Genome streamlining to improve performance of a fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973

Running Title: Genome streamlining in a photoautotroph

Annesha Sengupta[a*], Anindita Bandyopadhyay[a], Debolina Sarkar[b$], John I. Hendry[b^], Max G. Schubert[c], Deng Liu[a], George M. Church[c,d], Costas D. Maranas[b], Himadri B. Pakrasi[a#]

[a]*Department of Biology, Washington University, St. Louis, MO, USA*

[b]*Department of Chemical Engineering, Pennsylvania State University, PA, USA*

[c]*Wyss Institute for Biologically Inspired Engineering, Harvard University, MA, USA*

[d]*Department of Genetics, Harvard Medical School, MA, USA*

[#]Corresponding author: Himadri B. Pakrasi, pakrasi@wustl.edu.

Current addresses:

[*]*Department of Chemical Engineering, University of Toronto, Canada*

[$]*International Flavors and Fragrances, 925 Page Mill Road, Palo Alto, CA 94304*

^*Henry Jackson Foundation, 6720A Rock ledge Drive, Bethesda, Maryland 20817*

Abstract:

Cyanobacteria are photosynthetic organisms that have garnered significant recognition as potential hosts for sustainable bioproduction . However, their complex regulatory networks pose significant challenges to major metabolic engineering efforts, thereby limiting their feasibility as production hosts. Genome streamlining has been demonstrated to be a successful approach for improving productivity and fitness in heterotrophs but is yet to be explored to its full potential in phototrophs. Here we present the systematic reduction of the genome of the cyanobacterium exhibiting the fastest exponential growth, *Synechococcus elongatus* UTEX 2973. This work, the first of its kind in a photoautotroph, involved an iterative process using state-of-the-art genome-editing technology guided by experimental analysis and computational tools. CRISPR/Cas3 enabled large, progressive deletions of predicted dispensable regions and aided in the identification of essential genes. The large deletions were combined to obtain a strain with 55 kb genome reduction. The strains with streamlined genome showed improvement in growth (up to 23%) and productivity (by 22.7%) as compared to the WT. This streamlining strategy not only has the potential to develop cyanobacterial strains with improved growth and productivity traits but can also facilitate a better understanding of their genome to phenome relationships.

Importance:

Genome streamlining is an evolutionary strategy used by natural living systems to dispense unnecessary genes from their genome as a mechanism to adapt and evolve. While this strategy has been successfully borrowed to develop synthetic heterotrophic microbial systems with desired phenotype, it has not been extensively explored in photoautotrophs. Genome streamlining strategy incorporates both computational predictions to identify the dispensable regions and experimental validation using genome editing tool and in this study we have employed a modified strategy with

the goal to minimize the genome size to an extent that allows optimal cellular fitness under specified conditions. Our strategy has explored a novel genome-editing tool in photoautotrophs which, unlike other existing tools, enables large, spontaneous optimal deletions from the genome. Our findings demonstrate the effectiveness of this modified strategy in obtaining strains with streamlined genome, exhibiting improved fitness and productivity.

Keywords:CRISPR-Cas3, Genome minimization, cyanobacteria, large progressive deletion, essential gene identification

Introduction:

Cyanobacteria are the most ancient and abundant oxygenic photosynthetic organisms that are largely responsible for the Earth's viable environment (1). These photosynthetic prokaryotes have been identified as potential platforms for sustainable carbon-neutral bioproduction because of their unique ability to harvest sunlight as their sole energy source for converting greenhouse gases (carbon dioxide) into value-added chemicals. This bioprocess in theory is also economically sustainable as it enables free and ubiquitous substrates to enter the bioeconomy. In comparison to other photoautotrophs, cyanobacteria have several advantages and therefore, efforts are ongoing to understand and develop the cyanobacterial platform as sustainable bio-factories (2-5). However, relatively slower growth and limited knowledge of their genomic traits as compared to heterotrophs such as *E. coli* or yeast have restricted progress in this direction. Though the recent isolation of a few fast-growing strains have made cyanobacterial-based bioproduction more compelling and tractable than ever, concerted efforts are needed to get their productivity at par with their heterotrophic counterparts (6-10). Although most commonly studied cyanobacterial genomes are generally smaller in size than that of *E. coli*, they exhibit cryptic metabolic and regulatory features, owed likely to their photosynthetic lifestyle, unique evolutionary history, and adaptation to various unfavorable environmental conditions(11).

The main aim of this study was to test the feasibility of employing the genome reduction strategy as a means to shed excess biological complexity and simplify the genome of a fast-growing cyanobacterium without compromising its desirable traits. The clade *Synechococcus elongatus* hosts all of the fast-growing cyanobacterial strains identified to date (8) and a genome minimization approach will be beneficial for unravelling the genome level function and the overall metabolism of these strains which in turn will aid their development into cell factories. The goal

is not to obtain a truly minimum genome for a photosynthetic organism, but rather identify and remove genes dispensable under bioproduction-relevant conditions (high light and CO_2) without compromising growth and productivity.

Genome streamlining is a natural evolutionary process of eliminating non-beneficial genes, since a smaller genome reduces the metabolic burden on the cell and improves fitness (12-14). As an engineering strategy, it's been successfully employed in model heterotrophs leading to improved fitness and performance(15-18). A genome reduction of 25% in *E. coli* led to a 1.6-fold improvement in growth rate as well as improved recombinant protein production (19, 20). Similarly, genome streamlining in a *Pseudomonas* strain resulted in several appealing traits such as faster growth, increased biomass production, enhanced plasmid stability, and overall a more efficient energy metabolism (21, 22). These studies indicate that a chassis strain with a streamlined genome avoids the unnecessary burden of replicating and expressing genetic elements that are not useful under production conditions. Reducing this unnecessary genetic burden may ensure more cellular resources are available for expression of heterologous pathways. By removing genes of unknown function, it also creates a chassis organism that is more fully-understood, more amenable towards genetic engineering and synthetic biology efforts (16). Despite the favorable outcomes of genome minimization in heterotrophs (20-23), limited efforts have been made to implement such strategies in phototrophs. So far two reports explore genome streamlining. Removing ~2% (118 kb) of the genome of *Anabaena* PCC 7120 has been demonstrated using the CRISPR-Cas12a system for targeted deletions, but this study did not explore the effects of these edits on strain performance (24). Deletion of several large fragments of DNA from the genome of *Synechococcus elongatus* PCC 7942 has been performed, producing a septuple mutant with approximately 3.8% of its genome removed. These mutants were further studied to understand the changes in the

transcriptomics profile of the cells resulting from the deletions. The CRISPR-Cas12a editing tool was used to obtain the targeted, specified and markerless deletions in this study (25). Though Cas12a is a dynamic and versatile tool, it does not allow the flexibility to explore and identify unknown stretches of dispensable and indispensable regions in the genome of a strain. Recently, a novel Class I multi-Cas protein was commissioned for large deletions in heterotrophs. The dual helicase and exonuclease activity of Cas3 enabled large simultaneous bi-directional deletions, without the necessity of a repair template (26). Recently in one of our studies, we have successfully commissioned an inducible CRISPR Cas3 system in *Synechococcus elongatus* UTEX 2973 to truncate the light-harvesting antenna structure for maximizing fitness and productivity under specified condition (27). However, this tool has not been explored for genome streamlining in cyanobacteria. Therefore, it is of interest to investigate this Cas system and exploit its beneficial features for genome minimization of photoautotrophs.

In this study we investigate the effect of systematic reduction of dispensable regions from the genome of *Synechococcus* 2973, the fastest growing, high light thriving cyanobacterium that exhibits high sucrose production titer (6, 10). We first identified five large stretches of dispensable genomic regions in this strain using the MinGenome algorithm (28) and then commissioned the novel CRISPR-exonuclease system, CRISPR-Cas3 to achieve flexible progressive large deletions. CRISPR-Cas3 besides deleting large regions, enabled identification of non-dispensable genes in *Synechococcus* 2973 which otherwise were predicted as dispensable. We successfully deleted the optimal stretch of two of the five dispensable regions identified by our *in silico* analysis. The two large deletions were combined to obtain a strain with genome reduction of 55 kb. The strains with reduced genome showed improved growth and sucrose productivity. This proof-of concept study (Fig. 1) demonstrates that systematic minimization of cyanobacterial genomes has the potential to

develop these organisms as super-strains that might hold the potential to boost a carbon-neutral bio-economy and mitigates climate issues.

Result:

In silico prediction of probable dispensable regions from the genome of Synechococcus 2973

The first step towards streamlining the genome of *Synechococcus* 2973 was to identify the dispensable genomic regions (Fig. 1). Since *Synechococcus* 2973 shares >99% genomic identity with the model strain *Synechococcus* 7942, the list of essential genes already available for *Synechococcus* 7942 (29) was borrowed and mapped to *Synechococcus* 2973 using a bidirectional protein blast with a stringent e-value cut-off of 10^{-10} to avoid spurious hits. The MinGenome algorithm (28) was employed to predict large dispensable regions with three criteria: (1) The growth rate of the strain is fixed at the maximum, (2) Essential genes mapped from *Synechococcus* 7942 cannot be deleted, and (3) Find the longest stretch first and iterate to find the consecutive ones. This strategy predicts five large regions of the genome can be dispensed without affecting the strain fitness (Table 1). The list of gene annotations of each cluster are provided in Table S1-S5.

