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

Daminji Jaiswal, Annesha Sengupta, Shinjinee Sengupta, Swati Madhu, Himadri B. Pakrasi & Pramod P. Wangikar

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$_2$, 1000 µmole photons.m$^{-2}$.s$^{-1}$ and grows faster under high CO$_2$ 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$_2$ 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$_2$) 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. 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. 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. Moreover, organisms that are tolerant to high temperatures, light, and CO$_2$ levels are desirable as industrial strains considering the eventual outdoor utilization in biorefineries.

1Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, Mumbai, 400076, India. 2DBT-PAN IIT Centre for Bioenergy, Indian Institute of Technology Bombay, Powai, Mumbai, 400076, India. 3Department of Biology, Washington University, St. Louis, MO, 63130, USA. 4Department of Energy, Environmental and Chemical Engineering, Washington University, St. Louis, MO, 63130, USA. 5Wadhwani Research Centre for Bioengineering, Indian Institute of Technology Bombay, Powai, Mumbai, 400076, India. *email: 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. Their growth rates are higher than model cyanobacterial strains like PCC 6803, PCC 7942, and PCC 7002. Detailed growth, biochemical characterization, and systems biology data is now available for UTEX 2973 that was reported in 2015. Likewise, the physiological and proteome changes in PCC 11801 under elevated CO₂ conditions have been studied. 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. 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.

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. 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. 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 pathway across different strains. Likewise, there are limited reports on comparative metabolite profiling of different cyanobacterial strains under conditions that are expected to result in higher productivity.

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 CO₂ conditions, PCC 11802 has a least doubling time of 2.8 h at 38 °C, 1% CO₂ and a light intensity of 1000 μmol photons.m⁻².s⁻¹. The two strains thus have different CO₂ 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 CO₂ 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 studies 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. 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 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 (OD₇₅₀ ~ 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). It has been reported that the cell size of cyanobacteria can be affected by growth rate. 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 CO₂ 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), Integrated Microbial Genomes and Microbiomes, JGI (IMG) and NCBI Prokaryotic Genome Annotation Pipeline. 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. 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. 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,
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 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. Type II TAS consists of a stable toxin protein and an antitoxin protein that loses the stability under stress conditions. 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. Most TAS systems cause programmed cell death, growth cessation or reversible growth arrest during the...
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 µmole photons.m$^{-2}$.s$^{-1}$, 38°C and varied CO$_2$ levels in the chamber. (B) Biomass accumulation of PCC 11802 compared to that of PCC 11801 at 1% CO$_2$ in SF. (C) Growth of PCC 11802 at high light (HL, 1000 µmole photons.m$^{-2}$.s$^{-1}$), high CO$_2$ (HC, 1% CO$_2$) compared with growth at high light (HL) and low CO$_2$ (LC, 0.04% CO$_2$) at 38°C in MC. (D) Dependence of specific growth rate of PCC 11802 on the light intensity at 38°C and 1% CO$_2$ in MC. (E) Specific growth rates of PCC 11802 compared to PCC 11801 at high light (1000 µmole photons.m$^{-2}$.s$^{-1}$) and high CO$_2$ concentrations (1% and 3%) at 38°C in MC. (F) Comparison of growth of PCC 11802 to PCC 11801 at low light (LL, 200 µmole photons.m$^{-2}$.s$^{-1}$) and low CO$_2$ (LC, 0.04% CO$_2$) at 38°C. (G) Specific growth rates of PCC 11802 at varying temperatures at 1000 µmole photons.m$^{-2}$.s$^{-1}$ and 1% CO$_2$ in MC, and (H) Growth profile of PCC 11802 at 1000 µmole photons.m$^{-2}$.s$^{-1}$, 43°C and 5% CO$_2$ in MC.
under the optimal conditions in MC was found to be 2.8 h (Fig. 2C). The growth rates of PCC 11802 and 11801 declined (~25%) compared to PCC 11801 (~54%). The growth of PCC 11802 under low light declined when tested at 3% CO2 in MC (Fig. 2E). However, the decline in the growth of PCC 11802 was much greater biomass compared to PCC 11801 under high CO2 conditions (Fig. 2B), which may be due to its higher growth rate at elevated CO2.

