Design of an effective sgRNA for CRISPR/Cas9 knock-ins and full mutant segregation in polyploid *Synechocystis sp.* PCC 6803

María Isabel Nares-Rodriguez[1] and Esther Karunakaran[1,*]

1 Department of Chemical and Biological Engineering, University of Sheffield, Sheffield, United Kingdom

*Corresponding author: Email: e.karunakaran@sheffield.ac.uk

Abstract

Synechocystis sp. PCC 6803 (*Synechocystis*) is a highly promising organism for the production of diverse recombinant chemicals, including biofuels. However, conventional genetic engineering in *Synechocystis* is challenging due to its highly polyploid genome which not only leads to low product yields but also makes the recombinant organism less reliable for use in biomanufacturing. Due to its precision, effectiveness and reliability in a vast array of chassis, CRISPR/Cas9 has the potential of overcoming the drawbacks effected by a polyploid genome. Here we identified and developed an effective sgRNA for the knock-in of nucleotide sequences of varying lengths in the neutral site *slr*0168 of polyploid *Synechocystis* using CRISPR/Cas9. The gene encoding digeranylgeranylglycerophospholipid reductase from *Sulfolobus acidocaldarius* and the methyl ketone operon from *Solanum habrochaites* were chosen as the exemplar nucleotide sequences for incorporation into the chromosome of *Synechocystis*. It is demonstrated here that our sgRNA design was effective for both knock-ins and that CRISPR/Cas9 achieves complete mutant segregation after a single step of selection and induction.

Key words: CRISPR/Cas9; sgRNA design; mutant segregation; *Synechocystis sp.* PCC 6803; minimal induction

1. Introduction

An important aim of genetic engineering is the expression of heterologous proteins in a host organism that enables the generation of a target product. For robust industrial production, the genes encoding these heterologous proteins must be stable and homogenic in the host, that is, the genes must be borne on all chromosomal copies so that they are inherited and maintained

in every descendant in each cell-division cycle to prevent mutation dilution and mutant loss during manufacturing and ensure the high yields of the desired product (1-4).

The polyploid model cyanobacterium *Synechocystis sp.* PCC 6803 (hereafter *Synechocystis*) is a highly promising chassis for the production of many recombinant products relevant to a number of industries, including biofuels and personal care products (5-10). However, traditional genetic engineering of *Synechocystis* is laborious, time-consuming and not cost-effective due to its highly polyploid genome. To overcome this drawback, strategies such as phosphate deprivation, multiple counter-selection and combined random mutagenesis have been implemented (11-13). These strategies rely on the natural occurring mechanism of double homologous recombination (DHR) in *Synechocystis* (7, 12, 14, 15) and full mutant segregation remains difficult to achieve (11-13). Consequently, gene expression as well as target product yields are often compromised precluding use of *Synechocystis* as the host in the industrial-scale production of heterologous target products (4, 16-18). However, emerging technologies, such as CRISPR/Cas9 may alleviate this bottleneck.

The type II CRISPR system (clustered regularly interspaced short palindromic repeats) and its associated nuclease Cas9 (CRISPR/Cas9) is an adaptive immune system in Archaea and Bacteria that has been adapted for highly precise and effective genome editing. The CRISPR/Cas9 system implements a single guide RNA (sgRNA), a Cas9 nuclease and a donor DNA (dDNA) to introduce the latter with high precision in a target locus in the host's genome via its natural homology directed repair mechanism triggered by the double strand breaks (DSB) effected by Cas9 (19-23). Previous gene editing strategies are based on DNA-protein interactions which are considered highly complex. These have not reached the level of effectiveness and precision deemed to CRISPR/Cas9 (23-26), which acts upon nucleotide sequence complementarity and is considered to be simpler; this endows CRISPR/Cas9 with its highly elevated rate of success (20, 27, 28). Moreover, CRISPR/Cas9 is considered to be efficient, reliable and precise due to its chimeric and easily reprogrammable sgRNA, activity of Cas9 and the repair mechanism activated by DSB (29-31). CRISPR/Cas9 is able to perform multiplexed insertions, inactivation, deletions and editing of genes in multiple loci in a simultaneous manner (20, 32-34) and has successfully been implemented in a wide range of hosts, including humans (25, 29, 30, 35-38). Therefore, CRISPR/Cas9 may have the potential of overcoming bottlenecks such as the polyploid genome in *Synechocystis*.

Thus, with the hypothesis that CRISPR/Cas9 can favour mutant segregation, we engineered polyploid *Synechocystis* via CRISPR/Cas9 by separately knocking-in heterologous digeranylgeranylglycerophospholipid reductase (GGR) and a methyl ketone operon (ShMKS), to evaluate 1) if CRISPR/Cas9 can indeed improve mutant segregation, 2) if the size and complexity of the dDNA (GGR-dDNA (single gene) or ShMKS-dDNA (dual-gene operon)) has an impact on CRISPR/Cas9 and mutant segregation and 3) if the CRISPR/Cas9 effects would vary if applied upon markerless and selection marker-bearing dDNA. We implemented an empirical method based on available tools and principles to identify and design an effective sgRNA for successfully inserting the desired genes in the *slr*0168 neutral site, achieving full mutant segregation after a single step of induction and selection for both, markerless as well as mutants bearing a selection marker in the dDNA.

2. Materials and methods

2.1 Strains and media

E. coli 10-beta was used as a cloning strain. The cells were cultivated at 37 °C in sterile lysogeny broth (LB) medium at an orbital agitation of 150 rpm for aeration. The strain was used for all the molecular cloning experiments, plasmid propagation and maintenance. *E. coli* copy cutter cells bearing the pPMQK1 vector were cultivated at 37 °C in sterile LB medium at an orbital agitation of 150 rpm for aeration. The medium was supplemented with kanamycin, at a concentration of 50 µg/mL. The strain was used for the propagation of the plasmid bearing the Cas9 gene. *E. coli* MC1061 cells were cultivated at 37 °C in sterile LB medium at an orbital agitation of 150 rpm for aeration. The strain was used as a helper strain in all tri-parental mating events. The medium was supplemented with chloramphenicol at a concentration of 25 µg/mL and ampicillin at a concentration of 100 µg/mL for plasmid maintenance where appropriate.

The *Synechocystis* wild type (WT) strain (*Synechocystis sp.* PCC 6803) used for the CRISPR mutagenesis was purchased from the Pasteur Culture Collection of Cyanobacteria (PCC). All *Synechocystis* cells were grown photoautotrophically in sterile Blue-Green liquid medium (BG-11) at a temperature of 25°C, under continuous orbital agitation conditions of 100 rpm for aeration. Initially, the mutants were placed under white light, with an intensity of 125 µmol/m2/s; subsequently, this was changed to red light, with an intensity of 20 µmol/m2/s.