CRISPR Cas3-mediated genome editing enabled deletion of large genomic regions

A Rhamnose-Theophylline inducible Class I CRISPR system involving multi-Cas proteins was employed to streamline the genome of *Synechococcus* 2973 (Fig. 2a). In this CRISPR system, Cas3 is the key component exhibiting both helicase and nuclease activity, enabling large bidirectional deletions. The other Cas proteins (Cas5, 7, and 8) help in maintaining the fork structure required for these long deletions (26, 27). Unlike Cas12a-mediated editing where the length of deletion is pre-determined based on existing knowledge of the region of interest (designed repair template), the Cas3 system allows spontaneous deletion of DNA stretches on either side of the targeted region, as long as the deletions are not detrimental for the strain. Therefore, the Cas3 CRISPR system has the potential to not only identify previously unknown

dispensable regions in the genome but also uncover essential genes within a stretch of DNA that has been computationally determined as dispensable, thereby averting any adverse effects on the strain from large scale deletion experiments.

In our previous study, as a proof of concept, the inducible CRISPR-Cas3 system was tested to delete dispensable genes from the genomic cluster encoding the light-harvesting antenna complex of *Synechococcus* 2973. Interestingly, every deletion attempt led to a strain with a reduction of 4 kb. Investigation of the region beyond 4kb revealed that the genes immediately flanking the deletion on either side were designated as essential as per the gene essentiality data (27). These experiments revealed the potential of the Cas3 tool in identifying essential genes in a stretch of DNA while simultaneously identifying dispensable regions. Thus our next goal was to test the feasibility of using the Cas3 entourage for deletions of large genomic segments from cyanobacterial genomes, analogous to what has been achieved in heterotrophs. In this work, a strategy similar to a heterotrophic study (27) was implemented to target all the 5 regions predicted as dispensable by our *in silico* analysis (Table 1). We first commissioned CRISPR-Cas3 for the largest stretch identified, R1 (Table 1), in order to optimize the system for large deletion in *Synechococcus* 2973 (Fig. 2). A 34 bp long gRNA was designed for R1 that targeted the gene M744_13025 (which is approximately at the midpoint of the predicted region). Around 30 colonies out of 1000 were initially screened for deletions by using tiling PCR with 3 sets of primers (P1, P3 and P4) and around 20% of the colonies showed no deletions (Fig. S1). The probable position of the primers initially chosen are shown in Fig. 2b. A more extensive tiling PCR with multiple primers helped identify the range of deletions in different colonies, such as SG33 and SG14 (Fig. 2b). Similar analysis of all the rest 80% positive colonies indicated a wide range of deletions from 0 to >30kb (Fig. 2c). A frequency analysis showed around 40% of the deletions were in the range

of 14-15kb and ~10% colonies showed >30 kb deletion (Fig 2d). This demonstrated that the CRISPR-Cas3 tool has the potential to execute a large range of deletions in cyanobacteria, generating a library of random deletions.

This tested inducible CRISPR-Cas3 system was then used to create strains with streamlined genome by targeting the predicted regions (Table 1). Fig 3a briefly shows the workflow for generating these deletion mutants using this inducible CRISPR-Cas3 system. CRISPR-mediated targeting of the R1 region gave rise to a markerless strain 2973Δ33.5kb showing a deletion of 33.5 kb region (hereafter mentioned as SG33) (Fig. 3). The extent of deletion in the strain was initially confirmed by tiling PCR (Fig. 3b), and then the total length of deletion was determined by confirmation PCR using set of primers that were previously determined based on the tiling PCR results (Fig. 3b). Using primers flanking the 40kb targeted region, a band of ~7kb was obtained in the SG33 strain, indicating a large deletion of ~33kb (Fig.3c,d). Finally, whole genome sequencing confirmed the range of deletion to be 33.5kb (Fig. 3e). The WGS also confirmed no additional mutations in the SG33 strain.

Similar workflow was applied for the R2 region which led to its partial deletion. Though this region was predicted to be ~26.4 kb long (Table 1), experimentally the largest deletion obtained from was of 19.7 kb, giving rise to the strain 2973Δ19.7 kb (hereafter mentioned as SG20) (Fig 4). While the first two regions, R1 and R2 could be deleted (Fig. S2), the other identified regions (R3, R4, R5) could not, despite several attempts with different gRNA and PAM sequences (Fig. S3). We then attempted to delete these regions using the CRISPR-Cas12a system with 2 gRNAs (24, 30), but remained unsuccessful as well (data not shown). This suggested that the regions predicted as dispensable based on S7942 essentiality data are probably not accurate enough and indicated the need for *Synechococcus* 2973-specific essentiality data.

We then proceeded to combine the two large deletions (SG33 and SG20) to obtain a single strain with a large reduction in the genome. Initially, SG33, the strain with the largest deletion, was used as the base strain to tier the next largest deletion . However, the conjugation efficiency of this strain was very poor [1000 fold lower than the WT or the SG20 strain (data not shown)]. Therefore, we decided to use the SG20 strain as the base strain into which the plasmid carrying the Cas3 machinery for R1 deletion was introduced. Whole Genome Sequencing of one of the transformant colonies showed a total deletion of 55kb in this strain (henceforth referred to as SG55) (2973Δ19.7kbΔ34.3kb). Unlike SG33, where 33.5kb was deleted upon targeting R1, a 34.3 kb region could be deleted when the same region was targeted in the SG20 strain. The above results confirmed the efficacy of the CRISPR-Cas3 genome editing tool in creating spontaneous and large deletions in cyanobacteria.

Minimization of genome showed improved growth and productivity

The strains with streamlined genome were characterized to determine the fitness and photosynthetic productivity. Since our primary goal was to develop a reduced genome cyanobacterial host, that can thrive under high light and high CO_2, all studies were performed under high light (1500µmoles.m^{-2}.s^{-1}) and high CO_2 (1% CO_2 bubbling) conditions. A comparative analysis of growth rates revealed an increase of 23%, in the SG33 strain compared to the WT. In contrast, a growth rate increase of only 5% was observed in SG20, while SG55 showed an improvement of 9% (Fig. 5a, Table S6). However, this enhancement in growth rate was not evident under conditions where carbon was a limiting factor (Fig. S4). Interestingly, the SG55 strain showed a 10 fold higher transformation efficiency with RSF1010 plasmids compared to the WT. Analysis of the genome sequence of this strain indicated that this increase in efficiency could

likelybe due to a mutation it acquired in SeAgo, an argonaute protein known to reduce the efficiency of RSF1010 plasmid transformation (31).

In addition to the large chromosomal deletions, the two endogenous plasmids pANL (~50kb) and pANS (~7kb) were also deleted from the WT strain. Since these deletions did not show any added growth advantages (Fig. S5), these deletions were not tiered into the SG55 strain. Analysis of the photosynthetic efficiency of the strains with streamlined genomes revealed no difference in quantum yield (Fv/Fm 0.35±0.08) as compared to the WT, indicating no adverse impact of the deletions on key physiological processes. Thus, the streamlining strategy employed in this study reduced the metabolic burden in the WT *Synechococcus* 2973 strain resulting in faster growth but did not disrupt its basic cellular functions.

The final strain with streamlined genome (Δ55kb) and improved phenotypic characteristics was engineered to secrete sucrose by integrating *cscB,* a heterologous sucrose transporter gene (10, 32) into the genome. The engineered SG55 strain (SG55-cscB) showed 22.7% higher sucrose production as compared to the strain expressing the transporter alone (10) (Fig 5b). Our findings demonstrate that genome streamlining in cyanobacteria can be a promising strategy for improving both productivity and fitness, provided the correct genes are dispensed.

Analysis of the deleted regions revealed a large number of hypothetical genes.

Streamlining of genome is an effective strategy not only for developing a strain as a production host with beneficial features but analysis of the dispensable genes might also help deduce the role of certain genetic elements. In our analysis, we observed that in the predicted dispensable regions >50% of genes are annotated as hypothetical and majority of these hypothetical genes are in R1 and R2 (Table S1-S5), therefore deciphering the underlying reason for the observed phenotype of the deleted strain was not possible with certainty.

Since in this study the gene essentiality data, was borrowed from *Synechococcus* 7942 (29), there are some discrepancies between the experimentally obtained deletions and predicted dispensability data. To exemplify, in the R1 region there are 37 genes (M744_12940-M744_13135) and the genes flanking this region have been categorized as essential. Conversely, our sequencing results for the SG33 strain which has the R1 deletion, showed the deletion of M744_13140, a gene annotated as essential. Though for non-model strains, strain-specific essentiality data is important, the use of the novel Cas3 system provided the flexibility to dispense the optimal length of regions from the genome for improved fitness. Therefore, the genome streamlining with the CRISPR-Cas3 system is an advantageous strategy and a more high throughput streamlining effort might help create a library of strains exhibiting varied phenotypes.

Discussion:

Genome streamlining is a synthetic biology approach which allows strategic reduction of the genome for attaining a desirable strain phenotype and this approach has been demonstrated to be successful in heterotrophs. There are two approaches of genome streamlining, bottom-up and top-down (33). Though most minimized genome studies have employed a bottom-up strategy to create truly minimal genome from scratch (16, 34), this strategy can pose several challenges as it demands a vivid and thorough knowledgebase of all biological processes and interactions and requires efficient synthetic DNA synthesis and assembly tools. Moreover, cyanobacteria being a polyploid organism exponentially enhances the challenge of introducing and maintaining the synthetic chromosome. The other approach is top-down, which involves systemic streamlining of the existing genome based on prior information regarding the core essential genes. Most genome streamlining efforts have relied on traditional and more recently, the CRISPR-Cas12a/9-mediated genome editing tools (20, 21, 24, 25). For strains where the essentiality information is available, the strategy to delete fixed, specified regions of the genome is advantageous as it leaves less room for casualties such as drastic loss of fitness. However, even with prior knowledge, the experimental outcome might not correlate with *in silico* predictions. Like, in *E. coli* MG1655 a reduction of 29.7% exhibited severely impaired phenotype, while 7% reduction showed no retarded phenotype (20). Therefore, for known and more importantly for newly discovered strains, the spontaneity or randomness in the extent of deletion might be the key to obtain a strain with enhanced features, and CRISPR-Cas3 mediated genome editing has the potential (26, 27). The RNA-guided Cas3 protein has dual helicase-nuclease activity which allows large, progressive, random deletion of genomic region unless encountered with an essential gene. CRISPR-Cas3 has been demonstrated to be effective in heterotrophs, such as *Pseudomonas aeruginosa*, where deletion of genomic

276 regions as large as 424 kb with a mean of 92.9 kb and median of 58.2 kb deletion (26) was
277 observed. Therefore, its use in photoautotrophs for streamlining and other large scale editing
278 purposes is worth exploring.