Growth characteristics of PCC 11802. The growth characterization of PCC 11802 was performed under a range of light, CO2 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 CO2 levels tested for growth in SF at 38 °C, the minimum doubling time of 3.2 h was observed for PCC 11802 at 1% CO2 (0.04–15%) (Fig. 2A). PCC 11801, on the other hand, showed the least doubling time of 3.9 h at 0.5% CO2 (3.4 h for PCC 11802) under similar light and temperature conditions in SF. Note that PCC 11802 has higher growth rates compared to PCC 11801 under elevated CO2 levels (Fig. 2A,E). Both these strains have similar growth rates at 0.04% and 10% CO2 in SF, while PCC 11802 was not able to grow at 15% CO2 (Fig. 2A). Further, PCC 11802 accumulates greater biomass compared to PCC 11801 under high CO2 conditions (Fig. 2B), which may be due to its higher growth rate at elevated CO2.

It has been reported that the growth rate of PCC 11801 under ambient CO2 increased by ~4 fold when grown in MC with the bubbling of air compared to that in SF19. 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%) CO2 at 1000 µmole photons.m−2.s−1 and 41 °C, PCC 11802 has its fastest growth at 1% CO2, 38 °C and 1000 µmole photons.m−2.s−1 with bubbling in MC (Fig. 2C). 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% CO2 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−2.s−1) and low CO2 (0.04%) condition was ~1.7 times faster than PCC 11801 (Fig. 2F, Table S2)19. PCC 11802 maintains a better growth rate (doubling time of ~3.6 h) even at a further higher level of CO2 (5%) at 1000 µmole photons.m−2.s−1 and 1% CO2 (Fig. 2H). It is evident from these results that PCC 11802 has a different CO2 specificity and grows better at elevated CO2 compared to PCC 11801. It is also noticeable that the growth of PCC 11802 during the exponential phase is limited by CO2, 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 CO2.50. Under low light and high CO2 conditions, the abundance of carbon assimilatory proteins reduces to compensate for expanded light-harvesting proteins50. The differential responses of PCC 11801 and PCC 11802 to the availability of light and CO2 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% CO2 (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 CO2 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 °C19. 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. Thus, both the strains have optimal growth temperatures ranging between 38–41 °C.

Carbon storage and energy charge under elevated CO2 conditions. The synthesis of glycogen and carbon partitioning plays an important role in energy balancing during cyanobacterial growth. Consistent with our previous results in PCC 11801, PCC 11802 also has lower carbohydrate and glycogen content (Fig. 4A,B).

| Gene count | RAST | IMG | NCBI Prokaryotic Genome Annotation Pipeline |
|------------|------|-----|-------------------------------------------|
| 2924       | 2919 | 2752 |
| 54.8       | 54.8 | —   |
| 2883       | 2866 | 2708 |
| 2127       | 2201 | 2446 |
| 756        | 665  | 262  |
| 41         | 55   | 44   |
| 39         | 39   | 39   |
| 2          | 3    | 2    |
| —          | 11   | 3    |

Table 1. Genome annotation for Synechococcus elongatus PCC 11802 obtained using RAST, IMG, and NCBI prokaryotic genome annotation pipeline.
Type III restriction-modification system methylation

Proteins unique to PCC 11802

| Locus Tag† | Protein ID† | Putative Role | Remarks based on previous reports |
|------------|-------------|---------------|----------------------------------|
| peg.1952*  | 1823836–1821956 | DNA methylation | May be involved in epigenetic regulation of genes through differential methylation of genes. |

Retron-type RNA-directed DNA polymerase

Programmed cell death antitoxin MacE like

Programmed cell death toxan MacF like

RelB/StbD replication stabilization protein (antitoxin to RelE/StbE)

VapE protein (antitoxin to VapC)

VapC toxin protein

YefM protein (antitoxin to YoeB)

YoeB toxin protein

Putative ski2-type helicase MJ1124

HigA protein (antitoxin to HigB)