The media for *Synechocystis* mutants bearing chloramphenicol resistance was supplemented with chloramphenicol at a concentration of 12 µg/mL.

2.2 Insert and vector construction

The pPMQK1 vector, which was a kind gift from Dr. Paul Hudson and his group (Karolinska Institute; Solna, Sweden) carried the *Streptococcus pyogenes* Cas9 nuclease expression cassette and was transformed into *Synechocystis* for its expression. The pEERM4 Cm (Addgene plasmid # 64026) and the pEERM3 Km vectors (Addgene plasmid # 64025) were kind gifts from Pia Lindberg (39). The QIAGEN Miniprep Spin Kit was used following manufacturer's instructions for plasmid extraction from *E. coli* strains. For plasmid manipulation, appropriate enzymes were sourced from New England Biolabs (UK) and manufacturer's instructions were followed. Chemically competent *E. coli* cells were purchased from New England Biolabs (UK) and the heat shock method was method was performed following manufacturer's instructions for plasmid verification and maintenance. Once designed, the sgRNA and donor DNA sequences were synthesized as custom genes by Eurofins Genomics.

2.3 sgRNA selection

All available protospacer candidates in *slr*0168 were identified using the CasOT software (40). To enable the identification of suitable protospacer sequences, the sequence of *slr*0168 was submitted to CasOT as a FASTA file and the following parameters were set as the selection criteria. The -NGG Protospacer Adjacent Motif (PAM) only option (option "A-NGG only") was selected to generate only candidates to match this PAM. The "target and off-target" mode was selected to allow the generation of all the possible protospacer candidates followed by off-target analysis. A maximum of 2 mismatches were allowed in the seed region. The "non-5'G" requirement option was selected as the T7 promoter was not used in the sgRNA. And, finally, a "length" of 19 to 22 bp for the base-pairing region of the candidates was allowed. For the off-target analysis, each protospacer candidate was evaluated for the number of possible mismatches and off-targets they presented.

2.4 Cas9 handle secondary structure analysis

The RNAfold online tool from the ViennaRNA Package was used to evaluate if the Cas9 handle formed a proper secondary structure after the addition of the selected protospacer (41). The selected 20 nt protospacer was appended to the 42 nt sequence of the Cas9 handle, for a total of 62 nt sequence and subsequently input into the algorithm to evaluate its secondary structure formation. The minimum free energy (MFE) and partition function, dangling energies on both sides of a helix, the Turner model (2004) options for RNA parameters and the default criteria for the output options were selected.

2.5 Strain construction and CRISPR/Cas9 induction

The pPMQK1 vector was transformed into *Synechocystis* to generate *Synechocystis::*Cas9 through standard protocol for tri-parental mating in cyanobacteria (42). Subsequently, the pEERM4::sgRNA_GGR-dDNA and the pEERM3::ShMKS-dDNA vectors were transformed independently into *Synechocystis::*Cas9 following a standard protocol for tri-parental mating in cyanobacteria (42) to generate *Synechocystis:*:Cas9_sgRNA_GGR-dDNA and *Synechocystis*::Cas9_sgRNA_ShMKS-dDNA respectively.

The transformants were transferred to non-selective BG-11 plates for 24 hours for recovery and then subjected to induction and selection. Only ShMKS mutants were subjected to both, selection and induction while GGR mutants were only subjected to induction, effectuated on BG-11 plates supplemented with anhydrotetracycline (aTc) at a concentration of 200 ng/mL while using white light for the first attempt and a subsequent one with an aTc concentration of 400 ng/mL under red light.

2.6 Mutant segregation analysis

The segregation analysis was performed on genomic DNA extracts of *Synechocystis* and *Synechocystis* mutants obtained using the Gen-Elute Bacterial Genomic DNA Kit (Sigma-Aldrich) by effectuating a screen for GGR and ShMKS incorporation into *slr*0168 using PCR. The PCR analysis was performed using the GoTaq Green Master Mix (Promega). Segregation primers (Supplementary Table S1) amplifying GGR-dDNA and ShMKS-dDNA from the

homology regions were designed and used for analysis. The mutants were screened after a single round of selection/induction to determine their state of mutant segregation.

3. Results

3.1 Selection of the target site, Cas9 nuclease and protospacer

The gene *slr*0168 was selected as the target site for the incorporation of genes encoding heterologous proteins (cargo). It was chosen as it has been identified as a highly effective neutral site in *Synechocystis* (43) and therefore incorporation of the cargo at this site will avoid any undesired phenotypical changes to the host.

The Cas9 nuclease from *S. pyogenes* was chosen as it is the most widely employed nuclease for genome editing using CRISPR (44, 45) and recognizes an NGG PAM (46, 47), which has been reported to hold the highest activity when compared to other PAMs (48-50). However, the specificity of CRISPR elements, such as the sgRNA towards the target site is not ensured; hence, the protospacer region of the sgRNA could anchor to other sites instead of its target, leading to unintended editing by off-target effects (51, 52).

The protospacer is the sgRNA region that is complementary to the target site within the genome. Hence, this section is designed to anchor to the region adjacent to the PAM in order to allow the cleavage in the target to be edited (53). Within a protospacer, the first 12 nt closer to the PAM encompass the seed region, critical for proper cleavage. Mismatches in such region could significantly reduce the cleavage activity of the sgRNA/Cas9 complex while mismatches in the non-seed region (rest of protospacer) might have a less critical impact but can still increase the probabilities of off-target effects (48, 49). Hence, the fewer number of mismatches within the seed region would be less likely to have an impact on cleavage activity. Similarly, the fewer number of both, overall missmatches and off-targets would increase the probabilities of effective target-binding from a protospacer; which is why careful evaluation is crucial in order to identify the probable off-target activity of a given protospacer to increase its specificity and cleavage activity.

Therefore, using the CasOT software (40), the target gene was scanned in order to identify all the available protospacer candidates associated with the designated PAM and evaluate their off-target effects. For each candidate, a number of specific types of mismatches represented by codes (e.g. A205, A206) was given, followed by the number of off-targets for

each type of mismatch (shaded in blue) (Figure 1).The letter "A" on each code represents the selection of an NGG PAM. The first number after this letter indicates the number of mismatches in the seed region, the following two numbers represent the number of mismatches in the non-seed region (40). A total of 73 protospacer candidates were identified within *slr*0168. The selected protospacer (Supplementary Table S2) which was 20 nt long and within the suggested length for effective binding (22, 53), showed a total of 3 off-target sequences, one off-target generated by the A205 mismatch type and 2 off-targets generated by the A206 mismatch type (Figure 1). Hence, the one identified off-target sequence is shown for the A205 type and 2 sequences are shown for the A206 type (Figure 1). The A205 type indicates two mismatches in the seed region and five in the non-seed region while the A206 type represents 2 mismatches on the seed region and 6 in the non-seed region. The A000 code represents the original protospacer candidate, hence, it is not considered neither a type of mismatch nor an off-target. All identified off-target sequences for the selected protospacer are shown (Figure 1).