279 In this study, we focused on developing a genome streamlining strategy for a non-model
280 cyanobacterium *Synechococcus* 2973 to obtain a strain with reduced genome and improved fitness.
281 *Synechococcus* 2973 is the fastest growing cyanobacterium known so far with a doubling time
282 comparable to heterotroph model strains such as yeast (6, 35). This organism has the potential to
283 be developed as the next-generation production host, however, the complexity of the strain poses
284 challenges for major engineering efforts. We attempted to minimize the metabolic burden on this
285 strain by first identifying dispensable genomic regions using the MinGenome algorithm (28) and
286 then removing them under bioproduction-relevant conditions without compromising growth and
287 productivity (Fig. 1). This iterative integrated approach led to the creation of engineered
288 *Synechococcus* 2973 strains with minimized genomes exhibiting significant growth advantage (Fig
289 5a). Our results indicate that the extent of growth advantage is not dependent on the extent of
290 genome reduction but on the set of deleted genes. Our analysis revealed that some genes in R1 are
291 predicted as phage-associated proteins. Though SG55 strain has the largest range of deletion
292 (55kb), the growth improvement is more in SG33 (33.5kb deletion) and this might be due to the
293 deletion of prophage-like genes. A previous study in *Vibrio natriegens* revealed that deletion of
294 prophage-containing genomic regions is an effective engineering strategy for improving growth
295 (36).

296 Since a majority of the genes are annotated as hypothetical, further analysis to decipher the
297 genetic features responsible for the phenotype was not possible. We tested the effect of
298 minimization on productivity by engineering SG55 strain for sucrose overproduction and observed

299 a 22.7% improvement in sucrose titer (Fig. 5b) as compared to sucrose producing WT (10). This
300 strategy of streamlining the genome for improved growth and phenotype might come at the cost
301 of robustness under non-controlled conditions such as outdoor like conditions (carbon limited).
302 Under carbon limited condition, the engineered strains failed to show the improved phenotype
303 (Fig. S4). This decreased fitness in conditions mimicking natural environment (limited carbon
304 available) might be a boon as it lessens the risk of their release and the chance to overtake the WT
305 populations. However in under elevated CO_2, these strains might be advantageous to utilize more
306 available carbon.
307 A more high throughput understanding of the phenotype and genotype relationship would
308 provide a better insight into the complexity of the strains. The stochastic nature of the CRISPR-
309 Cas3 system offers the potential to discern essentiality of genes in the targeted regions and obtain
310 an optimal dispensable stretch while retaining or enhancing strain fitness. This hybrid strategy of
311 combining computational analysis with progressive deletion of genomic regions can generate a
312 library of engineered model as well as non-model photoautotrophic strains with streamlined
313 genomes, thereby paving the path for developing these remarkable green hosts into predictable
314 bio-systems with potential to mitigate climate problems and boost bio-economy.
315

Materials and Methods:

Computationally predicting longest dispensable regions

To identify regions in the *Synechococcus* 2973 genome that can be deleted without affecting organism growth, we used a genome reduction algorithm called minGenome (28). minGenome uses a mixed-integer linear programming based approach to systematically identify (in descending order) all contiguous dispensable regions. First, essential genes were identified using sequence homology with *Synechococcus* 7942 using a bidirectional protein BLAST. These can then be flagged and their deletion prohibited when implementing minGenome. Next, we used flux balance analysis on the published genome-scale metabolic model (GSM) for *Synechococcus* 2973 (37) to compute the maximum growth rate using 10 mmol/gDW hr CO_2 (as basis), and default model parameters. minGenome was then implemented using this metabolic model while fixing the growth rate at the determined maximal value, and constraining all essential genes to be preserved. Longest contiguous genome regions to be deleted were thus iteratively identified. It should be noted that as long as the GSM is used to simulate carbon-limited phototrophic growth, the deletion regions identified here should be invariant of the specific model parameters used.

Chemicals and Reagents

All enzymes were purchased from New England Biolabs (NEB, Ipswich, MA, USA) and ThermoFisher Scientific (Waltham, MA, USA). The molecular biology kits were obtained from Sigma-Aldrich (St. Louis, MO, USA). The chemicals, reagents and antibiotics used in this study were of analytical/HPLC grade and were procured from Sigma-Aldrich (St. Louis, MO, USA). The primers were ordered from Integrated DNA technologies (IDT, Coralville, IA, USA). The plasmids were sequenced by Genewiz® (South Plainfield, NJ, USA).

339 *Cultivation condition of the WT and mutants*

340 The WT *Synechococcus* 2973 strain and mutant strains were cultivated and maintained in Caron
341 chamber under 300μmoles.m^{-2}.s^{-1} light and 0.5% CO_2 at 38˚C and rpm of 250 rpm. *E. coli*
342 containing specific plasmids was cultivated overnight at 37˚C in LB supplemented with
343 appropriate antibiotic. Freezer stocks were maintained at -80˚C in 7% DMSO for cyanobacterial
344 strains and 25% glycerol for *E. coli* strains.

345 *Development of Cas3 mediated large random deletions*

346 The engineered strains were constructed using the CRISPR-Cas3 editing technique novel to the
347 cyanobacterial system. This is a multi-Cas system where Cas3 has the nuclease-helicase activity
348 (26). In our previous study we developed a an inducible CRISPR plasmid pSL3577 (pRhRB-
349 Cas3578), where the *cas* genes and the gRNA are controlled using the Rhamnose inducible
350 promoter and Theophylline inducible riboswitch (Fig. S6) (27, 38). . The BsaI restriction site
351 was used as the site for gRNA cloning in pSL3577 and this system does not require any
352 predetermined repair template. The recombinant plasmid (pSL3578) was conjugated into the
353 *Synechococcus* 2973 by triparental mating using a modified protocol. Briefly, the recombinant
354 plasmids were transformed into the competent cells of WM6026 containing the pRL623
355 plasmid. This strain was used for conjugation. 2 mL of *E. coli* cells were grown overnight in
356 SOB media supplemented with 30 μg/ml chloramphenicol and 250 mM DAP, which was
357 diluted in 25 mL SOB and DAP (without antibiotic) and grown to an OD of 0.1-0.2. This 25
358 mL of exponentially grown conjugal strain was mixed with 5 mL of exponentially grown
359 cyanobacterial culture and centrifuged at 2000 rpm. The mixed cells were washed gently and
360 resuspended in 400 μL of BG-11 medium and evenly spread on to a HATF nitrocellulose filter
361 paper (MilliporeSigma, St. Louis, MO, USA) that was placed on a BG-11 5% LB agar plate

without antibiotic, prior to the experiment. The plates were incubated for 24 h under 100 µmoles.m^{-2}.s^{-1} light and ambient air at 38˚C to obtain mat growth of cyanobacterial cells. The filters were then carefully transferred to BG-11 agar plate with 50 µg/mL kanamycin and incubated under 300 µmoles.m^{-2}.s^{-1} light and 0.05% CO_2 until the lawn growth disappeared and individual colonies appeared. A transformant colony was selected randomly and grown in 20ml BG11+Kan media. Then it was induced for 24 h with 2g/L Rhamnose and 1mM Theophylline. Post induction, the culture was diluted and plated on antibiotic plate to obtain single colonies, which were tested for deletion using tiling PCR (26), whole genome sequencing (WGS) as reported earlier (27). Primers are listed in Table S7.

Growth characteristics

The WT and the mutant strains were grown at 38 ˚C in 100 ml glass cultivation tubes of Multi-Cultivator photobioreactor (Photon Systems Instrument, Multi-Cultivator MC 1000, Czech Republic) containing 50 ml BG11 media under light intensity of 1500 (HL) µmoles.m^{-2}.s^{-1}. The aeration provided for growth was by bubbling either 1% CO_2 mixed air or atmospheric air through the culture. Cells were inoculated at an OD720nm of 0.025-0.1 and were cultivated to an OD of 1-1.5. The growth rate and doubling time was calculated in the exponential phase of growth. All experiments were performed in biological replicates (at least n=3).

The absorption spectra of the WT and mutant strains were obtained at room temperature using Shimadzu UV-1800 spectrophotometer. The whole cell chlorophyll contents were calculated from the absorption spectra using the formulae obtained from Arnon et al., 1974 (39). The cultivated cyanobacterial cells were normalized on the basis of chlorophyll content to determine the quantum efficiency (40) of strains grown under growth condition using the FL-200 dual modulation PAM fluorometer with blue light activation as per previous protocol (41).

Expressing sucrose transporter in the strain with minimized genome and sucrose measurement

The sucrose over-excreting strain was constructed as reported earlier (10, 27). Briefly, the cscB gene encoding for sucrose permease (source: *E. coli* (ATCC 700927)) was integrated at the neutral site 3 (NS3) of SG55 strain and streaked on Kanamycin plate for complete segregation. The sucrose productivity of SG55-cscB strain was compared with the highest sucrose producing strain of *Synechococcus* 2973 reported previously (10). The sucrose levels were determined following the previous protocol, briefly, cells were grown in 12 well-plate for a period of 3 days along with 1mM IPTG inducer added to the media. Post 3 days cells harvested (OD730nm reached ~1) and centrifuged to obtain the supernatant which was used for sucrose measurement using the sucrose/D-glucose assay kit (Megazyme) (10, 27). The standard curve for sucrose and glucose were performed (27). Experiment was performed in biological (n=3) and technical (n=3) replicates.