HigB toxin protein

**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. *The IDs are based on annotations obtained through NCBI Prokaryotic Genome Annotation Pipeline.** 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. **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% CO2 (Fig. 2A). PCC 11801 had the lowest glycogen content at 0.5% CO2, where its growth rate is maximum in SF (Fig. 2A). It is noticeable that the growth rate of PCC 11801 at 1% CO2 is less compared to 0.5% CO2 (Fig. 2A). The glycogen content of PCC 11801 at 1% CO2 is higher than at 0.5% CO2, 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% CO2 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 CO2 conditions (Fig. 4C). However, the glycogen content of PCC 11802 is ~5 times lower than PCC 11801 at 1% CO2. This suggests that PCC 11802 is more efficient than PCC 11801 in utilizing the available CO2 for growth rather than glycogen storage. This further strengthens our results on better growth and carbon assimilation efficiency of PCC 11802 under high CO2 conditions compared to PCC 11801.

PCC 11802 has elevated levels of nucleotides, ATP, ADP, and AMP at 1% CO2 compared to ambient CO2 (Fig. 4D). It is reported that the levels of these nucleotides increase before entering the exponential phase and decline steadily after the exponential phase41. Higher levels of these energy nucleotides at elevated CO2 might be indicative of better photosynthesis and ATP production. The higher growth rate under elevated CO2 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 CO2, the energy charge at ambient and 1% CO2 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 CO2 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 pathway32–34. 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 reactions30. 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 sensitivity39,55,56. 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 standard39,55–58.

We measured the relative metabolite pools of PCC 11801 and PCC 11802 by isotopic ratio method39,55–58. We utilized the intracellular metabolite extract of PCC 11801, which is fully labeled with isotopic 13C as an internal standard. This strategy allowed us to use 13C 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
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% CO2 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. CO2), 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 CO2 fixation was elevated at 1% CO2 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 bisphosphate (SBP) in PCC 11801 at 1% CO2, the levels of sedoheptulose-7-phosphate (S7P) showed the negligible change between ambient and 1% CO2. This indicates that the conversion of SBP to S7P catalyzed by the enzymes sedoheptulose 1, 7 bisphosphatase (SBPase)
might be rate-limiting under 1% CO₂. This might be a reason why increasing the CO₂ 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-bisphosphate (FDPase) 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. 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 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 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. 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 CO₂ conditions. 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, 2,3 butanediol, isobutanol, and ethylene 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. For enhanced production of biofuels using cyanobacterial hosts, strains capable of better growth under elevated CO₂ 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 CO₂ 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% CO₂. 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% CO₂ 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 (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 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 ± 10 cfu/µg of the plasmid (eYFP) using 7 mL culture of 0.6 OD₅₇₀.

For high-throughput characterization of biological parts, reporter systems such as enhanced yellow fluorescent protein (eYFP), mOrange, luciferase are commonly used. 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 (PEP) 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. Therefore, we demonstrate the heterologous expression of a reporter and a functional gene encoding for eYFP and PEP in PCC 11802 under cpcB promoter of PCC 11801 (Fig. 6B,C). 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% CO₂, 38 °C and 1000 µmole photons m⁻² s⁻¹ 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 CO₂, low light-high CO₂ and high light-high CO₂ conditions. PCC 11802 accumulates ~3 fold more and ~5 fold less glycogen than PCC 11801 under ambient and 1% CO₂, respectively. PCC 11802 thus diverts more carbon for growth at 1% CO₂ 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 CO₂ conditions. The results
Figure 5. LCMS based analysis of metabolic pool sizes in PCC 11802 and 11801 under elevated CO₂ conditions. The bar graphs represent the ratio of metabolite pools at 1% CO₂/ambient (0.04%) CO₂ 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 ¹³C 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 ¹²C monoisotopic peak to the respective ¹³C 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, S7P: sedoheptulose-7-bisphosphate, S7P: sedoheptulose-7-bisphosphate, UDPG: UDP-glucose, 2PGA: 2-phosphoglyceric acid, and 3PGA: 3-phosphoglyceric acid.
show that PCC 11802 has greater abundance of the CBB cycle intermediates and lesser triose phosphate utilization limitation under elevated CO2 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 CO2 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. The genomic DNA of PCC 11802 was isolated using a protocol described earlier 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. 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) and Rapid Annotation using Subsystem Technology (RAST), and NCBI Prokaryotic Genome Annotation Pipeline.
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. Concatenated sequences of these 29 proteins for 130 strains were obtained from the CyanobERA resource. The respective sequences were aligned using Clustal Omega, and the phylogenetic tree was constructed using MEGA7 and iTOL (Interactive Tree Of Life).