Compared to other candidates, the selected protospacer showed the fewest number of both off-targets as well as mismatches. The selected protospacer showed only two types of mismatches (A205 and A206) as opposed to other candidates such as 287f in which up to 11 types of mistatches were identified (A204 to A210). Similarly, only 3 off-targets were identified within the selected protospacer (1 for A205 and 2 for A206) whereas in other candidates such as 33r up to 28 off-targets were identified for a single type of mismatch (A207). The selected protospacer showed the lowest number of mismatches in both, the seed and non-seed region as well as the smallest number of off-targets. Therefore, it was considered the least probable to anchor to off-target sites, highly expected to present the highest activity and enable the knock-ins in the desired target editing site.

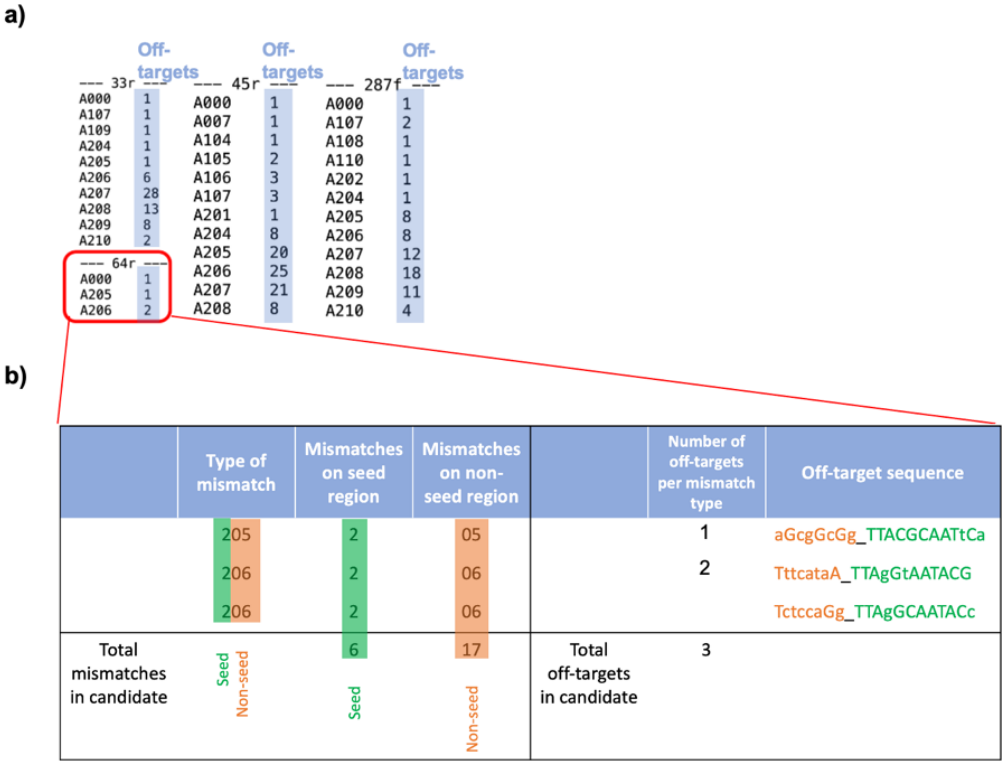

Figure 1 | Off-target analysis of the non-selected and selected (in red) protospacers showing the type and number of mismatches (codes; i.e. A206) and number of off-targets for each (shaded in blue) type of mismatch (a). Description of the off-targets and mismatches (lowercase) of the selected protospacer within the seed (green) and the non-seed (orange) region (b).

The selected protospacer, after integration with the Cas9 handle, must enable the formation of the proper secondary structure on the handle, which subsequently will aid in creating the sgRNA-Cas9 complex to effect the DSB at the target site (20, 41). The sequence of the protospacer contributes to whether the proper secondary structure will be formed within the handle. Moreover, the free energy between the interaction of such elements has, as well, been reported to have an impact in the formation of the proper hairpin. Protospacer-handle sequences showing a free energy value within 5% of the optimal folded structure have been deemed acceptable (41). Therefore, the secondary structure of the selected protospacer was then evaluated to determine if it would enable the formation of a proper handle loop. RNA secondary structure analysis suggested that the selected protospacer enabled the proper hairpin within the handle; moreover, it showed a free energy value of -14.61 kcal/mol and within 5% of the optimal folded structure (-14.30 kcal/mol) (Figure 2). Hence, an effective sgRNA-Cas9 complex is estimated to be formed intracellularly. Conjoined with the other selection factors

previously mentioned, the designed sgRNA (Figure 3d) (full sequences are provided in Supplementary Table S2) presented the highest probability of effectiveness for genome editing at *slr*0168.

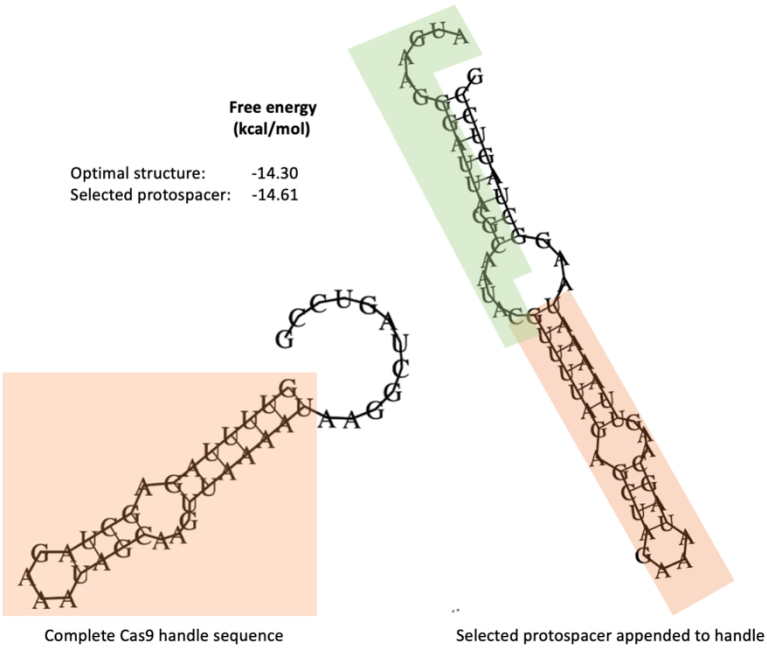

Figure 2 | Free energy values and Cas9-handle secondary structure analysis showing the proper structure (shaded in orange) within the complete Cas9 handle sequence and the proper structure also present (shaded in orange) with the selected protospacer appended (shaded in green) to the handle.

