

OPEN

# A Novel Cyanobacterium *Synechococcus elongatus* PCC 11802 has Distinct Genomic and Metabolomic Characteristics Compared to its Neighbor PCC 11801

Damini Jaiswal<sup>1</sup>, Annesha Sengupta<sup>1</sup>, Shinjinee Sengupta<sup>1,2</sup>, Swati Madhu<sup>1</sup>, Himadri B. Pakrasi<sup>3,4</sup> & Pramod P. Wangikar<sup>1,2,5\*</sup>

Cyanobacteria, a group of photosynthetic prokaryotes, are attractive hosts for biotechnological applications. It is envisaged that future biorefineries will deploy engineered cyanobacteria for the conversion of carbon dioxide to useful chemicals via light-driven, endergonic reactions. Fast-growing, genetically amenable, and stress-tolerant cyanobacteria are desirable as chassis for such applications. The recently reported strains such as *Synechococcus elongatus* UTEX 2973 and PCC 11801 hold promise, but additional strains may be needed for the ongoing efforts of metabolic engineering. Here, we report a novel, fast-growing, and naturally transformable cyanobacterium, *S. elongatus* PCC 11802, that shares 97% genome identity with its closest neighbor *S. elongatus* PCC 11801. The new isolate has a doubling time of 2.8 h at 1% CO<sub>2</sub>, 1000 μmole photons.m<sup>-2</sup>.s<sup>-1</sup> and grows faster under high CO<sub>2</sub> and temperature compared to PCC 11801 thus making it an attractive host for outdoor cultivations and eventual applications in the biorefinery. Furthermore, *S. elongatus* PCC 11802 shows higher levels of key intermediate metabolites suggesting that this strain might be better suited for achieving high metabolic flux in engineered pathways. Importantly, metabolite profiles suggest that the key enzymes of the Calvin cycle are not repressed under elevated CO<sub>2</sub> in the new isolate, unlike its closest neighbor.

Cyanobacteria are a group of prokaryotes that are capable of carrying out oxygenic photosynthesis. These microorganisms have gained attention in the field of biotechnology due to their efficient photoautotrophy, genetic amenability, and the potential for direct conversion of carbon dioxide (CO<sub>2</sub>) to useful products. Engineered cyanobacteria have been reported for the production of a wide variety of chemicals, albeit at titers that are well below those needed for commercial production<sup>1–9</sup>. Majority of these studies have employed model strains such as *Synechocystis* sp. PCC 6803, *Synechococcus elongatus* PCC 7942, and *Synechococcus* sp. PCC 7002 (henceforth referred to as PCC 6803, PCC 7942, and PCC 7002, respectively, for brevity). Substantial information is now available on transcriptome, proteome, metabolome, and synthetic biology tools of these model cyanobacteria<sup>10–15</sup>. Thus, these strains and other closely related cyanobacteria would be favorable choices as hosts for chemical production. Further, it is believed that fast-growing strains that afford more efficient photoautotrophy will be better suited as hosts for metabolic engineering<sup>16–18</sup>. Moreover, organisms that are tolerant to high temperatures, light, and CO<sub>2</sub> levels are desirable as industrial strains considering the eventual outdoor utilization in biorefineries.

<sup>1</sup>Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, 400076, India. <sup>2</sup>DBT-PAN IIT Centre for Bioenergy, Indian Institute of Technology Bombay, Powai, Mumbai, 400076, India. <sup>3</sup>Department of Biology, Washington University, St. Louis, MO, 63130, USA. <sup>4</sup>Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, 63130, USA. <sup>5</sup>Wadhvani Research Centre for Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, 400076, India. \*email: [wangikar@iitb.ac.in](mailto:wangikar@iitb.ac.in)

Thus, it would be of interest to expand the repertoire of cyanobacterial strains by adding strains that fulfill these criteria. Phylogenetic proximity to the widely studied model strains would be an added advantage.

Recently, fast-growing cyanobacterial strains *Synechococcus elongatus* UTEX 2973 and *S. elongatus* PCC 11801 (henceforth UTEX 2973 and PCC 11801, respectively) have been reported<sup>19,20</sup>. Their growth rates are higher than model cyanobacterial strains like PCC 6803, PCC 7942, and PCC 7002<sup>19,20</sup>. Detailed growth, biochemical characterization, and systems biology data is now available for UTEX 2973 that was reported in 2015<sup>18,21,22</sup>. Likewise, the physiological and proteome changes in PCC 11801 under elevated CO<sub>2</sub> conditions have been studied<sup>14,19</sup>. However, more comprehensive transcriptomic, proteomics, and metabolomics research is needed to characterize the strain PCC 11801 in detail. Furthermore, it would be of interest to develop a library of well-characterized synthetic parts like promoters, ribosomal binding sites (RBS), and terminators that can be used in the new hosts according to the requirement for a given product and in a particular environment<sup>10,23,24</sup>. For example, a gene for toxic metabolic intermediate can be expressed under a weak or tuneable promoter so that the metabolite does not accumulate inside the cell. Unlike the heterotrophic counterparts like *E. coli* and yeast, there are only a limited number of fast-growing cyanobacterial strains such as UTEX 2973, PCC 11801, and PCC 7002. Therefore, it is envisaged that some new cyanobacteria need to be isolated and characterized for metabolic engineering applications, and the present work is a step in that direction<sup>25</sup>.

Studies involving metabolomics and metabolic flux analysis are helpful in identifying the bottlenecks in the metabolic pathways that can be potential targets for strain engineering<sup>9,12,26,27</sup>. For example, a comparison of relative metabolic pool sizes across different conditions may provide valuable biological insights on the potential rate-limiting steps in the metabolic network under a particular condition<sup>28–30</sup>. Depending on the inherent metabolism of a cyanobacterial strain, the yield of desired product may differ among different strains. However, there are no reports of a head-on comparison of yields for a particular product by expressing a given pathway across different strains<sup>11</sup>. Likewise, there are limited reports on comparative metabolite profiling of different cyanobacterial strains under conditions that are expected to result in higher productivity<sup>29,31</sup>.

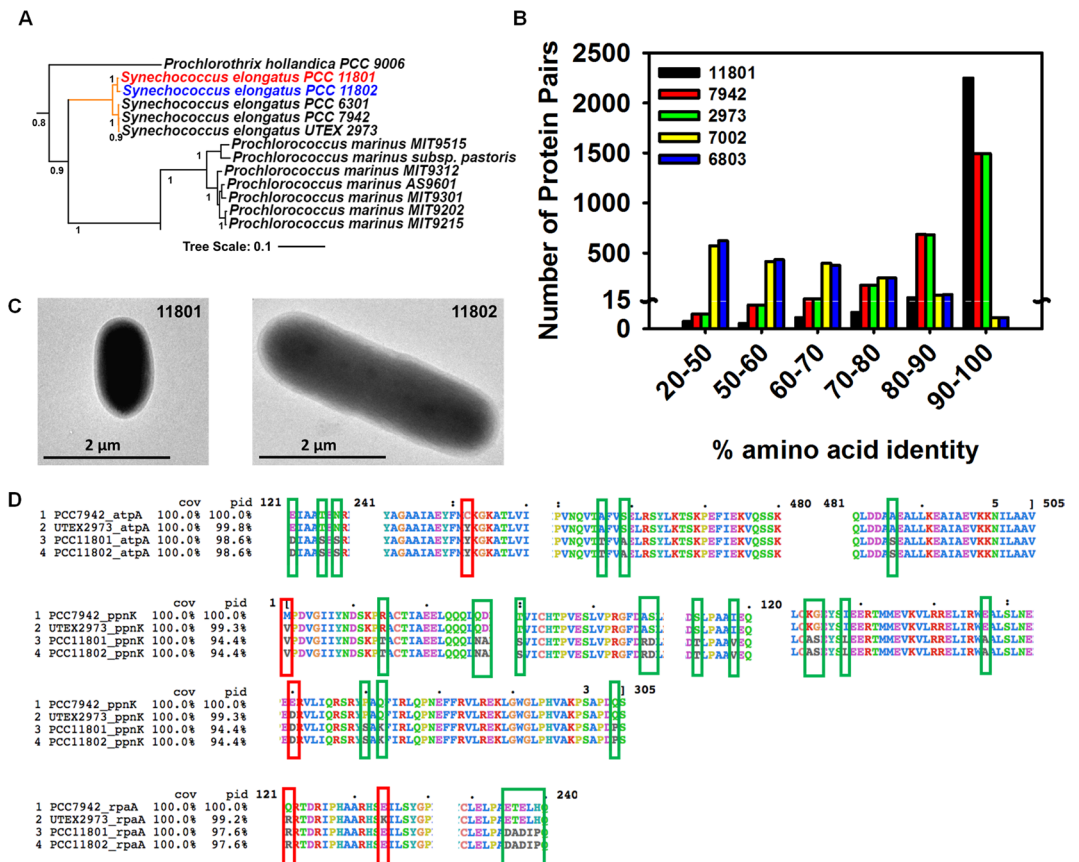
In this study, we report *Synechococcus elongatus* PCC 11802 (henceforth referred to as PCC 11802), which is a close relative of PCC 11801, also isolated from Powai Lake, Mumbai, India (19.1273°N, 72.9048°E). Detailed physiological, genomic, and metabolic characterization of PCC 11802 is presented here. Unlike its closest neighbor PCC 11801 that has its least doubling time of 2.3 h under ambient CO<sub>2</sub> conditions<sup>19</sup>, PCC 11802 has a least doubling time of 2.8 h at 38 °C, 1% CO<sub>2</sub> and a light intensity of 1000 μmole photons.m<sup>-2</sup>.s<sup>-1</sup>. The two strains thus have different CO<sub>2</sub> specificities when it comes to their optimal growth conditions. PCC 11802 is also naturally transformable similar to PCC 11801. This study reports detailed physiological and genomic characterization of PCC 11802 with a focus on differences between PCC 11802 and PCC 11801. Metabolite profiling of PCC 11802 and PCC 11801 under elevated CO<sub>2</sub> conditions uncovers the differences in their metabolic capabilities for carbon assimilation. This study adds one more cyanobacterial strain in the phylogenetic neighborhood of the widely studied PCC 7942 and a potential candidate for metabolic engineering.

## Results and Discussion

**Identification and genome analysis of PCC 11802.** PCC 11802 was isolated along with PCC 11801 and six other strains from Powai Lake, Mumbai, India (19.1273°N, 72.9048°E). All the eight strains showed growth rates higher than the closest model strain PCC 7942. The detailed characterization of PCC 11801 has been reported earlier<sup>19</sup>. Here, we report the characteristics of PCC 11802 and its comparison to its close neighbor PCC 11801 and the model strain PCC 7942. The phylogenetic analysis using 16S rRNA and concatenated sequences of a set of 29 house-keeping proteins<sup>32,33</sup> revealed that PCC 11802 belonged to *Synechococcus elongatus* clade (Figs. 1A, S1 and S2). The strain was deposited with Pasteur Culture Collection of Cyanobacteria as *Synechococcus elongatus* PCC 11802. The average length and width calculated by imaging the exponentially growing cells (OD<sub>730</sub> ~ 0.5–0.6) of PCC 11802 were found to be 3.8 and 1.2 μm, respectively. The cell length of PCC 11802 is thus significantly larger than PCC 11801 that has an average length and width of 2.5 and 1.4 μm, respectively (Fig. 1C)<sup>19</sup>. It has been reported that the cell size of cyanobacteria can be affected by growth rate<sup>34</sup>. However, we believe that the difference observed in the cell sizes of PCC 11801 and 11802 is not due to differences in growth rate as the imaging was performed under ambient CO<sub>2</sub> conditions at 38 °C where both the strains have comparable growth rates (Fig. 2A).