Comparative genome analysis. The function-based genome comparison was performed using RAST to identify the proteins unique to PCC 11802 compared to PCC 11801. The sequence-based comparison was performed using RAST 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, 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.

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 µL 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 ~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%) CO2, and light intensity of 300 µmol photons.m−2.s−1 in BG-11 medium (pH = 7.5) unless specified otherwise. The CO2 tolerance studies were performed in CO2 incubator shaker (Adolf Kuhner AG, LT-X, Birsfelden, Switzerland) at CO2 concentrations of 0.04%, 0.5%, 1%, 5%, 10% and 15% in the chamber at 38 °C and 120 rpm. The growth at different CO2 concentrations was measured as OD720 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 CO2 concentration was measured in terms of OD720 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 (μ) was estimated during the exponential growth phase from the slope of the semi-logarithmic plot between OD720 nm and time. Doubling time was calculated as ln (2)/μ.

Carbohydrate and glycogen estimation. The total carbohydrates and glycogen content were measured using PCC 11802 cells harvested at 0.6–0.7 OD720 nm at different CO2 concentrations (0.04%, 0.5%, and 1%), 38 °C, 400 µmol photons.m−2.s−1 and 120 rpm as described earlier.

Preparation of 13C 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 13C mono-isotopic peaks but no 12C mono-isotopic 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 13C 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 O2 concentration 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 12C present in the biomass from the inoculum. A stopper was used to prevent the exchange of 12CO2 from the environment. Initially, 13C-labeled sodium bicarbonate (NaH13CO3, 98 atom % 13C from Sigma-Aldrich, St. Louis, MO) was added at a concentration of 2g/L in the culture. Additional doses of NaH13CO3 were provided at 18, 19.5, 21, and 22.5 hours after inoculation at a final concentration of 1 g/L. The 13C 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 13C labeled metabolites were extracted from quenched cells as described earlier with a minor modification, viz., the use of extraction solvent volume that was twice as large as earlier. The 13C 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 13C 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 µmol photons.m−2.s−1.

Metabolite profiling using liquid chromatography-mass spectrometry (LCMS). PCC 11802 and 11801 cells were grown under 0.04% and 1% CO2 at 38 °C in shake flasks. Exponentially growing cells at an OD720 of ~0.6 were filtered rapidly on nylon membrane filters (Whatman, 0.8 µ) 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. The metabolite extract was lyophilized and stored at ~80 °C until ready for LCMS analysis. The metabolite extracts were reconstituted in 100 µL 50:50 methanol-water and filtered using nylon syringe filters to remove any particulate matter. Each sample was mixed with fully 13C-labeled biomass of PCC 11801 that acted as an internal standard.
obtained by transforming pSyn_11801 plasmids were also verified using gene-specific primers. The recombinant 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 BG-11 agar plate and incubated for 24 h at 38 °C and 150 μE light. Then the filter membrane was transferred to a BG-11 agar plate containing 50 μg/mL of antibiotic, spectinomycin. After 48 h of incubation, colonies appeared on the filter membrane, which was then patched on a plate containing 100 μg/mL of spectinomycin to achieve complete chromosomal segregation. Segregation was checked using confirmation primers, 5′CCAACGCCCTATTTCCAAGGGCGGC3′, and 5′TGGCAATGTCTCTCTAGGAGGATG3′. All the colonies obtained by transforming pSyn_11801 plasmids were also verified using gene-specific primers.