Acknowledgement:

The authors are grateful to the funding agency for this study National Science Foundation (NSF) MCB-2037887 to HBP, NSF MCB-2037995 to GMC and NSF MCB-2037829 to CDM

Authors Contribution

HBP, GMC and CDM conceived the study. AS, AB, MGS designed the experiment. AS and DL conducted the experiments. AS and AB analyzed the data. MSG performed the sequencing studies. DS and JIH performed the computational analysis. AS wrote the first draft of the manuscript. All the authors reviewed and revised the manuscript.

Conflict of Interest

The authors declare that they have no known competing financial interests and have reviewed and accepted the manuscript to be published in the journal.

References

1. Crockford PW, Bar On YM, Ward LM, Milo R, Halevy I. 2023. The geologic history of primary productivity. Current Biology 33:4741-4750.e5.
2. Sengupta A, Pakrasi HB, Wangikar PP. 2018. Recent advances in synthetic biology of cyanobacteria. Appl Microbiol Biotechnol 102:5457-5471.
3. Santos-Merino M, Singh AK, Ducat DC. 2019. New Applications of Synthetic Biology Tools for Cyanobacterial Metabolic Engineering. Front Bioeng Biotechnol 7:33.
4. Sun T, Li S, Song X, Diao J, Chen L, Zhang W. 2018. Toolboxes for cyanobacteria: Recent advances and future direction. Biotechnol Adv 36:1293-1307.
5. Berla BM, Saha R, Immethun CM, Maranas CD, Moon TS, Pakrasi HB. 2013. Synthetic biology of cyanobacteria: unique challenges and opportunities. Front Microbiol 4:246.
6. Yu J, Liberton M, Cliften PF, Head RD, Jacobs JM, Smith RD, Koppenaal DW, Brand JJ, Pakrasi HB. 2015. Synechococcus elongatus UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO(2). Sci Rep 5:8132.
7. Jaiswal D, Sengupta A, Sohoni S, Sengupta S, Phadnavis AG, Pakrasi HB, Wangikar PP. 2018. Genome Features and Biochemical Characteristics of a Robust, Fast Growing and Naturally Transformable Cyanobacterium Synechococcus elongatus PCC 11801 Isolated from India. Sci Rep 8:16632.
8. Jaiswal D, Sengupta A, Sengupta S, Madhu S, Pakrasi HB, Wangikar PP. 2020. A Novel Cyanobacterium Synechococcus elongatus PCC 11802 has Distinct Genomic and Metabolomic Characteristics Compared to its Neighbor PCC 11801. Sci Rep 10:191.
9. Sengupta A, Pritam P, Jaiswal D, Bandyopadhyay A, Pakrasi HB, Wangikar PP. 2020. Photosynthetic Co-Production of Succinate and Ethylene in A Fast-Growing Cyanobacterium, Synechococcus elongatus PCC 11801. Metabolites 10.
10. Lin PC, Zhang F, Pakrasi HB. 2020. Enhanced production of sucrose in the fast-growing cyanobacterium Synechococcus elongatus UTEX 2973. Sci Rep 10:390.
11. Mendonca AG, Alves RJ, Pereira-Leal JB. 2011. Loss of genetic redundancy in reductive genome evolution. PLoS Comput Biol 7:e1001082.
12. Stelzer C-P, Pichler M, Stadler P. 2023. Genome streamlining and clonal erosion in nutrient-limited environments: a test using genome-size variable populations. Evolution 77:2378-2391.
13. Giovannoni SJ, Cameron Thrash J, Temperton B. 2014. Implications of streamlining theory for microbial ecology. The ISME Journal 8:1553-1565.
14. Sela I, Wolf YI, Koonin EV. 2016. Theory of prokaryotic genome evolution. Proceedings of the National Academy of Sciences 113:11399-11407.
15. Sung BH, Choe D, Kim SC, Cho BK. 2016. Construction of a minimal genome as a chassis for synthetic biology. Essays Biochem 60:337-346.
16. Hutchison CA, 3rd, Chuang RY, Noskov VN, Assad-Garcia N, Deerinck TJ, Ellisman MH, Gill J, Kannan K, Karas BJ, Ma L, Pelletier JF, Qi ZQ, Richter RA, Strychalski EA, Sun L, Suzuki Y, Tsvetanova B, Wise KS, Smith HO, Glass JI, Merryman C, Gibson DG, Venter JC. 2016. Design and synthesis of a minimal bacterial genome. Science 351:aad6253.
17. Komatsu M, Uchiyama T, Omura S, Cane DE, Ikeda H. 2010. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. Proc Natl Acad Sci U S A 107:2646-51.

18. Reuss DR, Altenbuchner J, Mader U, Rath H, Ischebeck T, Sappa PK, Thurmer A, Guerin C, Nicolas P, Steil L, Zhu B, Feussner I, Klumpp S, Daniel R, Commichau FM, Volker U, Stulke J. 2017. Large-scale reduction of the Bacillus subtilis genome: consequences for the transcriptional network, resource allocation, and metabolism. Genome Res 27:289-299.
19. Park MK, Lee SH, Yang KS, Jung SC, Lee JH, Kim SC. 2014. Enhancing recombinant protein production with an Escherichia coli host strain lacking insertion sequences. Appl Microbiol Biotechnol 98:6701-13.
20. Pósfai G, Plunkett G, 3rd, Fehér T, Frisch D, Keil GM, Umenhoffer K, Kolisnychenko V, Stahl B, Sharma SS, de Arruda M, Burland V, Harcum SW, Blattner FR. 2006. Emergent properties of reduced-genome Escherichia coli. Science 312:1044-6.
21. Lieder S, Nikel PI, de Lorenzo V, Takors R. 2015. Genome reduction boosts heterologous gene expression in Pseudomonas putida. Microb Cell Fact 14:23.
22. Le S, Yao X, Lu S, Tan Y, Rao X, Li M, Jin X, Wang J, Zhao Y, Wu NC, Lux R, He X, Shi W, Hu F. 2014. Chromosomal DNA deletion confers phage resistance to Pseudomonas aeruginosa. Sci Rep 4:4738.
23. Redden H, Alper HS. 2015. The development and characterization of synthetic minimal yeast promoters. Nat Commun 6:7810.
24. Niu TC, Lin GM, Xie LR, Wang ZQ, Xing WY, Zhang JY, Zhang CC. 2019. Expanding the Potential of CRISPR-Cpf1-Based Genome Editing Technology in the Cyanobacterium Anabaena PCC 7120. ACS Synth Biol 8:170-180.
25. Hou F, Ke Z, Xu Y, Wang Y, Zhu G, Gao H, Ji S, Xu X. 2023. Systematic Large Fragment Deletions in the Genome of Synechococcus elongatus and the Consequent Changes in Transcriptomic Profiles. Genes (Basel) 14.
26. Csorgo B, Leon LM, Chau-Ly IJ, Vasquez-Rifo A, Berry JD, Mahendra C, Crawford ED, Lewis JD, Bondy-Denomy J. 2020. A compact Cascade-Cas3 system for targeted genome engineering. Nat Methods 17:1183-1190.
27. Sengupta A, Bandyopadhyay A, Schubert MG, Church GM, Pakrasi HB. 2023. Antenna Modification in a Fast-Growing Cyanobacterium Synechococcus elongatus UTEX 2973 Leads to Improved Efficiency and Carbon-Neutral Productivity. Microbiology Spectrum 11:e00500-23.
28. Wang L, Maranas CD. 2018. MinGenome: An In Silico Top-Down Approach for the Synthesis of Minimized Genomes. ACS Synth Biol 7:462-473.
29. Rubin BE, Wetmore KM, Price MN, Diamond S, Shultzaberger RK, Lowe LC, Curtin G, Arkin AP, Deutschbauer A, Golden SS. 2015. The essential gene set of a photosynthetic organism. Proc Natl Acad Sci U S A 112:E6634-43.
30. Ungerer J, Pakrasi HB. 2016. Cpf1 Is A Versatile Tool for CRISPR Genome Editing Across Diverse Species of Cyanobacteria. Sci Rep 6:39681.
31. Taton A, Gilderman TS, Ernst DC, Omaga CA, Cohen LA, Rey-Bedon C, Golden JW, Golden SS. 2023. Synechococcus elongatus Argonaute reduces natural transformation efficiency and provides immunity against exogenous plasmids. mBio 14:e0184323.
32. Santos-Merino M, Yun L, Ducat DC. 2023. Cyanobacteria as cell factories for the photosynthetic production of sucrose. Frontiers in Microbiology 14.
33. Kim K, Choe D, Lee DH, Cho BK. 2020. Engineering Biology to Construct Microbial Chassis for the Production of Difficult-to-Express Proteins. Int J Mol Sci 21.