The whole-genome sequencing of PCC 11802 revealed a genome size of 2.7 Mbp. The genome of PCC 11802 was annotated using three different annotation servers, viz., Rapid Annotation using Subsystem Technology (RAST)<sup>35–37</sup>, Integrated Microbial Genomes and Microbiomes<sup>38</sup>, JGI (IMG) and NCBI Prokaryotic Genome Annotation Pipeline<sup>39</sup>. The general characteristics of the annotated genome are presented in Table 1 (refer to the Supplementary File S-1 for annotation). PCC 11802 shares ~83% and ~97% genome identity with PCC 7942 and PCC 11801, respectively (Figs. S3–S6). There are 94,540 single nucleotide polymorphisms (SNPs) in PCC 11802 compared to PCC 11801 (Supplementary File S-2). Around 48,486 of these SNPs are present in the coding regions of the genome that contribute to 19,481 amino acid changes in the proteins. However, there are 338,422 SNPs in PCC 11802 compared to PCC 7942 (Supplementary File S-3). Next, it was of interest to analyze the amino acid identity between the corresponding homologous proteins of PCC 11802 and some of the model strains (Fig. 1B). As expected, PCC 11802 shares the highest identity with proteins of PCC 11801, followed by PCC 7942 and UTEX 2973 (Fig. 1B and Table S1). On the other hand, majority of the proteins of PCC 11802 showed fairly low identity with the corresponding homologs from PCC 7002 and PCC 6803 (Fig. 1B and Table S1).

It has been of significant interest to identify the genetic determinants of fast growth phenotype. To that end, Pakrasi and coworkers demonstrated that of the 53 single nucleotide polymorphisms (SNPs) between UTEX 2973 and PCC 7942, the SNPs in the genes, *atpA*, *ppnK*, and *rpaA* are responsible for the fast growth of UTEX 2973<sup>21,40–42</sup>. Of the five single amino acid polymorphisms (SAPs) present in these three genes of UTEX 2973, four SAPs were also found in PCC 11801 and PCC 11802 (Fig. 1D). In addition to these SAPs, we detected six, fifteen,

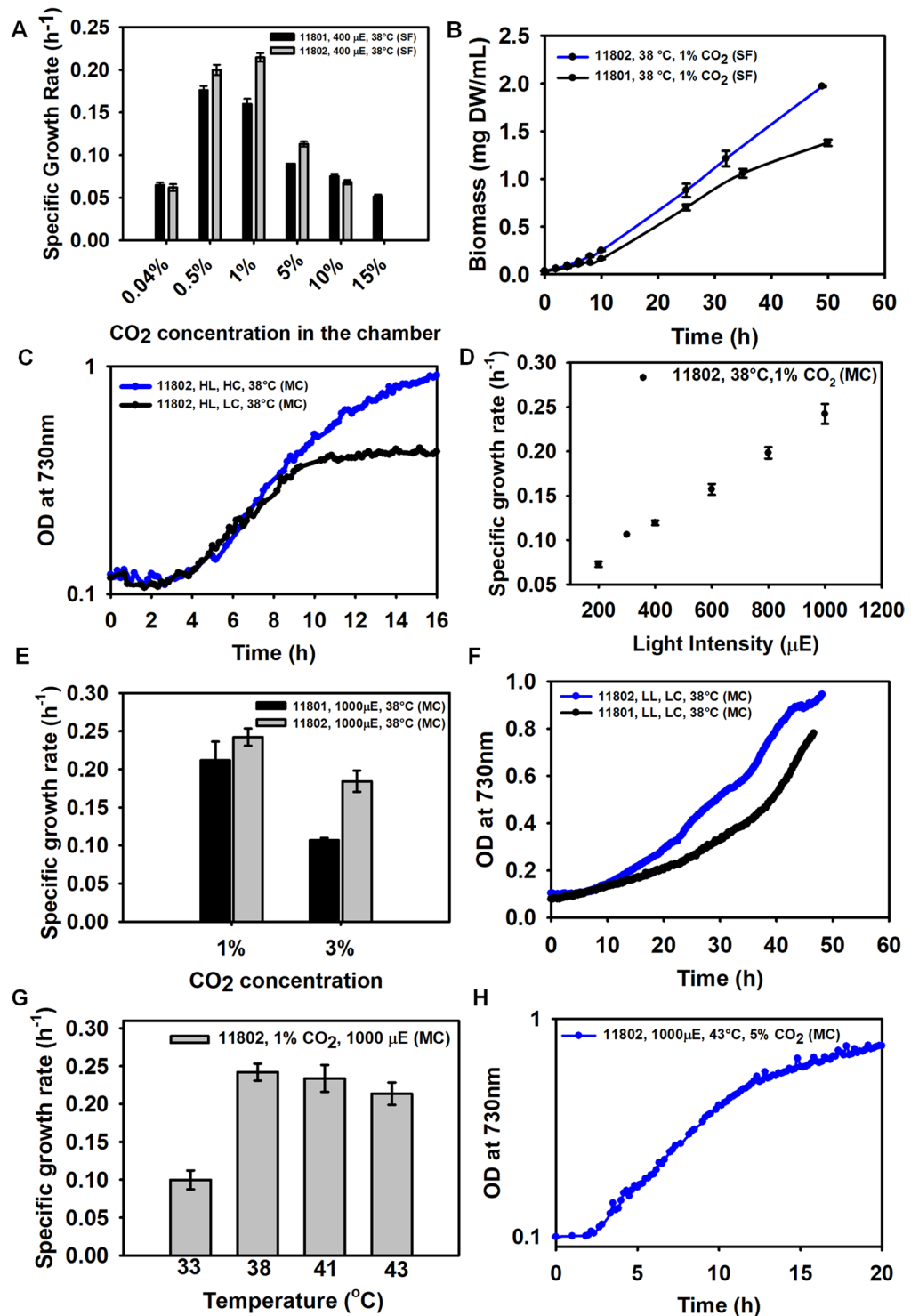


**Figure 1.** Identification, phylogenetic assignment, and genome comparison of PCC 11802. (A) A phylogenetic tree created using concatenated protein sequences of 29 constitutive genes showing the completely sequenced organisms in the phylogenetic neighborhood of PCC 11802. A complete tree with 128 cyanobacteria is shown in the supplementary information (Fig. S2). (B) A histogram depicting the level of amino acid identity between shared genes between PCC 11802 genome and the cyanobacterial strains PCC 11801, PCC 7942, UTEX 2973, PCC 7002 and PCC 6803. Only bidirectional best hits were included in the statistics. (C) Cell size comparison of exponentially growing cultures of PCC 11802 and PCC 11801 and (D) Alignment of sequences of proteins, AtpA, PpnK and RpaA showing the amino acid substitutions responsible for fast-growth phenotype and other single amino acid polymorphisms (SAPs) observed in PCC 11802 and 11801 with respect to PCC 7942 and UTEX 2973.

and five other SAPs in the protein sequences of genes, *atpA*, *ppnK*, and *rpaA*, respectively, in PCC 11801 and PCC 11802 compared to UTEX 2973 and PCC 7942. The functional significance of these additional SAPs in PCC 11802 and PCC 11801 cannot be explained presently. We further analyzed the 31 conserved house-keeping proteins<sup>32</sup> of PCC 11802 against the respective proteins of PCC 11801 and PCC 7942 (Figs. S7–34). The sequences of proteins, RpsJ, RpsS, and RplK do not have any mutation among these three strains. We found an average amino acid identity of ~97% between PCC 11802 and PCC 7942 for these 31 conserved proteins. On the other hand, PCC 11802 shares 99.5% amino acid identity with PCC 11801 for the 31 conserved proteins.

**Function-based genome comparison between PCC 11802 and 11801.** We further investigated the differences between the two closely related strains, PCC 11802 and PCC 11801. The function-based genome comparison resulted in a set of proteins that are unique in PCC 11802 and PCC 11801 compared to each other (Table 2). We observed that the major differences between the two strains were the presence of different kinds of type II toxin-antitoxin systems (TAS). Apart from the type II TAS proteins, there were a few other proteins specific to PCC 11802 or 11801. The potential role of these proteins, as described in literature, has been summarized in Table 2. However, the functional validation of these TAS proteins in PCC 11802 will be required before an exact role can be assigned.

The TAS has been poorly characterized in cyanobacteria with only a few reports describing their potential functions. Out of the six known TAS, type II is comparatively better characterized by well-defined genetic loci<sup>43</sup>. Type II TAS consists of a stable toxin protein and an antitoxin protein that loses the stability under stress conditions<sup>44</sup>. Under normal growth conditions, the antitoxin is synthesized in a concentration that is much higher than toxin forming a toxin-antitoxin complex, thereby inhibiting the toxicity caused by the toxin. Under stress conditions (heat, nutrient starvation and loss of plasmid), either the antitoxin production is repressed, or it becomes unstable, releasing the toxin into the cell that targets cellular machinery responsible for cell death or growth arrest<sup>45</sup>. Most TAS systems cause programmed cell death, growth cessation or reversible growth arrest during



**Figure 2.** Growth characterization of PCC 11802 and comparison with PCC 11801. (A) The growth rates of PCC 11802 and 11801 in shake flasks (SF) under 400  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ , 38°C and varied CO<sub>2</sub> levels in the chamber, (B) biomass accumulation of PCC 11802 compared to that of PCC 11801 at 1% CO<sub>2</sub> in SF, (C) growth of PCC 11802 at high light (HL, 1000  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ), high CO<sub>2</sub> (HC, 1% CO<sub>2</sub>) compared with growth at high light (HL) and low CO<sub>2</sub> (LC, 0.04% CO<sub>2</sub>) at 38°C in MC, (D) dependence of specific growth rate of PCC 11802 on the light intensity at 38°C and 1% CO<sub>2</sub> in MC, (E) specific growth rates of PCC 11802 compared to PCC 11801 at high light (1000  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ) and high CO<sub>2</sub> concentrations (1% and 3%) at 38°C in MC, (F) comparison of growth of PCC 11802 to PCC 11801 under low light (LL, 200  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ ) and low CO<sub>2</sub> (LC, 0.04% CO<sub>2</sub>) at 38°C, (G) specific growth rates of PCC 11802 at varying temperatures at 1000  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$  and 1% CO<sub>2</sub> in MC, and (H) growth profile of PCC 11802 at 1000  $\mu\text{mol photons}\cdot\text{m}^{-2}\cdot\text{s}^{-1}$ , 43°C and 5% CO<sub>2</sub> in MC.

	RAST	IMG	NCBI Prokaryotic Genome Annotation Pipeline
Gene count	2924	2919	2752
GC content	54.8	54.8	—
Protein coding genes	2883	2866	2708
Protein with known function	2127	2201	2446
Protein without known function	756	665	262
Total RNA coding genes	41	53	44
tRNA genes	39	39	39
rRNA genes	2	3	2
Other RNAs	—	11	3

**Table 1.** Genome annotation for *Synechococcus elongatus* PCC 11802 obtained using RAST, IMG, and NCBI prokaryotic genome annotation pipeline.

stressful conditions by targeting critical cellular functions like the integrity of cell membrane, cell wall synthesis, DNA replication, ribosome assembly, and translation<sup>43,45–48</sup>. Broadly, the function of TAS is to allow the cells to cope up with stress conditions. The presence of different type II TAS in PCC 11802 and PCC 11801 may be indicative of employment of different mechanisms for cell survival or growth arrest under a particular stress condition. These TAS can be taken up as potential targets in studies aimed at gaining a fundamental understanding of cellular behavior under unfavorable conditions. These systems should also be explored for biotechnological applications that may require growth arrest during product formation phases<sup>49</sup>.

