. https://doi.org/10.1186/s12934-016-0514-7

Zetsche, B., Gootenberg, Jonathan S., Abudayyeh, Omar O., Slaymaker, Ian M., Makarova, Kira S., Essletzbichler, P., Volz, Sara E., Joung, J., van der Oost, J., Regev, A., Koonin,

Eugene V., & Zhang, F. (2015). Cpf1 Is a Single RNA-Guided Endonuclease of a Class 2 CRISPR-Cas System. *Cell*, *163*(3), 759-771. https://doi.org/https://doi.org/10.1016/j.cell.2015.09.038