34. Richardson SM, Mitchell LA, Stracquadanio G, Yang K, Dymond JS, DiCarlo JE, Lee D, Huang CLV, Chandrasegaran S, Cai Y, Boeke JD, Bader JS. 2017. Design of a synthetic yeast genome. Science 355:1040-1044.
35. Ungerer J, Lin P-C, Chen H-Y, Pakrasi HB. 2018. Adjustments to Photosystem Stoichiometry and Electron Transfer Proteins Are Key to the Remarkably Fast Growth of the Cyanobacterium Synechococcus elongatus UTEX 2973. mBio 9.
36. Pfeifer E, Michniewski S, Gätgens C, Münch E, Müller F, Polen T, Millard A, Blombach B, Frunzke J. 2019. Generation of a Prophage-Free Variant of the Fast-Growing Bacterium Vibrio natriegens. Applied and Environmental Microbiology 85:e00853-19.
37. Hendry JI, Gopalakrishnan S, Ungerer J, Pakrasi HB, Tang YJ, Maranas CD. 2019. Genome-Scale Fluxome of Synechococcus elongatus UTEX 2973 Using Transient (13)C-Labeling Data. Plant Physiol 179:761-769.
38. Liu D, Johnson VM, Pakrasi HB. 2020. A Reversibly Induced CRISPRi System Targeting Photosystem II in the Cyanobacterium Synechocystis sp. PCC 6803. ACS Synth Biol 9:1441-1449.
39. Arnon DI, McSwain BD, Tsujimoto HY, Wada K. 1974. Photochemical activity and components of membrane preparations from blue-green algae. I. Coexistence of two photosystems in relation to chlorophyll a and removal of phycocyanin. Biochimica et Biophysica Acta (BBA) - Bioenergetics 357:231-245.
40. Murchie EH, Lawson T. 2013. Chlorophyll fluorescence analysis: a guide to good practice and understanding some new applications. J Exp Bot 64:3983-98.
41. Walker PL, Pakrasi HB. 2022. A Ubiquitously Conserved Cyanobacterial Protein Phosphatase Essential for High Light Tolerance in a Fast-Growing Cyanobacterium. Microbiology Spectrum 10:e01008-22.

Figures and Tables

Table 1: A list of large gene clusters that are predicted to be unessential for growth of *Synechococcus* 2973. The predicted dispensable region in between the start site of the first gene of the cluster and the start site of the end gene of the cluster.

Rank	Start	End	Length (bp)
1	M744_12940	M744_13140	33,952
2	M744_10800	M744_10895	26,399
3	M744_02780	M744_02900	25,699
4	M744_12500	M744_12615	22,544
5	M744_05410	M744_05555	22,235

Figure.1.

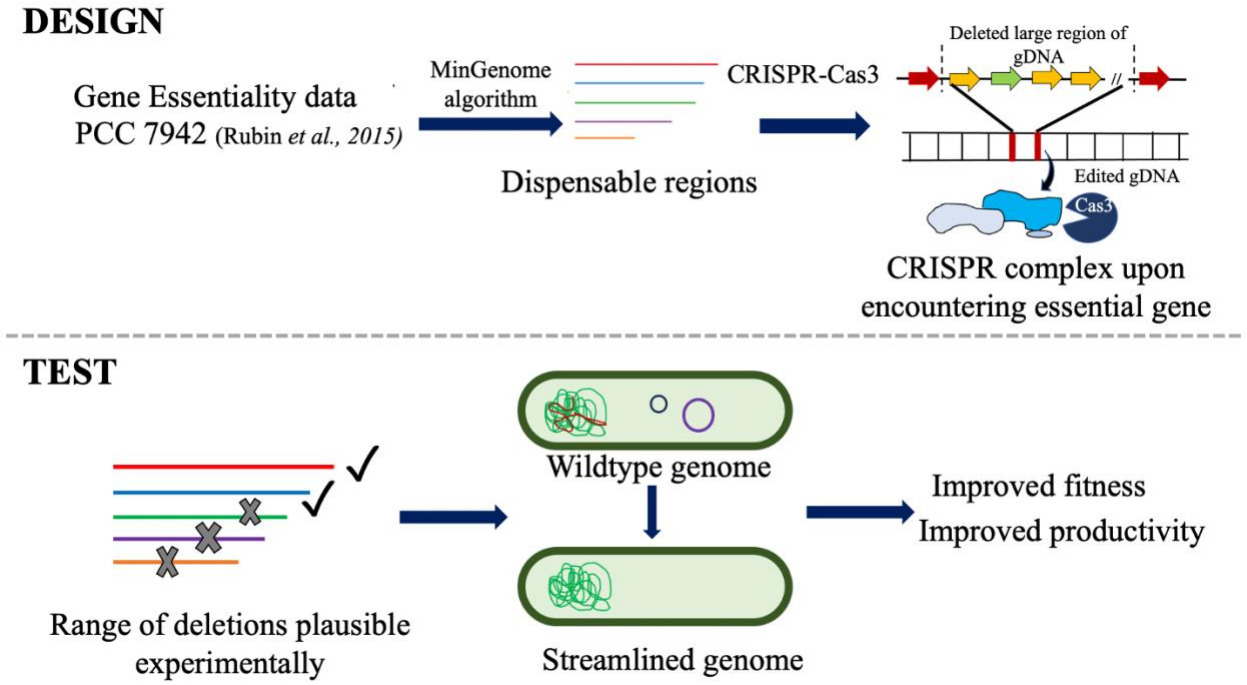
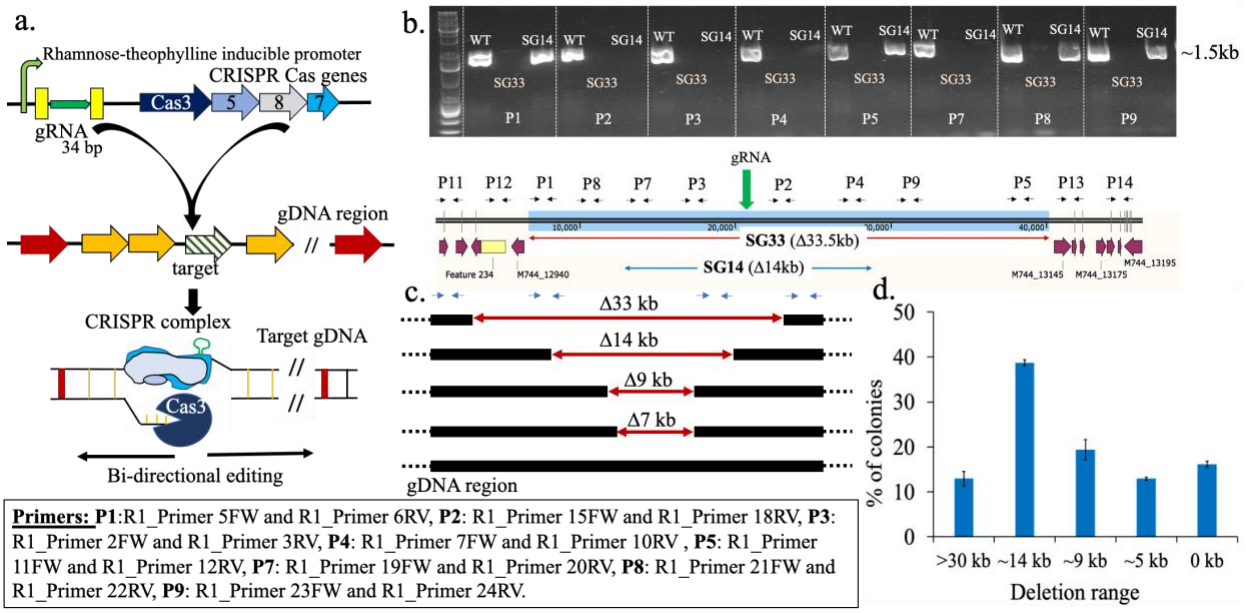


Fig. 1: The schematic showing the strategy used for systematic genome streamlining in cyanobacteria. A combinatorial approach integrating computational (design) and experimental (test) tools to first identify the dispensable regions in *Synechococcus elongatus* UTEX 2973 and further use novel CRISPR tool to validate the prediction and create a strain with improved fitness and productivity.

Figure.2.



Primers: **P1**:R1_Primer 5FW and R1_Primer 6RV, **P2**: R1_Primer 15FW and R1_Primer 18RV, **P3**: R1_Primer 2FW and R1_Primer 3RV, **P4**: R1_Primer 7FW and R1_Primer 10RV , **P5**: R1_Primer 11FW and R1_Primer 12RV, **P7**: R1_Primer 19FW and R1_Primer 20RV, **P8**: R1_Primer 21FW and R1_Primer 22RV, **P9**: R1_Primer 23FW and R1_Primer 24RV.

Fig. 2: Developing CRISPR-Cas3 system for spontaneous large deletions in cyanobacteria. (a) The schematic showing the functioning of the CRISPR-Cas3 system. (b) The result from Tiling PCR (gel picture) using different primer sets (P1-9) showing the absence or presence of 1.5 kb band for two colonies (SG33 and SG20). The absence of band indicate the region is deleted from the strain as compared to WT. (c) The length of deletions obtained when CRISPR system was targeted at the R1 region. (d) Frequency of occurrence of colonies of a particular length of deletion. The error bars represent the variation observed between the difference in the number of colonies obtained for a particular length of deletions between three separate experiments.

Figure.3.