3.2 Design of the donor DNA templates

Two donor DNA templates were designed in this study, the sgRNA_GGR-dDNA (Figure 3b). and the ShMKS-dDNA (Figure 3c). One of the templates encompassed the coding sequence corresponding to the enzyme digeranylgeranylglycerophospholipid reductase (GGR) from *Sulfolobus acidocaldarius* (Genbank accession # SACI_RS04710; 1,359 bp) whilst the other template encompassed the coding sequences corresponding to enzymes MKS1 and MKS2 belonging to the methyl ketone synthesis operon from *Solanum habrochaites* (Genbank accession/ENA coding Id # AAV87156.1; 798 bp and # ADK38536.1; 627 bp, respectively). The coding sequences in both templates were placed under the control of the anhydrotetracycline-inducible PL22 promoter, the B0034 RBS and the T7 terminator. PL22 was chosen as it has been reported to provide a suitable expression of CRISPRi elements (54, 55). In both templates, the assembled genetic circuits were placed in between homology arms

comprising approximately 400 bp for the GGR mutants and approximately 200 bp for the ShMKS mutants, upstream and downstream of the chosen protospacer sequence in *slr*0168 (Figure 3a, exact base positions on the genome are provided in Supplementary Table S3 and S4). Care was taken to omit the PAM sequence from inclusion in the homology arms to prevent unintended and continuous cleavage from Cas9 at the target site.

Once the basic design of the donor DNA was complete, two divergent approaches were taken with respect to the further design. The template harbouring the GGR was chosen arbitarily to generate a markerless knock-in mutant for the evaluation of Cas9-based genome editing in *Synechocystis.* Therefore an antibiotic selection marker was not included in the template. The total size of the cargo including the homology arms was 2,293 bp and was titled sgRNA_GGR-dDNA. The genetic circuit harbouring the coding sequence for the designed sgRNA under the control of the PL22 promoter and the *S. pyogenes* terminator was placed in the vicinity of the cargo. The assembly was placed in the pEERM4 plasmid backbone. Conversely, the template harbouring the MKS1 and MKS2 coding sequences was chosen to generate a knock-in mutant with chloramphenicol resistance as the selection marker. The assembly was placed in the pEERM3 plasmid backbone. The total size of the cargo including the homology arms was 2,936 bp and was titled ShMKS-dDNA. The genetic circuit harbouring the coding sequence for the designed sgRNA under the control of the PL22 promoter in this template was inserted into the BssHII site of the backbone.

Synechocystis bearing the gene encoding the *S. pyogenes* Cas9 on plasmid pPMQK1 under the control of the PL22 promoter was used as the host organism. This strain was independently transformed with either donor DNA template and subjected to a single round of induction with anhydrotetracycline.

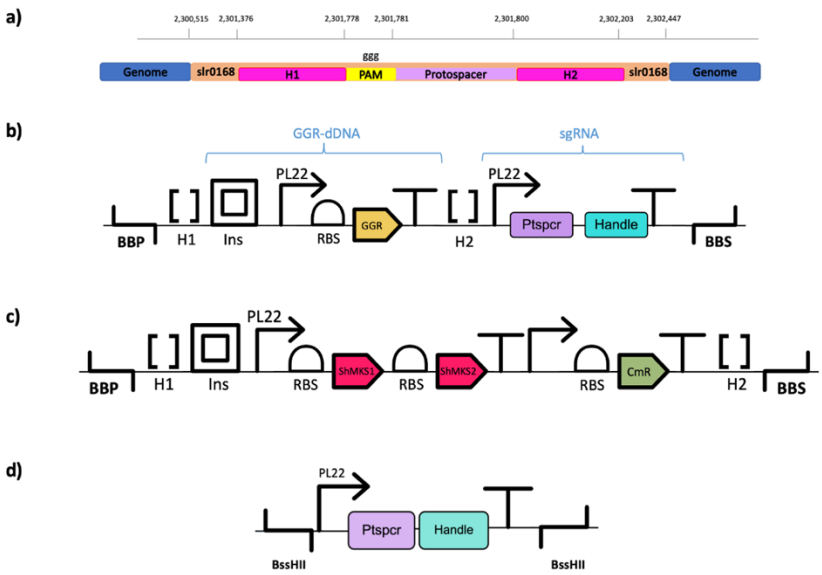

Figure 3 | Genetic context of the selected protospacer of the sgRNA, showing elements upstream and downstream of the protospacer in the genome of *Synechocystis* WT (a). Schematic of the sgRNA_GGR-dDNA, including flanking BioBrick prefix and suffix (BBP and BBS), upstream and downstream homology regions (H1 and H2), an insulator (ins), PL22 promoter, selected RBS, heterologous GGR gene (GGR), T7 terminator followed by the sgRNA (b). Schematic illustrating the ShMKS-dDNA similar elements to (b) but two heterologous genes (ShMKS1 and ShMKS2) encompassing the operon and a chloramphenicol resistance marker (CmR) (c). Schematic showing the designed sgRNA, with BssHII overhangs, PL22 promoter, selected protospacer (Ptspcr), Cas9 handle and *S. pyogenes* terminator (d).

3.3 Mutant segregation analysis

Induction of the CRISPR/Cas9 system was carried out in both strains with an aTc concentration of 200 ng/mL and white light for a first attempt and subsequently modified to 400 ng/mL of aTc and red light for a second attempt. The segregation state of the mutants was evaluated by PCR screening after a single round of selection/induction. For the first attempt, neither mutants showed full mutant segregation (data not shown). However, full mutant segregation was achieved on the second attempt after a single round of selection/induction for *Synechocystis::*Cas9_sgRNA_GGR-dDNA (Figure 4a) as well as for *Synechocystis::*Cas9_sgRNA_ShMKS-dDNA (Figure 4b). The mutant segregation analysis shows a band of approximately 350 base pairs for *Synechocystis* wild type strain (WT). This band corresponds to the homology regions (H1 and H2) flanking GGR-dDNA as well as ShMKS-dDNA and taken from the target neutral site *slr*0168 in the genome of *Synechocystis* wildtype. Therefore, this band, is expected to appear in the wildtype genome. Moreover, the amplicon from the pEERM4 vector bearing the GGR-dDNA shows a band of approximately 1,800 bp, corresponding to the sequence of the GGR-dDNA. Similarly, the amplicon from the pEERM3 vector bearing the ShMKS-dDNA shows a band of approximately 2,800 bp,

corresponding to the sequence of the ShMKS-dDNA. Both bands are not only expected since they are positive controls but also, represent the desired bands aimed to be present in the analysis of both mutants, respectively.