**Growth characteristics of PCC 11802.** The growth characterization of PCC 11802 was performed under a range of light, CO<sub>2</sub>, and temperature conditions, the three major factors determining cyanobacterial growth. We monitored growth in shake flasks (SF) as well as a multi-cultivator (MC), the latter offering a wider range of light intensities, a shorter light path, and the ability to bubble gases. Among the CO<sub>2</sub> levels tested for growth in SF at 38 °C, the minimum doubling time of 3.2 h was observed for PCC 11802 at 1% CO<sub>2</sub> (0.04–15%) (Fig. 2A). PCC 11801, on the other hand, showed the least doubling time of 3.9 h at 0.5% CO<sub>2</sub> (3.4 h for PCC 11802) under similar light and temperature conditions in SF (Fig. 2A). Note that PCC 11802 has higher growth rates compared to PCC 11801 under elevated CO<sub>2</sub> levels (Fig. 2A,E). Both these strains have similar growth rates at 0.04% and 10% CO<sub>2</sub> in SF, while PCC 11802 was not able to grow at 15% CO<sub>2</sub> (Fig. 2A). Further, PCC 11802 accumulates greater biomass compared to PCC 11801 under high CO<sub>2</sub> conditions (Fig. 2B), which may be due to its higher growth rate at elevated CO<sub>2</sub>.

It has been reported that the growth rate of PCC 11801 under ambient CO<sub>2</sub> increased by ~4 fold when grown in MC with the bubbling of air compared to that in SF<sup>19</sup>. Therefore, we investigated whether PCC 11802 behaves similarly when grown in MC. We observed that unlike PCC 11801 that has a least doubling time of 2.3 h under ambient (0.04%) CO<sub>2</sub> at 1000 μmole photons.m<sup>-2</sup>.s<sup>-1</sup> and 41 °C, PCC 11802 has its fastest growth at 1% CO<sub>2</sub>, 38 °C and 1000 μmole photons.m<sup>-2</sup>.s<sup>-1</sup> with bubbling in MC (Fig. 2C,D,G). The doubling time of PCC 11802 under the optimal conditions in MC was found to be 2.8 h (Fig. 2C). The growth rates of PCC 11802 and 11801 declined when tested at 3% CO<sub>2</sub> in MC (Fig. 2E). However, the decline in the growth of PCC 11802 was much less (~25%) compared to PCC 11801 (~54%). The growth of PCC 11802 under low light (200 μmole photons.m<sup>-2</sup>.s<sup>-1</sup>) and low CO<sub>2</sub> (0.04%) condition was ~1.7 times faster than PCC 11801 (Fig. 2F, Table S2)<sup>19</sup>. PCC 11802 maintains a better growth rate (doubling time of ~3.6 h) even at a further higher level of CO<sub>2</sub> (5%) at 1000 μmole photons.m<sup>-2</sup>.s<sup>-1</sup> and 1% CO<sub>2</sub> (Fig. 2H). It is evident from these results that PCC 11802 has a different CO<sub>2</sub> specificity and grows better at elevated CO<sub>2</sub> compared to PCC 11801. It is also noticeable that the growth of PCC 11802 during the exponential phase is limited by CO<sub>2</sub>, whereas that of PCC 11801 is limited by light (Fig. 3). It has been demonstrated in PCC 6803 that the proteome is significantly altered in response to light compared to CO<sub>2</sub><sup>50</sup>. Under low light and high CO<sub>2</sub> conditions, the abundance of carbon assimilatory proteins reduces to compensate for expanded light-harvesting proteins<sup>50</sup>. The differential responses of PCC 11801 and PCC 11802 to the availability of light and CO<sub>2</sub> might be due to differential regulatory controls, abundance, and utilization of light and carbon assimilatory proteins. We speculate that under low light conditions, PCC 11801 might allocate significant proteome fraction for light-harvesting proteins resulting in shrinking of carbon-assimilatory proteins and thus affecting the overall growth. However, a detailed comparative proteomics study on these two strains under these conditions will be required to pinpoint the differential regulatory controls.

The growth of PCC 11802 was also monitored under different temperatures at 1% CO<sub>2</sub> (Fig. 2G). Both PCC 11802 and 11801 are able to tolerate a temperature of 20 °C with a doubling time of 18–19 h under ambient CO<sub>2</sub> conditions (Fig. S35). We observed that PCC 11802 has similar growth at 38 °C and 41 °C (Fig. 2G, Table S2), whereas the growth of PCC 11801 increased ~16% at 41 °C<sup>19</sup>. PCC 11802 grows better at 43 °C with only a ~9% decline in the growth rate, whereas the growth of PCC 11801 shows a more drastic, 60% decline at 43 °C<sup>19</sup>. Thus, both the strains have optimal growth temperatures ranging between 38–41 °C.

**Carbon storage and energy charge under elevated CO<sub>2</sub> conditions.** The synthesis of glycogen and carbon partitioning plays an important role in energy balancing during cyanobacterial growth<sup>51</sup>. Consistent with our previous results in PCC 11801<sup>19</sup>, PCC 11802 also has lower carbohydrate and glycogen content (Fig. 4A,B)

	Locus Tag <sup>†</sup>	Protein ID <sup>‡</sup>	Putative Role	Remarks based on previous reports
<b>Proteins unique to PCC 11802</b>				
<i>Type III restriction-modification system methylation subunit/site-specific DNA-methyltransferase</i>	1823836–1821905*	peg.1952*	DNA methylation	May be involved in epigenetic regulation of genes through differential methylation of genes <sup>57,78</sup>
<i>Retron-type RNA-directed DNA polymerase</i>	EKO22_06085	QFZ92002	Reverse transcription	Uses RNA as a template as well as the primer for reverse transcription <sup>79,80</sup>
<i>Programmed cell death antitoxin MazE like</i>	EKO22_08965	QFZ92458	Programmed cell death/Reversible Growth Arrest	Toxin-antitoxin system (TAS) responsible for heat-induced programmed cell death in <i>Synechocystis</i> sp. PCC 6803 <sup>81</sup>  Involved in the adjustment of nutrient consumption under starvation conditions <sup>44–46</sup>
<i>Programmed cell death toxin MazF like</i>	EKO22_08970	QFZ92459		
<i>RelB/StbD replicon stabilization protein (antitoxin to RelE/StbE)</i>	EKO22_03050	QFZ91496		
<i>VapB protein (antitoxin to VapC)</i>	EKO22_08890	QFZ92444		
<i>VapC toxin protein</i>	EKO22_08885	QFZ92443		
<i>YefM protein (antitoxin to YoeB)</i>	EKO22_07885	QFZ92286	Survival of cells in prolonged exposure to stress	This TAS system has been shown to be involved in cell survival under prolonged exposure to antibiotics by entering a dormant state <sup>82</sup>
<i>YoeB toxin protein</i>	EKO22_07880	QFZ92285		
<b>Proteins unique to PCC 11801</b>				
<i>Putative ski2-type helicase MJ1124</i>	DOP62_06670	AZB72445	RNA degradation, processing, and splicing	Demonstrated to activate RNA degradation in eukaryotes <sup>83</sup>
<i>HigA protein (antitoxin to HigB)</i>	DOP62_07875	AZB72637	Reversible Growth Arrest	These are found to be overexpressed in <i>Mycobacterium tuberculosis</i> under different stress conditions such as heat shock, nutrient starvation, DNA damage and hypoxia <sup>84</sup>
<i>HigB toxin protein</i>	DOP62_07880	AZB72638		

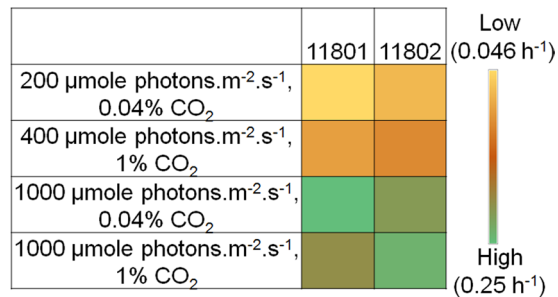
**Table 2.** The function-based comparison between *Synechococcus elongatus* PCC 11802 and 11801 identified using RAST tool. The genes annotated only in PCC 11802 compared to PCC 11801 genome and vice-versa are represented. The obtained results were also verified by comparing the obtained protein sequence against the whole genome sequence using the NCBI BLAST. <sup>†</sup>The IDs are based on annotations obtained through NCBI Prokaryotic Genome Annotation Pipeline. <sup>\*</sup>The coordinates of gene annotated as a single protein by RAST have been provided instead of NCBI locus tag since a translated protein sequence was not obtained using NCBI Prokaryotic Genome Annotation Pipeline. <sup>‡</sup>RAST protein ID is provided instead of NCBI protein ID. The protein sequence can be obtained from Supplementary File S-1 (Table S3).

under fast growth conditions of 0.5 and 1% CO<sub>2</sub> (Fig. 2A). PCC 11801 had the lowest glycogen content at 0.5% CO<sub>2</sub><sup>19</sup>, where its growth rate is maximum in SF (Fig. 2A). It is noticeable that the growth rate of PCC 11801 at 1% CO<sub>2</sub> is less compared to 0.5% CO<sub>2</sub> (Fig. 2A). The glycogen content of PCC 11801 at 1% CO<sub>2</sub> is higher than at 0.5% CO<sub>2</sub><sup>19</sup>, signifying that the excess carbon that could not be utilized for biomass production is stored as glycogen. On the other hand, the growth of PCC 11802 is highest at 1% CO<sub>2</sub> in SF (Fig. 2A) with the least glycogen content (Fig. 4B). PCC 11802 has ~3 times higher glycogen content and ~2 times less ADP-glucose (the precursor for glycogen synthesis) than PCC 11801 under ambient CO<sub>2</sub> conditions (Fig. 4C). However, the glycogen content of PCC 11802 is ~5 times lower than PCC 11801 at 1% CO<sub>2</sub>. This suggests that PCC 11802 is more efficient than PCC 11801 in utilizing the available CO<sub>2</sub> for growth rather than glycogen storage. This further strengthens our results on better growth and carbon assimilation efficiency of PCC 11802 under high CO<sub>2</sub> conditions compared to PCC 11801.