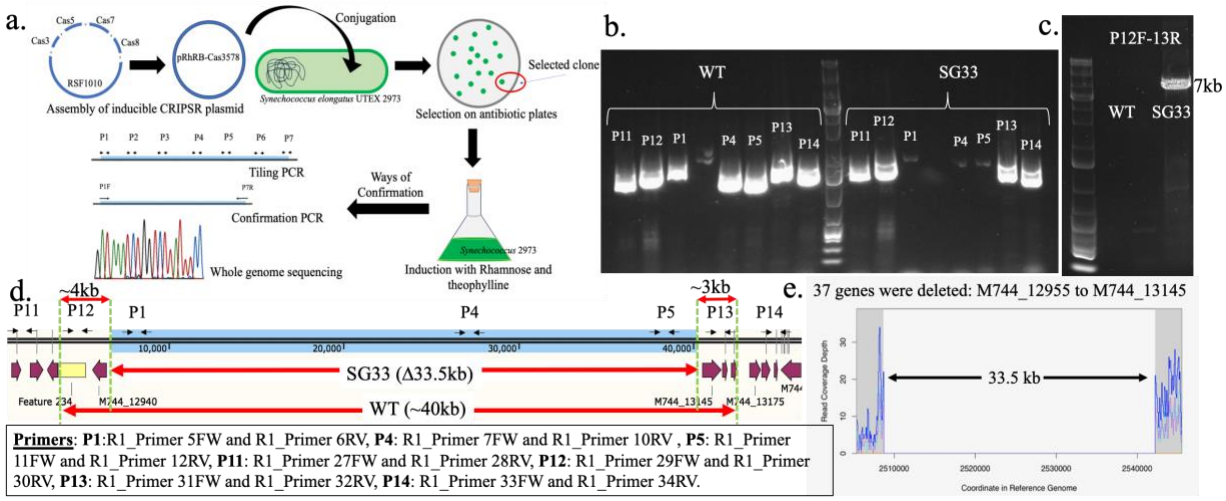
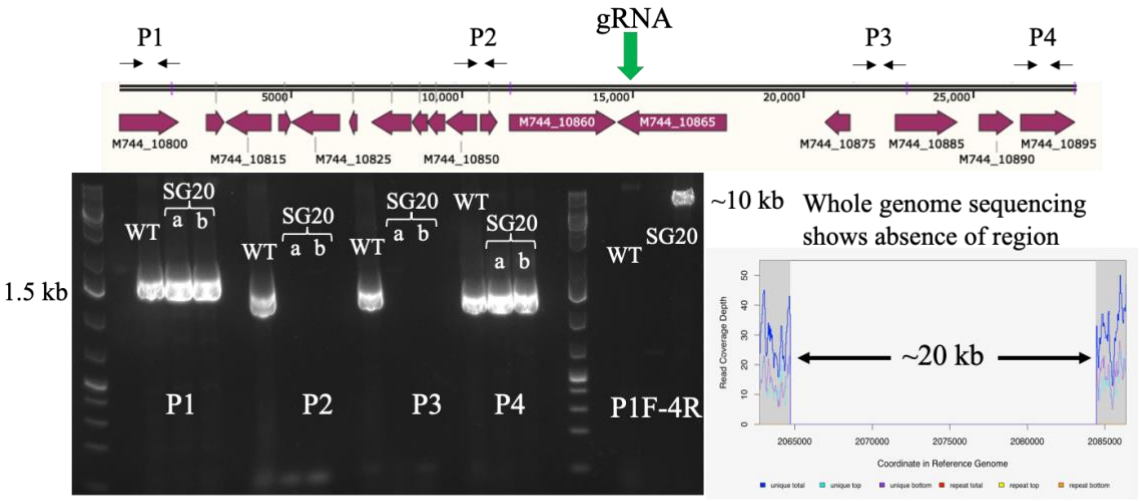


Fig. 3: Creating the SG33 strain using CRISPR-Cas3 system. (a) The workflow for obtaining the markerless strain with minimized genome. (b) The gel picture showing the absence or presence of 15 kb band when amplified with different primer sets in a tiling PCR to identify the deleted regions. (c) The Confirmation PCR showing the presence of 7kb band in SG33 strain as opposed to the 40 kb band. No band in WT as the 40 kb band is difficult to obtain. (d) The schematic showing the R1 region and the range of deletions and the region of primers binding. (e) The whole genome sequencing of the SG33 strain showed a clean and segregated deletion.

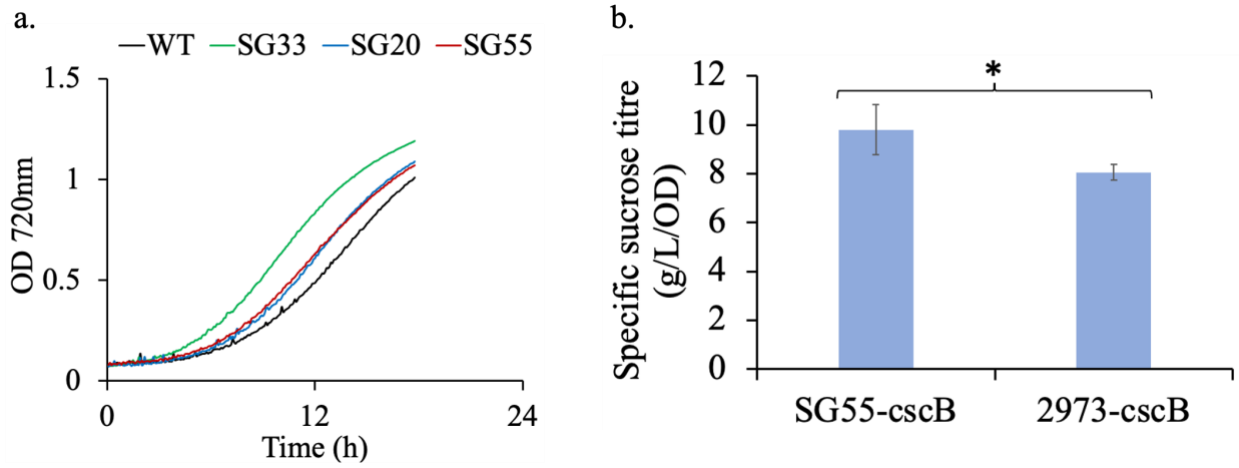
Figure.4.



Primers: **P1**:R2_Primer 7FW and R2_Primer 8RV, **P2**: R2_Primer 5FW and R2_Primer 6RV , **P3**: R2_Primer 3FW and R2_Primer 4RV, **P4**: R2_Primer 1FW and R2_Primer 2RV.

Fig. 4: Generation of the SG20 strain. The schematic showing the R2 region and the primer binding sites. The gel picture is of the tiling PCR performed using the different primer sets (P1-P4) and the confirmation gel using the primers P1Forward and P4 Reverse. A 10 kb band is visible instead of the 30kb band in SG20 as compared to the WT. No band is visible as 30kb is difficult to amplify and visualize. A snap of the whole genome sequencing chromatogram shows the absence of reads around the R2 region. Two different colonies with deletion were tested (a, b) as shown in the gel picture.

573 **Figure.5.**



574

575 Fig. 5: Characterizing the engineered strains with minimized genome. (a) Growth comparison of
576 WT, SG33, SG20 and SG55 under high light and high CO_2. Experiments were performed in
577 triplicates (n=3) and the graphs are representative growth profile. (b) Sucrose productivity of SG55
578 as compared to WT. Experiments were performed in triplicates (n=3) and the error bars correspond
579 to standard deviations from at least 3 biological replicates. Asterisk (*) denote statistically different
580 values of μ ($p < 0.05$).
581

Supplementary Information

Genome streamlining to improve performance of a fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973

Running Title: Genome streamlining in a photoautotroph

Annesha Sengupta[a*], Anindita Bandyopadhyay[a], Debolina Sarkar[b$], John I. Hendry[b^], Max G. Schubert[c], Deng Liu[a], George M. Church[c,d], Costas D. Maranas[b], Himadri B. Pakrasi[a#]

[a]*Department of Biology, Washington University, St. Louis, MO, USA*

[b]*Department of Chemical Engineering, Pennsylvania State University, PA, USA*

[c]*Wyss Institute for Biologically Inspired Engineering, Harvard University, MA, USA*

[d]*Department of Genetics, Harvard Medical School, MA, USA*

[#]Corresponding author: Himadri B. Pakrasi, pakrasi@wustl.edu.

Current addresses:

[*]*Department of Chemical Engineering, University of Toronto, Canada*

[$]*International Flavors and Fragrances, 925 Page Mill Road, Palo Alto, CA 94304*

^*Henry Jackson Foundation, 6720A Rock ledge Drive, Bethesda, Maryland 20817*

Fig.S1: Colony screening of SG33 strain by Tiling PCR using primers P1, P3, P4 (Primers: P1:R1_Primer 5FW and R1_Primer 6RV, P3: R1_Primer 2FW and R1_Primer 3RV, P4: R1_Primer 7FW and R1_Primer 10RV)

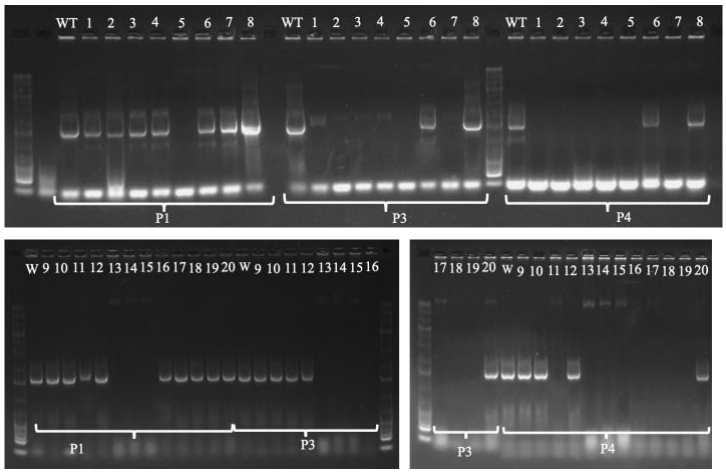


Fig.S2: Colony screening to identify positive clones for Region 3,4,5 by Tiling PCR