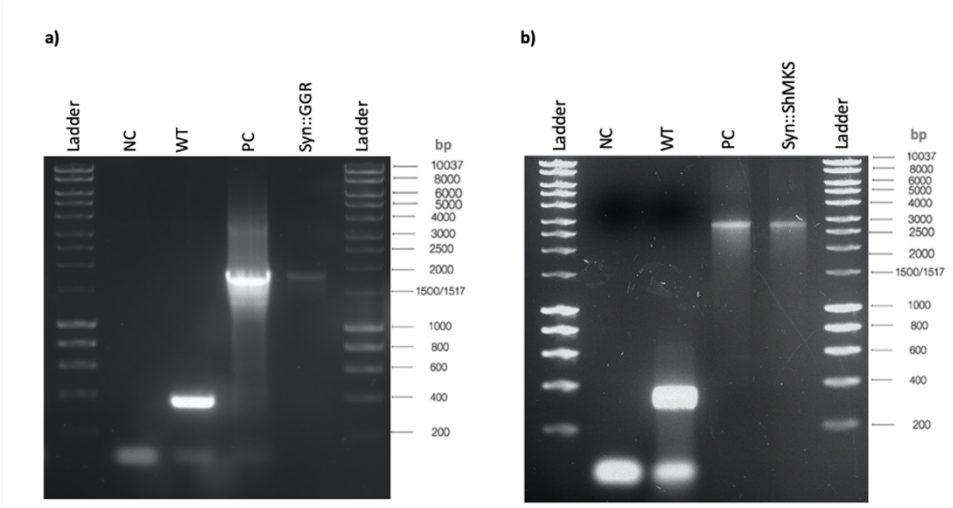

Figure 4 | Mutant segregation analysis of *Synechocystis*::Cas9_sgRNA_GGR-dDNA (a) and *Synechocystis::*Cas9_sgRNA_ShMKS-dDNA (b). Showing a negative control (NC); *Synechocystis* wild type (WT) displaying a band corresponding to the homology regions present in the GGR-dDNA, and ShMKS-dDNA taken from the target site in the genome of *Synechocystis* WT; a positive control (PC), displaying a amplicon of the desired size; *Synechocystis*::Cas9_sgRNA_GGR-dDNA for (a) and *Synechocystis*::Cas9_sgRNA_shMKS-dDNA for (b), both showing a band of the desired size that coincides with the desired band on their respective positive control as well as the lack of additional bands for both, (a) and (b).

Furthermore, the mutant segregation analysis for *Synechocystis*::Cas9_sgRNA_GGR-dDNA (Figure 4a) showed a band of the desired size which coincides with the desired band on the positive control. This suggests that the GGR-dDNA sequence is present in the genome of such mutants and, hence, that knock-in of the GGR-dDNA via CRISPR/Cas9 was effective. Furthermore, the desired band is present as a single band, highlighting the absence of additional bands. This indicates that only complete sequences of the GGR-dDNA were found in this mutant population and that the insertion of the GGR-dDNA into all chromosomal copies of all the individuals of the *Synechocystis*::Cas9_sgRNA_GGR-dDNA population was successful. It must be highlighted that this screening was made after a single round of induction. According to these results, the design and selected sgRNA was effective in the knock-in of a single gene up to 1,455 bp. Placing the sgRNA in close vicinity to the cargo did not affect the incorporation of the cargo at *slr*0168. Furthermore, mutant segregation does not seem to be affected although the template DNA did not bear a selection marker.

Comparably, the mutant segregation analysis for *Synechocystis*::Cas9_sgRNA_ShMKS-dDNA (Figure 4b), shows, as well, that the expected

band of the desired size is present for the CRISPR/Cas9 mutant which coincides with the desired band on the positive control. Furthermore, the desired band is present as a single band in the lane and there are no other additional bands, similar to the GGR mutant. The results indicate the insertion of the ShMKS-dDNA into all chromosomal copies of *Synechocystis* was successful. Thus, only complete sequences of the ShMKS-dDNA are present in all chromosomal copies within each cell and that all of these mutants in the culture bear the insert in their genome. It must be highlighted that this screening was made after a single round of antibiotic selectection/induction and that the ShMKS mutant bears a dual operon construct and a resistance marker gene, considerably larger than the GGR mutant. According to these results, the size and complexity of the dDNA, up to 2,399 bp, appears to have no effects on CRISPR/Cas9's ability of favouring mutant segregation after a single round of selection/induction. Furthermore, mutant segregation does not seem to be affected when the sgRNA and dDNA are constructed as distinct inserts in the same vector for inserts bearing a selection marker.

4. Discussion

Genetic engineering aims to express heterologous desired traits to ultimately generate industrial scale yields of a target product in a pre-determined chassis. To achieve this, the mutant culture must be fully segregated. The introduced traits must be homogenic and stable, they must be present in all chromosome copies of an individual and in all individuals. Failing to ensure the presence of the desired mutation in all chromosomal copies of all individuals will lead to mutation dilution, ultimately leading to loss of the mutants and low to null yields of the desired product (1, 3). Therefore, full mutation segregation is pivotal for industrial scale biomanufacturing. Furthermore, the achievement of full mutant segregation is more complicated in polyploid genomes such as *Synechocystis* (3, 56, 57). Nevertheless, CRISPR/Cas9 has the potential of overcoming this drawback given it has been deemed highly effective and successful (27, 30).

In this study, an efficacious sgRNA was designed for the successful knock-in of heterologous genes up to 2,399 bp long in the neutral site *slr*0168 of the highly polyploidy *Synechocystis* chassis with complete mutant segregation after a single round of selection/induction through CRISPR/Cas9. The selected protospacer showed the least number of both, mismatches as well as off-targets, resulting in effective insertions at the target site by

the designed sgRNA. Similarly, the selection of a sgRNA with the proper secondary structure formation of the Cas9handle constitutes another important factor to be considered in order to allow the correct formation of the sgRNA-Cas9 complex intracellularly and enable the DSB.