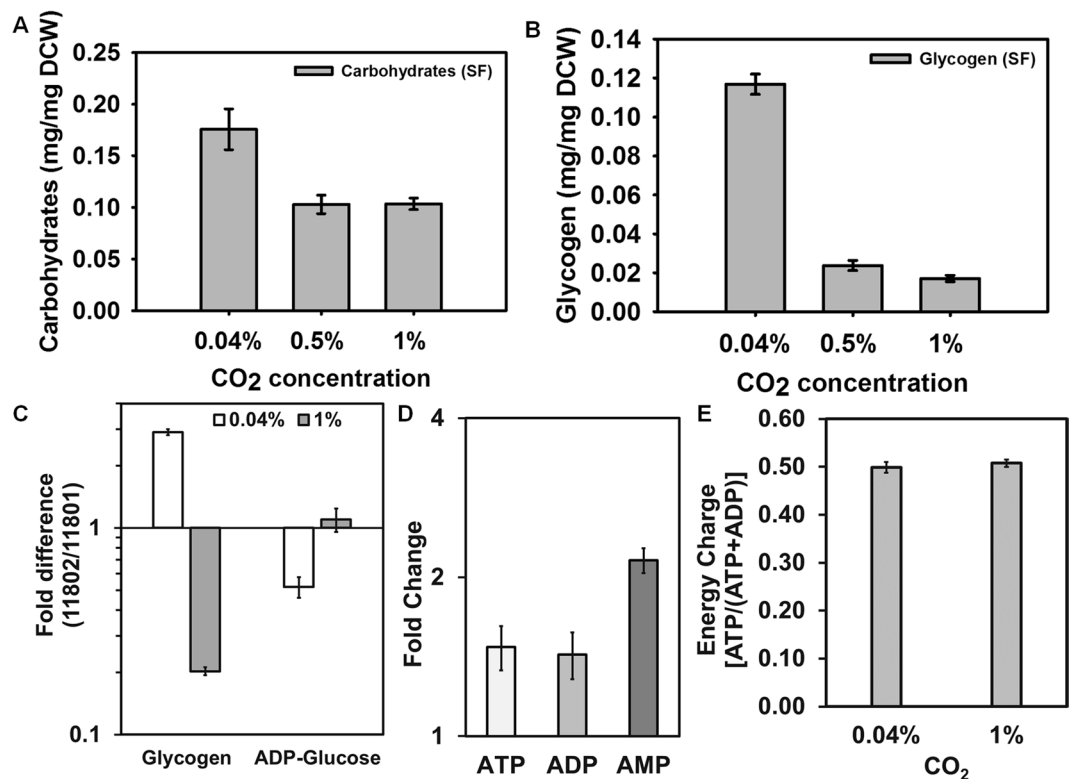
PCC 11802 has elevated levels of nucleotides, ATP, ADP, and AMP at 1% CO<sub>2</sub> compared to ambient CO<sub>2</sub> (Fig. 4D). It is reported that the levels of these nucleotides increase before entering the exponential phase and decline steadily after the exponential phase<sup>51</sup>. Higher levels of these energy nucleotides at elevated CO<sub>2</sub> might be indicative of better photosynthesis and ATP production. The higher growth rate under elevated CO<sub>2</sub> might be facilitated by higher pools of these energy nucleotides that participate in anabolic reactions during cell growth. Despite higher levels of these energy nucleotides under high CO<sub>2</sub>, the energy charge at ambient and 1% CO<sub>2</sub> was found to be constant (Fig. 4E). This suggests that the cells try to balance the energy charge even when the metabolite pools changes under different conditions.

**Metabolic changes under elevated CO<sub>2</sub> conditions.** It is of interest to identify the metabolic flux control points or the reactions that exert a high degree of control over the flux through a pathway. Among the omics tools that can be used for this purpose, metabolomics can provide useful information regarding the potential bottlenecks in a pathway<sup>52–54</sup>. For example, accumulation of a particular metabolite can result either from a higher rate of production or a lower rate of utilization by the downstream reactions<sup>30</sup>. Importantly, a rate-limiting reaction in a particular pathway can be readily identified based on the accumulation of its substrate and depletion of its product. Measurement of absolute metabolite concentrations is challenging due to matrix effects, inefficient extraction, degradation during extraction and variation in the detector sensitivity<sup>28,55,56</sup>. Therefore, to achieve even relative quantitation, the use of internal standards is necessary. Isotopic dilution mass spectrometry (IDMS) is a technique that can correct for artifacts in quantitation by using the peak area ratio of an analyte and its isotopic internal standard<sup>28,55–58</sup>.

We measured the relative metabolite pools of PCC 11801 and PCC 11802 by isotopic ratio method<sup>29,55,56</sup>. We utilized the intracellular metabolite extract of PCC 11801, which is fully labeled with isotopic <sup>13</sup>C as an internal standard. This strategy allowed us to use <sup>13</sup>C isotopologue of each metabolite as its respective internal reference. Although the levels of individual metabolites varied between PCC 11801 and PCC 11802 (Fig. S36), this data alone may not suffice for the identification of flux control points. The individual carbon flux control points might



**Figure 3.** Comparison of growth rates of PCC 11802 and PCC 11801 under different combinations of light and CO<sub>2</sub>. The color map is created based on their growth rate values in MC at 38 °C under the respective condition without any normalization. The scale bar from low to high growth rate is represented.



**Figure 4.** Comparison of carbon storage and energy balance between ambient and elevated CO<sub>2</sub> condition. (A) Total carbohydrate and (B) glycogen content of PCC 11802 under ambient (0.04%) and elevated (0.5% and 1%) CO<sub>2</sub> levels, (C) fold difference of glycogen and ADP-glucose (precursor for glycogen synthesis) between PCC 11802 and 11801 under ambient and 1% CO<sub>2</sub>, (D) fold change of the nucleotides ATP, ADP and AMP at 1% CO<sub>2</sub> compared to ambient CO<sub>2</sub> in PCC 11802 and (E) cellular energy charge of PCC 11802 under ambient and 1% CO<sub>2</sub> condition.

differ in various strains. Therefore, to understand the distinct metabolic changes and differential flux control that occur in PCC 11802 and PCC 11801, resulting in their different growth and biochemical phenotypes, we assessed the fold changes in metabolite levels while shifting from ambient to 1% CO<sub>2</sub> conditions. A proportionate fold change in the levels of metabolites of a particular pathway is expected with the change in the external conditions (e.g. CO<sub>2</sub>), the absence of which will indicate the presence of a regulatory node or a rate-limiting reaction. The principal component analysis (PCA) performed using the fold change values showed PCC 11802 and 11801 as distinct groups (Fig. S37). In general, we observed that the abundance of metabolites involved in the Calvin-Benson-Bassham (CBB) cycle and participating in CO<sub>2</sub> fixation was elevated at 1% CO<sub>2</sub> in both the strains (Fig. 5). However, the enhancement of metabolite levels was greater for PCC 11802 compared to PCC 11801. Despite an increase in abundance of sedoheptulose 1,7 biphosphate (SBP) in PCC 11801 at 1% CO<sub>2</sub>, the levels of sedoheptulose-7-phosphate (S7P) showed the negligible change between ambient and 1% CO<sub>2</sub>. This indicates that the conversion of SBP to S7P catalyzed by the enzymes sedoheptulose 1,7 biphosphatase (SBPase)

might be rate-limiting under 1% CO<sub>2</sub>. This might be a reason why increasing the CO<sub>2</sub> concentration does not result in a similar enhancement of growth rate in PCC 11801 compared to PCC 11802 (Fig. 2A). In fact, cyanobacterial fructose-1,6-bisphosphatase (FBP/SBPase) has been identified as one of the carbon flux control enzymes, and its overexpression has been shown to increase the growth rate and biomass accumulation in PCC 6803<sup>59,60</sup>. PCC 11802, on the other hand, has a higher fold increase of both SBP as well as S7P suggesting efficient carbon assimilation and regeneration through the CBB cycle (Fig. 5).

The higher fold increase of ribulose 1,5 bisphosphate (RuBP) and 3-phosphoglycerate (3PGA) in 11801 may be indicative of slower conversions to downstream metabolites as seen by the comparative lower fold increase of fructose 1,6 bisphosphate (FBP) and fructose-6-phosphate (F6P). This might be indicative of probable triose phosphate utilization (TPU) limitation<sup>11</sup> in 11801, which might be negligible in PCC 11802. TPU limitation arises when the inherent metabolism of the organism is insufficient to completely utilize the triose phosphates generated from the CBB cycle, or the rate of utilization of triose phosphates is less than the production rate<sup>11</sup>. The extent of TPU limitation may vary in different cyanobacterial strains, and it may be hypothesized that TPU utilization will be more evident under elevated CO<sub>2</sub> conditions<sup>61</sup>. Although there are no direct reports of overcoming TPU limitation in cyanobacteria, a few reports suggest that the photosynthetic efficiency of recombinant strains of cyanobacteria producing sucrose<sup>62</sup>, 2,3 butanediol<sup>5</sup>, isobutanol<sup>63</sup>, and ethylene<sup>64</sup> was found to be increased compared to the native wild type strains. The decline in the substrates for carboxylation reactions like RuBP and phosphoenolpyruvate (PEP) in PCC 11802 might be indicative of faster conversion to products, 3PGA, and aspartate (ASP), respectively (Fig. 5).

The spontaneity of a particular reaction depends on the Gibbs' free energy change and the concentrations of reactants and the products. Appropriate thermodynamic evaluation coupled with the overexpression of the enzymes that replenish the rate-limiting metabolites may increase product titers<sup>65</sup>. For enhanced production of biofuels using cyanobacterial hosts, strains capable of better growth under elevated CO<sub>2</sub> are preferred so that a greater carbon flux can be rerouted towards the desired products. The hypothesis that PCC 11802 utilizes more carbon for growth rather than for storage compared to PCC 11801 at elevated CO<sub>2</sub> is further strengthened by monitoring the fold increase of another storage molecule, sucrose. We observe that apart from making more glycogen, PCC 11801 also makes more sucrose as shown by higher levels of UDP-glucose (UDPG), sucrose-6-phosphate (SUC-6-P) and sucrose (SUC) at 1% CO<sub>2</sub>. We also observed a higher fold increase of succinate and glutamate in PCC 11801. The higher fold increase of the rate-limiting the CBB cycle metabolites and less TPU limitation in PCC 11802 coupled with its higher growth rate at 1%, CO<sub>2</sub> makes it an interesting candidate for heterologous expression and production of biofuels/biochemical.

**Genetic modification of PCC 11802.** The ability to carry out genetic modifications is an important pre-requisite for biotechnological applications. Cyanobacterial genomes can be readily engineered via homologous recombination and several vectors are readily available for model cyanobacteria. We first tried the integrative vector pSyn\_1 that has been widely used for integration into the neutral site I (NSI) of PCC 7942. This was based on the presence of a putative NSI in PCC 11802 that shares 82% homology with that of PCC 7942. However, the transformation efficiency was too low to be used on a regular basis. To that end, we replaced the homology arms of pSyn\_1 with those for NSI of PCC 11801 using polymerase incomplete primer extension (PIPE) cloning<sup>66</sup> method to obtain the plasmid pSyn\_11801 (Fig. 6A). Note that the neutral site regions of PCC 11801, NS1A and NS1B share 100% and 95% identity with the respective regions of PCC 11802. As expected, the modified plasmid shows satisfactory rates of natural transformation in both PCC 11801 and 11802. The transformation rate was found to be  $54 \pm 10$  cfu/ $\mu$ g of the plasmid (eYFP) using 7 mL culture of 0.6 OD<sub>730</sub>.