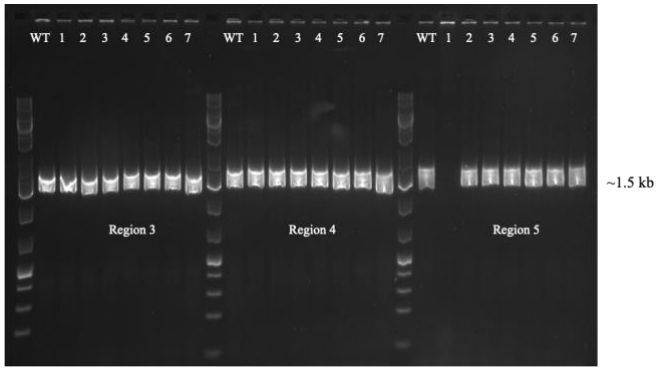


Fig.S3: Confirmation PCR for SG33 and SG20

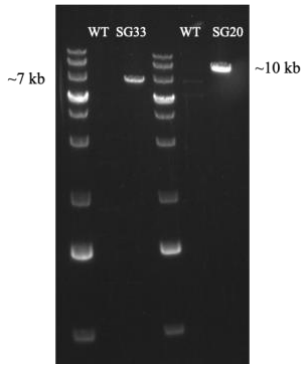
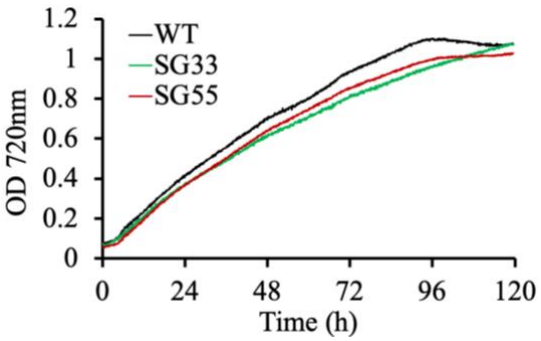
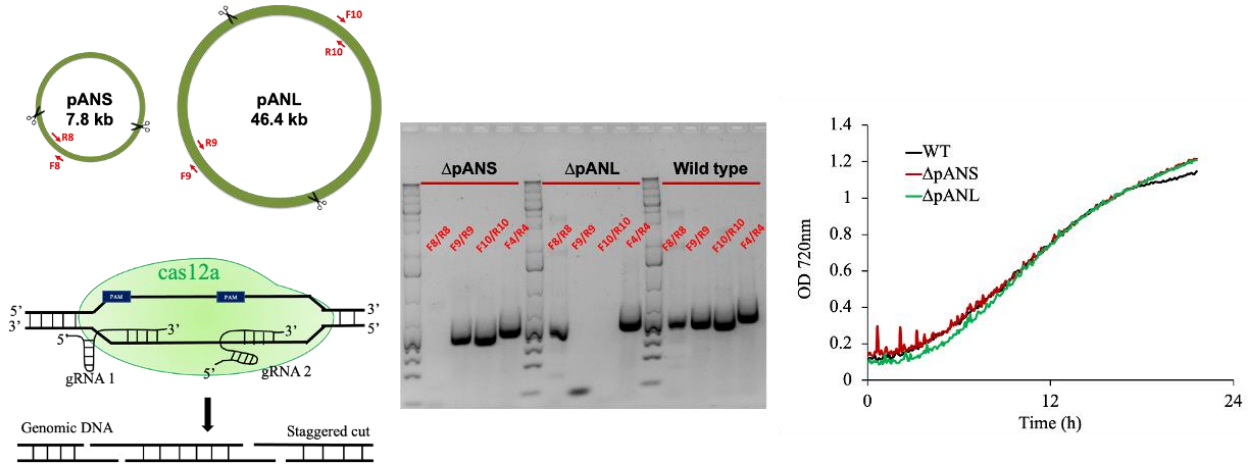


Fig.S4: Growth comparison for WT, SG33 and SG20 under high light and ambient CO_2



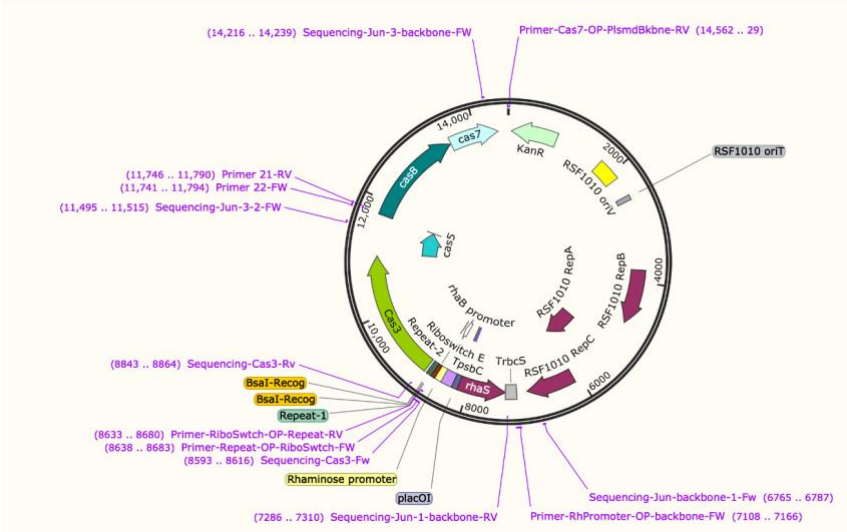
620

621 **Fig. S5:** Deletion of the large (pANL) and small (pANS) plasmids: CRISPR-Cas12a was
622 employed for the task and the gel picture shows the absence of band indicating the small and large
623 plasmids were deleted in the strains ΔpANS and ΔpANL respectively as compared to WT. Growth
624 profile of WT and plasmid deleted strains (ΔpANS and ΔpANL) under high light and high CO_2.
625



626

627 **Fig. S6:** The assembled plasmid map of pSL3578 showing the primer binding sites.



628

Table S1: The gene id and annotations of genes in the R1 region

Gene	Function from Cyanobase
M744_12940	hypothetical protein
M744_12945	lysozyme
M744_12955	hypothetical protein
M744_12960	hypothetical protein
M744_12965	hypothetical protein
M744_12970	late control protein D
M744_12975	tail protein
M744_12980	hypothetical protein
M744_12985	hypothetical protein
M744_12995	hypothetical protein
M744_13000	major tail tube protein
M744_13005	hypothetical protein
M744_13010	hypothetical protein
M744_13015	hypothetical protein
M744_13020	hypothetical protein
M744_13025	hypothetical protein
M744_13030	hypothetical protein
M744_13035	phage tail protein
M744_13040	hypothetical protein
M744_13045	baseplate assembly protein W
M744_13050	hypothetical protein
M744_13055	baseplate assembly protein V
M744_13060	hypothetical protein
M744_13065	hypothetical protein
M744_13070	electron transfer flavoprotein
M744_13075	hypothetical protein
M744_13080	hypothetical protein
M744_13085	phage portal protein
M744_13090	hypothetical protein
M744_13095	hypothetical protein
M744_13100	hypothetical protein
M744_13105	hypothetical protein
M744_13110	hypothetical protein
M744_13115	hypothetical protein
M744_13120	hypothetical protein
M744_13125	helicase
M744_13135	hypothetical protein

633 **Table S2:** The gene id and annotations of genes in the R2 region

Genes	Function from Cyanobase
M744_10800	peptide ABC transporter substrate-binding protein
M744_10805	1-(5-phosphoribosyl)-5-amino-4-imidazole- carboxylate carboxylase
M744_10810	chorismate-binding protein
M744_10815	hypothetical protein
M744_10820	phosphorylase
M744_10825	hypothetical protein
M744_10830	hypothetical protein
M744_10835	hypothetical protein
M744_10840	hypothetical protein
M744_10845	hypothetical protein
M744_10850	hypothetical protein
M744_10855	transcriptional regulator
M744_10860	diguanylate cyclase
M744_10865	hypothetical protein
M744_10870	hypothetical protein
M744_10875	hypothetical protein
M744_10880	hypothetical protein
M744_10885	hypothetical protein
M744_10890	lipoprotein

634

635

636 **Table S3:** The gene id and annotations of genes in the R3 region

Genes	Function from Cyanobase
M744_02780	protein phosphatase
M744_02790	glucose-1-phosphate cytidylyltransferase
M744_02795	CDP-glucose 4,6-dehydratase
M744_02800	hypothetical protein
M744_02805	hypothetical protein
M744_02810	dTDP-4-dehydrorhamnose 3,5-epimerase
M744_02815	SAM-dependent methyltransferase
M744_02820	glutamine--scyllo-inositol aminotransferase
M744_02825	hypothetical protein
M744_02830	cephalosporin hydroxylase
M744_02835	hypothetical protein
M744_02840	methyltransferase type 11
M744_02845	hypothetical protein
M744_02850	hypothetical protein
M744_02855	pili assembly chaperone
M744_02860	pili assembly chaperone

M744_02865	hypothetical protein
M744_02870	ubiquinone biosynthesis protein
M744_02875	hypothetical protein
M744_02880	FAD-binding protein
M744_02885	hypothetical protein
M744_02890	50S ribosomal protein L28
M744_02895	transporter
M744_02900	dihydroneopterin aldolase

637

638

639 **Table S4:** The gene id and annotations of genes in the R4 region

Genes	Function from Cyanobase
M744_12500	excinuclease ABC subunit A
M744_12505	twitching motility protein
M744_12510	hypothetical protein
M744_12515	phosphotransferase
M744_12520	hypothetical protein
M744_12525	glutathione reductase
M744_12530	FAD-dependent oxidoreductase
M744_12535	hypothetical protein
M744_12540	aliphatic nitrilase
M744_12545	radical SAM protein
M744_12550	GNAT family acetyltransferase
M744_12555	selenophosphate synthetase
M744_12565	ATP-dependent DNA helicase RuvB
M744_12570	hypothetical protein
M744_12580	hypothetical protein
M744_12585	hypothetical protein
M744_12590	transposase
M744_12595	phosphohydrolase
M744_12600	asparaginyl-tRNA synthetase
M744_12605	hypothetical protein
M744_12610	hypothetical protein
M744_12615	cobalamin biosynthesis protein CobW