**Genetic manipulation.** Seven mL of 0.6 OD730 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 μL BG-11 medium. Two μg 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 μm filter membrane placed on a 1% BG-11 agar plate and incubated for 24 h at 38 °C and 150 μE light. Then the filter membrane was transferred to a BG-11 agar plate containing 50 μg/mL of antibiotic, spectinomycin. After 48 h of incubation, colonies appeared on the filter membrane, which was then patched on a plate containing 100 μg/mL of spectinomycin to achieve complete chromosomal segregation. Segregation was checked using confirmation primers, 5′CCAACGCCCTATTTCCAAGGGCGGC3′, and 5′TGGCAATGTCTCTCTAGGAGGATG3′. All the colonies obtained by transforming pSyn_11801 plasmids were also verified using gene-specific primers.

**Fluorescence microscopy.** Fluorescence microscopy was performed as described earlier19. Briefly, the wild-type and eYFP expressing PCC 11802 cells (OD730 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 AxioiCam 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 wild-type and recombinant PCC 11802 cells containing a gene encoding PEPC protein were grown till OD730 of 0.6–0.7 at 1% CO2 and 200 μE 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 μL of lysis buffer containing 50 mM TRIS HCl (pH 8), 10 mM MgCl2, 10% (v/v) glycerol, 1 mM EDTA, 1 mM DTT and 1 mg/L of lysozyme. One mM PMSF was added to minimize proteolysis18. 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 is catalyzed by the added malate dehydrogenase, which was overexpressed in E. coli and purified. The reaction mixture thus contained 100 mM Tris-HCl (pH 8), 10 mM NaHCO3, 5 mM MgCl2, 200 μM NADH, 0.13 μg 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 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).
11. Santos-Merino, M., Singh, A. K. & Ducat, D. C. New applications of synthetic biology tools for cyanobacterial metabolic engineering.

12. Tan, X., Liang, Y. & Lu, X. The potential of Synechococcus sp. strain PCC 7002 to convert CO2 to ethanol.

13. Hendry, J. I. & Wei, C. T. Metabolic engineering of cyanobacteria for the photosynthetic production of succinate.

14. Mehta, K. & Bryant, D. A. The role of systems biology in developing non-model cyanobacteria as hosts for chemical production.

15. Ludwig, M. & Bryant, D. A. Synechococcus sp. strain PCC 7002 transcriptome: Accretion to temperature, salinity, oxidative stress, and mixotrophic growth conditions.

16. Vasudevan, R. et al. CyanGate: A modular cloning suite for engineering cyanobacteria based on the plant MoClo system.

17. Knoot, C. J., Khatri, Y., Hohliman, R. M., Sherman, D. H. & Pakrasi, H. B. Engineered production of haptalinidole alkaloids in the cyanobacterium Synechocystis sp. UTEX 2973. ACS Synth. Biol., 1053–1058 (2019).

18. Song, K., Tan, X., Liang, Y. & Lu, X. The potential of Synechococcus sp. strain PCC 7002 to develop a synthetic toolbox for heterologous protein expression.

19. Sengupta, A., Pakrasi, H. B. & Wangikar, P. P. Recent advances in synthetic biology of cyanobacteria.

20. Yu, J. et al. Synechocystis PCC 7002, a fast growing cyanobacterial chassis for biomass synthesis using light and CO2.

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 Synechocystis PCC 7002. Proc. Natl. Acad. Sci. 115, E17761–E17770 (2018).

22. Ungerer, J. et al. The RAST Server: Rapid annotations using subsystems technology.

23. Calteau, A. et al. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in cyanobacteria. BMC Genomics 15, 1–14 (2014).

24. Zheng, X. & O’Shea, E. K. Cyanobacteria maintain constant protein concentration despite genome copy-number variation.

25. Burtin, T. et al. RASTK: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5, 8365 (2015).

26. Arora, S. K. et al. The RAST Server: Rapid annotations using subsystems technology.

27. Overbeek, R. et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST).

28. Ackermann, M. H. et al. Deciphering cyanobacterial phenotypes for fast phototrophic growth via isotopically nonstationary metabolic flux analysis. Biotechnol. Biofuels 11, 218 (2018).