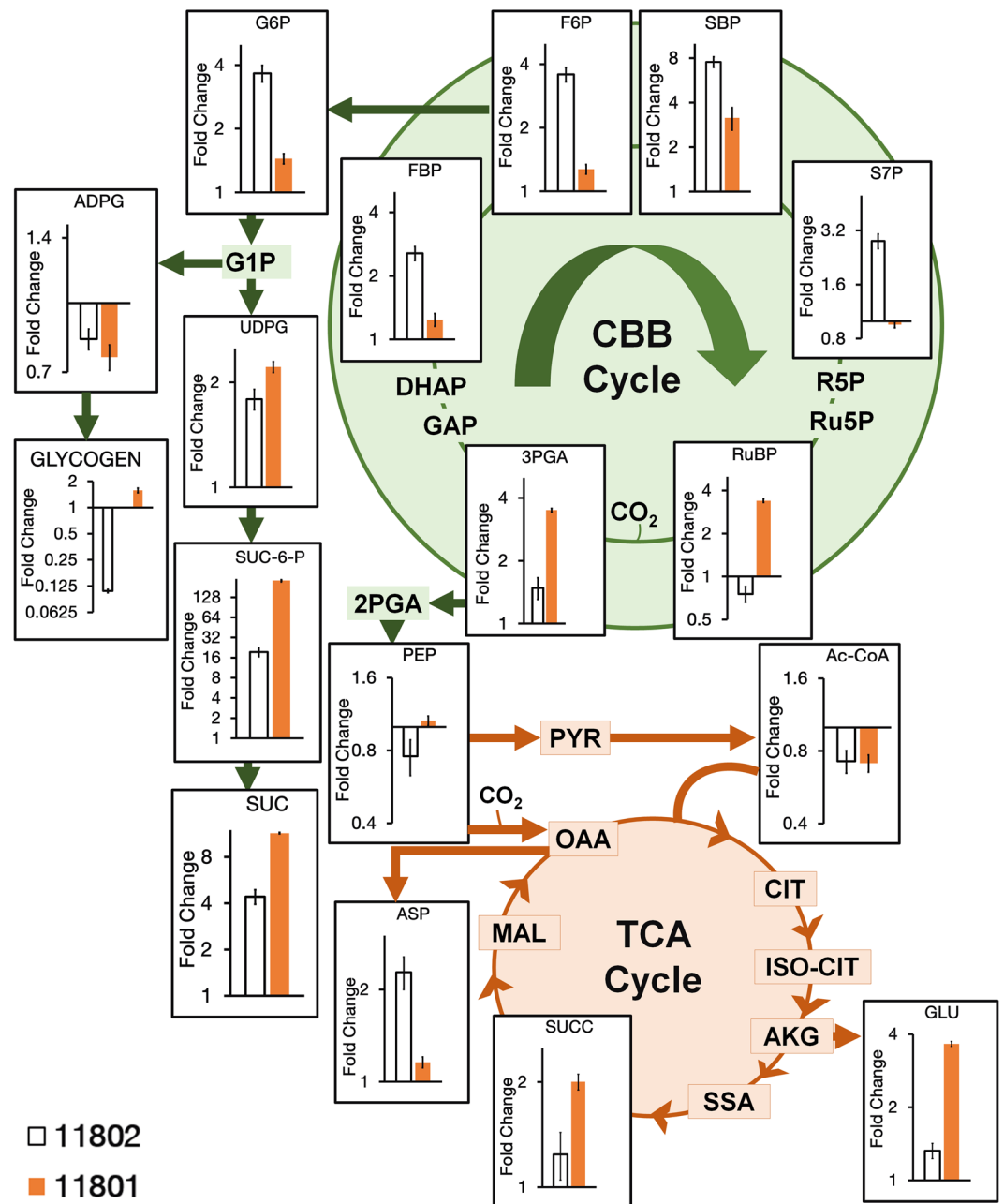
For high-throughput characterization of biological parts, reporter systems such as enhanced yellow fluorescent protein (eYFP), mOrange, luciferase are commonly used<sup>11,67</sup>. However, metabolic engineering necessitates the expression of heterologous functional genes (enzymes) for the production of value-added chemicals. To enhance the flux towards the desired product, engineering approaches like knockout of storage molecules and overexpression of enzymes crucial for the replenishment of rate-limiting metabolites are essential. Phosphoenolpyruvate carboxylase (PEPC) is a crucial enzyme for an anaplerotic pathway that replenishes the TCA cycle intermediates and is usually overexpressed for enhanced production of biochemical derived from TCA cycle<sup>8,27</sup>. Therefore, we demonstrate the heterologous expression of a reporter and a functional gene encoding for eYFP and PEPC in PCC 11802 under cpcB promoter of PCC 11801 (Fig. 6B,C respectively). These results also show the portability of integrative plasmid and the cpcB promoter of PCC 11801 in PCC 11802.

## Conclusion

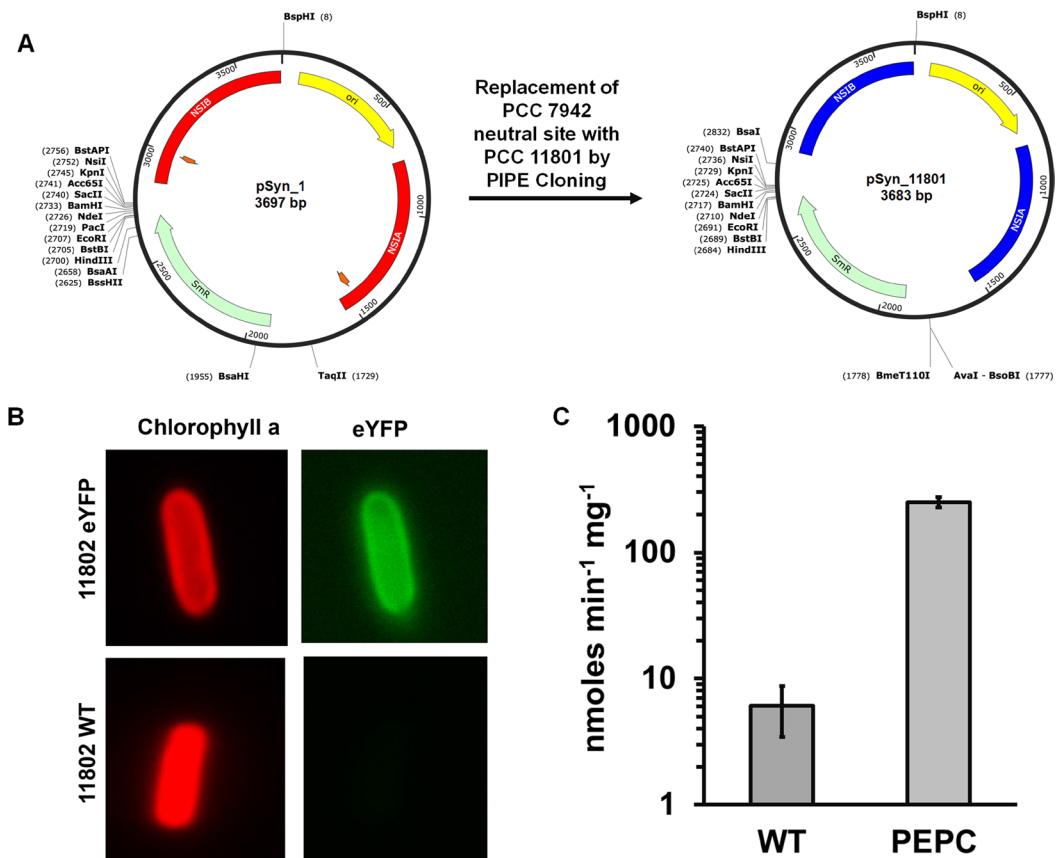
We report the physiological, genomic, biochemical, and metabolic characterization of a novel fast-growing and naturally transformable cyanobacterium, *Synechococcus elongatus* PCC 11802. The strain shows a doubling time of 2.8 h under the optimal growth conditions of 1% CO<sub>2</sub>, 38 °C and 1000  $\mu$ mol photons.m<sup>-2</sup>.s<sup>-1</sup> and without the addition of any vitamin supplement. PCC 11802 is phylogenetically close to PCC 11801 with ~97% genome identity, and ~97% average protein identity. Both these strains were isolated from the same geographical location (Powai lake, Mumbai, India). Function-based genome comparison shows that these strains differ majorly in toxin-antitoxin systems that are responsible for programmed cell death or reversible growth arrest under unfavorable conditions.

The physiological and biochemical characterization of PCC 11802 showed striking differences compared to PCC 11801. Our results show that the exponential growth of PCC 11802 is better than PCC 11801 under low light-low CO<sub>2</sub>, low light-high CO<sub>2</sub>, and high light-high CO<sub>2</sub> conditions. PCC 11802 accumulates ~3 fold more and ~5 fold less glycogen than PCC 11801 under ambient and 1% CO<sub>2</sub>, respectively. PCC 11802 thus diverts more carbon for growth at 1% CO<sub>2</sub> compared to PCC 11801. We performed metabolomics studies to assess possible reasons for the higher growth rate of PCC 11802 compared to PCC 11801 under high CO<sub>2</sub> conditions. The results





**Figure 5.** LCMS based analysis of metabolic pool sizes in PCC 11802 and 11801 under elevated CO<sub>2</sub> conditions. The bar graphs represent the ratio of metabolite pools at 1% CO<sub>2</sub>/ambient (0.04%) CO<sub>2</sub> in PCC 11802 and 11801. The metabolites represented in the figure show the difference between PCC 11802 and 11801 at a p-value < 0.01 using the t-test with the exception of phosphoenolpyruvate (PEP), acetyl coenzyme A (Ac-CoA) and ADP-glucose (ADPG) which show p-values of 0.02, 0.09 and 0.5 respectively. The relative metabolite pool sizes in each condition were estimated by extracting the metabolites as described previously and mixing with an extract of PCC 11801 which is fully labeled with isotopic <sup>13</sup>C carbon. High-resolution LCMS data was collected by achieving separation of metabolites using a reverse phase, ion-pairing chromatography method. The ratio of areas of the <sup>12</sup>C monoisotopic peak to the respective <sup>13</sup>C monoisotopic peak was quantified for each metabolite, which served as a measure of relative pool size in a given condition. Abbreviations used for metabolites, ADPG: ADP-glucose, Ac-CoA: acetyl coenzyme A, AKG: alpha-ketoglutaric acid, ASP: aspartic acid, CIT: citric acid, DHAP: dihydroxyacetone phosphate, FBP: fructose 1,6 bisphosphate, F6P: fructose-6-phosphate, GAP: glyceraldehyde-3-phosphate, GLU: glutamate, G1P: glucose-1-phosphate, G6P: glucose-6-phosphate, ISO-CIT: isocitric acid, MAL: malic acid, OAA: oxaloacetic acid, PEP: phosphoenolpyruvate, PYR: pyruvic acid, RuBP: ribulose 1,5 bisphosphate, Ru5P: ribulose-5-phosphate, R5P: ribulose-5-phosphate, SBP: sedoheptulose 1,7 bisphosphate, SSA: succinyl semialdehyde, SUC: sucrose, SUCC: succinate, SUC-6-P: sucrose-6-phosphate, S7P: sedoheptulose-7-bisphosphate, UDPG: UDP-glucose, 2PGA: 2-phosphoglyceric acid, and 3PGA: 3-phosphoglyceric acid.



**Figure 6.** Genetic Modification of PCC 11802. (A) Construction of modified vector by replacing the neutral site I (NSI) homology arms of PCC 7942 in the pSyn\_1 plasmid with those of PCC 11801 by PIPE cloning method. (B) Microscopic images of wildtype and recombinant cells of PCC 11802 showing chlorophyll a fluorescence (left panels) and eYFP fluorescence (right panels), and (C) The activity of phosphoenolpyruvate carboxylase (PEPC) protein in wildtype and recombinant cells overexpressing PEPC protein grown under 1% CO<sub>2</sub> in shake flasks. The activity data is plotted on log<sub>10</sub> scale.

show that PCC 11802 has greater abundance of the CBB cycle intermediates and lesser triose phosphate utilization limitation under elevated CO<sub>2</sub> condition. The metabolic bottlenecks like repression of SBPase and RuBisCO that are evident in PCC 11801 seem to be alleviated in PCC 11802. Moreover, the reactions involving CO<sub>2</sub> assimilation catalyzed by RuBisCO and PEPC appear to be more active in PCC 11802.

In terms of metabolic engineering applications, both PCC 11802 and PCC 11801 appear to be promising candidates owing to their faster growth and genetic amenability. However, the choice of strain will depend on the product of interest. We believe that PCC 11802 may be a good candidate for the products that primarily derived from the intermediates of the CBB cycle. PCC 11801, on the other hand, might be a good candidate for the products derived from the TCA cycle. However, a detailed fluxomics study along with a head-on-comparison of product titers in each strain will be required to verify this hypothesis. The unique insights gained from the studies presented in this report strengthen the fundamental understanding of cyanobacterial carbon partitioning and metabolism under carbon limited and sufficient conditions. Furthermore, our study draws attention to the fact that metabolic differences can exist between two very closely related fast-growing cyanobacteria. Also, the metabolic characterization of potential host cyanobacterial strains is necessary to design strain-specific rational engineering efforts rather than a generalized approach.