640

641

642 **Table S5:** The gene id and annotations of genes in the R5 region

Genes	Function from Cyanobase
M744_05410	hypothetical protein

M744_05415	sodium:proton antiporter
M744_05420	hypothetical protein
M744_05425	hypothetical protein
M744_05430	ribonuclease H
M744_05435	hypothetical protein
M744_05440	Zn-dependent protease
M744_05445	alkyl hydroperoxide reductase
M744_05450	lipid kinase
M744_05455	membrane protein
M744_05460	branched-chain amino acid ABC transporter permease
M744_05465	hypothetical protein
M744_05470	iron deficiency-induced protein A
M744_05475	Crp/Fnr family transcriptional regulator
M744_05480	hypothetical protein
M744_05485	hypothetical protein
M744_05495	DNA polymerase III subunit beta
M744_05500	Fur family transcriptional regulator
M744_05505	hypothetical protein
M744_05510	ATPase AAA
M744_05515	hypothetical protein
M744_05520	hypothetical protein
M744_05525	hypothetical protein
M744_05530	cell death suppressor protein Lls1
M744_05535	hypothetical protein
M744_05540	hypothetical protein
M744_05550	hypothetical protein
M744_05555	alanine--glyoxylate aminotransferase

Table S6: Doubling time of the strains used in this study under the tested conditions 1500 µmoles.m^{-2}.s^{-1}, 1% CO_2 and 38˚C.

Strain	**Doubling time (h)**
Synechococcus 2973	2.4±0.07
SG33	1.9±0.1
SG20	2.3±0.2
SG55	2.2±0.12

649

650 **Table S7:** List of primers used in this study

Primer Name	Primer Sequence
Construction of plasmid	
Primer-Repeat-OP-RiboSwtch-FW	ctgctaaggaggtaacaacaagATGAAATTCAACTAGGTCGCGCCC
Primer21-RV	tgataatagtcattgagggcCGAAAGGATCATGCCTTGTCCTCTG
Primer22-FW	ggcggcagaggacaaggcatgatcctttCGGCCCTCAATGACTATTATCAGCGA
Primer-Cas7-OP-PlsmdBkbne-RV	cagagcattacgctgacttgacgggacacGGCTGACGCCGTTGGATA
Primer-RhPromoter-OP-backbone-FW	ataggccgctttcctggctttgcttcccacTAATTGACAATTGACAATTCCCCACTTAG
Primer-RiboSwtch-OP-Repeat-RV	cgcgacctagttgaatttcatCTTGTTGTTACCTCCTTAGCAGGGTGC
Sequencing-Cas3-Fw	gaataccggtgataccagcatcgt
Sequencing-Cas3-Rv	ctgcgatctgcctttaaggtgg
Guide RNA	
R1-gRNA1-FW	GAAACtatcccgtttacacgatcgtcaacggtggcctgaG
R1-gRNA1-RV	GCGACtcaggccaccgttgacgatcgtgtaaacgggataG
R2-gRNA1-FW	GAAACcgtcgcaaagcagctagctaactgcagtgcctgcG
R2-gRNA1-RV	GCGACgcaggcactgcagttagctagctgctttgcgacgG
R3-gRNA1-FW	GAAACaacgtcccccaatagctttgctggatcagctcagG
R3-gRNA1-RV	GCGACctgagctgatccagcaaagctattgggggacgttG
R4-gRNA1-FW	GAAACagcagacgcgctgtgcgctggacaatatccgtcaG
R4-gRNA1-RV	GCGACtgacggatattgtccagcgcacagcgcgtctgctG
R5-gRNA1-FW	GAAACaaatcctttgccaaacagcagattgaaatgcagaG
R5-gRNA1-RV	GCGACtctgcatttcaatctgctgtttggcaaaggatttG
R5-gRNA2-FW	GAAACacaaattttgattgggacaagactgtggactcG
R5-gRNA2-RV	GCGACgagtccacagtcttgtcccaatcaaaatttgtG
R2-gRNA2-FW	GAAACaaataatgagcactcactcatctgccgcgaccggG
R2-gRNA2-RV	GCGACccggtcgcggcagatgagtgagtgctcattatttG
R3-gRNA2-FW	GAAACcaccaagagcaagctctatcaattgcaatcgcggG
R3-gRNA2-RV	GCGACccgcgattgcaattgatagagcttgctcttggtgG
R4-gRNA2-FW	GAAACcacaccactcacggggtcgaatcttaatgctgccG
R4-gRNA2-RV	GCGACggcagcattaagattcgaccccgtgagtggtgtgG
Tiling and Confirmation PCR	
R1_TilePrimer 1FW	catagcttaggggggtgtcagg
R1_TilePrimer 2RV	gctgctgttcgatggcatcaa
R1_TilePrimer 3FW	gttgccatgcggttagatcgtc
R1_TilePrimer 4RV	gacgatctaaccgcatggcaac
R1_TilePrimer 5FW	cttgattactgcccatcatgcgg
R1_TilePrimer 6RV	ggccagcatcatcgaaatcaagg
R1_TilePrimer 7FW	gatcagcatctgccatcgcca
R1_TilePrimer 8RV	tggcgatggcagatgctgatc
R1_TilePrimer 9FW	tggactactcactcagccagcc
R1_TilePrimer 10RV	ggctggctgagtgagtagtcca
R1_TilePrimer 11FW	acccgaagcaatcagtccatg
R1_TilePrimer 12RV	Cttcttgtctggaagcagcagc

R1_TilePrimer 13FW	cgatccggtcgattgtgtccg
R1_TilePrimer 14RV	cggacacaatcgaccggatcg
R1_TilePrimer 15FW	cttcagctccttgccactgtca
R1_TilePrimer 16RV	tgacagtggcaaggagctgaag
R1_TilePrimer 17FW	gatcgtcaggtgctcgagctcg
R1_TilePrimer 18RV	gagctcgagcacctgacgatc
R1_TilePrimer 19FW	gtccagaagttttctcgccgg
R1_TilePrimer 20RV	ctcgtccgtgtcgaggaagg
R1_TilePrimer 21FW	gaaactgataggcgacctcgcg
R1_TilePrimer 22RV	cgctactcagtgaggccaagc
R1_TilePrimer 23FW	cagctctcgctgcagacacag
R1_TilePrimer 24RV	agcatcagtactcgcagcgtc
R1_TilePrimer 25FW	cgacctcagtggttgcgaag
R1_TilePrimer 26RV	gactctgccctggcagct
R1_TilePrimer 27FW	ggcgatcgcaactggcttag
R1_TilePrimer 28RV	ccatgacccgccttgcattaag
R1_TilePrimer 29FW	ctcccactctagcctcctaaatccg
R1_TilePrimer 30RV	acctgaacctgaagcagctaagc
R1_TilePrimer 31FW	cattcacaccgagcccgtcg
R1_TilePrimer 32RV	ccattcttcctctgcagcccaac
R1_TilePrimer 33FW	atctatgacgcaccaccgacg
R1_TilePrimer 34RV	ttggaagcgatcgatcggctttg
R2_TilePrimer 1FW	gcccaagtacgcctagacgttc
R2_TilePrimer 2RV	catccaagactcaaaggactcggtg
R2_TilePrimer 3FW	gactcggctgcttctgaaccg
R2_TilePrimer 4RV	ccctgaactagctccagacagcg
R2_TilePrimer 5FW	cctgctcaatcgctgcgaca
R2_TilePrimer 6RV	ctggcgtgaagattgttccatcg
R2_TilePrimer 7FW	gtggcgctcttggttcgtga
R2_TilePrimer 8RV	cagcttcccaaggggcaac
R3_TilePrimer 1FW	gatgctgaatcgccgtcacc
R3_TilePrimer 2RV	gcgtgctctcggctttcattc
R3_TilePrimer 3FW	gcgatccgtaatacggccagt
R3_TilePrimer 4RV	cgaatgaccgctaggctgc
R3_TilePrimer 5FW	ccgatacagcaactcagtttatggc
R3_TilePrimer 6RV	cggtccagatacgtattgctcagc
R3_TilePrimer 7FW	gtgattccgggacgactggc
R3_TilePrimer 8RV	cctcctgtctgactcatctcgc
R4_TilePrimer 1FW	gccctagcttgtcgagatcgag
R4_TilePrimer 2RV	gcgcagggcgctttttc
R4_TilePrimer 3FW	ggcaatcatccaatcgatgctgg
R4_TilePrimer 4RV	ggcgcttggctcaagaacg
R4_TilePrimer 5FW	gggcgatcgcacctcaag
R4_TilePrimer 6RV	tgctcgagggtcagctcc
R4_TilePrimer 7FW	atgcccaagcgcgcttc
R4_TilePrimer 8RV	caggcgcgcttgaatttccc

R5_TilePrimer 1FW	gcgagtcattatgtacaaccgcg
R5_TilePrimer 2RV	ctggctagcattgacgcgat
R5_TilePrimer 3FW	caaatttatgcagacctcagtgccg
R5_TilePrimer 4RV	gccgagggtcaaacccgttt
R5_TilePrimer 5FW	ctaccctgaatccagcagtgagg
R5_TilePrimer 6RV	ccagccaagtcaccagtaaatcc
R5_TilePrimer 7FW	gccaaaagcttggatttgcacg
R5_TilePrimer 8RV	cagcagtacctacagtcaaagcc

651
652
653
654