In the same fashion, it is important to highlight the use of red light, as opposed to white, when selecting the PL22 promoter for driving expression. The expression of all of the CRISPR/Cas9 elements was controlled by the aTc-inducible PL22. In the first attempt, white light was implemented and full mutant segregation was not achieved. This might have probably been caused by the sensitivity of aTc to white light (58). And even though PL22 has been deemed to drive reasonable expression with CRISPRi, it has as well described as weak under white light in some instances (59). Therefore, it is evidenced here that red light drives the best expression of PL22 in *Synechocystis*. Moreover, we designed all of the CRISPR/Cas9 elements to be driven by the same promoter in order to tune their induction under a single inducer and minimize interference. On the same token, we have as well demonstrated that the increased concentration of aTc to 400 ng/mL is more effective than the original of 200 ng/mL, which is why we recommend such concentration when working with PL22, red light, CRISPR/Cas9 and *Synechocystis*.

For both knock-ins, our designed sgRNA was able to achieve full mutant segregation after a single round of selection/induction in the mutagenesis mediated by CRISPR/Cas9. According to these results, the size and complexity of the dDNA appears to have no effects on the ability of the constructed CRISPR/Cas9 systems to achieve complete mutant segregation. Operons up to 2 genes and a resistance marker gene, with an insert size close to 2,399 bp appears to be easily managed by the CRISPR/Cas9 machinery regarding full mutant segregation after a single round of selection/induction. Furthermore, mutant segregation is highly favoured and is achieved to a full state after a single round of selection/induction whether the CRISPR/Cas9 elements are in close proximity to the cargo on the same plasmid or not. Moreover, the efficiency of our designed sgRNA is not affected by the generation of either markerless or mutants bearing a selection marker as it was effective in both instances. Furthermore, the selected length of homology regions of 400 bp and 200 bp seem to have facilitated well the recombination events (60, 61) and favoured CRISPR/Cas9. Similarly, tri-parental mating technique (42) for all CRISPR/Cas9 transformation events with *Synechocystis* facilitates well the delivery of all the genetic elements.

In conclusion, this work has described the development of an effective sgRNA, able to effectuate knock-ins of heterologous genes up to 2,399 bp in the neutral site *slr0168* of *Synechocystis* driven by CRISPR/Cas9 and achieve full mutant segregation after a single step of selection/induction. This sgRNA will be an effective addition to the Synthetic Biology toolkit for the genome editing of the highly promising but highly polyploid host organism, *Synechocystis*.

Funding

The authors would like to thank the National Mexican Council of Humanities, Science and Technology (CONAHCYT) and The Secretariat of Energy (SENER), México for the scholarship granted to Dr. Maria Isabel Nares Rodriguez (scholarship ID 406735). We also thank the European Molecular Biology Organization (EMBO) (Funding ID 7456) and PHYCONET (Funding ID PHYCSF-02) for their fellowship and funding. Finally, we would like to thank Dr. Paul Hudson and his group from the Karolinska Institute in Solna, Sweden for all of their support.

References

1. Caicedo-Burbano, P., T. Smit, H. Pineda Hernández, W. Du, and F. Branco dos Santos. (2020) Construction of fully segregated genomic libraries in polyploid organisms such as Synechocystis sp. PCC 6803. *ACS synthetic biology*, 9(10): p. 2632-2638.
2. Hitchcock, A., C.N. Hunter, and D.P. Canniffe. (2020) Progress and challenges in engineering cyanobacteria as chassis for light-driven biotechnology. *Microbial biotechnology*, 13(2): p. 363-367.
3. Pope, M.A. (2020) Improving Synechocystis sp. PCC 6803 as a model organism.
4. Taton, A., A.T. Ma, M. Ota, S.S. Golden, and J.W. Golden. (2017) NOT Gate Genetic Circuits to Control Gene Expression in Cyanobacteria. *ACS Synthetic Biology*, 6(12): p. 2175-2182. 10.1021/acssynbio.7b00203
5. Gao, X., T. Sun, G. Pei, L. Chen, and W. Zhang. (2016) Cyanobacterial chassis engineering for enhancing production of biofuels and chemicals. *Applied Microbiology and Biotechnology*, 100(8): p. 3401-3413. 10.1007/s00253-016-7374-2
6. Harrison, K.W. and B.G. Harvey. (2018) High cetane renewable diesel fuels prepared from bio-based methyl ketones and diols. *Sustainable Energy and Fuels*, 2(2): p. 367-371. 10.1039/c7se00415j
7. Heidorn, T., D. Camsund, H.-H. Huang, P. Lindberg, P. Oliveira, K. Stensjö, and P. Lindblad, *Synthetic Biology in Cyanobacteria*. Vol. 497. 2011. 539-579.
8. Oliver, N.J., C.A. Rabinovitch-Deere, A.L. Carroll, N.E. Nozzi, A.E. Case, and S. Atsumi. (2016) Cyanobacterial metabolic engineering for biofuel and chemical