## Methods

### Isolation, Genome Sequencing, and Genome Annotation of *Synechococcus elongatus* PCC 11802.

*Synechococcus elongatus* PCC 11802 was isolated along with PCC 11801 and six other cyanobacterial strains from Powai Lake, Mumbai, India (19.1273°N, 72.9048°E) as described previously<sup>19</sup>. The genomic DNA of PCC 11802 was isolated using a protocol described earlier<sup>19</sup> and treated with RNAase for removal of RNA contamination before sequencing. The genome sequencing of PCC 11802 was outsourced to Life Technologies (Thermo Fischer Scientific, Waltham, MA, USA) and was sequenced using Ion Torrent Personal Genome Machine (PGM). Other quality checks for raw reads and assembly were as described previously<sup>19</sup>. PCR was performed for filling additional gaps in the genome by designing primers specifically to amplify the gap regions. The whole-genome sequence of PCC 11802 was submitted to GenBank, NCBI (National Center for Biotechnology Information) under accession number, CP034671. The assembled genome was annotated using Integrated Microbial Genomes and Microbiomes, JGI (IMG)<sup>38</sup>, Rapid Annotation using Subsystem Technology (RAST)<sup>35–37</sup>, and NCBI Prokaryotic Genome Annotation Pipeline<sup>39</sup>.

**Phylogenetic analysis.** The cDNA of the 16S rRNA sequence was amplified using specific PCR primers. The 16S rRNA sequence of PCC 11802 was submitted to NCBI under accession number, MH666134. The phylogenetic analysis was performed using 16S rRNA and protein sequences of a set of 29 housekeeping genes from PCC 11802 and other cyanobacterial strains as described previously<sup>32,33</sup>. Concatenated sequences of these 29 proteins for 130 strains were obtained from the *CyanoGEBAs*<sup>32</sup> resource. The respective sequences were aligned using Clustal Omega<sup>68</sup>, and the phylogenetic tree was constructed using MEGA7<sup>69</sup> and iTOL (Interactive Tree Of Life)<sup>70</sup>.

**Comparative genome analysis.** The function-based genome comparison was performed using RAST<sup>35–37</sup> to identify the proteins unique to PCC 11802 compared to PCC 11801. The sequence-based comparison was performed using RAST<sup>35–37</sup> by using PCC 11802 as a reference against its closest neighbor strains PCC 11801, PCC 7942, UTEX 2973, and distantly related model strains PCC 7002 and PCC 6803. The percent identity between each protein of PCC 11802 and the proteins of the abovementioned strains was calculated. The proteins that gave bi-directional best hits were identified and represented in this study.

The whole-genome of PCC 11802 was used as the reference and aligned against PCC 11801 and PCC 7942 using Mauve tool<sup>71</sup>, and single nucleotide polymorphisms (SNPs) in PCC 11802 were calculated in comparison to PCC 11801 and PCC 7942. The differences between sequences of house-keeping proteins of PCC 11802 to that of respective protein sequences in neighbor strains PCC 11801 and PCC 7942 were identified by multiple sequence alignment using Clustal Omega<sup>68</sup>.

**Transmission electron microscopy (TEM).** One mL of an exponentially growing culture of PCC 11802 and PCC 11801 was centrifuged at 8000 g for 5 min, and the pellet was washed with MilliQ to remove traces of BG-11 medium components. The pellet was re-suspended in MilliQ water, and approx 10  $\mu$ L of the re-suspended culture was spotted on copper-coated formvar grids (Electron Microscopy Sciences, Hatfield, PA, USA), washed with Milli Q water and air-dried for 20 minutes. These grids were imaged using an EM (Philips CM-200, Amsterdam, Netherlands) at 200 kV with a magnification of  $\text{\AA} \sim 6600$ . Images were recorded digitally by using the Keen View Soft imaging system (Olympus, Tokyo, Japan).

**Growth conditions.** The strain was maintained in a shaker at a temperature of 38 °C, 120 rpm, ambient (0.04%) CO<sub>2</sub>, and light intensity of 400  $\mu$ mole photons. $\text{m}^{-2}.\text{s}^{-1}$  in BG-11 medium (pH = 7.5) unless specified otherwise. The CO<sub>2</sub> tolerance studies were performed in CO<sub>2</sub> incubator shaker (Adolf Kuhner AG, LT-X, Birsfelden, Switzerland) at CO<sub>2</sub> concentrations of 0.04%, 0.5%, 1%, 5%, 10% and 15% in the chamber at 38 °C and 120 rpm. The growth at different CO<sub>2</sub> concentrations was measured as OD<sub>720</sub> using one mL of cells in a UV-visible spectrophotometer (Shimadzu, UV-2600, Singapore).

The growth of PCC 11802 under a range of conditions including light intensity, temperature, and CO<sub>2</sub> concentration was measured in terms of OD<sub>720</sub> using a multi-cultivator (Photon Systems Instruments, MC 1000-OD, Czech Republic) with an inlet gas flow rate of 500 SCCM to bubble all eight tubes of multi-cultivators. The specific growth rate ( $\mu$ ) was estimated during the exponential growth phase from the slope of the semi-logarithmic plot between OD<sub>720</sub> nm and time. Doubling time was calculated as  $\ln(2)/\mu$ .

**Carbohydrate and glycogen estimation.** The total carbohydrates and glycogen content were measured using PCC 11802 cells harvested at 0.6–0.7 OD<sub>720</sub> nm at different CO<sub>2</sub> concentrations (0.04%, 0.5%, and 1%), 38 °C, 400  $\mu$ mole photons. $\text{m}^{-2}.\text{s}^{-1}$  and 120 rpm as described earlier<sup>19</sup>.

**Preparation of <sup>13</sup>C isotopically labeled biomass of *synechococcus elongatus* PCC 11801 for use as internal standard.** After several trials, a protocol was developed to obtain metabolite extract that shows dominant <sup>13</sup>C monoisotopic peaks but no <sup>12</sup>C monoisotopic peaks for all metabolites. Modified BG-11 medium that does not contain any organic carbon source (henceforth referred to as BG11-C) such as sodium carbonate, citric acid, and ferric ammonium citrate was used to prepare fully <sup>13</sup>C labeled metabolite extracts of PCC 11801. Iron sulfate heptahydrate was used to provide an iron source. The exponentially growing culture of PCC 11801 pre-adapted to BG11-C medium was used for inoculation with an OD<sub>730</sub> nm of 0.05 with a culture volume of 20 mL in 100 mL Erlenmeyer flask. Lower biomass was used for inoculation to minimize dilution with <sup>12</sup>C present in the biomass from the inoculum. A stopper was used to prevent the exchange of <sup>12</sup>CO<sub>2</sub> from the environment. Initially, <sup>13</sup>C-labeled sodium bicarbonate (NaH<sup>13</sup>CO<sub>3</sub>, 98 atom % <sup>13</sup>C from Sigma-Aldrich, St. Louis, MO) was added at a concentration of 2 g/L in the culture. Additional doses of NaH<sup>13</sup>CO<sub>3</sub> were provided at 18, 19.5, 21, and 22.5 hours after inoculation at a final concentration of 1 g/L. The <sup>13</sup>C labeled biomass was harvested at 23 h by fast filtration in the presence of light followed by rapid quenching in 80:20 methanol-water (precooled to –80 °C). The <sup>13</sup>C labeled metabolites were extracted from quenched cells as described earlier<sup>72</sup> with a minor modification, viz., the use of extraction solvent volume that was twice as large as earlier. The <sup>13</sup>C labeled metabolite extracts were filtered, and multiple aliquots of equal volume were dispensed in Eppendorf tubes, lyophilized and stored at –80 °C until ready for use. The <sup>13</sup>C labeling procedure was carried out in New Brunswick Innova 44 R shaker (Eppendorf, Hamburg, Germany) maintained at 38 °C, 120 rpm, and a light intensity of 300  $\mu$ mole photons. $\text{m}^{-2}.\text{s}^{-1}$ .

**Metabolite profiling using liquid chromatography-mass spectrometry (LCMS).** PCC 11802 and 11801 cells were grown under 0.04% and 1% CO<sub>2</sub> at 38 °C in shake flasks. Exponentially growing cells at an OD<sub>720</sub> of ~0.6 were filtered rapidly on nylon membrane filters (Whatman, 0.8  $\mu$ ) in the presence of light. The cells on the membrane filters were quickly transferred to 80:20 methanol-water (precooled to –80 °C) to quench the metabolism. The metabolites were extracted using a protocol as reported previously<sup>73</sup>. The metabolite extract was lyophilized and stored at –80 °C until ready for LCMS analysis. The metabolite extracts were reconstituted in 100  $\mu$ L 50:50 methanol-water and filtered using nylon syringe filters to remove any particulate matter. Each sample was mixed with fully <sup>13</sup>C-labeled biomass of PCC 11801 that acted as an internal

standard. An equal volume of labeled internal standard was added to all the samples. The data was acquired using an information-dependent acquisition (IDA) method on Triple-TOF 5600+ instrument (SCIEX, Framingham, MA) interfaced with Shimadzu Ultra Performance- Liquid Chromatography (UPLC) system Nexera LC – 30 AD (Shimadzu, Singapore). The instrument was operated under negative ion mode to acquire data in the  $m/z$  range of 50–1000 Da. The cycle and accumulation times were 1 s and 250 ms, respectively. Six  $\mu\text{L}$  sample was analyzed on reverse-phase ion-pairing chromatography using a C18 Synergi 4  $\mu\text{m}$  Hydro-RP LC column  $150 \times 2 \text{ mm}$  (Phenomenex Inc, Torrance, CA). The gradient program and other instrument parameters were as reported earlier<sup>72</sup>. Peak areas corresponding to the  $^{12}\text{C}$  monoisotopic peak and  $^{13}\text{C}$  monoisotopic peak for the metabolites of interest were quantified using the proprietary software MultiQuant (SCIEX, Framingham, MA).

The relative quantification of targeted metabolites under 0.04% and 1%  $\text{CO}_2$  conditions was performed by the normalizing area under the peak for  $^{12}\text{C}$  monoisotopic peak of a particular metabolite by its respective  $^{13}\text{C}$  monoisotopic peak to obtain area ratios. The fold change in metabolite pools under 1% compared to 0.04%  $\text{CO}_2$  was then calculated (area ratio under 1%  $\text{CO}_2$ /area ratio under 0.04%  $\text{CO}_2$ ). Statistical analysis and principal component analysis (PCA) were performed, and a heat map constructed using MetaboAnalyst 4.0<sup>74,75</sup> (Figs. S37 and 38).

**Genetic manipulation.** Seven mL of 0.6  $\text{OD}_{730}$  culture of PCC 11802 was centrifuged at 4,000 g for 5 minutes at room temperature. The pellet was washed and re-suspended in 100  $\mu\text{L}$  BG-11 medium. Two  $\mu\text{g}$  pSyn\_11801 plasmid, containing the gene of interest (eYFP or PEPC), was added to the resuspended cells. The cell-DNA mixture was incubated at 34°C in the dark for 12 h. The mixture was then spread on a 0.22  $\mu\text{m}$  filter membrane placed on a 1% BG-11 agar plate and incubated for 24 h at 38°C and 150  $\mu\text{E}$  light. Then the filter membrane was transferred to a BG-11 agar plate containing 50  $\mu\text{g}/\text{mL}$  of antibiotic, spectinomycin. After 48 h of incubation, colonies appeared on the filter membrane, which was then patched on a plate containing 100  $\mu\text{g}/\text{mL}$  of spectinomycin to achieve complete chromosomal segregation. Segregation was checked using confirmation primers, 5'CCAACGCCTATTCCAAGGGCGGC3', and 5'TGGCAATGTCTCTCTGAGGGGATG3. All the colonies obtained by transforming pSyn\_11801 plasmids were also verified using gene-specific primers.