29. Aerts, A. K. & Mooy, R. T. Metabolomics and isotope tracing. Cell 173, 822–837 (2018).

30. Will, S. E. et al. Day and night: Metabolic profiles and evolutionary relationships of six axenic non-marine cyanobacteria.

31. 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).

32. Calteau, A. et al. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in cyanobacteria. BMC Genomics 15, 1–14 (2014).

33. Zheng, X. & O’Shea, E. K. Cyanobacteria maintain constant protein concentration despite genome copy-number variation. Cell Rep. 19, 497–504 (2017).

34. Brettn, T. et al. RASTK: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5, 8365 (2015).

35. Arora, S. K. et al. The RAST Server: Rapid annotations using subsystems technology.

36. Overbeek, R. et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST).

37. Ackermann, M. H. et al. Deciphering cyanobacterial phenotypes for fast phototrophic growth via isotopically nonstationary metabolic flux analysis. Biotechnol. Biofuels 11, 1–13 (2017).

38. Jang, C., Chen, L. & Rabenowit, J. D. Metabolomics and isotope tracing. Cell 173, 822–837 (2018).

39. 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).

40. 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).

41. Calteau, A. et al. Phylum-wide comparative genomics unravel the diversity of secondary metabolism in cyanobacteria. BMC Genomics 15, 1–14 (2014).

42. Zheng, X. & O’Shea, E. K. Cyanobacteria maintain constant protein concentration despite genome copy-number variation. Cell Rep. 19, 497–504 (2017).

43. Brettn, T. et al. RASTK: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and annotating batches of genomes. Sci. Rep. 5, 8365 (2015).

44. Arora, S. K. et al. The RAST Server: Rapid annotations using subsystems technology.

45. Overbeek, R. et al. The SEED and the rapid annotation of microbial genomes using subsystems technology (RAST).

46. Ackermann, M. H. et al. Deciphering cyanobacterial phenotypes for fast phototrophic growth via isotopically nonstationary metabolic flux analysis. Biotechnol. Biofuels 11, 1–13 (2017).
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 pSYA 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. Guitiás, C., Montenegro-Burke, J. R., Warth, R., Spilker, M. E. & Szudak, G. Metabolomics activity screening for identifying metabolites that modulate phenotype. *Nat. Biotechnol.* 36, 316–320 (2018).

55. Qui, Y. et al. Isotypic ratio outlier analysis of the *S. cerevisiae* metabolism using accurate mass 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. & Mejia, 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 CO2 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 *Methods in Enzymology*, vol. 519, pp. 133–142 (Elsevier, 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., Prasanna, C. B., Hendry, J. I. & Wangikar, P. P. SWATH tandem mass spectrometry workflow for quantification of mass isotopologue distribution of intracellular metabolites and fragements labeled with isotopic 13C Carbon. *Anal. Chem.* 90, 6486–6493 (2018).

73. Prasanna, 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., Echevarria, 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. & Bheeanaik, S. Type III restriction-modification enzymes: a historical perspective. *Nucleic Acids Res.* 42, 15–55 (2014).

78. Hu, L. et al. Transgenerational epigenetic inheritance under environmental stress by genome-wide DNA methylation profiling in cyanobacterium. *Front. Microb.* 14, 1–11 (2018).

79. Toro, N. & Nisa-Martinez, R. Comprehensive phylogenetic analysis of bacterial reverse transcriptases. *PLoS One* 9, 1–16 (2014).

80. Matiasiovicova, J. et al. Retrotranscribe transcriptase rRTIs is ubiquitous in strains of *Salmonella enterica serovar typhimurium*. *FEMS Microbiol. Lett.* 223, 281–286 (2003).

81. Srikrumar, A. et al. The Sda2245-SBII130 toxin-antitoxin system mediates heat-induced programmed cell death in *Synechocystis* sp. PCC 6803. *J. Biol. Chem.* 292, 4222–4234 (2017).

82. Chen, 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 *Streptococcus pneumoniae* strains. *RNA Biol.* 192, 3438–4356 (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/ICERDC-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.

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