production. *Current Opinion in Chemical Biology*, 35(p. 43-50. 10.1016/j.cbpa.2016.08.023
9. Pade, N., S. Erdmann, H. Enke, F. Dethloff, U. Dühring, J. Georg, J. Wambutt, J. Kopka, W.R. Hess, R. Zimmermann, D. Kramer, and M. Hagemann. (2016) Insights into isoprene production using the cyanobacterium Synechocystis sp. PCC 6803. *Biotechnology for Biofuels*, 9(1): p. 1-16. 10.1186/s13068-016-0503-4
10. Shabestary, K. and E.P. Hudson. (2016) Computational metabolic engineering strategies for growth-coupled biofuel production by Synechocystis. *Metabolic Engineering Communications*, 3(p. 216-226. 10.1016/j.meteno.2016.07.003
11. Deshpande, A., J. Vue, and J. Morgan. (2020) Combining random mutagenesis and metabolic engineering for enhanced tryptophan production in Synechocystis sp. strain PCC 6803. *Applied and environmental microbiology*, 86(9): p. e02816-19.
12. Pope, M.A., J.A. Hodge, and P.J. Nixon. (2020) An improved natural transformation protocol for the cyanobacterium Synechocystis sp. PCC 6803. *Frontiers in Plant Science*, 11(p. 372.
13. Ranade, S. and Q. He. (2021) Escherichia coli AraJ boosts utilization of arabinose in metabolically engineered cyanobacterium Synechocystis sp. PCC 6803. *AMB Express*, 11(1): p. 1-14.
14. Gao, E.-B., K. Kyere-Yeboah, J. Wu, and H. Qiu. (2021) Photoautotrophic production of p-Coumaric acid using genetically engineered Synechocystis sp. Pasteur Culture Collection 6803. *Algal Research*, 54(p. 102180.
15. Reinsvold, R.E., R.E. Jinkerson, R. Radakovits, M.C. Posewitz, and C. Basu. (2011) The production of the sesquiterpene β-caryophyllene in a transgenic strain of the cyanobacterium Synechocystis. *Journal of Plant Physiology*, 168(8): p. 848-852. 10.1016/j.jplph.2010.11.006
16. Begemann, M.B., E.K. Zess, E.M. Walters, E.F. Schmitt, A.L. Markley, and B.F. Pfleger. (2013) An Organic Acid Based Counter Selection System for Cyanobacteria. *PLoS One*, 8(10): p. e76594-e76594. 10.1371/journal.pone.0076594
17. Berla, B.M., R. Saha, C.M. Immethun, C.D. Maranas, T.S. Moon, and H.B. Pakrasi. (2013) Synthetic biology of cyanobacteria: Unique challenges and opportunities. *Frontiers in Microbiology*, 4(AUG): p. 1-14. 10.3389/fmicb.2013.00246
18. Cheah, Y.E., S.C. Albers, and C.A.M. Peebles. (2013) A novel counter-selection method for markerless genetic modification in Synechocystis sp. PCC 6803. *Biotechnology Progress*, 29(1): p. 23-30. 10.1002/btpr.1661
19. Doudna, J.A. and E. Charpentier. (2014) The new frontier of genome engineering with CRISPR-Cas9. *Science*, 346(6213).
20. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. (2012) A Programmable Dual-RNA – Guided. *Science*, 337(August): p. 816-822. 10.1126/science.1225829
21. Miura, H., R.M. Quadros, C.B. Gurumurthy, and M. Ohtsuka. (2018) Easi-CRISPR for creating knock-in and conditional knockout mouse models using long ssDNA donors. *Nature Protocols*, 13(1): p. 195-215. 10.1038/nprot.2017.153
22. Ran, F.A., D.H. Patrick, W. Jason, A. Vineeta, A.S. David, and Z. Feng. (2013) Genome engineering using the CRISPR- Cas9 system. *Nature Protocols*, 8(11): p. 2281-2281. 10.1038/nprot.2013.143
23. Sternberg, S.H., S. Redding, M. Jinek, E.C. Greene, and J.A. Doudna. (2014) DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature*, 507(7490): p. 62-67. 10.1038/nature13011
24. Davis, L. and N. Maizels. (2014) Homology-directed repair of DNA nicks via pathways distinct from canonical double-strand break repair. *Proceedings of the National*

Academy of Sciences of the United States of America, 111(10): p. 1-9. 10.1073/pnas.1400236111
25. Deltcheva, E., K. Chylinski, C.M. Sharma, K. Gonzales, Y. Chao, Z.A. Pirzada, M.R. Eckert, J. Vogel, and E. Charpentier. (2011) CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. *Nature*, 471(7340): p. 602-607. 10.1038/nature09886
26. Overballe-Petersen, S., K. Harms, L.A.A. Orlando, J.V. Moreno Mayar, S. Rasmussen, T.W. Dahl, M.T. Rosing, A.M. Poole, T. Sicheritz-Ponten, S. Brunak, S. Inselmann, J. De Vries, W. Wackernagel, O.G. Pybus, R. Nielsen, P.J. Johnsen, K.M. Nielsen, and E. Willerslev. (2013) Bacterial natural transformation by highly fragmented and damaged DNA. *Proceedings of the National Academy of Sciences of the United States of America*, 110(49): p. 19860-19865. 10.1073/pnas.1315278110
27. Anzalone, A.V., L.W. Koblan, and D.R. Liu. (2020) Genome editing with CRISPR–Cas nucleases, base editors, transposases and prime editors. *Nature biotechnology*, 38(7): p. 824-844.
28. Ran, F.A., P.D. Hsu, C.Y. Lin, J.S. Gootenberg, S. Konermann, A.E. Trevino, D.A. Scott, A. Inoue, S. Matoba, Y. Zhang, and F. Zhang. (2013) Double nicking by RNA-guided CRISPR cas9 for enhanced genome editing specificity. *Cell*, 154(6): p. 1380-1389. 10.1016/j.cell.2013.08.021
29. Barrangou, R. and L.A. Marraffini. (2014) CRISPR-cas systems: Prokaryotes upgrade to adaptive immunity. *Molecular Cell*, 54(2): p. 234-244. 10.1016/j.molcel.2014.03.011
30. Jinek, M., A. East, A. Cheng, S. Lin, E. Ma, and J. Doudna. (2013) RNA-programmed genome editing in human cells. *eLife*, 2013(2): p. 1-9. 10.7554/eLife.00471
31. Ramey, C.J., Á. Barón-Sola, H.R. Aucoin, and N.R. Boyle. (2015) Genome Engineering in Cyanobacteria: Where We Are and Where We Need to Go. *ACS Synthetic Biology*, 4(11): p. 1186-1196. 10.1021/acssynbio.5b00043
32. Hsu, P.D., E.S. Lander, and F. Zhang. (2014) Development and applications of CRISPR-Cas9 for genome engineering. *Cell*, 157(6): p. 1262-1278. 10.1016/j.cell.2014.05.010
33. Labuhn, M., F.F. Adams, M. Ng, S. Knoess, A. Schambach, E.M. Charpentier, A. Schwarzer, J.L. Mateo, J.-H. Klusmann, and D. Heckl. (2018) Refined sgRNA efficacy prediction improves large- and small-scale CRISPR-Cas9 applications. *Nucleic Acids Res*, 46(3): p. 1375-1385. 10.1093/nar/gkx1268
34. Wright, A.V., J.K. Nuñez, and J.A. Doudna. (2016) Biology and Applications of CRISPR Systems: Harnessing Nature's Toolbox for Genome Engineering. *Cell*, 164(1-2): p. 29-44. 10.1016/j.cell.2015.12.035
35. Biot-Pelletier, D. and V.J.J. Martin. (2016) Seamless site-directed mutagenesis of the Saccharomyces cerevisiae genome using CRISPR-Cas9. *Journal of Biological Engineering*, 10(1): p. 1-5. 10.1186/S13036-016-0028-1
36. Choi, K.R. and S.Y. Lee. (2016) CRISPR technologies for bacterial systems: Current achievements and future directions. *Biotechnology Advances*, 34(7): p. 1180-1209. 10.1016/j.biotechadv.2016.08.002
37. Heidi, L. (2019) CRISPR BABIES. *Nature (London)*, 570(7761): p. 293-296. 10.1038/d41586-019-01906-z
38. Li, H., C.R. Shen, C.H. Huang, L.Y. Sung, M.Y. Wu, and Y.C. Hu. (2016) CRISPR-Cas9 for the genome engineering of cyanobacteria and succinate production. *Metabolic Engineering*, 38(August): p. 293-302. 10.1016/j.ymben.2016.09.006
39. Englund, E., J. Andersen-Ranberg, R. Miao, B. Hamberger, and P. Lindberg. (2015) Metabolic Engineering of Synechocystis sp. PCC 6803 for Production of the Plant