**Fluorescence microscopy.** Fluorescence microscopy was performed as described earlier<sup>19</sup>. Briefly, the wild-type and eYFP expressing PCC 11802 cells ( $\text{OD}_{720}$  of ~1.0) were centrifuged at 8000 g for 3 min and washed with Milli-Q water twice. The cells were then re-suspended in 4% paraformaldehyde for 30 minutes at 4°C. The fluorescence images of fixed WT and eYFP mutant cells were acquired using a Zeiss Axio Observer Z1 (100X objectives, NA = 1.40; Carl Zeiss MicroImaging Inc., Oberkochen, Germany) equipped with AxioCam camera controlled by Axiovision software [Axio Vision Release 4.8.3 SP1 (06–2012)]. Exposure time for imaging was 300 ms.

**Measurement of phosphoenolpyruvate carboxylase (PEPC) activity.** The wildtype and recombinant PCC 11802 cells containing a gene encoding PEPC protein were grown till  $\text{OD}_{730}$  of 0.6–0.7 at 1%  $\text{CO}_2$  and 200  $\mu\text{E}$  of light intensity in a shaker (Adolf Kuhner AG, LT-X, Birsfelden, Switzerland). One hundred mL of the culture was harvested by centrifugation at 8000 g for 10 minutes. The pellet was then resuspended in 500  $\mu\text{L}$  of lysis buffer containing 50 mM TRIS HCl (pH 8), 10 mM  $\text{MgCl}_2$ , 10% (v/v) glycerol, 1 mM EDTA, 1 mM DTT and 1 mg/L of lysozyme. One mM PMSF was added to minimize proteolysis<sup>76</sup>. The cells were lysed using tissue lyser LT (Qiagen, Hilden, Germany) with 0.1 mm glass beads in 30 cycles of 1 min of bead beating with 1 minute of cooling on the ice between two cycles.

The cell debris was separated by centrifugation at 20,000 g for 30 min at 4°C, and the soluble crude extract collected. The protein concentration in the extract was determined via Bradford assay using a standard curve for bovine serum albumin (BSA). PEPC activity was measured using a coupled assay where PEPC from crude extracts converts phosphoenolpyruvate (PEP) to oxaloacetic acid (OAA), which is then converted to malate. The second step concomitantly converts NADH to  $\text{NAD}^+$  and is catalyzed by the added malate dehydrogenase, which was over-expressed in *E. coli* and purified. The reaction mixture thus contained 100 mM Tris-HCl (pH 8), 10 mM  $\text{NaHCO}_3$ , 5 mM  $\text{MgCl}_2$ , 200  $\mu\text{M}$  NADH, 0.13  $\mu\text{g}$  of purified *E. coli* malate dehydrogenase, 2 mM PEP, and appropriate amounts of the crude extract. The reaction was initiated by addition of the crude extract at 30°C. The decrease in absorbance at 340 nm due to the conversion of NADH to  $\text{NAD}^+$  was monitored. The PEPC activity was calculated from the absorbance versus time graph and normalized to the total concentration of protein in the crude extract.

## Data availability

All data generated or analyzed during this study are included in this article (and its Supplementary Information Files). The complete genome and 16S rRNA gene of *Synechococcus elongatus* PCC 11802 is available at GenBank under accession numbers, CP034671 and MH666134, respectively. The data files for the metabolomics study of PCC 11802 presented in this article are deposited to the Metabolomics Workbench repository (<http://www.metabolomicsworkbench.org/>), <https://doi.org/10.21228/M89M4D>.

Received: 16 September 2019; Accepted: 20 December 2019;

Published online: 13 January 2020

## References

1. Wang, W., Liu, X. & Lu, X. Engineering cyanobacteria to improve photosynthetic production of alka (e) nes. *Biotechnol. Biofuels* **6**, 1–9 (2013).
2. Qian, X., Zhang, Y., Lun, D. S. & Dismukes, G. C. Rerouting of metabolism into desired cellular products by nutrient stress: Fluxes reveal the selected pathways in cyanobacterial photosynthesis. *ACS Synth. Biol.* **7**, 1465–1476 (2018).
3. Gao, Z., Zhao, H., Li, Z., Tan, X. & Lu, X. Photosynthetic production of ethanol from carbon dioxide in genetically engineered cyanobacteria. *Energy Environ. Sci.* **5**, 9857–9865 (2012).
4. Lan, E. I. & Liao, J. C. ATP drives direct photosynthetic production of 1-butanol in cyanobacteria. *Proc. Natl. Acad. Sci.* **109**, 6018–6023 (2012).

5. Oliver, J. W. K., Machado, I. M. P., Yoneda, H. & Atsumi, S. Cyanobacterial conversion of carbon dioxide to 2,3-butanediol. *Proc. Natl. Acad. Sci.* **110**, 1249–1254 (2013).
6. Zhou, J., Zhang, H., Zhang, Y., Li, Y. & Ma, Y. Designing and creating a modularized synthetic pathway in cyanobacterium *Synechocystis* enables production of acetone from carbon dioxide. *Metab. Eng.* **14**, 394–400 (2012).
7. Liu, X., Sheng, J. & Curtiss, R. III Fatty acid production in genetically modified cyanobacteria. *Proc. Natl. Acad. Sci.* **108**, 6899–6904 (2011).
8. Lan, E. I. & Wei, C. T. Metabolic engineering of cyanobacteria for the photosynthetic production of succinate. *Metab. Eng.* **38**, 483–493 (2016).
9. Xiong, W. *et al.* The plasticity of cyanobacterial metabolism supports direct CO<sub>2</sub> conversion to ethylene. *Nat. Plants* **1**, 15053 (2015).
10. Sengupta, A., Pakrasi, H. B. & Wangikar, P. P. Recent advances in synthetic biology of cyanobacteria. *Appl. Microbiol. Biotechnol.* **102**, 5457–5471 (2018).
11. Santos-Merino, M., Singh, A. K. & Ducat, D. C. New applications of synthetic biology tools for cyanobacterial metabolic engineering. *Front. Bioeng. Biotechnol.* **7**, 1–24 (2019).
12. Hendry, J. I., Prasannan, C. B., Joshi, A., Dasgupta, S. & Wangikar, P. P. Metabolic model of *Synechococcus* sp. PCC 7002: Prediction of flux distribution and network modification for enhanced biofuel production. *Bioresour. Technol.* **213**, 190–197 (2016).
13. Hendry, J. I. *et al.* Rerouting of carbon flux in a glycogen mutant of cyanobacteria assessed via isotopically non-stationary <sup>13</sup>C metabolic flux analysis. *Biotechnol. Bioeng.* **114**, 2298–2308 (2017).
14. Mehta, K. *et al.* Elevated carbon dioxide levels lead to proteome-wide alterations for optimal growth of a fast-growing cyanobacterium, *Synechococcus elongatus* PCC 11801. *Sci. Rep.* **9**, 6257 (2019).
15. Ludwig, M. & Bryant, D. A. *Synechococcus* sp. strain PCC 7002 transcriptome: Acclimation to temperature, salinity, oxidative stress, and mixotrophic growth conditions. *Front. Microbiol.* **3**, 354 (2012).
16. Vasudevan, R. *et al.* CyanoGate: A modular cloning suite for engineering cyanobacteria based on the plant MoClo syntax. *Plant Physiol.* **180**, 39–55 (2019).
17. Knoot, C. J., Khatri, Y., Hohlman, R. M., Sherman, D. H. & Pakrasi, H. B. Engineered production of hapalindole alkaloids in the cyanobacterium *Synechococcus* sp. UTEX 2973. *ACS Synth. Biol.*, <https://doi.org/10.1021/acssynbio.9b00229> (2019).
18. Song, K., Tan, X., Liang, Y. & Lu, X. The potential of *Synechococcus elongatus* UTEX 2973 for sugar feedstock production. *Appl. Microbiol. Biotechnol.* **100**, 7865–7875 (2016).
19. Jaiswal, D. *et al.* 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 (2018).
20. Yu, J. *et al.* *Synechococcus elongatus* UTEX 2973, a fast growing cyanobacterial chassis for biosynthesis using light and CO<sub>2</sub>. *Sci. Rep.* **5**, 8132 (2015).
21. Ungerer, J., Wendt, K. E., Hendry, J. I., Maranas, C. D. & Pakrasi, H. B. Comparative genomics reveals the molecular determinants of rapid growth of the cyanobacterium *Synechococcus elongatus* UTEX 2973. *Proc. Natl. Acad. Sci.* **115**, E11761–E11770 (2018).
22. Tan, X. *et al.* The primary transcriptome of the fast-growing cyanobacterium *Synechococcus elongatus* UTEX 2973. *Biotechnol. Biofuels* **11**, 218 (2018).
23. Knoot, C. J., Ungerer, J., Wangikar, P. P. & Pakrasi, H. B. Cyanobacteria: Promising biocatalysts for sustainable chemical production. *J. Biol. Chem.* **293**, 5044–5052 (2018).
24. Sengupta, A., Sunder, A. V., Sohoni, S. V. & Wangikar, P. P. Fine-tuning native promoters of *Synechococcus elongatus* PCC 7942 to develop a synthetic toolbox for heterologous protein expression. *ACS Synth. Biol.* **8**, 1219–1223 (2019).
25. Mukherjee, B., Madhu, S. & Wangikar, P. P. The role of systems biology in developing non-model cyanobacteria as hosts for chemical production. *Curr. Opin. Biotechnol.* **64**, 62–69 (2020).
26. Jazmin, L. J. *et al.* Isotopically nonstationary <sup>13</sup>C flux analysis of cyanobacterial isobutyraldehyde production. *Metab. Eng.* **42**, 9–18 (2017).
27. Hasunuma, T., Matsuda, M., Kato, Y., Vavricka, C. J. & Kondo, A. Temperature enhanced succinate production concurrent with increased central metabolism turnover in the cyanobacterium *Synechocystis* sp. PCC 6803. *Metab. Eng.* **48**, 109–120 (2018).
28. Schatschneider, S. *et al.* Quantitative isotope-dilution high-resolution-mass-spectrometry analysis of multiple intracellular metabolites in *Clostridium autoethanogenum* with uniformly <sup>13</sup>C-labeled standards derived from *Spirulina*. *Anal. Chem.* **90**, 4470–4477 (2018).
29. Abernathy, M. H. *et al.* Deciphering cyanobacterial phenotypes for fast photoautotrophic growth via isotopically nonstationary metabolic flux analysis. *Biotechnol. Biofuels* **10**, 1–13 (2017).
30. Jang, C., Chen, L. & Rabinowitz, J. D. Metabolomics and isotope tracing. *Cell* **173**, 822–837 (2018).
31. Will, S. E. *et al.* Day and night: Metabolic profiles and evolutionary relationships of six axenic non-marine cyanobacteria. *Genome Biol. Evol.* **11**, 270–294 (2019).
32. Shih, P. M. *et al.* Improving the coverage of the cyanobacterial phylum using diversity-driven genome sequencing. *Proc. Natl. Acad. Sci.* **110**, 1053–1058 (2013).
33. Calteau, A. *et al.* Phylum-wide comparative genomics unravel the diversity of secondary metabolism in cyanobacteria. *BMC Genomics* **15**, 1–14 (2014).
34. Zheng, X. & O’Shea, E. K. Cyanobacteria maintain constant protein concentration despite genome copy-number variation. *Cell Rep.* **19**, 497–504 (2017).
35. Brettin, T. *et al.* RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. *Sci. Rep.* **5**, 8365 (2015).
36. Aziz, R. K. *et al.* The RAST Server: Rapid annotations using subsystems technology. *BMC Genomics* **9**, 75 (2008).
37. Overbeek, R. *et al.* The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST). *Nucleic Acids Res.* **42**, D206–D214 (2014).
38. Markowitz, V. M. *et al.* IMG: the integrated microbial genomes database and comparative analysis system. *Nucleic Acids Res.* **40**, D115–D122 (2012).
39. Tatusova, T. *et al.* NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* **44**, 6614–6624 (2016).
40. Jaiswal, D. Alleles. In *Encyclopedia of Animal Cognition and Behavior* 1–4, [https://doi.org/10.1007/978-3-319-47829-6\\_29-1](https://doi.org/10.1007/978-3-319-47829-6_29-1) (Springer International Publishing, 2017).
41. Zhou, J. & Li, Y. SNPs deciding the rapid growth of cyanobacteria are alterable. *Proc. Natl. Acad. Sci.* **116**, 3945–3945 (2019).
42. Lou, W. *et al.* A specific single nucleotide polymorphism in the ATP synthase gene significantly improves environmental stress tolerance of *Synechococcus elongatus* PCC 7942. *Appl. Environ. Microbiol.* **84**, 1–16 (2018).
43. Kopfmann, S., Roesch, S. & Hess, W. Type II toxin–antitoxin systems in the unicellular cyanobacterium *Synechocystis* sp. PCC 6803. *Toxins (Basel)*. **8**, 228 (2016).
44. Makarova, K. S., Wolf, Y. I. & Koonin, E. V. Comprehensive comparative-genomic analysis of type 2 toxin–antitoxin systems and related mobile stress response systems in prokaryotes. *Biol. Direct* **4**, 19 (2009).
45. Ning, D. *et al.* Transcriptional and proteolytic regulation of the toxin–antitoxin locus vapBC10 (ssr2962/slr1767) on the chromosome of *Synechocystis* sp. PCC 6803. *PLoS One* **8**, e80716 (2013).
46. Fei, Q., Gao, E.-B., Liu, B., Wei, Y. & Ning, D. A toxin–antitoxin system VapBC15 from *Synechocystis* sp. PCC 6803 shows distinct regulatory features. *Genes (Basel)*. **9**, 173 (2018).

47. Erental, A., Sharon, I. & Engelberg-Kulka, H. Two programmed cell death systems in *Escherichia coli*: An apoptotic-like death is inhibited by the mazEF-mediated death pathway. *PLoS Biol.* **10**, e1001281 (2012).
48. Kopfmann, S. & Hess, W. R. Toxin-antitoxin systems on the large defense plasmid pSYSYA of *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **288**, 7399–409 (2013).
49. Shabestary, K. *et al.* Targeted repression of essential genes to arrest growth and increase carbon partitioning and biofuel titers in cyanobacteria. *ACS Synth. Biol.* **7**, 1669–1675 (2018).
50. Jahn, M. *et al.* Growth of cyanobacteria is constrained by the abundance of light and carbon assimilation proteins. *Cell. Rep.* **25**, 478–486.e8 (2018).
51. Cano, M. *et al.* Glycogen synthesis and metabolite overflow contribute to energy balancing in cyanobacteria. *Cell. Rep.* **23**, 667–672 (2018).
52. Mathew, A. K. & Padmanaban, V. C. Metabolomics: The apogee of the omics trilogy. *Int. J. Pharm. Pharm. Sci.* **5**, 45–48 (2013).
53. Fiehn, O. Metabolomics—the link between genotypes and phenotypes. *Plant Mol. Biol.* **48**, 155–171 (2002).
54. Guijas, C., Montenegro-Burke, J. R., Warth, B., Spilker, M. E. & Siuzdak, G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat. Biotechnol.* **36**, 316–320 (2018).
55. Qiu, Y. *et al.* Isotopic ratio outlier analysis of the *S. cerevisiae* metabolome using accurate mass gas chromatography/time-of-flight mass spectrometry: A new method for discovery. *Anal. Chem.* **88**, 2747–2754 (2016).
56. Stupp, G. S. *et al.* Isotopic ratio outlier analysis global metabolomics of *Caenorhabditis elegans*. *Anal. Chem.* **85**, 11858–11865 (2013).
57. Xu, Y.-F., Lu, W. & Rabinowitz, J. D. Avoiding misannotation of in-source fragmentation products as cellular metabolites in liquid chromatography–mass spectrometry-based metabolomics. *Anal. Chem.* **87**, 2273–2281 (2015).
58. Pagliano, E., Mester, Z. & Meija, J. Calibration graphs in isotope dilution mass spectrometry. *Anal. Chim. Acta* **896**, 63–67 (2015).
59. Liang, F. & Lindblad, P. Effects of overexpressing photosynthetic carbon flux control enzymes in the cyanobacterium *Synechocystis* PCC 6803. *Metab. Eng.* **38**, 56–64 (2016).
60. Janasch, M., Asplund-Samuelsson, J., Steuer, R. & Hudson, E. P. Kinetic modeling of the Calvin cycle identifies flux control and stable metabolomes in *Synechocystis* carbon fixation. *J. Exp. Bot.* **70**, 1017–1031 (2018).
61. Paul, M. J. & Foyer, C. H. Sink regulation of photosynthesis. *J. Exp. Bot.* **52**, 1383–1400 (2001).
62. Ducat, D. C., Avelar-Rivas, J. A., Way, J. C. & Silver, P. A. Rerouting carbon flux to enhance photosynthetic productivity. *Appl. Environ. Microbiol.* **78**, 2660–2668 (2012).
63. Li, X., Shen, C. R. & Liao, J. C. Isobutanol production as an alternative metabolic sink to rescue the growth deficiency of the glycogen mutant of *Synechococcus elongatus* PCC 7942. *Photosynth. Res.* **120**, 301–310 (2014).
64. Ungerer, J. *et al.* Sustained photosynthetic conversion of CO<sub>2</sub> to ethylene in recombinant cyanobacterium *Synechocystis* 6803. *Energy Environ. Sci.* **5**, 8998 (2012).
65. Asplund-Samuelsson, J., Janasch, M. & Hudson, E. P. Thermodynamic analysis of computed pathways integrated into the metabolic networks of *E. coli* and *Synechocystis* reveals contrasting expansion potential. *Metab. Eng.* **45**, 223–236 (2018).
66. Stevenson, J., Krycer, J. R., Phan, L. & Brown, A. J. A practical comparison of ligation-independent cloning techniques. *PLoS One* **8**, e83888 (2013).
67. Liu, D. & Pakrasi, H. B. Exploring native genetic elements as plug-in tools for synthetic biology in the cyanobacterium *Synechocystis* sp. PCC 6803. *Microb. Cell Fact.* **17**, 48 (2018).
68. Sievers, F. & Higgins, D. G. Clustal Omega, accurate alignment of very large numbers of sequences. In *Multiple Sequence Alignment Methods* (ed. Russell, D. J.) 105–116 (Humana Press, 2014).
69. Kumar, S., Stecher, G. & Tamura, K. MEGA7: Molecular evolutionary genetics analysis version 7.0 for bigger datasets. *Mol. Biol. Evol.* **33**, 1870–1874 (2016).
70. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL): An online tool for phylogenetic tree display and annotation. *Bioinformatics* **23**, 127–128 (2007).
71. Darling, A. C. E., Mau, B., Blattner, F. R. & Perna, N. T. Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* **14**, 1394–403 (2004).
72. Jaiswal, D., Prasannan, C. B., Hendry, J. I. & Wangikar, P. P. SWATH tandem mass spectrometry workflow for quantification of mass isotopologue distribution of intracellular metabolites and fragments labeled with isotopic <sup>13</sup>C Carbon. *Anal. Chem.* **90**, 6486–6493 (2018).
73. Prasannan, C. B., Jaiswal, D., Davis, R. & Wangikar, P. P. An improved method for extraction of polar and charged metabolites from cyanobacteria. *PLoS One* **13**, e0204273 (2018).
74. Xia, J. & Wishart, D. S. Web-based inference of biological patterns, functions and pathways from metabolomic data using MetaboAnalyst. *Nat. Protoc.* **6**, 743–760 (2011).
75. Chong, J. *et al.* MetaboAnalyst 4.0: towards more transparent and integrative metabolomics analysis. *Nucleic Acids Res.* **46**, W486–W494 (2018).
76. González, M.-C., Osuna, L., Echevarría, C., Vidal, J. & Cejudo, F. J. Expression and localization of phosphoenolpyruvate carboxylase in developing and germinating wheat grains. *Plant Physiol.* **116**, 1249–1258 (1998).
77. Rao, D. N., Dryden, D. T. F. & Bheemanaik, S. Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* **42**, 45–55 (2014).
78. Hu, L. *et al.* Transgenerational epigenetic inheritance under environmental stress by genome-wide DNA methylation profiling in cyanobacterium. *Front. Microbiol.* **9**, 1–11 (2018).
79. Toro, N. & Nisa-Martínez, R. Comprehensive phylogenetic analysis of bacterial reverse transcriptases. *PLoS One* **9**, 1–16 (2014).
80. Matiasovicova, J. *et al.* Retron reverse transcriptase rrtT is ubiquitous in strains of *Salmonella enterica* serovar typhimurium. *FEMS Microbiol. Lett.* **223**, 281–286 (2003).
81. Srikumar, A. *et al.* The Ssl2245-Sll1130 toxin-antitoxin system mediates heat-induced programmed cell death in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* **292**, 4222–4234 (2017).
82. Chan, W. T. *et al.* Genetic regulation of the yefM-yoeB toxin-antitoxin locus of *Streptococcus pneumoniae*. *J. Bacteriol.* **193**, 4612–4625 (2011).
83. Johnson, S. J., Jackson, R. N. & Rna, K. Ski2-like RNA helicase structures: common themes and complex assemblies. *RNA Biol.* **10**, 33–43 (2013).
84. Fivian-hughes, A. S. & Davis, E. O. Analyzing the regulatory role of the HigA antitoxin within *Mycobacterium tuberculosis*. *J. Bacteriol.* **192**, 4348–4356 (2010).

## Acknowledgements

The authors acknowledge grants from Department of Biotechnology (DBT), Government of India, awarded to PPW towards DBT-PAN IIT Center for Bioenergy (Grant No: BT/EB/PAN IIT/2012), the Indo-US Science and Technology Forum (IUSSTF) awarded to PPW and HBP for Indo-US Advanced Bioenergy Consortium (IUABC) (Grant No: IUSSTF/JCERDC-SGB/IUABC-IITB/2016) and Office of Science, Department of Energy-BER awarded to HBP. The authors are thankful to Prof. Santanu Kumar Ghosh, IIT Bombay for providing microscope facility.

### Author contributions

P.P.W. and H.B.P. conceived and supervised the research and acquired the funding. D.J., P.P.W. and H.B.P. designed the research. D.J., A.S., S.S. and S.M. performed the research. D.J., A.S. and P.P.W. analyzed the data. D.J. and P.P.W. wrote the manuscript.

### Competing interests

The authors declare no competing interests.

### Additional information

**Supplementary information** is available for this paper at <https://doi.org/10.1038/s41598-019-57051-0>.

**Correspondence** and requests for materials should be addressed to P.P.W.

**Reprints and permissions information** is available at [www.nature.com/reprints](http://www.nature.com/reprints).

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020