Diterpenoid Manoyl Oxide. *ACS Synthetic Biology*, 4(12): p. 1270-1278. 10.1021/acssynbio.5b00070
40. Xiao, A., Z. Cheng, L. Kong, Z. Zhu, S. Lin, G. Gao, and B. Zhang. (2014) CasOT: a genome-wide Cas9/gRNA off-target searching tool. *Bioinformatics*, 30(8): p. 1180-1182.
41. Larson, M.H., L.A. Gilbert, X. Wang, W.A. Lim, J.S. Weissman, and L.S. Qi. (2013) CRISPR interference (CRISPRi) for sequence-specific control of gene expression. *Nature Protocols*, 8(11): p. 2180-2196. 10.1038/nprot.2013.132
42. Gale, G.A., A.A. Schiavon Osorio, A. Puzorjov, B. Wang, and A.J. McCormick, *Genetic modification of cyanobacteria by conjugation using the cyanogate modular cloning toolkit*. 2016: MyJoVE Corporation.
43. Pinto, F., C.C. Pacheco, P. Oliveira, A. Montagud, A. Landels, N. Couto, P.C. Wright, J.F. Urchueguía, and P. Tamagnini. (2015) Improving a Synechocystis-based photoautotrophic chassis through systematic genome mapping and validation of neutral sites. *DNA Research*, 22(6): p. 425-437. 10.1093/dnares/dsv024
44. Alonso-Lerma, B., Y. Jabalera, S. Samperio, M. Morin, A. Fernandez, L.T. Hille, R.A. Silverstein, A. Quesada-Ganuza, A. Reifs, and S. Fernández-Peñalver. (2023) Evolution of CRISPR-associated endonucleases as inferred from resurrected proteins. *Nature Microbiology*, 8(1): p. 77-90.
45. Takasugi, P.R., S. Wang, K.T. Truong, E.P. Drage, S.N. Kanishka, M.A. Higbee, N. Bamidele, O. Ojelabi, E.J. Sontheimer, and J.A. Gagnon. (2022) Orthogonal CRISPR-Cas tools for genome editing, inhibition, and CRISPR recording in zebrafish embryos. *Genetics*, 220(1): p. iyab196.
46. Kenney, C.T. and L.A. Marraffini. (2023) Rarely acquired type II-A CRISPR-Cas spacers mediate anti-viral immunity through the targeting of a non-canonical PAM sequence. *Nucleic Acids Research*: p. gkad501.
47. Wang, J., Y. Teng, X. Gong, J. Zhang, Y. Wu, L. Lou, M. Li, Z.-R. Xie, and Y. Yan. (2023) Exploring and engineering PAM-diverse Streptococci Cas9 for PAM-directed bifunctional and titratable gene control in bacteria. *Metabolic Engineering*, 75(p. 68-77.
48. Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, and L.A. Marraffini. (2013) Multiplex genome engineering using CRISPR/Cas systems. *Science*, 339(6121): p. 819-823.
49. Jiang, W., D. Bikard, D. Cox, F. Zhang, and L.A. Marraffini. (2013) RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nature Biotechnology*, 31(3): p. 233-239. 10.1038/nbt.2508
50. Liu, D., Z. Wang, A. Xiao, Y. Zhang, W. Li, Y. Zu, S. Yao, S. Lin, and B. Zhang. (2014) Efficient gene targeting in zebrafish mediated by a zebrafish-codon-optimized cas9 and evaluation of off-targeting effect. *Journal of genetics and genomics= Yi chuan xue bao*, 41(1): p. 43-46.
51. Guo, C., X. Ma, F. Gao, and Y. Guo. (2023) Off-target effects in CRISPR/Cas9 gene editing. *Frontiers in Bioengineering and Biotechnology*, 11(p. 1143157.
52. Rao, X., H. Zhao, C. Shao, and C. Yi. (2023) Characterizing off-target effects of genome editors. *Current Opinion in Biomedical Engineering*: p. 100480.
53. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier. (2012) A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity. *science*, 337(6096): p. 816-821.
54. Kojima, S. and Y. Okumura. (2020) Outer membrane-deprived cyanobacteria liberate periplasmic and thylakoid luminal components that support the growth of heterotrophs. *BioRxiv*: p. 2020.03. 24.006684.

55. Sarnaik, A., K. Sawant, J. Khadilkar, G. Pillai, R. Pandit, and A. Lali, *Cyanobacterial Cell Factories for Improved Carotenoid Biosynthesis through a Synthetic Biology Approach*, in *Next Generation Biomanufacturing Technologies*. 2019, ACS Publications. p. 23-39.
56. Pecoraro, V., K. Zerulla, C. Lange, and J. Soppa. (2011) Quantification of ploidy in proteobacteria revealed the existence of monoploid, (mero-)oligoploid and polyploid species. *PLoS ONE*, 6(1). 10.1371/journal.pone.0016392
57. Zerulla, K., K. Ludt, and J. Soppa. (2016) The ploidy level of synechocystis sp. PCC 6803 is highly variable and is influenced by growth phase and by chemical and physical external parameters. *Microbiology (United Kingdom)*, 162(5): p. 730-739. 10.1099/mic.0.000264
58. Huang, H.H. and P. Lindblad. (2013) Wide-dynamic-range promoters engineered for cyanobacteria. *Journal of Biological Engineering*, 7(1): p. 1-11. 10.1186/1754-1611-7-10
59. Nayak, M., *Metabolic engineering of Synechocystis PCC 6803 with the aim to enhance isobutanol/1-butanol production using CRISPR interference*. 2017.
60. Ikeuchi, M. and S. Tabata. (2001) Synechocystis sp. PCC 6803 - a useful tool in the study of the genetics of cyanobacteria. *Photosynthesis research*, 70(1): p. 73-83. 10.1023/A:1013887908680
61. Marbouty, M., C. Saguez, C. Cassier-Chauvat, and F. Chauvat. (2009) ZipN, an FtsA-like orchestrator of divisome assembly in the model cyanobacterium Synechocystis PCC6803. *Molecular microbiology*, 74(2): p. 409-420.